

International Journal of

Computational Linguistics & Chinese Language Processing

中文計算語言學期刊

A Publication of the Association for Computational Linguistics and Chinese Language Processing

This journal is included in THCI, Linguistics Abstracts, and ACL Anthology.

Special Issue on “Selected Papers from ROCLING XXVII”

Guest Editors: Hung-Yu Kao, Yih-Ru Wang, and Jen-Tzung Chien

易繫辭曰上古結繩而
治後世聖人易之以書
契百官以治萬民以察
說文敘曰蓋文字者經
藝之本宣教明化之始
前人所以垂後後人所
以識古故曰本立而道
生知天下之至蹟而不
可亂也教化既萌文心
雕龍則謂人之立言因
字而生句積句而成章
積章而成篇篇之彪炳

Vol.20

No.2

December 2015

ISSN: 1027-376X

International Journal of Computational Linguistics & Chinese Language Processing

Advisory Board

- Jason S. Chang*
National Tsing Hua University, Hsinchu
- Hsin-Hsi Chen*
National Taiwan University, Taipei
- Keh-Jiann Chen*
Academia Sinica, Taipei
- Sin-Horng Chen*
National Chiao Tung University, Hsinchu
- Eduard Hovy*
University of Southern California, U. S. A.
- Chu-Ren Huang*
The Hong Kong Polytechnic University, H. K.
- Jian-Yun Nie*
University of Montreal, Canada
- Richard Sproat*
University of Illinois at Urbana-Champaign, U. S. A.
- Keh-Yih Su*
Behavior Design Corporation, Hsinchu
- Chiu-Yu Tseng*
Academia Sinica, Taipei
- Jhing-Fa Wang*
National Cheng Kung University, Tainan
- Kam-Fai Wong*
Chinese University of Hong Kong, H.K.
- Chung-Hsien Wu*
National Cheng Kung University, Tainan

Editorial Board

- Yuen-Hsien Tseng (Editor-in-Chief)*
National Taiwan Normal University, Taipei
- Kuang-hua Chen (Editor-in-Chief)*
National Taiwan University, Taipei
- Speech Processing**
- Yuan-Fu Liao (Section Editor)*
National Taipei University of Technology, Taipei
- Berlin Chen*
National Taiwan Normal University, Taipei
- Hung-Yan Gu*
National Taiwan University of Science and Technology, Taipei
- Hsin-Min Wang*
Academia Sinica, Taipei
- Yih-Ru Wang*
National Chiao Tung University, Hsinchu
- Linguistics & Language Teaching**
- Shu-Kai Hsieh (Section Editor)*
National Taiwan University, Taipei
- Hsun-Huei Chang*
National Chengchi University, Taipei
- Hao-Jan Chen*
National Taiwan Normal University, Taipei
- Huei-ling Lai*
National Chengchi University, Taipei
- Meichun Liu*
National Chiao Tung University, Hsinchu
- James Myers*
National Chung Cheng University, Chiayi
- Shu-Chuan Tseng*
Academia Sinica, Taipei
- Information Retrieval**
- Ming-Feng Tsai (Section Editor)*
National Chengchi University, Taipei
- Chia-Hui Chang*
National Central University, Taoyuan
- Chin-Yew Lin*
Microsoft Research Asia, Beijing
- Shou-De Lin*
National Taiwan University, Taipei
- Wen-Hsiang Lu*
National Cheng Kung University, Tainan
- Shih-Hung Wu*
Chaoyang University of Technology, Taichung
- Natural Language Processing**
- Richard Tzong-Han Tsai (Section Editor)*
Yuan Ze University, Chungli
- Lun-Wei Ku*
Academia Sinica, Taipei
- Chuan-Jie Lin*
National Taiwan Ocean University, Keelung
- Chao-Lin Liu*
National Chengchi University, Taipei
- Jyi-Shane Liu*
National Chengchi University, Taipei
- Liang-Chih Yu*
Yuan Ze University, Chungli

Executive Editor: *Abby Ho*

English Editor: *Joseph Harwood*

The Association for Computational Linguistics and Chinese Language Processing, Taipei

International Journal of

Computational Linguistics & Chinese Language Processing

Aims and Scope

International Journal of Computational Linguistics and Chinese Language Processing (IJCLCLP) is an international journal published by the Association for Computational Linguistics and Chinese Language Processing (ACLCLP). This journal was founded in August 1996 and is published four issues per year since 2005. This journal covers all aspects related to computational linguistics and speech/text processing of all natural languages. Possible topics for manuscript submitted to the journal include, but are not limited to:

- Computational Linguistics
- Natural Language Processing
- Machine Translation
- Language Generation
- Language Learning
- Speech Analysis/Synthesis
- Speech Recognition/Understanding
- Spoken Dialog Systems
- Information Retrieval and Extraction
- Web Information Extraction/Mining
- Corpus Linguistics
- Multilingual/Cross-lingual Language Processing

Membership & Subscriptions

If you are interested in joining ACLCLP, please see appendix for further information.

Copyright

© The Association for Computational Linguistics and Chinese Language Processing

International Journal of Computational Linguistics and Chinese Language Processing is published four issues per volume by the Association for Computational Linguistics and Chinese Language Processing. Responsibility for the contents rests upon the authors and not upon ACLCLP, or its members. Copyright by the Association for Computational Linguistics and Chinese Language Processing. All rights reserved. No part of this journal may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical photocopying, recording or otherwise, without prior permission in writing form from the Editor-in Chief.

Cover

Calligraphy by Professor Ching-Chun Hsieh, founding president of ACLCLP

Text excerpted and compiled from ancient Chinese classics, dating back to 700 B.C.

This calligraphy honors the interaction and influence between text and language

Contents

Special Issue Articles:

Selected Papers from ROCLING XXVII

Forewords.....	i
<i>Hung-Yu Kao, Yih-Ru Wang, and Jen-Tzung Chien</i> <i>Guest Editors</i>	
Papers	
Designing a Tag-Based Statistical Math Word Problem Solver with Reasoning and Explanation.....	1
<i>Yi-Chung Lin, Chao-Chun Liang, Kuang-Yi Hsu, Chien-Tsung Huang, Shen-Yun Miao, Wei-Yun Ma, Lun-Wei Ku, Churn-Jung Liau, and Keh-Yih Su</i>	
Explanation Generation for a Math Word Problem Solver.....	27
<i>Chien-Tsung Huang, Yi-Chung Lin, and Keh-Yih Su</i>	
Word Co-occurrence Augmented Topic Model in Short Text.....	45
<i>Guan-Bin Chen and Hung-Yu Kao</i>	
節錄式語音文件摘要使用表示法學習技術 [Extractive Spoken Document Summarization with Representation Learning Techniques].....	65
<i>施凱文(Kai-Wun Shih), 陳冠宇(Kuan-Yu Chen), 劉士弘 (Shih-Hung Liu), 王新民(Hsin-Min Wang), 陳柏琳(Berlin Chen)</i>	
調變頻譜分解技術於強健語音辨識之研究 [Investigating Modulation Spectrum Factorization Techniques for Robust Speech Recognition].....	87
<i>張庭豪(Ting-Hao Chang), 洪孝宗(Hsiao-Tsung Hung), 陳冠宇 (Kuan-Yu Chen), 王新民(Hsin-Min Wang), 陳柏琳(Berlin Chen)</i>	
透過語音特徵建構基於堆疊稀疏自編碼器演算法之婚姻治療 中夫妻互動行為量表自動化評分系統 [Automating Behavior Coding for Distressed Couples Interactions Based on Stacked Sparse Autoencoder Framework using Speech-acoustic Features].	107
<i>陳柏軒(Po-Hsuan Chen), 李祈均(Chi-Chun Lee)</i>	
Reviewers List & 2015 Index.....	121

Forewords

The 27th Conference on Computational Linguistics and Speech Processing (ROCLING 2017) was held at National Chiao Tung University, Hsinchu, Taiwan on Oct. 1-2, 2015. ROCLING, which sponsored by the Association for Computational Linguistics and Chinese Language Processing (ACLCLP), is the leading and most comprehensive conference on computational linguistics and speech processing in Taiwan, bringing together researchers, scientists and industry participants from fields of computational linguistics, information understanding, and speech processing, to present their work and discuss recent trends in the field. This special issue presents extended and reviewed versions of six papers meticulously selected from ROCLING 2015, including 3 natural language processing papers and 3 speech processing papers.

The first two papers from Academia Sinica focused the math word problem solver. The first one paper proposes a tag-based statistical framework to solve math word problems with understanding and reasoning. It analyzes the body and question texts into their associated tag-based logic forms, and then performs inference on them. The proposed statistical approach alleviates rules coverage and ambiguity resolution problems, and their tag-based approach also provides the flexibility of handling various kinds of related questions with the same body logic form. This paper is also awarded as the best paper of ROCLING 2015. The second paper proposes a math operation oriented approach to explain how the answers are obtained for math word problems. They adopt a specific template to generate the text for each kind of math operator. This is also the first explanation generation that is specifically tailored to the math word problems. The third paper from National Cheng Kung University focused the problem of the frequent bi-term in BTM. This paper proposed an improvement of word co-occurrence method to enhance the topic models. They apply the word co-occurrence information to the BTM. The experimental result that show the enhanced PMI- β -BTM gets better results in the both of regular short news title text and the noisy tweet text.

The last three papers are spoken language processing papers. The first two of them are co-works from National Taiwan Normal University and Academia Sinica. The first one explores a novel use of both word and sentence representation techniques for extractive spoken document summarization. In this paper, three variants of sentence ranking models building on top of such representation techniques are also proposed. The second one attempts to obtain noise-robust speech features through modulation spectrum processing of the original speech features. They explore the use of nonnegative matrix factorization (NMF) and its extensions on the magnitude modulation spectra of speech features so as to distill the most important and noise-resistant information cues that can benefit the ASR performance. The last paper from Nation Tsing Hua University aims at using machine learning approach to automate the observations of human behaviors, and by using signal processing technique. This paper proposes to use stacked sparse

autoencoder (SSAE) to reduce the dimensionality of the acoustic-prosodic features used in order to identify the key higher-level features.

The Guest Editors of this special issue would like to thank all of the authors and reviewers for sharing their knowledge and experience at the conference. We hope this issue provide for directing and inspiring new pathways of NLP and spoken language research within the research field.

Guest Editors

Hung-Yu Kao

Department of Computer Science and Information Engineering, National Cheng Kung University, Taiwan

Yih-Ru Wang

Department of Electrical and Computer Engineering, National Chiao Tung University, Taiwan

Jen-Tzung Chien

Department of Electrical and Computer Engineering, National Chiao Tung University, Taiwan

Designing a Tag-Based Statistical Math Word Problem Solver with Reasoning and Explanation

Yi-Chung Lin*, Chao-Chun Liang*, Kuang-Yi Hsu*,

Chien-Tsung Huang*, Shen-Yun Miao*, Wei-Yun Ma*,

Lun-Wei Ku*, Churn-Jung Liao*, and Keh-Yih Su*

Abstract

This paper proposes a tag-based statistical framework to solve math word problems with understanding and reasoning. It analyzes the body and question texts into their associated tag-based logic forms, and then performs inference on them. Comparing to those rule-based approaches, the proposed statistical approach alleviates rules coverage and ambiguity resolution problems, and our tag-based approach also provides the flexibility of handling various kinds of related questions with the same body logic form. On the other hand, comparing to those purely statistical approaches, the proposed approach is more robust to the irrelevant information and could more accurately provide the answer. The major contributions of our work are: (1) proposing a tag-based logic representation such that the system is less sensitive to the irrelevant information and could provide answer more precisely; (2) proposing a unified statistical framework for performing reasoning from the given text.

Keywords: Math Word Problem Solver, Machine Reading, Natural Language Understanding.

1. Introduction

Since *Big Data* mainly aims to explore the correlation between surface features but not their underlying causality relationship, the *Big Mechanism*¹ program was initiated by DARPA

* Institute of Information Science, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 11529, Taiwan

E-mail: {lyc; celiang; ianhsu; joeeth; jackymiu; ma; lwku; liaucj; kysu}@iis.sinica.edu.tw

¹ http://www.darpa.mil/Our_Work/I2O/Programs/Big_Mechanism.aspx

(from July 2014) to find out “*why*” behind the “Big Data”. However, the pre-requisite for it is that the machine can read each document and learn its associated knowledge, which is the task of *Machine Reading* (MR) (Strassel *et al.*, 2010). Therefore, the Natural Language and Knowledge Processing Group, under the Institute of Information Science of Academia Sinica, formally launched a 3-year MR project (from January 2015) to attack this problem.

As a domain-independent MR system is complicated and difficult to build, the *math word problem* (MWP) (Mukherjee & Garain, 2008) is chosen as the first task to study MR for the following reasons: (1) Since the answer for the MWP cannot be extracted by simply performing keyword matching (as Q&A usually does), MWP thus can act as a test-bed for understanding the text and then drawing the answer via inference. (2) MWP usually possesses less complicated syntax and requires less amount of domain knowledge. It can let the researcher focus on the task of understanding and reasoning, not on how to build a wide-coverage grammar and acquire domain knowledge. (3) The body part of MWP (which mentions the given information for solving the problem) usually consists of only a few sentences. Therefore, the understanding and reasoning procedure could be checked more efficiently. (4) The MWP solver could have its own standalone applications, such as computer tutor, etc. It is not just a toy test case.

According to the framework of making the decision while there are several candidates, previous MWP algebra solvers can be classified into: (1) Rule-based approaches with logic inference (Bobrow, 1964; Slagle, 1965; Charniak, 1968, 1969; Dellarosa, 1986; Bakman, 2007), which apply rules to get the answer (via identifying entities, quantities, operations, etc.) with a logic inference engine. (2) Rule-based approaches without logic inference (Gelb, 1971; Ballard & Biermann, 1979; Biermann & Ballard, 1980; Biermann *et al.*, 1982; Fletcher, 1985; Hosseini *et al.*, 2014), which apply rules to get the answer without a logic inference engine. (3) Purely statistics-based approaches (Kushman *et al.*, 2014; Roy *et al.*, 2015), which use statistical models to identify entities, quantities, operations, and get the answer without conducting language analysis or inference.

The main problem of the rule-based approaches mentioned above is that the coverage rate problem is serious, as rules with wide coverage are difficult and expensive to construct. Also, it is awkward in resolving ambiguity problems. Besides, since they adopt Go/No-Go approach (unlike statistical approaches which can adopt a large Top-N to have high including rates), the error accumulation problem would be severe. On the other hand, the main problem of those approaches not adopting logic inference is that they usually need to implement a new handling procedure for each new type of problems (as the general logic inference mechanism is not adopted). Also, as there is no inference engine to generate the reasoning chain, additional effort would be required for generating the explanation. In contrast, the main problem of those purely statistical approaches is that they are sensitive to irrelevant

information (Hosseini *et al.*, 2014) (as the problem is solved without first understanding the text). Also, the performance deteriorates significantly when they encounter complicated problems due to the same reason.

To avoid the problems mentioned above, a *tag-based statistical framework* which is able to perform understanding and reasoning is proposed in this paper. For each body statement (which specifies the given information), the text will be first analyzed into its corresponding semantic tree (with its anaphora/ellipses resolved and semantic roles labeled), and then converted into its associated logic form (via a few mapping rules). The obtained logic form is then mapped into its corresponding domain dependent generic concepts (also expressed in logic form). The same process also goes for the question text (which specifies the desired answer). Finally, the inference (based on the question logic form) is performed on the logic statements derived from the body text. Please note that a statistical model will be applied each time when we have choices.

Furthermore, to reply any kind of questions associated with the given information, we keep all related semantic roles (such as agent, patient, etc.) and associated *specifiers* (which restrict the given quantity, and is freely exchangeable with the term *tag*) in the logic form (such as verb(q1,進貨), agent(q1,文具店), head(n1_p,筆), color(n1_p,紅), etc.), which are regarded as various *tags* (or conditions) for selecting the appropriate information related to the given question. Therefore, the proposed approach can be regarded as a *tag-based statistical approach with logic inference*. Since extra-linguistic knowledge would be required for bridging the gap between the linguistic semantic form and the desired logic form, we will extract the desired background knowledge (ontology) from E-HowNet (Chen *et al.*, 2005) for verb-entailment.

In comparison with those rule-based approaches, the proposed approach alleviates the ambiguity resolution problem (i.e., selecting the appropriate semantic tree, anaphora/co-reference, domain-dependent concepts, inference rules) via a statistical framework. Furthermore, our tag-based approach provides the flexibility of handling various kinds of possible questions with the same body logic form. On the other hand, in comparison with those purely statistical approaches, the proposed approach is more robust to the irrelevant information (Hosseini *et al.*, 2014) and could provide the answer more precisely (as the semantic analysis and the tag-based logic inference are adopted). In addition, with the given reasoning chain, the explanation could be more easily generated. Last, since logic inference is a general problem solving mechanism, the proposed approach can solve various types of problems that the inference engine could handle (i.e., not only arithmetic or algebra as most approaches aim to).

The contributions of our work are: (1) Proposing a semantic composition form for abstracting the text meaning to perform semantic reasoning; (2) Proposing verb entailment via

E-HowNet for bridging the lexical gap (Moldovan & Rus, 2001); (3) Proposing a tag-based logic representation to adopt one body logic form for handling various possible questions; (4) Proposing a set of domain dependent (for math algebra) generic concepts for solving MWP; (5) Proposing a statistical solution type classifier to indicate the way for solving MWP; (6) Proposing a semantic matching method for performing unification; (7) Proposing a statistical framework for performing reasoning from the given text.

2. Design Principles

Since we will have various design options in implementing a math word problem solver, we need some guidelines to judge which option is better when there is a choice. Some principles are thus proposed as follows for this purpose:

- (1) Solutions should be given via understanding and inference (versus the template matching approach proposed in (Kushman *et al.*, 2014), as the math word problem is just the first case for our text understanding project and we should be able to explain how the answer is obtained.
- (2) The expressiveness of the adopted body logical form should be powerful enough for handling various kinds of possible questions related to the body, which implies that logic form transformation should be information lossless. In other words, all the information carried by the semantic representation should be kept in the corresponding logical form. It also implies that the associated body logical form should be independent on the given question (as we don't know which question will be asked later).
- (3) The dynamically constructed knowledge should not favor any specific kind of problem/question. This principle suggests that the *Inference Engine* (IE) should regard logic statements as a flat list, instead of adopting a pre-specified hierarchical structure (e.g., the container adopted in (Hosseini *et al.*, 2014), which is tailored to some kinds of problems/questions). Any desired information will be located from the list via the same mechanism according to the specified conditions.
- (4) The *Logic Form Converter* (LFC) should be compositional (Moldovan & Rus, 2001) after giving co-reference and solution type², which implies that each sub-tree (or nonterminal node) should be independently transformed regardless of other nodes not under it, and the logic form of a given nonterminal node is formed by concatenating the corresponding logic forms of its associated child-nodes.
- (5) The IE should only deal with domain dependent generic concepts (instead of complicated

² Solution Type specifies the desired mathematic utility/operation that LFC should adopt (see Section 3.3 for details).

problem dependent concepts); otherwise, it would be too tedious. Take the problem “100 顆糖裝成 5 盒糖, 1 盒糖裝幾顆糖? (If 100 candies are packed into 5 boxes, how many candies are there in a box?)” as an example. Instead of using a problem-dependent *First Order Logic* (FOL) predicate like “裝成(100,顆,糖,5,盒,糖)”, the problem-independent FOL functions/predicates like “quan(q1,顆,糖) = 100”, “quan(q2,盒,糖) = 5”, “qmap(m1,q1,q2)”, and “verb(m1,裝成)” are adopted to represent the facts provided by problem description³.

- (6) The LFC should know the global skeleton of the whole given text (which is implicitly implied by the associated semantic segments linked via the given co-reference information) to achieve a reasonable balance between it and the IE.
- (7) The IE should separate the knowledge from the reasoning procedures to ease porting, which denotes that those domain dependent concepts and inference rules should be kept in a declarative form (and could be imported from some separated files); and the inference rules should not be a part of the IE’s source code.

3. System Framework

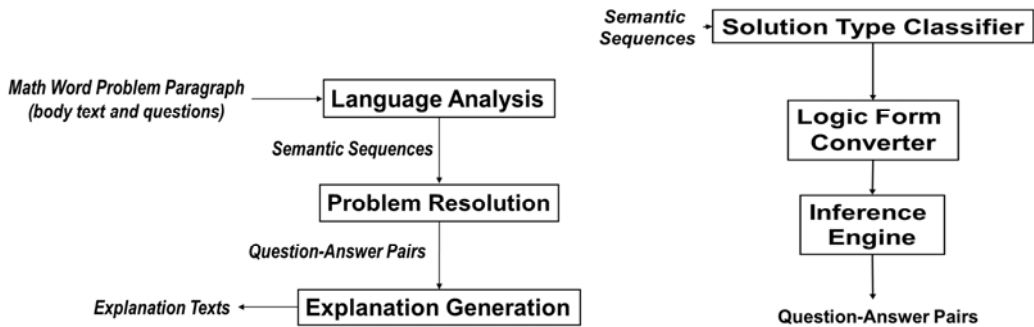


Figure 1. The block diagram of the proposed Math Word Problem Solver.

The block diagram of the proposed MWP solver is shown in Figure 1. First, every sentence in the MWP, including both body text and the question text, is analyzed by the *Language Analysis* module, which transforms each sentence into its corresponding *Semantic Representation (SR) tree*. The sequence of SR trees is then sent to the *Problem Resolution* module, which adopts logic inference approach to obtain the answer for each question. Finally,

³ “quan(…)” is an FOL function to describe quantity facts. “quan(q1,顆,糖)=100” and “quan(q2,盒,糖)=5” describe two quantity facts about “100 顆糖” and “5 盒糖”, respectively. “qmap(m1,q1,q2)” is an FOL predicate to describe that there is a relation (denoted by “m1”) between two quantity facts “q1” and “q2”. “verb(m1,裝成)” indicates that the verb “裝成” is associated with the quantity relation “m1”.

the *Explanation Generation* module will explain how the answer is obtained (in natural language text) according to the given reasoning chain.

As the figure depicted, the Problem Resolution module in our system consists of three components: *Solution Type Classifier* (STC), LFC and IE. The STC suggests a scenario to solve the problem for every question in an MWP. In order to perform logic inference, the LFC first extracts the related facts from the given SR tree and then represents them as FOL predicates/functions (Russell & Norvig, 2009). It also transforms each question into an FOL-like utility function according to the assigned solution type. Finally, according to inference rules, the IE derives new facts from the old ones provided by the LFC. Besides, it is also responsible for providing utilities to perform math operations on related facts.

The entities (like noun phrases) or events (like verb phrases) described in the given sentence may be associated with modifiers, which usually restrict the scope (or specify the property) of the entities/events that they are associated. Since the system does not know which kind of questions will be asked when it reads the body sentences, modifiers should be also included in logic expressions (act as specifiers) and involved in binding. Therefore, the reification technique (Jurafsky & Martin, 2000) is employed to map the nonterminals in the given semantic tree, including verb phrases and noun phrases, into quantified objects which can be related to other objects via specified relations. For example, the logic form of the noun phrase “紅筆(red pens)” would be “color(n1,紅)&head(n1,筆)”, where “n1” is an identified object referring to the noun phrase. Usually, the specifiers in the *Body Logic Form* (BLF) are optional in *Question Logic Form* (QLF), as the body might contain irrelevant text. On the contrary, the specifiers in the QLF are NOT optional (at least in principle) in BLF (i.e., the same (or corresponding) specifier must exist in BLF). This restriction is important as we want to make sure that each argument (which will act as a filtering-condition) in the QLF will be exactly matched to keep irrelevant facts away during the inference procedure.

Take the MWP “文具店進貨 2361 枝紅筆和 1587 枝藍筆(A stationer bought 2361 red pens and 1587 blue pens), 文具店共進貨幾枝筆(How many pens did the stationer buy)?” as an example. The STC will assign the operation type “Sum” to it. The LFC will extract the following facts from the first sentence:

```

quan(q1,枝,n1,p)=2361&verb(q1,進貨)&agent(q1,文具店)&head(n1,p,筆)&color(n1,p,紅)
quan(q2,枝,n2,p)=1587&verb(q2,進貨)&agent(q2,文具店)&head(n2,p,筆)&color(n2,p,藍)

```

The quantity-fact “2361 枝紅筆(2361 red pens)” is represented by “ $\text{quan}(q1, \text{枝}, n1_p)=2361$ ”, where the argument “ $n1_p$ ”⁴ denotes “紅筆(red pens)” due to the facts “ $\text{head}(n1_p, \text{筆})$ ” and “ $\text{color}(n1_p, \text{紅})$ ”. Also, those specifiers “ $\text{verb}(q1, \text{進貨})\&\text{agent}(q1, \text{文具店})\&\text{head}(n1_p, \text{筆})\&\text{color}(n1_p, \text{紅})$ ” are regarded as various tags which will act as different conditions for selecting the appropriate information related to the question specified later. Likewise, the quantity-fact “1587 枝藍筆(1587 blue pens)” is represented by “ $\text{quan}(q2, \text{枝}, n2_p)=1587$ ”. The LFC also issues the utility call “ $\text{ASK Sum}(\text{quan}(?q, \text{枝}, \text{筆}), \text{verb}(?q, \text{進貨})\&\text{agent}(?q, \text{文具店}))$ ” (based on the assigned solution type) for the question. Finally, the IE will select out two quantity-facts “ $\text{quan}(q1, \text{枝}, n1_p)=2361$ ” and “ $\text{quan}(q2, \text{枝}, n2_p)=1587$ ”, and then perform “Sum” operation on them to obtain “3948”.

If the question in the above example is “文具店共進貨幾枝紅筆(How many red pens did the stationer buy)?”, the LFC will generate the following facts and utility call for this new question:

$\text{head}(n3_p, \text{筆})\&\text{color}(n3_p, \text{紅})$
 $\text{ASK Sum}(\text{quan}(?q, \text{枝}, n3_p), \text{verb}(?q, \text{進貨})\&\text{agent}(?q, \text{文具店}))$

As the result, the IE will only select the quantity-fact “ $\text{quan}(q1, \text{枝}, n1_p)=2361$ ”, because the specifier in QLF (i.e., “ $\text{color}(n3_p, \text{紅})$ ”) cannot match the associated specifier “藍(blue)” (i.e., “ $\text{color}(n2_p, \text{藍})$ ”) of “ $\text{quan}(q2, \text{枝}, n2_p)=1587$ ”. After performing “Sum” operation on it, we thus obtain the answer “2361”. Each module will be described in detail as follows (We will skip Explanation Generation due to space limitation. Please refer to (Huang *et al.*, 2015) for the details).

3.1 Language Analysis (Jurafsky & Martin, 2000)

Since the Chinese sentence is a string of characters with no delimiters to mark word boundaries, the first step for analyzing the MWP text is to segment each given sentence string into its corresponding word sequence. Our Chinese word segmentation system (Chen & Ma, 2002; Ma & Chen, 2003) adopts a modularized approach. Independent modules were designed to solve the problems of segmentation ambiguities and identifying unknown words. Segmentation ambiguities are resolved by a hybrid method of using heuristic and statistical rules. Regular-type unknown words are identified by associated regular expressions, and

⁴ The subscript “p” in “ $n1_p$ ” indicates that “ $n1_p$ ” is a pseudo nonterminal derived from the nonterminal “ $n1$ ”, which has four terminals “2361”, “枝”, “紅” and “筆”. More details about pseudo nonterminal will be given at Section 3.3.

irregular types of unknown words are detected first by their occurrence and then extracted by morphological rules with statistical and morphological constraints. Part-of-Speech tagging is also included in the segmentation system for both known and unknown words by using HMM models and morphological rules. Please refer to (Tseng & Chen, 2002; Tsai & Chen, 2004) for the details.

In order to design a high precision and broad coverage Chinese parser, we had constructed a Chinese grammar via generalizing and specializing the grammar extracted from Sinica Treebank (Hsieh *et al.*, 2013; Hsieh *et al.*, 2014) to achieve this goal. The designed F-PCFG (Feature-embedded Probabilistic Context-free Grammar) parser was based on the probabilities of the grammar rules. It evaluates the plausibility of each syntactic structure to resolve parsing ambiguities. We refine the probability estimation of a syntactic tree (for tree-structure disambiguation) by incorporating word-to-word association strengths. The word-to-word association strengths were self-learned from parsing the CKIP corpus (Hsieh *et al.*, 2007). A semantic-role assignment capability is also incorporated into the system.

3.1.1 Semantic Composition

Once the syntactic structure (with semantic roles) for a sentence is obtained, its semantic representation can be further derived through a process of semantic composition (from lexical senses) and achieved near-canonical representations. To represent lexical senses, we had implemented a universal concept-representation mechanism, called E-HowNet (Chen *et al.*, 2005; Huang *et al.*, 2014). It is a frame-based entity-relation model where word senses are expressed by both primitives (or well-defined senses) and their semantic relations. We utilize E-HowNet to disambiguate word senses by referencing its ontology and the related synsets of the target words.

To solve math word problems, it is crucial to know who or what entity is being talked about in the descriptions of problems. This task is called reference resolution, and it can be classified into two types – anaphora resolution and co-reference resolution. Anaphora resolution is the task of finding the antecedent for a single pronoun while co-reference is the task of finding referring expressions (within the problem description) that refer to the same entity. We attack these two types of resolution mainly based on assessing whether a target pronoun/entity coincides its referent candidate in E-HowNet definition. For example, the definition of “她(he)” is “{3rdPerson|他人:gender={female|女}}”. Therefore, it would restrict that the valid referent candidates must be a female human, and result in a much fewer number of candidates for further consideration.

In the following example, the semantic composition, anaphora resolution and co-reference resolution are shown in the table.

小豪有 62 張貼紙，哥哥再給他 56 張，小豪現在共有幾張貼紙？
(Xiaohao had 64 stickers, and his brother gave him 56 more. How many stickers does Xiahao have now?)

小豪有 62 張貼紙， {有(2): theme={ [x1]小豪(1)}, range={貼紙(4): quantifier={ 6 2 張(3)} } } 小豪(1): {human 人:name={"小豪"}} 有(2): {own 有} 6 2 張(3): quantifier={張.null 無 義:quantity={62}} 貼紙(4): {paper 紙張: qualification = {sticky 黏}}	哥哥再給他 56 張， {給(3): agent={哥哥(1)}, time={再(2)}, goal={ [x1]他(4)}, theme={貼紙(5.1): quantifier={ 5 6 張(5)} } } 哥哥(1): {哥哥 ElderBrother} 再(2): frequency={again 再} 給(3): {give 給} 他(4): {3rdPerson 他人} 5 6 張(5): quantifier={張.null 無 義:quantity={56}} 貼紙(5.1): {paper 紙 張:qualification={sticky 黏}}	小豪現在共有幾張貼紙？ {有(4): theme={ [x1]小豪(1)}, time={現在(2)}, quantity={共(3)}, range={貼紙(6): quantifier={幾張(5)} } } 小豪(1): {human 人:name={"小豪 "}} 現在(2): {present 現在} 共(3): {all 全} 有(4): {own 有} 幾張(5): quantifier={張.null 無義: 幾.quantity={Ques 疑問}} 貼紙(6): {paper 紙 張:qualification={sticky 黏}}
--	--	---

We use numbers following words to represent words' positions in a sentence. For instance, “有(2)” is the second word in the first sentence. The semantic representation uses a near-canonical representation form, where semantic role labels, such as “agent”, “theme” and “range”, are marked on each word, and each word is identified with its sense, such as “有(2): {own|有}”.

The co-referents of all sentences in a math problem are marked with the same “x[#]”. For example, we mark the proper noun “小豪(1)” with “[x1]” to co-refer with the pronoun “他(4)” and the second occurrence of the proper noun “小豪(1)”. In the second sentence of the example, the head of the quantifier “5 6 張” is omitted in the text but it is recovered in the semantic representation and annotated with a decimal point in its word position. The missing head is recovered as “貼紙(5.1)”, which is an extra word with its constructed position based on decimal point.

3.2 Solution Type Identification

However, even we know what the given math word problem means, we still might not know how to solve it if we have not been taught for solving the same type of problems in a math class before (i.e., without enough math training/background). Therefore, we need to collect various types of math operations (e.g., addition, subtraction, multiplication, division, sum, etc.), aggregative operations (e.g., Comparison, Set-Operation, etc.) and specific problem types (e.g., Algebra, G.C.D., L.C.M., etc.) that have been taught in the math class. And the LFC needs to know which math operation, aggregative operation or specific problem type should be adopted to solve the given problem. Therefore, we need to map the given semantic representation to a specific problem type. However, this mapping is frequently decided based on the global information across various input sentences (even across body text and question text). Without giving the corresponding mathematic utility/operation, the logic form transformation would be very complicated. A *Solution Type Classifier* (STC) is thus proposed to decide the desired utility/operation that LFC should adopt (i.e., to perform the mapping).

Currently, 16 different solution types are specified (in Table 1; most of them are self-explained with their names) to cover a wide variety of questions found in our elementary math word corpus. They are listed according to their frequencies found in 75 manually labeled questions. The STC is similar to the *Question Type Classifier* commonly adopted at Q&A (Loni, 2011). For mathematic operation type, it will judge which top-level math operation is expected (based on the equation used to get the final answer). For example, if the associated equation is “Answer = $q1 - (q2 \times q3)$ ”, then “Subtraction” will be the assigned math operation type, which matches human reasoning closely.

Table 1. Various solution types for solving elementary school math word problems with frequency in the training set (75 questions in total).

Multiply (24%)	Utility (6%)	Surplus (4%)	L.C.M (2%)
Sum (14%)	Algebra (5%)	Difference (4%)	G.C.D (2%)
Subtraction (12%)	Comparison (5%)	Ceil-Division (3%)	Addition (1%)
Floor-Division (7%)	Ratio (5%)	Common-Division (3%)	Set-Operation (1%)

Take the following math word problem as an example, “一艘輪船 20 分鐘可以行駛 25 公里(A boat sails 25 kilometers in 20 minutes), 2.5 小時可以行駛多少公里(How far can it sail in 2.5 hours)?”. Its associated equation is “Answer = $150 \times (25 \div 20)$ ”. Therefore, the top-level operation is “Multiplication”, and it will be the assigned solution type for this example. However, for the problem “某數乘以 11(Multiply a number with 11), 再除以 4 的答案是 22(then divide it by 4. The answer is 22), 某數是多少(What is the number)?”, its

associated equation is “Answer $\times 11 \div 4 = 22$ ”; since there is no specific natural top-level operation, the “Algebra” solution type will be assigned⁵.

The STC will check the SR trees from both the body and the question to make the decision. Therefore, it provides a kind of global decision, and the LFC will perform logic transformation based on it (i.e., the statistical model of the LFC is formulated to condition on the solution type). Currently, a SVM classifier with linear kernel functions (Chang & Lin, 2011) is used, and it adopted four different kinds of feature-sets: (1) all word unigrams in the text, (2) head word of each nonterminal (inspired by the analogous feature adopted in (Huang *et al.*, 2008) for question classification), (3) E-HowNet semantic features, and (4) pattern-matching indicators (currently, patterns/rules are manually created).

3.3 Logic Form Transformation

A two-stage approach is adopted to transform the SR tree of an input sentence to its corresponding logic forms. In the first stage, the syntactic/semantic relations between the words are deterministically transformed into their domain-independent logic forms. Afterwards, crucial generic math facts and the possible math operations are non-deterministically generated (as domain-dependent logic forms) in the second stage. Basically, logic forms are expressed with the first-order logic (FOL) formalism (Russell & Norvig, 2009)

In the first stage, FOL predicates are generated by traversing the input SR tree which mainly depicts the syntactic/semantic relations between its words (with associated word-senses). For example, the SR tree of the sentence “100 顆糖裝成 5 盒(If 100 candies are packed into 5 boxes)” is shown as follows:

$$\{\text{裝成}(t1); \quad \text{theme}=\{\text{糖}(t2); \text{quantity}=100(t3); \text{unit}=\text{顆}(t4)\}; \\ \text{result}=\{\text{糖}(t5); \text{quantity}=5(t6); \text{unit}=\text{盒}(t7)\} \}$$

Where “theme” and “result” are semantic roles, and information within brace are their associated attributes. Also, the symbols within parentheses are the identities of the terminals in the SR tree. Note that the terminal t5 is created via zero anaphora resolution in the language analysis phase. The above SR tree is transformed into the following FOL predicates separated by the logic AND operator &.

⁵ However, the “Algebra” solution type in this case is useless to LFC because the body text has already mentioned how to solve it, and the LFC actually does not need STC to tell it how to solve the problem.

$\text{verb}(v1,t1)\&\text{theme}(v1,n1)\&\text{result}(v1,n2)\&$
 $\text{head}(n1,t2)\&\text{quantity}(n1,t3)\&\text{unit}(n1,t4)\&$
 $\text{head}(n2,t5)\&\text{quantity}(n2,t6)\&\text{unit}(n2,t7)$

All the first arguments of the above FOL predicates (i.e., $v1$, $n1$ and $n2$) are the identities to the nonterminals in the SR tree. To ease reading, the terminal identities in logic forms are replaced with their corresponding terminal strings in the rest of this paper. After replacement, the above logic forms become more readable as follows:

$\text{verb}(v1,\text{裝成})\&\text{theme}(v1,n1)\&\text{result}(v1,n2)\&\text{head}(n1,\text{糖})\&\text{quantity}(n1,100)\&$
 $\text{unit}(n1,\text{顆})\&\text{head}(n2,\text{糖})\&\text{quantity}(n2,5)\&\text{unit}(n2,\text{盒})$

The above FOL predicates are also called logic-form-1 (LF1) facts. The predicate names of LF1 facts are just the domain-independent syntactic/semantic roles of the constituents in a sub-tree. Therefore, the LF1 facts are also domain-independent.

The domain-dependent logic-form-2 (LF2) facts are generated in the second stage. The LF2 facts are derived from some crucial generic math facts associated with quantities and relations between quantities. The FOL function “ $\text{quan}(\text{quan_id}, \text{unit_id}, \text{object_id}) = \text{number}$ ” is used to describe the facts about quantities. The first argument is a unique identity to represent this quantity-fact. The other arguments and the function value describe the meaning of this fact. For example, “ $\text{quan}(q1, \text{顆}, \text{糖}) = 100$ ” means “100 顆糖(100 candies)” and “ $\text{quan}(q2, \text{盒}, \text{糖}) = 5$ ” means “5 盒糖(five boxes of candies)”. The FOL predicate “ $\text{qmap}(\text{map_id}, \text{quan_id}_1, \text{quan_id}_2)$ ” (denotes the mapping from quan_id_1 to quan_id_2) is used to describe a relation between two quantity-facts, where the first argument is a unique identity to represent this relation. For example, “ $\text{qmap}(m1, q1, q2)$ ” indicates that there is a relation between “100 顆糖” and “5 盒糖”. Now, LF2 facts are transformed by rules with a predefined set of lexico-semantic patterns as conditions. When more cases are exploited, a nondeterministic approach would be required.

In addition to domain-dependent facts like “ $\text{quan}(\dots)$ ” and “ $\text{qmap}(\dots)$ ”, some auxiliary domain-independent facts associated with quan_id and map_id are also created in this stage to help the IE find the solution. The auxiliary facts of the quan_id are created by 4 steps: First, locate the nonterminal (said n_q) which quan_id is coming from. Second, traverse upward from n_q to find the nearest nonterminal (said n_v) which directly connects to a verb terminal. Third, duplicate all LF1 facts whose first arguments are n_v , except the one whose second argument is n_q . Finally, replace the first arguments of the duplicated facts with quan_id . In the above

example, for the quantity-fact q_1 , n_q is n_1 and n_v is v_1 in the first and second steps. “verb(v_1 , 裝成)” and “result(v_1 , n_2)” will be copied at the third step. Finally, “verb(q_1 , 裝成)” and “result(q_1 , n_2)” are created. Likewise, “verb(q_2 , 裝成)” and “theme(q_2 , n_1)” are created for q_2 . The auxiliary facts of “qmap(map_id , $quan_id_1$, $quan_id_2$)” are created by copying all facts of the forms “verb($quan_id_1$, *)” and “verb($quan_id_2$, *)” (where “*” is a wildcard), and then replace all the first arguments of the copied facts with map_id . So, “verb(m_1 , 裝成)” is created for m_1 .

Sometimes, the third argument of a quantity-fact (i.e., *object_id*) is a pseudo nonterminal identity created in the second stage. For example, the LF1 facts of the phrase “2361 枝紅筆 (2361 red pens)” are “quantity(n_1 , 2361)”, “unit(n_1 , 枝)”, “color(n_1 , 紅)” and “head(n_1 , 筆)”, where n_1 is the nonterminal identity of the phrase. A pseudo nonterminal identity, said n_{1p} , is created to carry the terminals “紅(red)” and “筆(pen)” so that the quantity-fact “2361 枝紅筆(2361 red pens)” can be expressed as “quan(q_1 , 枝, n_{1p}) = 2361”. The subscript “p” in n_{1p} indicates that n_{1p} is a pseudo nonterminal derived from the n_1 . To express that fact that n_{1p} carries the terminals “紅(red)” and “筆(pen)”, two auxiliary facts “color(n_{1p} , 紅)” and “head(n_{1p} , 筆)” are also generated.

The questions in an MWP are transformed into FOL-like utility functions provided by the IE. One utility function is issued for each question to find the answer. For example, the question “文具店共進貨幾枝筆(How many pens did the stationer buy)” is converted into “ASK Sum(quant(?q,枝,筆), verb(?q,進貨)&agent(?q,文具店))”. This conversion is completed by two steps. First, select an IE utility (e.g., “Sum(⋯)”) to be called. Since the solution type of the question is “Sum”, the IE utility “Sum(*function*, *condition*) = value” is selected. Second, instantiate the arguments of the selected IE utility. In this case, the first argument *function* is set to “quant(?q, 枝, 筆)” because an unknown quantity fact is detected in the phrase “幾枝筆 (how many pens)”. Let the FOL variable “?q” play the role of *quan_id* in the steps of finding the auxiliary facts. The auxiliary facts “verb(?q, 進貨)” and “agent(?q, 文具店)” are obtained to compose the second argument *condition*.

To sum up, the LFC transforms the semantic representation obtained by language analysis into domain dependent FOL expressions on which inference can be performed. In contrast, most researches of semantic parsing (Jurcicek *et al.*, 2009; Das *et al.*, 2014; Berant *et al.*, 2013; Allen, 2014) seek to directly map the input text into the corresponding logic form. Therefore, across sentences deep analysis of the input text (e.g., anaphora and co-reference resolution) cannot be handled. The proposed two-stage approach (i.e., language analysis and then logic form transformation) thus provides the freedom to enhance the system capability for handling complicated problems which require deep semantic analysis.

3.4 Logic Inference

3.4.1 Basic Operation

In our design, an IE is used to find the solution for an MWP. It is responsible for providing utilities to select desired facts and then obtaining the answer by taking math operations on those selected facts. In addition, it is also responsible for using inference rules to derive new facts from the facts directly provided from the description of the MWP. Facts and inference rules are represented in first-order logic (FOL) (Russell & Norvig, 2009).

In some simple cases, the desired answer can be calculated from the facts directly derived from the MWP. For those cases, the IE only needs to provide a utility function to calculate the answer. In the example of Figure 2, quantities 300, 600, 186 and 234 are mentioned in the MWP. The LFC transforms the question into “ASK Sum(quant(?q, 朵, 百合), verb(?q, 賣出)&agent(?q, 花店)” to ask the IE to find the answer, where “Sum(⋯)” is a utility function provided by the IE. The first argument of “Sum(⋯)” is an FOL function to indicate which facts should be selected. In this case, the unification procedure of the IE will successfully unify the first argument “quant(?q, 朵, 百合)” with three facts “quant(q2, 朵, 百合)”, “quant(q3, 朵, 百合)” and “quant(q4, 朵, 百合)”. When unifying “quant(?q, 朵, 百合)” with “quant(q2, 朵, 百合)”, the FOL variable “?q” will be bound/substituted with q2. The second argument of “Sum(⋯)” (i.e., “verb(?q, 賣出)&agent(?q, 花店)”) is the condition to be satisfied. Since “quant(q2, 朵, 百合)” is rejected by the given condition, “Sum(⋯)” will sum the values of the remaining facts (i.e., q3 and q4) to obtain the desired answer “420”.

<p>花店進貨 300 朵玫瑰和 600 朵百合(A flower store bought 300 roses and 600 lilies), 上午賣出 186 朵百合(It sold 186 lilies in the morning), 下午賣出 234 朵(It sold 234 lilies in the afternoon) , 問花店共賣出幾朵百合(How many lilies did the flower store sell)?</p>
<p>quant(q1, 朵, 玫瑰)=300&verb(q1, 進貨)&agent(q1, 花店)&... quant(q2, 朵, 百合)=600&verb(q2, 進貨)&agent(q2, 花店)&... quant(q3, 朵, 百合)=186&verb(q3, 賣出)&agent(q3, 花店)&... quant(q4, 朵, 百合)=234&verb(q4, 賣出)&agent(q4, 花店)&... ASK Sum(quant(?q, 朵, 百合), verb(?q, 賣出)&agent(?q, 花店))</p>

Figure 2. A simple problem and its essential corresponding logic forms.

Table 2 lists the utilities provided by the IE. The first one, as we have just described, returns the sum of the values of FOL function instances which can be unified with the function argument and satisfy the condition argument. The *Addition* utility simply returns the value of “value₁+value₂”, where value_i is either a constant number, or an FOL function value, or a value returned by a utility. Likewise, *Subtraction* and *Multiplication* utilities return

“value₁-value₂” and “value₁×value₂” respectively. *Difference* returns the absolute value of *Subtraction*. *CommonDiv* returns the value of “value₁÷value₂”. *FloorDiv* returns the largest integer value not greater than “value₁÷value₂” and *CeilDiv* returns the smallest integer value not less than “value₁÷value₂”. *Surplus* returns the remainder after division of value₁ by value₂.

Table 2. The utilities provided by the IE.

Sum(function, condition)=value	CommonDiv(value ₁ , value ₂)=value
Addition(value ₁ , value ₂)=value	FloorDiv(value ₁ , value ₂)=value
Subtraction(value ₁ , value ₂)=value	CeilDiv(value ₁ , value ₂)=value
Difference(value ₁ , value ₂)=value	Surplus(value ₁ , value ₂)=value
Multiplication(value ₁ , value ₂)=value	

Solving MWP's may require deriving new facts according to common sense or domain knowledge. In Figure 3, the MWP provides the facts that “爸爸(Papa)” bought something but it does not provide any facts associated to the money that “爸爸(Papa)” must pay. As a result, we are not able to obtain the answer from the question logic form “Sum(quant(?q,元,#), verb(?q,付)&agent(?q,爸爸))”. However, it is common sense that people must pay some money to buy something. The following inference rule implements this common-sense implication.

$$\begin{aligned} & \text{quant}(\text{?q}, \text{?u}, \text{?o}) \& \text{verb}(\text{?q}, \text{買}) \& \text{agent}(\text{?q}, \text{?a}) \& \text{price}(\text{?o}, \text{?p}) \\ \rightarrow & \text{quant}(\text{\$q}, \text{元}, \text{\#}) = \text{quant}(\text{?q}, \text{?u}, \text{?o}) \times \text{?p} \& \text{verb}(\text{\$q}, \text{付}) \& \text{agent}(\text{\$q}, \text{?a}) \end{aligned}$$

In the above implication inference rule, “quant(?q,?u,?o)&…&price(?o,?p)” is the premise of the rule and “quant(\$q,元,#)=…&agent(\$q,?a)” is the consequence of the rule. Please note that “\$q” indicates a unique ID generated by the IE.

爸爸買了 3 本 329 元的故事書和 2 枝 465 元的鋼筆(Papa bought three \$329 books and two \$465 pens) · 爸爸共要付幾元(How much money did Papa pay)?
$\text{quant}(\text{q1}, \text{本}, \text{n1}_p) = 3 \& \text{verb}(\text{q1}, \text{買}) \& \text{agent}(\text{q1}, \text{爸爸}) \& \text{head}(\text{n1}_p, \text{故事書}) \& \text{price}(\text{n1}_p, 329)$ $\text{quant}(\text{q2}, \text{枝}, \text{n2}_p) = 2 \& \text{verb}(\text{q2}, \text{買}) \& \text{agent}(\text{q2}, \text{爸爸}) \& \text{head}(\text{n2}_p, \text{鋼筆}) \& \text{price}(\text{n2}_p, 465)$ ASK Sum(quant(?q,元,#),verb(?q,付)&agent(?q,爸爸))

Figure 3. An example for deriving new facts.

After unifying this inference rule with the facts in Figure 3, we can get two possible bindings (for $q1$ and $q2$, respectively). The following shows the binding of $q1$.

$$\begin{aligned} & \text{quan}(q1, \text{本}, n1) \& \text{verb}(q1, \text{買}) \& \text{agent}(q1, \text{爸爸}) \& \text{price}(n1, 329) \\ \rightarrow & \text{quan}(q3, \text{元}, \#) = \text{quan}(q1, \text{本}, n1) \times 329 \& \text{verb}(q3, \text{付}) \& \text{agent}(q3, \text{爸爸}) \end{aligned}$$

Since “ $\text{quan}(q1, \text{本}, n1) \times 329 = 3 \times 329 = 987$ ”, the consequence of the above inference will generate three new facts “ $\text{quan}(q3, \text{元}, \#) = 987$ ”, “ $\text{verb}(q3, \text{付})$ ” and “ $\text{agent}(q3, \text{爸爸})$ ”. The semantics of the consequence is “爸爸付 987 元(Papa pays 987 dollars)”. Likewise, the consequence of another binding of this inference rule will also generate three new facts “ $\text{quan}(q4, \text{元}, \#) = 930$ ”, “ $\text{verb}(q4, \text{付})$ ” and “ $\text{agent}(q4, \text{爸爸})$ ”. By taking these new facts into account, the utility call “ $\text{Sum}(\text{quan}(?q, \text{元}, \#), \text{verb}(?q, \text{付}) \& \text{agent}(?q, \text{爸爸}))$ ” can thus return the correct answer “1917”.

Furthermore, the unification process in a conventional IE is based on string-matching. The expression “ $\text{quan}(?q, \text{枝}, \text{筆})$ ” can be unified with a fact “ $\text{quan}(q1, \text{枝}, \text{筆})$ ”. However, it cannot be unified with the fact “ $\text{quan}(q2, \text{朵}, \text{花})$ ”. String-matching guarantees that the IE will not operate on undesired quantities. But, it sometimes prevents the IE from operating on desired quantities. For instance, in Figure 4, two quantity-facts “ $\text{quan}(q1, \text{枝}, n1_p) = 2361$ ” and “ $\text{quan}(q2, \text{枝}, n2_p) = 1587$ ” are converted from “2361 枝紅筆(2361 red pens)” and “1587 枝藍筆(1587 blue pens)”, respectively. The first argument of “ $\text{Sum}(\dots)$ ” is “ $\text{quan}(?q, \text{枝}, \text{筆})$ ” because “幾枝筆(how many pens)” is concerned in the question. The conventional unification is not able to unify “ $\text{quan}(?q, \text{枝}, \text{筆})$ ” to either “ $\text{quan}(q1, \text{枝}, n1_p)$ ” or “ $\text{quan}(q2, \text{枝}, n2_p)$ ” due to different strings of the third arguments. However, from the semantic point of view, “ $\text{quan}(?q, \text{枝}, \text{筆})$ ” should be unified with both “ $\text{quan}(q1, \text{枝}, n1_p)$ ” and “ $\text{quan}(q2, \text{枝}, n2_p)$ ”, because $n1_p$ and $n2_p$ represent “紅筆(red pens)” and “藍筆(blue pens)” respectively (and either one is a kind of “筆(pen)”).

文具店進貨2361 枝紅筆和1587 枝藍筆(A stationer bought 2361 red pens and 1587 blue pens), 文具店共進貨幾枝筆(How many pens did the stationer buy)?

$\text{quan}(q1, \text{枝}, n1_p) = 2361 \& \text{verb}(q1, \text{進貨}) \& \text{agent}(q1, \text{文具店}) \& \text{head}(n1_p, \text{筆}) \& \text{color}(n1_p, \text{紅})$ $\text{quan}(q2, \text{枝}, n2_p) = 1587 \& \text{verb}(q2, \text{進貨}) \& \text{agent}(q2, \text{文具店}) \& \text{head}(n2_p, \text{筆}) \& \text{color}(n2_p, \text{藍})$ ASK Sum($\text{quan}(?q, \text{枝}, \text{筆}), \text{verb}(?q, \text{進貨}) \& \text{agent}(?q, \text{文具店})$)
--

Figure 4. An example for requiring semantic-matching

Therefore, a *semantic matching* method is proposed to be incorporated into the unification procedure. The idea is to match the semantic constituent sets of the two arguments

involved in unification. For example, while matching the third arguments of two functions during unifying the *request*⁶ “quan(?q, 枝, 筆)” with the fact “quan(q1, 枝, n1_p)”, IE will construct and compare two semantic constituent sets, one is for “筆” and the other is for “n1_p”. Let SCS denote “semantic constituent set” and SCS(x) denote the semantic constituent set of x. In our approach, “SCS(筆) = {筆}”⁷ and “SCS(n1_p) = {筆, color(紅)}”⁸. Since “SCS(筆)” is covered by the “SCS(n1_p)”, “quan(?q, 枝, 筆)” can be unified with “quan(q1, 枝, n1_p)”. Likewise, “quan(?q, 枝, 筆)” can be unified with “quan(q2, 枝, n2_p)” because “SCS(n2_p) = {筆, color(藍)}” covers “SCS(筆)”. As the result, the utility call “Sum(quant(?q,枝,筆), verb(?q,進貨)&agent(?q,文具店))” will obtain the correct answer “3948”. On the other hand, if the question is “文具店共進貨幾枝紅筆(How many red pens did the stationer buy)?”, the request will become “quan(?q, 枝, n3_p)”, where n3_p is a pseudo nonterminal consisting of the terminals “紅(red)” and “筆(pen)” under the noun phrase “幾枝紅筆(how many red pens)”. Since “SCS(n3_p) = {筆, color(紅)}”, “quan(?q, 枝, n3_p)” can be unified only with “quan(q1, 枝, n1_p)”. It cannot be unified with “quan(q2, 枝, n2_p)” because SCS(n3_p) cannot be covered by SCS(n2_p). Therefore, the quantity of “藍筆(blue pens)” will not be taken into account for the question “文具店共進貨幾枝紅筆(How many red pens did the stationer buy)?”.

3.4.2 Verb Entailment (Jurafsky & Martin, 2000)

Since we might adopt the verb “買(buy)” in the body text “爸爸買了 3 本 329 元的故事書 (Papa bought three \$329 books)”, but adopt the verb “付(pay)” in the question text “爸爸共要付幾元(How much money did Papa pay) ? ” (as illustrated in the previous section), we need the knowledge that “buy” implies “pay” to perform logic binding (Moldovan & Rus, 2001). Verb entailment is thus required to identify whether there is an entailment relation between these two verbs (Hashimoto *et al.*, 2009). Verb entailment detection is an important function for the IE (de Salvo Braz *et al.*, 2006), as it can indicate the event progress and the status changing. In the math problem “Bill had no money. Mom gave Bill two dollars, and Dad gave Bill three dollars. How much money Bill had then?”, the entailment between “give (給)” and “have (有)” can update the status of Bill from “no money”, then “two dollars”, and to the final

⁶ An FOL predicate/function in an IE utility or in the premise of an inference rule is called a *request*. A request usually consists of FOL variables.

⁷ The SCS of a terminal consists of the terminal string only (e.g., “SCS(筆) = {筆}”).

⁸ SCS(n1_p) is constructed by two steps. First, enumerate all facts whose first arguments are n1_p. Second, for each enumerated fact, denote the predicate name as Child-Role and the SCS of the second argument as Child-SCS. If Child-Role is “head”, put the elements of Child-SCS into SCS(n1_p). Otherwise, for each string s in Child-SCS, put the string “Child-Role(s)” into SCS(n1_p). In the first step, the facts “head(n1_p, 筆)” and “color(n1_p, 紅)” are picked out. In the second step, the strings “筆” and “color(紅)” are put into SCS(n1_p).

answer “five dollars”.

We define the verb entailment problem as follows: given an ordered verb pair “(v1, v2)” as input, we want to detect whether the entailment relation ‘v1 \rightarrow v2’ holds for this pair. E-HowNet (Chen *et al.*, 2009; Huang *et al.*, 2014) is adopted as the knowledge base for solving this problem. For the previous example verb “give (給)”, we can find its conflation of events, which has been described as the phenomenon involved in predicates where the verb expresses a co-event or accompanying event, rather than the main event (Talmy, 1972; Haugen, 2009; Mateu, 2012), from E-HowNet as shown in Figure 5. The conflations of events are defined by predicates and their arguments (Huang *et al.*, 2015), as shown in Figure 5.

Conflation of events:	lose \rightarrow agent({give 給})=theme({lose 失去}); lose \rightarrow theme({give 給})=possession({lose 失去}); obtain \rightarrow theme({give 給})=possession({obtain 得到}); obtain \rightarrow target({give 給})=theme({obtain 得到}); receive \rightarrow target({give 給})=agent({receive 收受}); receive \rightarrow theme({give 給})=possession({receive 收受})
-----------------------	---

Figure 5. The conflation events of the verb “give (給)”.

Verb entailment is vital for solving the elementary school math problem. Consider the following math problem as a simple example:

老師原有 9 枝鉛筆, 送給小朋友 5 枝後, 老師還有幾枝筆? (The teacher has 9 pencils. After giving his students 5 pencils, how many pencils he has?)

The verbs are “有(have)” and “送給(give as a gift)” in this problem. If we want to derive the concept of “有(have)” from “送給(give as a gift)”, we can follow the direction of their definitions in E-HowNet: “送給(give as a gift)” is a hyponym of “給(give)”, and one of its implication from the conflation of events is “得到(obtain)”, which is a hyponym of “有(have)”.

However, for the four verbs in this derivation, implications are defined only in the verb “給(give)”. As we can see, given all those definitions of words in E-HowNet, we need to find a valid path (which may involve word sense disambiguation) to determine whether there is an entailment between two verbs. Therefore, we need a model to automatically build the relations of these verbs by finding paths from E-HowNet or other resources, and then rank or validate these paths to find the verb entailment. The conflation of events also indicates that when the entailed verb pair is detected, we may further map semantic roles of these two verbs to

proceed the inference and find the solution (Wang & Zhang, 2009).

4. Proposed Statistical Framework

Since the accuracy rate of the Top-1 SR tree cannot be 100%, and the decisions made in the following phases (i.e., STC, LFC and IE) are also uncertain, we need a statistical framework to handle those non-deterministic phenomena. Under this framework, the problem of getting the desired answer for a given WMP can be formulated as follows:

$$\widehat{Ans} = \underset{Ans}{\operatorname{arg\,max}} P(Ans | Body, Qus) \quad (1)$$

Where \widehat{Ans} is the obtained answer, Ans denotes a specific possible answer, $Body$ denotes the given body text of the problem, and Qus denotes the question text of the problem.

The probability factor in the above equation can be further derived as follows via introducing some related intermediate/latent random variables:

$$\begin{aligned} & P(Ans | Body, Qus) \\ &= \sum P(Ans, IR, LF_B, LF_Q, SM_B, SM_Q, ST | Body, Qus) \\ &\approx \max P(Ans, IR, LF_B, LF_Q, SM_B, SM_Q, ST | Body, Qus) \quad (2) \\ &\approx \max P(Ans | IR, LF_B, LF_Q) \times P(IR | LF_B, LF_Q, ST) \times P(LF_B | SM_B, ST) \\ &\quad \times P(LF_Q | SM_Q, ST) \times P(ST | SM_B, SM_Q) \times P(SM_B | Body) \times P(SM_Q | Qus) \end{aligned}$$

IR : Inference Rules Applied.

LF_B : Logic Form of Body text.

LF_Q : Logic Form of Question text.

SM_B : Semantic Representation of Body text.

SM_Q : Semantic Representation of Question text.

ST : Solution Type.

In the above equation, we will further assume that $P(Ans | IR, LF_B, LF_Q) \approx P(Rm)$, where Rm is the remaining logic factors in LF_Q after the IE has bound it with LF_B (with referring to the knowledge-base adopted). Last, *Viterbi* decoding (Seshadri & Sundberg, 1994) could be used to search the most likely answer with the above statistical model.

To obtain the associated parameters of the model, we will first get the initial parameter-set from a small seed corpus annotated with various intermediate/latent variables involved in the model. Afterwards, we perform *weakly supervised learning* (Artzi & Zettlemoyer, 2013) on a partially annotated training-set (in which only the answer is annotated with each question). That is, we iteratively conduct beam-search (with the parameter-set obtained from the last iteration) on this partially annotated training-set starting from the given

body text (and question text) to the final obtained answer. If the annotated answer match some of the obtained answers (within the search-beam), simply pick up the matched path with the maximal likelihood value. We then re-estimate the parameter-set (of the current iteration) from those picked up paths. If the annotated answer cannot match any of the obtained answers (within the search-beam), we simply drop that case, and then repeat the above re-estimation procedure.

5. Current Status and Future Work

Currently, we have completed all the associated modules (including Word Segmenter, Syntactic Parser, Semantic Composer, STC, LFC, IE, and Explanation Generation), and have manually annotated 75 samples (from our elementary school math corpus) as the seed corpus (with syntactic tree, semantic tree, logic form, and reasoning chain annotated). Besides, we have cleaned the original elementary school math corpus and encoded it into the appropriate XML format. There are total 23,493 problems from six different grades; and the average number of words of the body text is 18.2 per problem. Table 3 shows the statistics of the converted corpus.

Table 3. MWP corpus statistics and Average length per problem.

<table border="1"> <thead> <tr> <th>Corpus</th> <th>Num. of problems</th> </tr> </thead> <tbody> <tr> <td>Training Set</td> <td>20,093</td> </tr> <tr> <td>Develop Set</td> <td>1,700</td> </tr> <tr> <td>Test Set</td> <td>1,700</td> </tr> <tr> <td>Total</td> <td>23,493</td> </tr> </tbody> </table>		Corpus	Num. of problems	Training Set	20,093	Develop Set	1,700	Test Set	1,700	Total	23,493	<table border="1"> <thead> <tr> <th>Corpus</th> <th>Avg. Chinese Chars.</th> <th>Avg. Chinese Words</th> </tr> </thead> <tbody> <tr> <td>Body</td> <td>27</td> <td>18.2</td> </tr> <tr> <td>Question</td> <td>9.4</td> <td>6.8</td> </tr> </tbody> </table>		Corpus	Avg. Chinese Chars.	Avg. Chinese Words	Body	27	18.2	Question	9.4	6.8
Corpus	Num. of problems																					
Training Set	20,093																					
Develop Set	1,700																					
Test Set	1,700																					
Total	23,493																					
Corpus	Avg. Chinese Chars.	Avg. Chinese Words																				
Body	27	18.2																				
Question	9.4	6.8																				
MWP corpus statistics		Average length per problem																				

We have completed a prototype system which is able to solve 11 different solution types (including *Multiplication*, *Summation*, *Subtraction*, *Floor-Division*, *Algebra*, *Comparison*, *Surplus*, *Difference*, *Ceil-Division*, *Common-Division* and *Addition*), and have tested it on the seed corpus. The success of our pilot run has demonstrated the feasibility of the proposed approach. We plan to use the next few months to perform weakly supervised learning, as mentioned above, and fine tune the system.

6. Related Work

To the best of our knowledge, all those MWP solvers proposed before year 2014 adopted the rule-based approach (Mukherjee & Garain, 2008). For example, Bobrow’s STUDENT

(Bobrow, 1964; Slagle, 1965) used format matching to map the input English sentence into the corresponding logic statement (all start with predicate “EQUAL”). Another system, WORDPRO, was developed by Fletcher (1985) to understand and solve simple one-step addition and subtraction arithmetic word problems designed for third-grade children. It did not accept the surface representation of text as input. Instead it begins with a set of propositions (manually created) that represent the text's meaning. Afterwards, the problem was solved with a set of rules (also called schemas), which matched the given proposition and then took the corresponding actions. Besides, it adopted key word match to obtain the answer.

Solving the problem with schemata was then adopted in almost every later system (Mukherjee & Garain, 2008). In 1986, ARITHPRO was designed with an inheritance network in which word classes inherit attributes from those classes above them on a verb hierarchy (Dellarosa, 1986). The late development of ROBUST (Bakman, 2007) demonstrated how it could solve free format word problems with multi-step arithmetic through splitting one single sentence into two formula propositions. In this way, transpositions of problem sentences or additional irrelevant data to the problem text do not affect the problem solution. However, it only handles state change scenario. In 2010, Ma *et al.* (Ma *et al.*, 2010) proposed a MSWPAS system to simulate people's arithmetic multi-step addition and subtraction word problems behavior. It uses frame-based calculus and means-end analysis (AI planning) to solve the problem with pre-specified rules. In 2012, Liguda and Pfeiffer (Liguda & Pfeiffer, 2012) proposed a model based on augmented semantic networks to represent the mathematical structure behind word problems. It read and solved mathematical text problems from German primary school books. With more attributes associated with the semantic network, it claimed that the system was able to solve multi-step word problems and complex equation systems and was more robust to irrelevant information. Also, it was declared that it was able to solve all classes of problems that could be solved by the schema-based systems, and could solve around 20 other classes of word problems from a school book which were in most cases not solvable by other systems.

Recently, Hosseini *et al.* (2014) proposed a Container-Entity based approach, which solved the math word problem with a state transition sequence. Each state consists of a set of containers, and each container specifies a set of entities identified by a few heuristic rules. How the quantity of each entity type changes depends on the associated verb category. Each time a verb is encountered, it will be classified (via a SVM, which is the only statistical module adopted) into one of the seven categories which pre-specify how to change the states of associated entities. Therefore, logic inference is not adopted. Furthermore, the anaphora and co-reference are left un-resolved, and it only handles addition and subtraction.

Kushman *et al.* (2014) proposed the first statistical approach, which used a few heuristic rules to extract the algebra equation templates (consists of variable slots and number slots)

from a set of problems annotated with equations. For a given problem, all possible variable/number slots are identified first. Afterwards, they are aligned with those templates. The best combination of the template and alignment (scored with a statistical model) is then picked up. Finally, the answer is obtained from those equations instantiated from the selected template. However, without really understanding the problem (i.e., no semantic analysis is performed), the performance that this approach can reach is limited; also, it is sensitive to those irrelevant statements (Hosseini *et al.*, 2014). Furthermore, it can only solve algebra related problems. Last, it cannot explain how the answer is obtained.

The most recent statistical approach was proposed by Roy *et al.* (2015), which used 4 cascade statistical classifiers to solve the elementary school math word problems: *quantity identifier* (used to find out the related quantities), *quantity pair classifier* (used to find out the operands), *operation classifier* (used to pick an arithmetic operation), and *order classifier* (used to order operands for subtraction and division cases). It not only shares all the drawbacks associated with Kushman *et al.* (2014), but also limits itself for allowing only one basic arithmetic operation (i.e., among addition, subtraction, multiplication, division) with merely 2 or 3 operand candidates.

Our proposed approach differs from those previous approaches by combining the statistical framework with logic inference. Besides, the tag-based approach adopted for selecting the appropriate information also distinguishes our approach from that of others.

7. Conclusion

A tag-based statistical framework is proposed in this paper to perform understanding and reasoning for solving MWP. It first analyzes the body and question texts into their corresponding semantic trees (with anaphora/ellipse resolved and semantic role labeled), and then converted them into their associated tag-based logic forms. Afterwards, the inference (based on the question logic form) is performed on the logic facts derived from the body text. The combination of the statistical frame and logic inference distinguishes the proposed approach from other approaches. Comparing to those rule-based approaches, the proposed statistical approach alleviates the ambiguity resolution problem; also, our tag-based approach provides the flexibility of handling various kinds of related questions with the same body logic form. On the other hand, comparing to those purely statistical approaches, the proposed approach is more robust to the irrelevant information and could more accurately provide the answer.

The contributions of our work mainly lie in: (1) proposing a tag-based logic representation which makes the system less sensitive to the irrelevant information and could provide answer more precisely; (2) proposing a statistical framework for performing reasoning from the given text.

Acknowledgment

We would like to thank Prof. Wen-Lian Hsu for suggesting this research topic and making the original elementary school math corpus available to us, and Prof. Keh-Jiann Chen for providing the resources and supporting this project. Besides, our thanks should be extended to Dr. Yu-Ming Hsieh and Dr. Ming-Hong Bai for implementing the syntactic parser and the semantic composer, respectively. Also, we would like to thank Prof. Chin-Hui Lee for suggesting the solution type. Last, our thanks should also go to Ms. Su-Chu Lin for manually annotating the corpus.

References

- Allen, J. F. (2014). Learning a Lexicon for Broad-Coverage Semantic Parsing. In the *Proceedings of the ACL 2014 Workshop on Semantic Parsing*, 1-6.
- Artzi, Y., & Zettlemoyer, L. (2013). Weakly supervised learning of semantic parsers for mapping instructions to actions. *Transactions of the Association for Computational Linguistics*, 1(2013), 49-62.
- Bakman, Y. (2007). *Robust Understanding of Word Problems With Extraneous Information*. Retrieved from arXiv:math/0701393.
- Ballard, B. & Biermann, A. (1979). PROGRAMMING IN NATURAL LANGUAGE : "NLC" AS A PROTOTYPE. *ACM-Webinar*, 1979, DOI: 10.1145/800177.810072.
- Berant, J., Chou, A., Frostig, R., & Liang, P. (2013). Semantic Parsing on Freebase from Question-Answer Pairs. *Conference on Empirical Methods in Natural Language Processing (EMNLP)2013*, 1533-1544.
- Biermann, A. W., & Ballard, B. W. (1980). Toward Natural Language Computation. *American Journal of Computational Linguistic*, 6(2), 71-86.
- Biermann, A., Rodman, R., Ballard, B., Betancourt, T., Bilbro, G., Deas, H., Fineman, L., Fink, P., Gilbert, K., Gregory, D., & Heidlage, F. (1982). INTERACTIVE NATURAL LANGUAGE PROBLEM SOLVING:A PRAGMATIC APPROACH. In *Proc. of the first conference on applied natural language processing*, 180-191.
- Bobrow, D. G. (1964). *Natural language input for a computer problem solving system*. Ph.D. Dissertation, Massachusetts Institute of Technology.
- Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3). Doi:10.1145/1961189.1961199.
- Charniak, E. (1968). *CARPS, a program which solves calculus word problems*. Report MAC-TR-51, Project MAC, MIT.
- Charniak, E. (1969). Computer solution of calculus word problems. In *IJCAI'69 Proc. of International Joint Conference on Artificial Intelligence*, 303-316.

- Chen, K.-J., Huang, S.-L., Shih, Y.-Y., & Chen, Y.-J. (2005). Extended-HowNet- A Representational Framework for Concepts. *OntoLex 2005 - Ontologies and Lexical Resources IJCNLP-05 Workshop*, Jeju Island, South Korea.
- Chen, K.J., & Ma, W.Y. (2002). Unknown Word Extraction for Chinese Documents. In *Proceedings of Coling 2002*, 169-175.
- Das, D., Chen, D., Martins, A. F. T., Schneider, N., & Smith, N. A. (2014). Frame-Semantic Parsing. *Computational Linguistics*, 40(1), 9-56.
- Dellarosa, D. (1986). A computer simulation of children's arithmetic word-problem solving. *Behavior Research Methods, Instruments, & Computers*, 18(2), 147-154.
- Fletcher, C. R. (1985). COMPUTER SIMULATION -- Understanding and solving arithmetic word problems: A computer simulation. *Behavior Research Methods, Instruments, & Computers*, 17(5,) 565-571.
- Gelb, J. P. (1971). Experiments with a natural language problem solving system. In *Proc. of IJCAI-71*, 455-462.
- Hashimoto, C., Torisawa, K., Kuroda, K., De Saeger, S., Murata, M., & Kazama, J. J. (2009). Large-scale verb entailment acquisition from the web. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 3, 1172-1181.
- Haugen, J. D. (2009). Hyponymous objects and Late Insertion. *Lingua*, 119, 242-262.
- Hosseini, M. J., Hajishirzi, H., Etzioni, O., & Kushman, N. (2014). Learning to Solve Arithmetic Word Problems with Verb Categorization. *EMNLP(2014)*, 523-533.
- Hsieh, Y.-M., Chang, J. S., & Chen, K.-J. (2014). Ambiguity Resolution for Vt-N Structures in Chinese. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, 928-937.
- Hsieh, Y.-M., Lin, S.-C., Chang, J. S., & Chen, K.-J. (2013). Improving Chinese Parsing with Special-Case Probability Re-estimation. In *Proceedings of 2013 International Conference on Asian Language Processing (IALP)*, 177-180.
- Hsieh, Y.-M., Yang, D.-C., & Chen, K.-J. (2007). Improve Parsing Performance by Self-Learning. *International Journal of Computational Linguistics and Chinese Language Processing*, 12(2), 195-216.
- Huang, S.-L., Hsieh, Y.-M., Lin, S.-C., & Chen, K.-J. (2014). Resolving the Representational Problems of Polarity and Interaction between Process and State Verbs. *International Journal of Computational Linguistics and Chinese Language Processing (IJCLCLP)*, 19(2), 33-52.
- Huang, S.-L., Lin, S.-C., Ma, W.-Y., & Chen, K.-J. (2015). *Semantic Roles and Semantic Role Labeling*. (CKIP technical report no. 2015-01). Institute of Information Science, Academia Sinica.
- Huang, C. T., Lin, Y. C., & Su, K. Y. (2015). Explanation Generation for a Math Word Problem Solver. *International Journal of Computational Linguistics and Chinese Language Processing (IJCLCLP)*, 20(2), 27-44.

- Huang, Z., Thint, M., & Qin, Z.(2008). Question classification using head words and their hypernyms. In *Proceeding of EMNLP '08 Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 927-936.
- Jurafsky, D., & Martin, J. H. (2000). *Speech and Language Processing*. New Jersey: Prentice Hall.
- Jurcicek, F., Mairesse, F., Gašić, M., Keizer, S., Thomson, B., Yu, K., Young, S., & Gasic, M. (2009). Transformation-based Learning for Semantic parsing. In *Proceedings of INTERSPEECH 2009*, 2719-2722.
- Kushman, N., Artzi, Y., Zettlemoyer, L., & Barzilay, R. (2014). Learning to Automatically Solve Algebra Word Problems. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 271-281.
- Liguda, C., & Pfeiffer, T. (2012). Modeling math word problems with augmented semantic networks. *NLDB 2012*, 247-252.
- Loni, B. (2011). A survey of State-of-the-Art Methods on Question Classification. *Literature Survey*, Published on TU Delft Repository, 2011 Jun.
- Ma, W.-Y., & Chen, K.-J. (2003). Introduction to CKIP Chinese Word Segmentation System for the First International Chinese Word Segmentation Bakeoff. In *Proceedings of ACL, Second SIGHAN Workshop on Chinese Language Processing*, 168-171.
- Ma, Y. H., Zhou, Y., Cui, G. Z., Ren, Y., & Huang, R. H. (2010). Frame-Based Calculus of solving Arithmetic MultiStep Addition and Subtraction word problems. In *2010 Second International Workshop on Education Technology and Computer Science*, 476-479.
- Mateu, J. (2012). *Conflation and incorporation processes in resultative constructions*. In Violeta Demonte & Louise McNally (eds.), *Telicity, Change, and State: A Cross-Categorial View of Event Structure*, Oxford: Oxford University Press, 252-278.
- Moldovan, D., & Rus, V. (2001). Logic Form Transformation of WordNet and Its Applicability to Question Answering. In *ACL '01 Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, 402-409.
- Mukherjee, A., & Garain, U. (2008). A review of methods for automatic understanding of natural language mathematical problems. *Artif Intell Rev*, 29(2), 93-122.
- Roy, S. I., Vieira, T. J. H., & Roth, D. I.(2015). Reasoning about Quantities in Natural Language. *TACL*, 3, 1-13.
- Russell, S. J. & Norvig, P. (2009). *Artificial Intelligence : A Modern Approach*(3rd Edition), Prentice Hall.
- de Salvo Braz, R., Girju, R., Punyakanok, V., Roth, D., & Sammons, M. (2006). An inference model for semantic entailment in natural language. In *Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Textual Entailment.*, Springer Berlin Heidelberg, 2006, 261-286.
- Seshadri, N., Sundberg, C.-E.W. (1994). List Viterbi Decoding Algorithms with Applications. *IEEE Transactions on Communications*, 42(234), 313-323.

- Slagle, J. R. (1965). Experiments with a deductive question-answering program. *J-CACM*, 8(12), 792-798.
- Strassel, S., Adams, D., Goldberg, H., Herr, J., Keesing, R., Oblinger, D., Simpson, H., Schrag, R., & Wright, J. (2010). The DARPA Machine Reading Program - Encouraging Linguistic and Reasoning Research with a Series of Reading Tasks. *LREC 2010*.
- Talmy, L. (1972). *Semantic Structures in English and Atsugewi*. PhD thesis, Berkeley: University of California at Berkeley.
- Tsai, Y.-F., & Chen, K.-J. (2004). Reliable and Cost-Effective Pos-Tagging. *International Journal of Computational Linguistics & Chinese Language Processing*, 9(1), 83-96.
- Tseng, H. H., & Chen, K.-J. (2002). Design of Chinese Morphological Analyzer. In *Proceedings of SIGHAN 2002*, 49-55.
- Wang, R., & Zhang, Y. (2009). Recognizing textual relatedness with predicate-argument structures. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 2, 784-792.

Explanation Generation for a Math Word Problem Solver

Chien-Tsung Huang*, Yi-Chung Lin* and Keh-Yih Su*

Abstract

This paper proposes a math operation (e.g., *Summation*, *Addition*, *Subtraction*, *Multiplication*, *Division*, etc.) oriented approach to explain how the answers are obtained for *math word problems*. Based on the reasoning chain given by the inference engine, we search each math operator involved. For each math operator, we generate one sentence. Since explaining math operation does not require complicated syntax, we adopt a specific template to generate the text for each kind of math operator. To the best of our knowledge, this is the first explanation generation that is specifically tailored to solving the math word problem.

Keywords: Explanation Generation, Math Word Problem Explanation, Machine Reading

1. Introduction

Since *Big Data* mainly aims to explore the correlation between surface features but not their underlying causality relationship (Mayer-Schönberger & Cukier, 2013), the “*Big Mechanism*” program¹ has been proposed by DARPA to find out “why” behind the big data. However, the pre-requisite for it is that the machine can read each document and learn its associated knowledge, which is the task of *Machine Reading* (MR) (Strassel *et al.*, 2010). Therefore, the Natural Language and Knowledge Processing Group (under the Institute of Information Science) of Academia Sinica formally launched a 3-year MR project (from January 2015) to attack this problem.

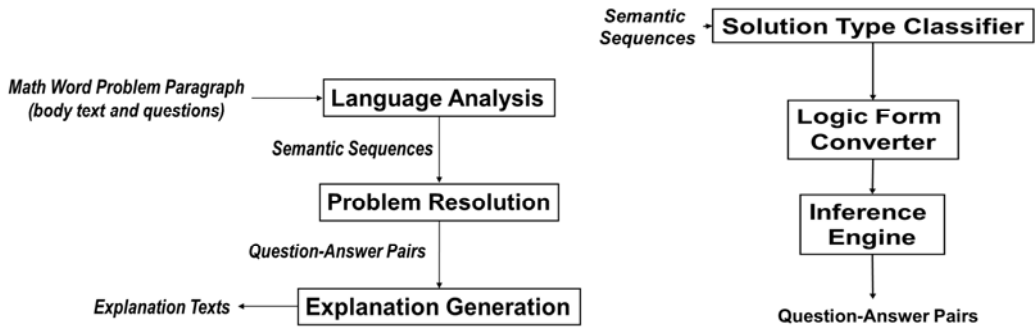
Since a domain-independent MR system is difficult to build, the *Math Word Problem* (MWP) (Mukherjee & Garain, 2008) is chosen as our first test case to study MR. The main reason for that is that it not only adopts less complicated syntax but also requires less amount of domain knowledge; therefore, the researcher can focus more on text understanding and

* Institute of Information Science, Academia Sinica
128 Academia Road, Section 2, Nankang, Taipei 11529, Taiwan
E-mail: { joeth; lyc; kysu }@iis.sinica.edu.tw

¹ http://www.darpa.mil/Our_Work/I2O/Programs/Big_Mechanism.aspx

reasoning (instead of looking for a wide coverage parser and acquiring considerable amount of domain knowledge). We thus also choose it as the goal of the first year for studying the MR problem, and propose a tag-based statistical approach (Lin *et al.*, 2015) to find out the answer.

The architecture of this proposed approach is shown in Figure 1. First, every sentence in the MWP, including both body text and the question text, is analyzed by the *Language Analysis* module, which transforms each sentence into its corresponding semantic representation tree. The sequence of semantic representation trees is then sent to the *Problem Resolution* module, which adopts logic inference approach, to obtain the answer of each question in the MWP. Finally, the *Explanation Generation* module will explain how the answer is found (in natural language text) according to the given *reasoning chain* (Russell & Norvig, 2009) (which includes all related logic statements and inference steps to reach the answer).



(a) Math Word Problem Solver Diagram

(b) Problem Resolution Diagram

Figure 1. The block diagram of the proposed Math Word Problem Solver.

As depicted in Figure 1(b), the *Problem Resolution* module in the proposed system consists of three components: *Solution Type Classifier* (TC), *Logic Form Converter* (LFC) and *Inference Engine* (IE). The TC is responsible to assign a math operation type for every question of the MWP. In order to perform logic inference, the LFC first extracts the related facts from the given semantic representation tree and then represents them in *First Order Logic* (FOL) *predicates/functions* form (Russell & Norvig, 2009). In addition, it is also responsible for transforming every question into an FOL-like utility function according to the assigned solution type. Finally, according to inference rules, the IE derives new facts from the old ones provided by the LFC. Besides, it is also responsible for providing utilities to perform math operations on related facts.

In addition to understanding the given text and then performing inference on it, a very desirable characteristic of an MWP solver (also an MR system) is being able to explain how the answer is obtained in a human comprehensible way. This task is done by the *Explanation*

Generator (EG) module, which is responsible to explaining the associated reasoning steps in fluent natural language from the given *reasoning chain* (Russell & Norvig, 2009). In other words, explanation generation is the process of constructing natural language outputs from a non-linguistic input, and is a task of *Natural Language Generation* (NLG).

Various applications of NLG (such as weather report) have been proposed before (Halliday, 1985; Goldberg *et al.*, 1994; Paris & Vander Linden, 1996; Milosavljevic, 1997; Paris *et al.*, 1998; Coch, 1998; Reiter *et al.*, 1999). However, to the best of our knowledge, none of them discusses how to generate the explanation for WMP, which possesses some special characteristics (e.g., math operation oriented description) that are not shared with other tasks.

A typical architecture for NLG is shown at Figure 2, which is re-drawn from Jurafsky and Martin (Jurafsky & Martin, 2000). Under this architecture, *Communicative Goal*, which specifies the purpose for communication, and *Knowledge Base*, which specifies the content to be generated, are fed as the inputs to *Discourse Planner*. The Discourse Planner will then output a hierarchy form to the *Surface Realizer*, which further solves the issues of selecting lexicons, functional words, lexicon order in the sentence, syntactic form, subject-verb agreement (mainly required for English), tense (mainly required for English), and so on for the texts to be generated.

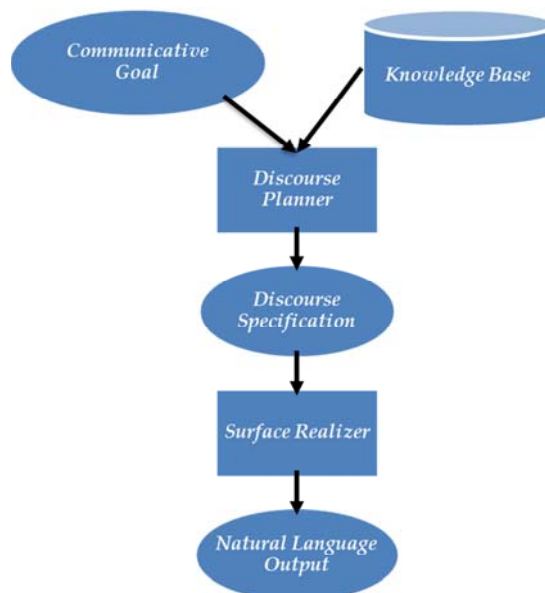


Figure 2. A typical architecture for NLG systems (Jurafsky & Martin, 2000)

To implement the Discourse Planner, D. Jurafsky (Jurafsky & Martin, 2000) proposed to adopt text schemata and rhetorical structure planning to implement the Discourse Planner. On

the other hand, Kay proposed to implement the Surface Realizer with both Systemic Grammar, which is a part of Systemic Functional Linguistic proposed by Halliday (Halliday, 1985), and Functional Unification Grammar (Kay, 1979).

Since the description for math operation centering on an operator is in a relatively fixed textual format, which is disparate from other kinds of NLG tasks, those approaches mentioned above might be over-killed for the task of MWP explanation generation (and thus introduce unnecessary complexity). Therefore, we propose an operator oriented approach to search each math operator involved in the reasoning chain. For each math operator, we generate one sentence. Since explaining math operation does not require complicated syntax, a specific template is adopted to generate the text for each kind of math operator. To the best of our knowledge, this is the first approach that is specifically tailored to the MWP task.

Our main contributions are listed as following,

1. We proposed a *math operation oriented Explanation Tree* for facilitating the discourse work on MWP.
2. We propose an *operator oriented algorithm* to segment the Explanation Tree into various sentences, which makes our Discourse Planner universal for MWP and independent to the language adopted.
3. We propose using *operator-based templates* to generate the natural language text for explaining the associated math operation.

The remainder of this paper is organized as follows: Section 2 introduces the framework of our Explanation Generator. Afterwards, various templates of more operators (other than SUM used in Section 2) are introduced in Section 3. Section 4 discusses the future work of our explanation system. Section 5 then reviews the related works. Finally, the conclusions are drawn in Section 6.

2. Proposed Framework for MWP Explanation Generator (EG)

Figure 3 shows the block diagram of our proposed EG. First, the Inference Engine generates the answer and its associated reasoning chain for the given MWP. First, to ease the operation of the EG, we convert the given reasoning chain into its corresponding *Explanation Tree* (shown at Figure 5) to center on each operator appearing in the reasoning chain (such that it is convenient to perform sentence segmentation later). Next, the Explanation Tree will be fed as input to the *Discourse Planner*, which divides the given Explanation Tree into various subtrees such that each subtree will generate one explanation sentence later. Finally, the *Function Word Insertion & Ordering Module* will insert the necessary functional words and order them with those extracted content words (from the segmented Explanation Subtree) to generate the *Explanation Texts*.

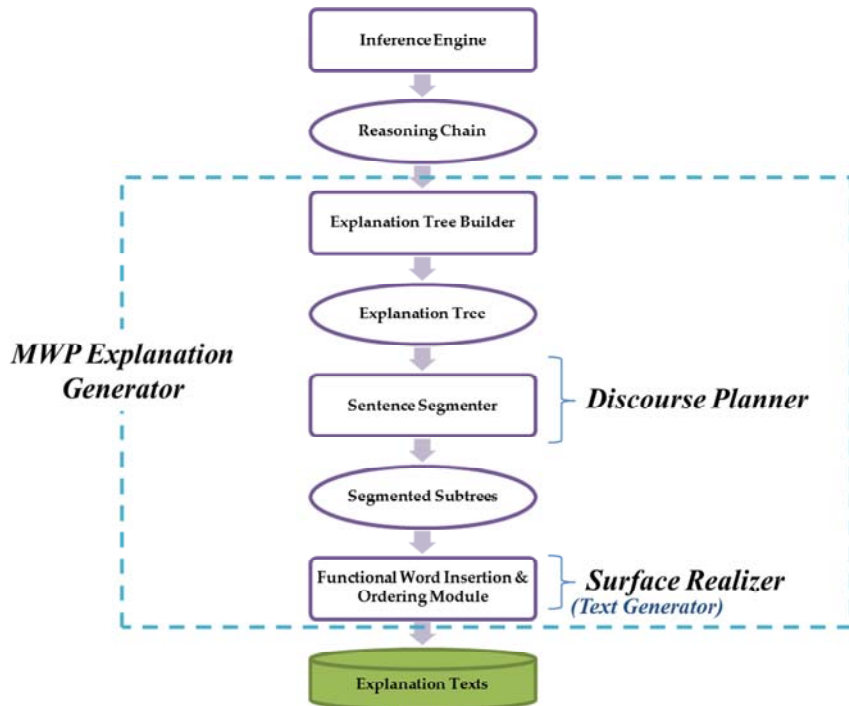


Figure 3. Block Diagram of the proposed MWP Explanation Generator

Following example demonstrates how the framework works. And Figure 4 (a) reveals more details for each part illustrated in Figure 3.

[Sample-1] 阿志買一臺冰箱和一臺電視機，付 2 疊一萬元鈔票、6 張千元鈔票和 13 張百元鈔票，阿志共付了幾元？

(A-Zhi bought a refrigerator and a TV. He paid 2 stacks of ten-thousand-dollar bill, six thousand-dollar bills and 13 hundred-dollar bills. How many dollars did A-Zhi pay in total?)

Facts Generation in Figure 4(a) shows how the body text is transformed into meaningful logic facts to perform inference. In math problems, the facts are mostly related to quantities. The generated facts are either the quantities explicitly appearing in the sentence of the problem or the implicit quantities deduced by the IE. Those generated facts are linked together within the reasoning chain constructed by the IE as shown in Figure 4(b). Within this framework, the discourse planner is responsible for selecting the associated content for each sentence to be generated.

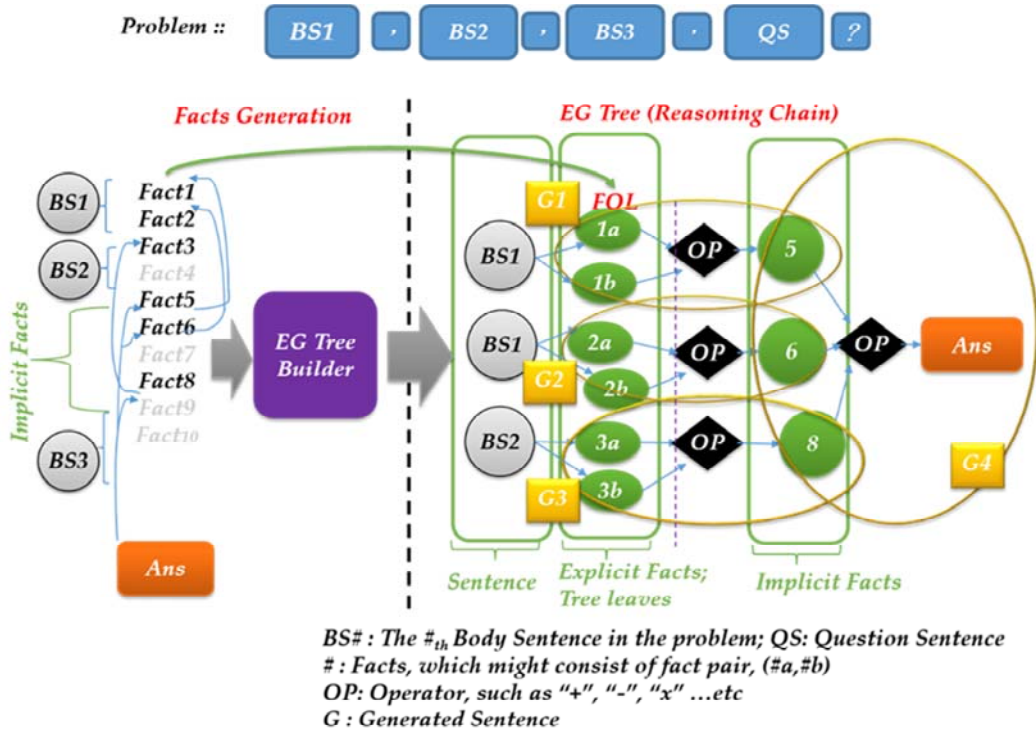


Figure 4(a). Facts Generation and EG Tree Builder

Figure 4(b). Reasoning Chain (represented as an Explanation Tree for illustration)

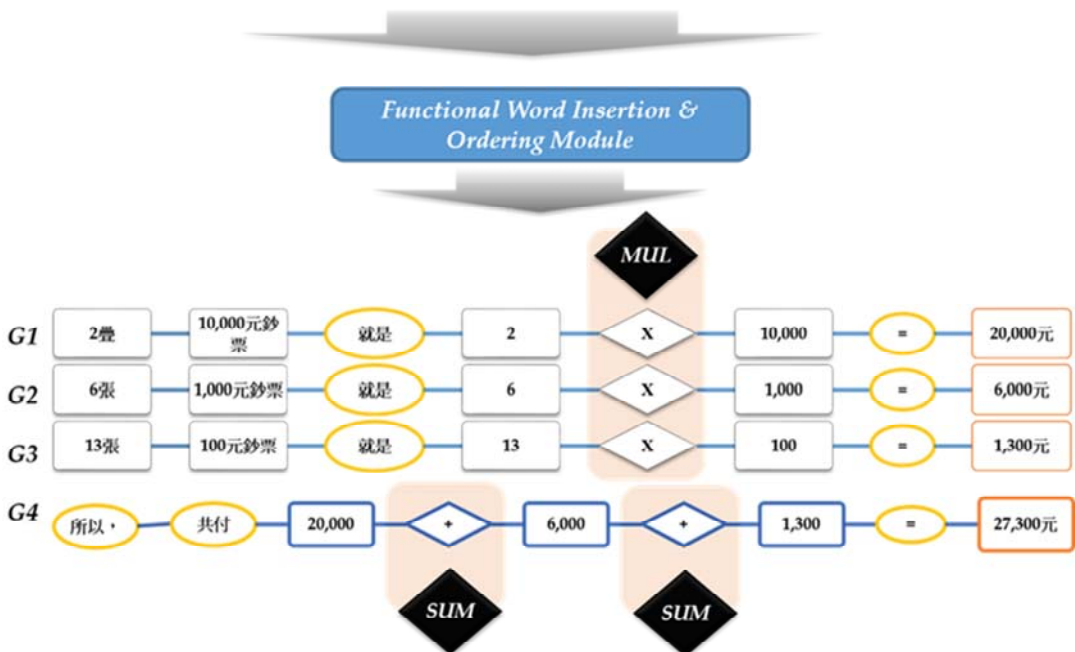



Figure 4(c). Function Word Insertion & Ordering Module, serving as the Surface Realizer. It shows how surface realization is done with pre-specified function words (circled by ellipses) and extracted slot-fillers (enclosed by diamond for operator, and rectangle for quantities).

Figure 4. (a) Facts Generated from the Body Text. (b) The associated Reasoning Chain, where “G#” shows the facts grouped within the same sentence. (c) Explanation texts generated by the TG for this example (labeled as G1~G4). Except those ellipses which symbolize pre-specified function words, other shapes denote extracted slot-fillers. Furthermore, Diamond symbolizes OP_node while Rectangle symbolizes Quan_node.

A typical reasoning chain, represented with an Explanation Tree structure, is shown at Figure 4(b). The *operator-node* (OP_node) layers and *quantity-node* (Quan_node) layers are interleaved within the Explanation Tree, and serving as the input data structure to *OP Oriented Algorithm* in Discourse Planner, which will be further presented as pseudo code in Section 2.2 (*Algorithm 1*). As shown at Figure 4(b), the (#a, #b) pair denotes facts derived from the body sentences. The OP means the operator used to deduce implicit facts and represented as non-leaf circle nodes. Each “G?” expresses a sentence to be generated. Given the reasoning chain, the first step is to decide how many sentences will be generated, which corresponds to the *Discourse Planning phase* (Jurafsky & Martin, 2000) of the traditional NLG task. Currently, we will generate one sentence for each operator shown in the reasoning chain. For the above example, since there are four operators (three IE-Multiplications² and one LFC-Sum), we will have four corresponding sentences; and the associated nodes (i.e., content) are circled by “G?” for each sentence in the figure.

Furthermore, Figure 5 shows that three sets of facts are originated from the 2nd body sentence (indicated by three *S2* nodes). Each set contains a corresponding quantity-fact (i.e., $q1(\text{疊})$, $q2(\text{張})$, and $q3(\text{張})$) and its associated object (i.e., $n1$, $n2$, and $n3$). For example, the first set (the left most one) contains $q1(\text{疊})$ (for “2 疊”) and $n1$ (for “一萬元鈔票”). This figure also shows that the outputs of three IE-Multiplication operators (i.e., “20,000 元”, “6,000 元”, and “1,300 元”) will be fed into the last LFC-Sum to get the final desired result “27,300 元” (denoted by the “Ans(SUM)” node in the figure).

After having given the corresponding content (associated with those nodes within the big circle), we need to generate the corresponding sentence with appropriate function words added. This step corresponds to the *Surface Realization phase* (Jurafsky & Martin, 2000) in NLG. Currently, since the syntax of the explanation text of our task is not complicated, we use various templates to take into account the pre-specified fillers (“

² Prefixes “IE-“ and “LFC-“ denote that those operators are issued by IE and LFC, respectively.

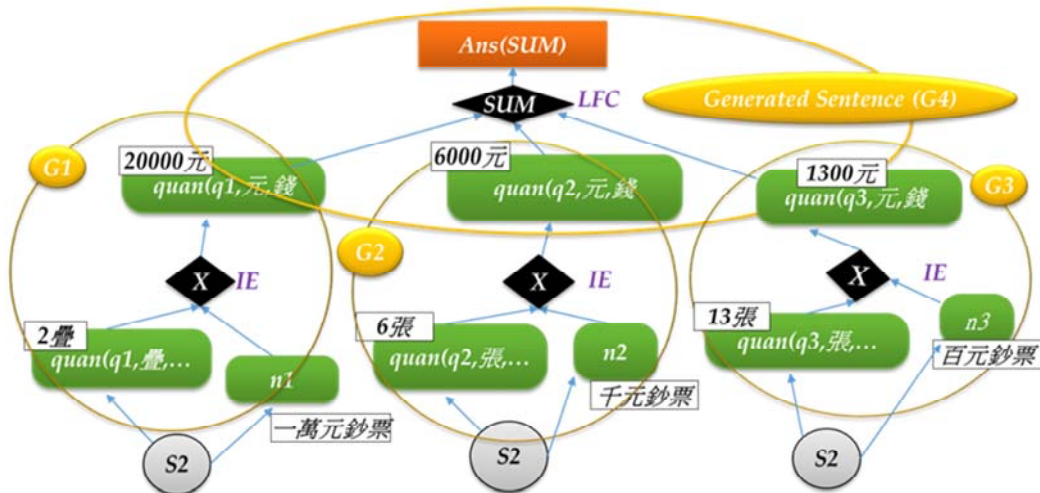


Figure 5. Explanation Tree for Discourse Planning, where S2 means that those facts are from the 2nd body sentence.

2.1 Explanation Tree Builder

The original reasoning chain resulted from the IE is actually a stream of chunks (as shown in Figure 4(a)), in which the causal chain is *implicitly* embedded. Therefore, it is not suitable for explaining inference steps. The *Explanation Tree Builder* is thus adopted to build up the *Explanation Tree*, which centers on the math operations involved in the inference process, to *explicitly* express the causal chain implied.

The Explanation Tree Builder first receives various facts, as a stream of chunks, from the IE. It then creates the nodes of the Explanation Tree according to the content of those chunks. After the Explanation Tree is created, it serves as the corresponding reasoning chain for the following process since then.

With the *root node* serving as the Answer, which is a Quan_node, the Explanation Tree is interleaved with Quan_node layers and OP_node layers, as shown in Figure 4(b). Each OP_node has one Quan_node as its parent node, and has at least one Quan_node as its child node. On the other hand, each Quan_node (except the root node) serves as the input to an OP_node. With the Explanation Tree, the work of discourse planning can be simply done via traversing those OP_nodes, which will be described in the following section.

2.2 Sentence Segmenter (Discourse Planner)

In NLG, the discourse planner selects the content from the knowledge base according to what should be presented in the output text, and then structures them coherently. To facilitate the explanation process, we first convert the given reasoning chain to its corresponding

Explanation Tree, as shown at Figure 4(b) to ease the following operations. The Explanation Tree is adopted because its structure allows us to regard the OP as a basis to do sentence segmentation for the deductive steps adopted in MWP. Within the Explanation Tree, the layers of OP nodes are interleaved with the layers of quantity nodes, and the root-node is the quantity node which denotes the desired Answer.

Algorithm 1: OP_Oriented_ExplanationGenerator

Input : (i) Directed Tree $g = (V, E)$; where V are either *OP_node* or *Quan_node*

(ii) source node **root**;

Every node records *node_number* & *depth* as it's functional data member

Besides,

OP_node records operators as its content data member

Quan_node records values as its content data member

Output : Sequence of Explanation Sentences

```

1  Initialize L to an empty List;
2  Initialize ExpSet to an empty List.
3
4  for each vertex  $j \in V$  do
5      if  $j \in OP\_node$ 
6          L.EnList(j) /* Add j into L list */
7      else /*  $j \in Value$  node */
8          pass
9
10 while L is not empty do
11      $l = DeList(L)$  /* pop-out the node with largest depth; */
12         /* if more than one, the one with smallest number is selected */
13      $s = FunctionWordInsertionAndOrderingModule(l)$  /* Algorithm 2 */
14     ExpSet.EnList(s) /* Add s into ExpSet list */
15
16 Output(ExpSet)

```

After having constructed the Explanation Tree, we need to know how to group the nodes within the tree to make a sentence. As one can imagine, there are various ways to combine different quantities and operators (within the Explanation Tree) into a sentence: you can either explain several operations within one complicated sentence, or explain those operations with several simple sentences. Discourse planner therefore controls the process for generating the discourse structure, which mainly decides how to group various Explanation Tree nodes into different discourse segments. The proposed *OP Oriented Algorithm*, as shown above, is




introduced to organize various Explanation Tree nodes into different groups (each of them will correspond to a sentence to be generated). Basically, it first locates the lowest operation node, and then traverses each operation node from *left to right* (with the same parent node) and *bottom to top*. For each operation node found, it will group the related nodes around that operation node into one discourse segment (i.e., one sentence). For each group, it will call the *Surface Realizer* module to generate the final sentence. It is named “*OP oriented*” because every generated sentence in the explanation text is based on one operator, which serves as a central hub to associate all quantities directly linked with it. Also, the template for building up a sentence is selected based on the associated operator, which will be further introduced in Section 2.3.

Figure 6 shows three grouped explanation subtrees within the original explanation tree. The arrows between SUM node and its children show the sequence of those subtrees to be presented, and the numbers imposed on tree nodes indicate the indexes of the corresponding sentence to be generated.

2.3 Function Word Insertion and Ordering Module (Surface Realizer)

The sentence segmenter module discussed previously only partitions the explanation tree into various Explanation Subtrees. It has no control over how the components within an explanation subtree should be positioned. Also, we frequently need to insert extra functional words (sometimes even verbs) such as “就是”、”共是”、”等同於” (“are”, “equal”, “mean”) and the like to have a fluent sentence. For example, in Sample-1, to explain what “2 疊一萬元” (2 stacks of 10-thousand-dollar bill) means, we need an extra functional word “就是” (“are”) (or ”共是”、”等同於” (“equal”, “mean”) and the like) to make the sentence readable. Furthermore, people prefer to add “所以” (“Thus”), to explicitly hint that the following text is closely related to the answer.

Since the syntax for explaining math operation is not complicated, we adopt the template approach to accomplish both tasks mentioned above in the same time. Currently, for each math operator, a corresponding template is manually created, which contains various slots that will be filled with contents from the nodes in Explanation Tree.

Figure 6 shows the connection between a template and its associated Explanation Tree for Sample-1. It comprises three kinds of nodes: the answer-node (shown by the rectangle ) which denotes the final answer and is basically a Quan_node; the OP_nodes (shown by the diamond ) which denote associated operators; and the quantity-nodes (shown by the rounded-corner rectangle ) which represent the values extracted by the LFC or inferred by the IE.

Take the last explanation sentence of the above sample 1 as an example,

所以，共付了 $20000 + 6000 + 1300 = 27300$ 元

Since its associated operator is “SUM”, the template of “SUM” is first fetched and there are four slots to be filled. The arrow then directs the flow to ① for “20,000” to be printed out and then SUM for the “+”. Next on, the flow is directed to the middle child node, ②, and “6,000” is therefore outputted as the subsequent component in this sentence, and then it directs back to SUM again to print “+”. Finally, the flow directs to the most right-hand-side node, ③, then goes back to SUM; the “1,300” is then popped out accordingly. We don’t print out the “+” for the SUM this time since we know there’s no more child node below the SUM node that hasn’t been traversed. After all the child nodes are traversed and their contents are copied into the associated slots, the parent node, ④, is traversed and the text “=27,300 元” is printed out to complete the explanation sentence.

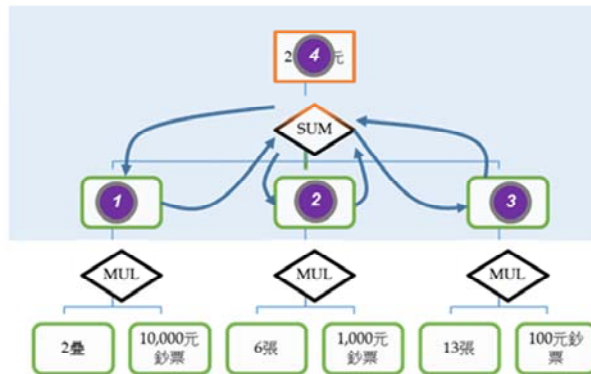


Figure 6(a). Surface Realizer – OP_SUM template

Benchmark:

2疊 10,000元鈔票 就是 $2 \times 10,000 = 20,000$ 元
 6張 1,000元鈔票 就是 $6 \times 1,000 = 6,000$ 元
 13張 100元鈔票 就是 $13 \times 100 = 1,300$ 元



Figure 6(b). Benchmark for the output of Surface Realizer

Figure 6. The template for OP_SUM (“SUM” in Figure 6 (a), and the explanation sentences for Sample-1)

Algorithm 2 shows the *Function Word Insertion and Ordering* algorithm, which illustrates how the surface realizer is implemented. After the list *S* is initiated at Line 4, the operation type of the *OP_node* is checked at Line 7 to select a corresponding template, which is assigned to *OPtemplate* at Line 8 (each kind of operator has its own template). Take Sample-1 for example, the template shown in Figure 6 (a) is selected for the “SUM” operator. Following the “Arrow” notation mentioned above, contents of the *OP_node* and its connecting nodes are put into List *S* at Line 9. Later on, the nodes in List *S* are filled into the template described above at Line 10, which corresponds to the Benchmark shown in Figure 6(b). Finally, at Line 12, the slots of *OPtemplate* are all filled with appropriate contents. It then returns them as an explanation sentence string.

Algorithm 2: FunctionWordInsertionAndOrderingModule

Input: (i) Directed Tree $g=(V,E)$; where V are either *OP_node* or *Quan_node*
(ii) one specific node $v \in V$

Output: One sentence string instantiated with components of neighboring nodes of an *OP_node*

```

1  if v.type != OP_node    /* Quan_node is returned here */
2      return NULL
3
4  Initial S to an empty list for a generated sentence
5
6  /* Select the template for Surface Realizer according to the type of operator */
7  switch(v.content)    /* for OP_node, v.content shows what kind of operator the OP_node is */
8      OPtemplate = the specified template for the OP_node
9      S.EnList(the contents of “v”, its children, and its parent)
10     Fill contents of nodes in S into the OPtemplate
11
12  return OPtemplate as a String to represent this sentence

```

Since each question will be processed separately and a reasoning chain will be associated with only one question, there is no restriction for the number of allowable question sentences (as the proposed algorithm only handles one reasoning chain each time).

3. Some Other Associated Templates

As described in the previous section, the template adopted is closely related to the associated math operation. However, various templates share a meta-form with some common characteristics:

- (1) Each operator generates a sentence.
- (2) Each sentence is generated from the operator and the quantities connected to it.
- (3) The operators and the quantities are inserted into the slots specified in the template.
- (4) The instantiated template serves as the corresponding explanation sentence string.

Apart from the OP_SUM, this section introduces a few other templates associated with OP_MUL, OP_COMMON_DIVISION, and OP_UNIT_TRANS as follows. OP_MUL is related to Sample-1 mentioned above (Figure 7). OP_COMMON_DIV is associated with Sample-2 (Figure 8). Also, Figure 9 shows the template associated with “OP_UNIT_TRANS” adopted in Sample-3.

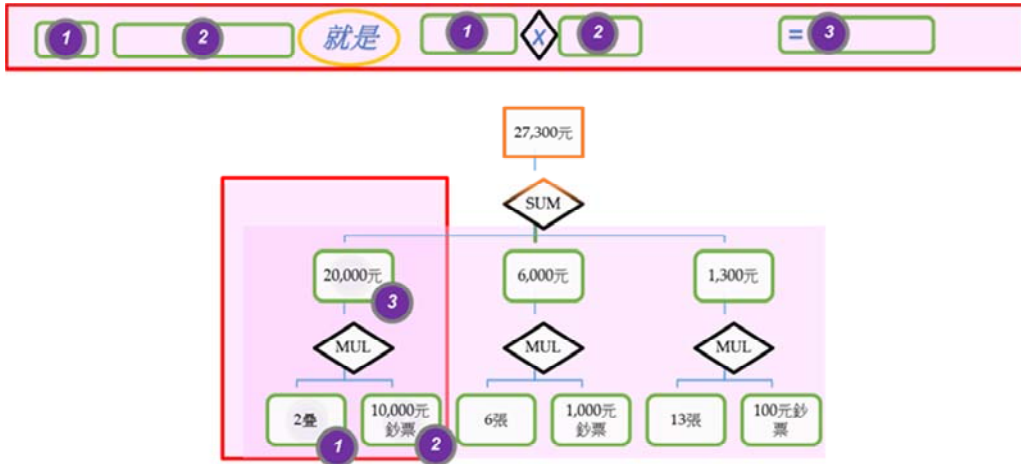


Figure 7(a)

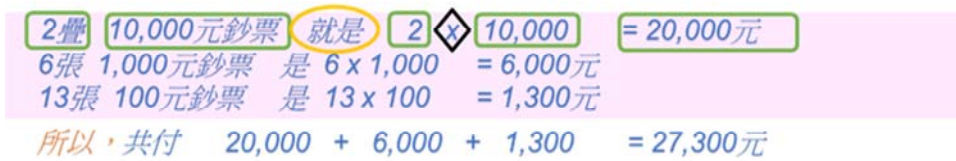


Figure 7(b)

Figure 7. The template for OP_MUL (“MUL” in Figure 7 (a)) and the explanation sentences for Sample-1.

[Sample-2] 1 個平年有 365 天，3 個平年共有幾天？

(One common-year (non-leap year) has 365 days. How many days do 3 common-year have?)

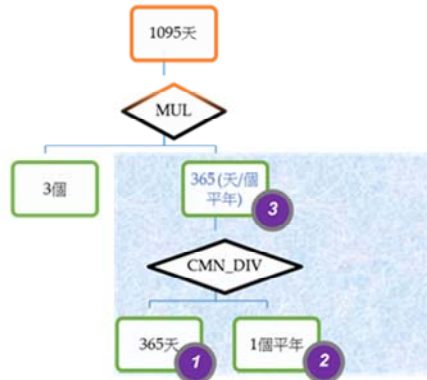


Figure 8 (a)



所以， 3個 365(天/個平年) 就是 3 x 365 = 1095天

Figure 8 (b)

Figure 8. The template for *OP_COMMON_DIV* (“*CMN_DIV*” in Figure 8 (a)) and the explanation sentences for Sample-2.

[Sample-3] 一艘輪船 20 分鐘可以行駛 25 公里，2.5 小時可以行駛多少公里？

(A ship can travel 25 km in 20 minutes. How many kilometers can it travel for 2.5 hours?)

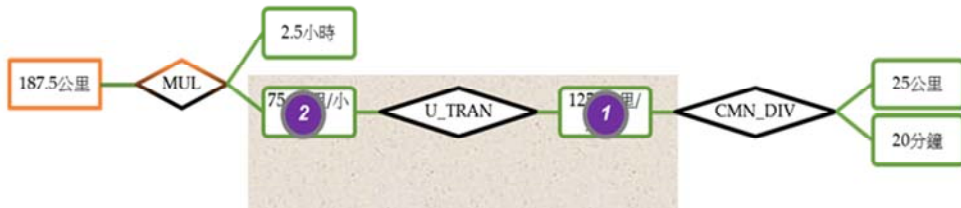


Figure 9 (a)



Figure 9 (b)

Figure 9. The template for OP_UNIT_TRANS (“U_TRAN” in Figure 9 (a)), which performs unit conversions, and the explanation sentences for Sample-3.

4. Current Status

Currently, 11 types of operators are supported. They are shown at Figure 10. After having manually checked 37 MWP problems with their associated operations specified in Figure 10,

Operation Utilities

- Sum(function[,condition])=value
- Add(value1,value2)=value
- Subtract(value1,value2)=value
- Diff(value1,value2)=value
- Multiply(value1,value2)=value
- FloorDiv(value1,value2)=value
- CeilDiv(value1,value2)=value
- Surplus(value1,value2)=value
- ArgMin(arg,function,condition)=value
- ArgMax(arg,function,condition)=value
- UnitTrans(Old-Fact, New-Fact)=value

Figure 10. Supported Operators by EG

it is observed that the proposed approach could generate fluent explanation for all of them.

5. Related Work

Earlier reported NLG applications include generating weather reports (Goldberg *et al.*, 1994; Coch, 1998), instructions (Paris *et al.*, 1998; Wahlster *et al.*, 1993), encyclopedia-like descriptions (Milosavljevic, 1997; Dale *et al.*, 1998), letters (Reiter *et al.*, 1999), and an alternative to machine translation (Hartley & Paris, 1997) which adopts the techniques of connectionist (Ward, 1994) and statistical techniques (Langkilde & Knight, 1998). However, none of them touched the problem of generating explanation for MWP.

Previous approaches of natural language generation typically consist of a discourse planner that plans the structure of the discourse, and a surface realizer that generates the real sentences (Jurafsky & Martin, 2000). D. Jurafsky adopted the model of text schemata and rhetorical relation planning for discourse planning. Approaches for surface realizer include Systemic Grammar, which is a part of Systemic Functional Linguistic proposed by Halliday (Halliday, 1985), and Functional Unification Grammar (FUG) by Kay (Kay, 1979).

Different from those previous approaches for Discourse Planner (Reiter *et al.*, 1999), we solved the EG for MWP problem through first building the Explanation Tree, which is particularly suitable for representing math based problems. The OP oriented algorithm is then proposed for solving the discourse planning work in MWP. Furthermore, different from the FUG proposed by Kay (Kay, 1979), the Function Word Insertion and Ordering Module adopts the OP based template for our Surface Realizer.

6. Conclusion

Since the EG for MWP differs from that of other NLG applications in that the inference process centers on the mathematical operation, an operator oriented algorithm is required. In the proposed framework, we first introduce the Explanation Tree to explicitly show how the answer of a math problem is acquired. Afterwards, an OP Oriented Algorithm performs sentence segmentation (act as Discourse Planner) for MWP. Lastly, for each operator, a corresponding template is adopted to achieve surface string realization.

Our Explanation Generator of MWP solver is able to explain how the answer is obtained in a human comprehensible way, where the related reasoning steps can be systematically explained with fluent natural language. The main contributions of this paper are:

1. *Proposing the Explanation Tree for facilitating the discourse planning on MWP.*
2. *Proposing an Operator oriented algorithm for structuring output sentence sequence.*
3. *Proposing the OP oriented templates for generating final explanation strings.*

References

- Coch, J. (1998). Interactive generation and knowledge administration in MultiMétéo. In *Proceedings of the Ninth International Workshop on Natural Language Generation*, 300-303.
- Dale, R., Oberlander, J., Milosavljevic, M., & Knott, A. (1998). Integrating natural language generation and hypertext to produce dynamic documents. *Interacting with Computers*, 11(2), 109-135.
- Goldberg, E., Driedger, N., & Kittredge, R. (1994). Using natural-language processing to produce weather forecasts. *IEEE Expert: Intelligent Systems and Their Applications*, 9(2), 45-53.
- Halliday, M. A. K. (1985). *An Introduction to Functional Grammar*. London, England: Edward Arnold.
- Hartley, A., & Paris, C. (1997). Multilingual document production: From support for translating to support for authoring. *Machine Translation*, 12(1), 109-128.
- Jurafsky, D., & Martin, J. H. (2000). *Speech and Language Processing*. New Jersey: Prentice Hall.
- Kay, M. (1979). Functional Grammar. In *BLS-79*, Berkeley, CA, 142-158.
- Langkilde, I., & Knight, K. (1998). The practical value of n-grams in generation. In *Proceedings of the Ninth International Workshop on Natural Language Generation*, Niagara-on-the-Lake, Ontario, Canada, 248-255.
- Lin, Y. C., Liang, C. C., Hsu, K. Y., Huang, C. T., Miao, S. Y., Ma, W. Y., Ku, L. W., Liao, C. J., & Su, K. Y. (2015). Designing a Tag-Based Statistical Math Word Problem Solver with Reasoning and Explanation. *International Journal of Computational Linguistics and Chinese Language Processing (IJCLCLP)*, 20(2), 1-26.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data – A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt Publishing Company.
- Milosavljevic, M. (1997). Content selection in comparison generation. In *Proceedings of the 6th European Workshop on Natural Language Generation*, Duisburg, Germany, 72-81.
- Mukherjee, A., & Garain, U. (2008). A review of methods for automatic understanding of natural language mathematical problems. *Artif Intell Rev*, 29(2), 93-122.
- Paris, C., & Vander Linden, K. (1996). DRAFTER: An interactive support tool for writing multilingual instructions. *IEEE Computer*, 29(7), 49-56.
- Paris, C., Vander Linden, K., & Lu, S. (1998). Automatic document creation from software specifications. In *Proceedings of the 3rd Australian Document Computing Symposium (ADCS-98)*, 26-31.
- Reiter, E., Robertson, R., & Osman, L. (1999). Types of knowledge required to personalise smoking cessation letters. In *Proceedings of the Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making*. Springer-Verlag, 389-399.

- Russell, S. J. & Norvig, P. (2009). *Artificial Intelligence : A Modern Approach*(3rd Edition), Prentice Hall.
- Strassel, S., Adams, D., Goldberg, H., Herr, J., Keesing, R., Oblinger, D., Simpson, H., Schrag, R., & Wright, J. (2010). The DARPA Machine Reading Program - Encouraging Linguistic and Reasoning Research with a Series of Reading Tasks. *LREC 2010*.
- Wahlster, W., André, E., Finkler, W., Profitlich, H.-J., & Rist, T. (1993). Plan based Integration of Natural Language and Graphics Generation. *Artificial Intelligence*, 63(1993) 387-428.
- Ward, N. (1994). *A Connectionist Language Generator*. New Jersey: Ablex Publishing Corporation.

Word Co-occurrence Augmented Topic Model in Short Text

Guan-Bin Chen* and Hung-Yu Kao*

Abstract

The large amount of text on the Internet cause people hard to understand the meaning in a short limit time. Topic models (e.g. LDA and PLSA) has been proposed to summarize the long text into several topic terms. In the recent years, the short text media such as tweet is very popular. However, directly applies the transitional topic model on the short text corpus usually gating non-coherent topics. Because there is no enough words to discover the word co-occurrence pattern in a short document. The Bi-term topic model (BTM) has been proposed to improve this problem. However, BTM just consider simple bi-term frequency which cause the generated topics are dominated by common words. In this paper, we solve the problem of the frequent bi-term in BTM. Thus, we proposed an improvement of word co-occurrence method to enhance the topic models. We apply the word co-occurrence information to the BTM. The experimental result that show our PMI- β -BTM gets well result in the both of regular short news title text and the noisy tweet text. Moreover, there are two advantages in our method. We do not need any external data and our proposed methods are based on the original topic model that we did not modify the model itself, thus our methods can easily apply to some other existing BTM based models.

Keywords: Short Text, Topic Model, Document Clustering, Document Classification

1. Introduction

With the advancement of information and communication technology, the information we obtained is very abundant and multivariate. Especially, in the recent 15 years, many type of the Internet media grow up so that people can get large amount of the information in a short time. These internet media include Wikipedia, blogs and the recently popular social medial

* Department of Computer Science and Information Engineering, National Cheng Kung University
E-mail: gbchen@ikmlab.csie.ncku.edu.tw; hykao@mail.ncku.edu.tw

such as Twitter, Facebook et.al. Generally, the articles/documents in the Wikipedia, and blogs are usually the long text and have the complete content. While the short text social media, such as Twitter, become very popular in the recent years. The reason is that these short text social media provide a very convenient way to share the people feeling and thinking.

Generally, these Internet media deliver the people thinking by using the text. However, the large amount of text on the Internet cause people hard to understand the meaning in a short limit time. To solve the problem, many document summarization technologies have been proposed. Among them, topic models summarize the context in large amount of documents into several topic terms. By reading these topic terms, people will understand the content in a short time. Topic model can be performed by the vector space model or the probability model. In the recent years, the probability models such as Probabilistic Latent Semantic Analysis (pLSA) (Hofmann, 1999) and Latent Dirichlet Allocation (LDA) (Blei *et al.*, 2003) are very popular because the probability models base on the document generation process. The inspirations of the document generation process come from the human written articles. When a person writes an article, he or she will inspire some thinking in mind, then extend these thinking into some related words. Finally, they write down these words to complete an article. Probability topic models simulate the behavior of above document generating process. In the view of the vectorization of the probability topic models, when we have a text corpus, we have known the documents and its words distribution by statistic the word vector. Then, the probability topic models split the document-word matrix into the document-topic and topic-word matrices. The distribution of the document-topic matrix describes that the degree of each document belongs each topic while the topic-word matrix describes the degree of each word belongs each topic. The “topic” in these two matrices is the latent factor as the human thinking.

In essence, the topic models capture the word co-occurrence information and these highly co-occurrence words are put together to compose a topic (Divya *et al.*, 2013; Mimno *et al.*, 2011). So, the key to find out high quality topics is that the corpus must contain a large amount of word co-occurrence information and the topic model has the ability to correctly capture the amount of the word co-occurrence. However, the traditional topic models work well in the long text corpus but work poorly in short text corpus. The reason is that the original intention of LDA is designed to model the long text corpus. Exactly, LDA capture the word co-occurrence in document-level (Divya *et al.*, 2013; Yan *et al.*, 2013), but there are no enough words to well judge the word co-occurrence in document-level in a short text document. Figure 1 is an example which shows the difference of the topic model in between the long text and short text corpus. In the long text corpus, each document provides a lot of word co-occurrence information, so that LDA can well capture these information to discover the high quality topics. While in the short text document, there are no enough words in a

single document to discover the word co-occurrence information.

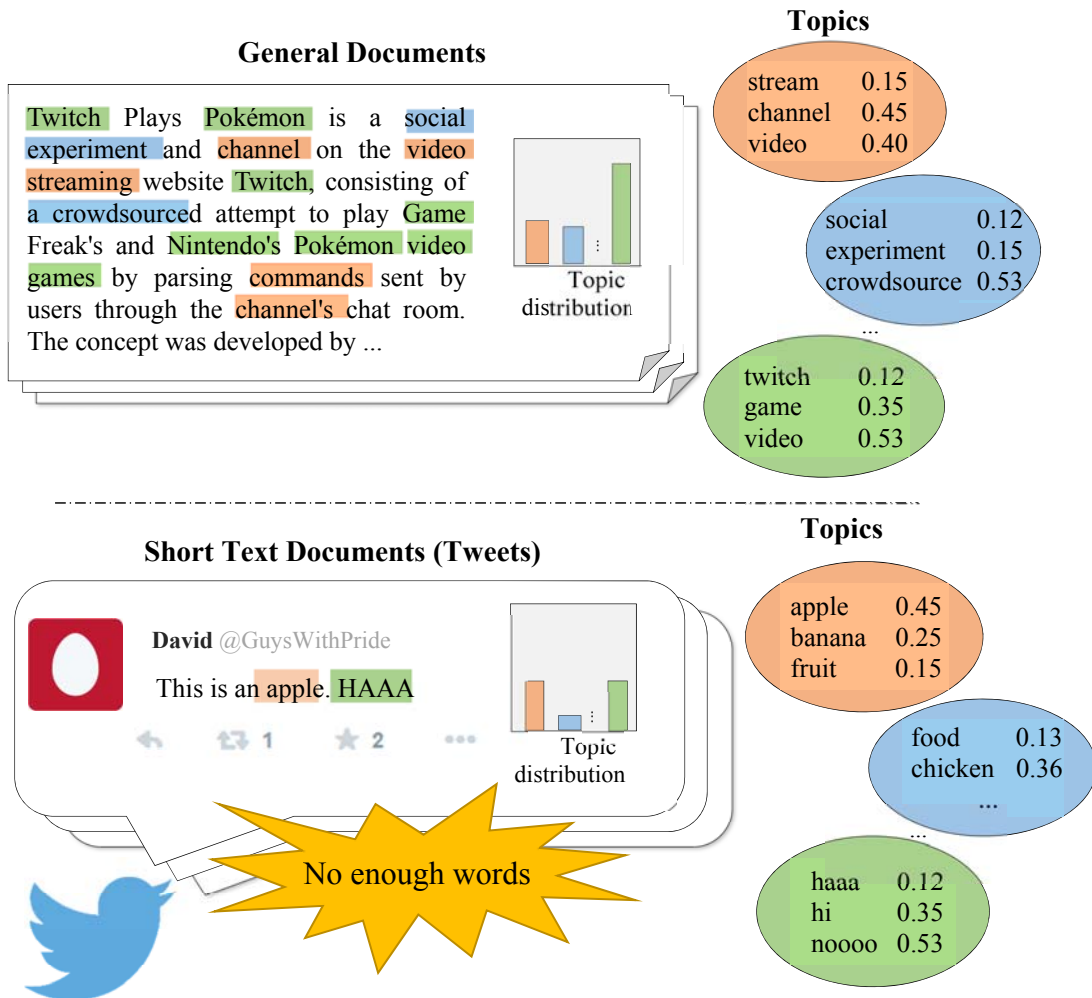


Figure 1. An example of LDA in the long text and short text corpus

To overcome above problems in short text, many researchers consider a simpler topic model, mixture of unigrams model. Mixture of unigrams model samples topics in global corpus level (Nigam *et al.*, 2000; Zhao *et al.*, 2011). More specifically, the word co-occurrence in document-level means that the amount of the word co-occurrence relation comes from a single document. On the contrary, the word co-occurrence in corpus-level means that the amount of the word co-occurrence relation comes from a full corpus which contains many documents. Mixture of unigrams overcomes the lack of words in the short text documents. Further, Xiaohui Yan *et al.* proposed the Bi-term Topic Model (BTM) (Yan *et al.*,

2013; Cheng *et al.*, 2014) which directly model the word co-occurrence and use the corpus-level bi-term to overcome the lack of the text information problem. A bi-term is an unordered word pair co-occurring in a short text document. The major advantage of BTM is that 1) BTM model the word co-occurrence by using the explicit bi-term, and 2) BTM aggregate these word co-occurrence patterns in the corpus for topic discovering (Yan *et al.*, 2013; Cheng *et al.*, 2014). BTM abandons the document-level directly. A topic in BTM contains several bi-term and a bi-term crosses many documents. BTM emphasizes that the co-occurrence information comes from all bi-terms in whole corpus. However, BTM will make the common words be performed excessively because the frequency of bi-term comes from the whole corpus instead of a short document.

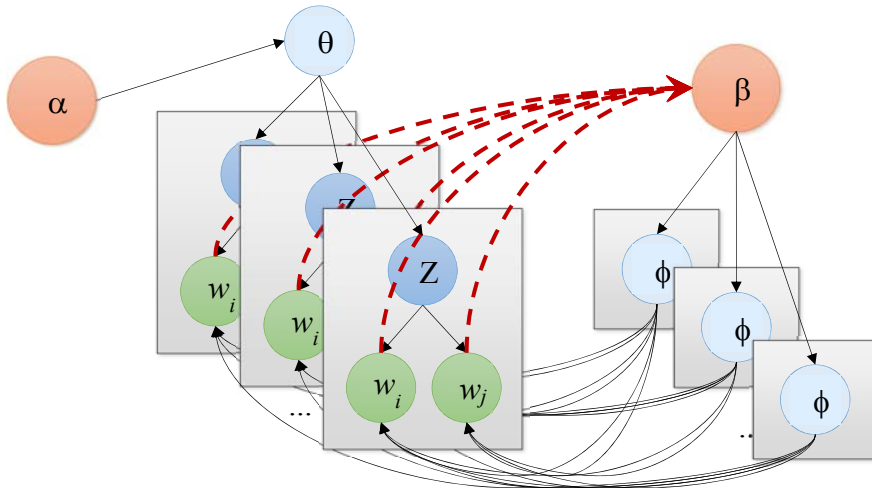


Figure 2. The graphical representation of the PMI- β -BTM

In this paper, we solve the frequent bi-term problem in BTM. We propose an approach base on BTM. For the problem in BTM, a simple and intuitive solution is to use pointwise mutual information (PMI) (Church & Hanks, 1990) to decrease the statistical amount of the frequent words in whole corpus. With respect to the frequency of bi-term, the PMI can normalize the score by each single word frequency in the bi-term. Otherwise, the priors in the topic models usually set symmetric. This symmetric priors mean that there is not any preference of words in any specific topic (Wallach *et al.*, 2009). An intuitive idea is that why not adopt some word co-occurrence information in priors to restrict the generated topics. Base on above two ideas, we propose a novel prior adjustment method, PMI- β priors, which first use the PMI to mine the word co-occurrence from the whole corpus. Then, we transform such PMI scores to the priors of BTM. Figure 2 shows the graphical representation of the PMI- β -BTM.

In summary, the proposed approach enhance the amount of the word co-occurrence and

also based on the original topic model. Basing on the original topic model means we did not modify the model itself, thus our methods can easily apply to some other existing BTM based models, to overcome the short text problem without any modification. To test the performance of our two methods completely, we prepare two different types of short text corpus for the experiments. One is the tweet text and another is the news title. The context of news title dataset is regular and formal while the text in tweet usually contain many noise. Experimental results show our PMI- β priors method is better than the BTM in both tweet and news title datasets.

The remaining of this paper shows below. In Section 2, we show the survey of some traditional topic models and the previous works of topic model to overcome the short text. Section 3 shows our proposed PMI- β priors and the re-organized document methods. The experiment results show in Section 4. Finally, we conclude this research in Section 5.

2. Related Work

2.1 The Survey of the Traditional Topic Models for Normal Text

Topic Model is a method to find out the hidden semantic topics from the observed documents in the text corpus. Topic Models have been researched several years. Generally, topic model can be performed by the vector space model or the probability model. The early one of the vector space topic model, Latent Semantic Analysis (LSA) (Landauer *et al.*, 1998), uses the singular value decomposition (SVD) to find out the latent topic. However, LSA does not model the polysemy well and the cost of SVD is very high (Hofmann, 1999; Blei *et al.*, 2003). Afterward, Thomas Hofmann proposed the one-document-multi-topics model, probabilistic Latent Semantic Analysis (pLSA) (Hofmann, 1999). pLSA bases on the document generation process which like the human writing. However, the numerous parameters of pLSA cause the overfitting problem and pLSA does not define the generation of the unknown documents. In 2003, Blei *et al.* proposed a well-known Latent Dirichlet Allocation (LDA) (Blei *et al.*, 2003), LDA use the prior probability in Bayes theory to extents pLSA and simplify the parameters estimate process in pLSA. Also, the non-zero priors let LDA have the ability to infer the new documents.

However, there are some drawbacks in LDA. First, LDA works under the bag-of-word model hypothesis. In the bag-of-word model, each word of the document is no order and independent of others (Wallach, 2006). The hypothesis compared with the human writing behavior is unreasonable (Divya *et al.*, 2013). Second, LDA emphasizes the relations between topics are weak, but actually, the topics may have hierarchical structure. Third, LDA requires the large number of articles and well-structured long articles to get the high quality topics. Apply LDA on the short text or uncompleted sentences corpus usually get poor results. The

fourth drawback is that in spite of the LDA has the concept of the prior probabilities but LDA priors generally set the symmetric values in each prior vector, like $\langle 0.1 \rangle$ or $\langle 0.01 \rangle$. The symmetric prior means no bias of each words in the specific topic (Wallach *et al.*, 2009). In this situation, the priors only provide the smooth technology to avoid the zero probability and the model only use the statistical information from the data to discover the hidden topics.

To overcome above four drawbacks, many researchers propose new modify models. Such as N-gram Topic Model (Wang *et al.*, 2007) and HMM-LDA (Griffiths *et al.*, 2004) provide the context modeling. Wei Li *et al.* proposed the Pachinko Allocation Model (PAM) (Li & McCallum, 2006) which adds the super topic concept and make the topic have the hierarchical structure. Otherwise, Zhiyuan Chen *et al.* apply the must-link and cannot-link information to guide the document generation process which words must or not to be put into a topic (Chen & Liu, 2014).

2.2 Topic Models for Short Text

With the rise of social media in recent years, topic models have been utilized for social media analysis. For example, some researches apply topic models in social media for event tracking (Lin *et al.*, 2010), content characterizing (Zhao *et al.*, 2011; Ramage *et al.*, 2010), and content recommendation (Chen *et al.*, 2010; Phelan *et al.*, 2009). However, to share people thinking conveniently, the context is usually short. These short text contexts make topic models hard to discover the amount of word co-occurrence. For the short text corpus, there are three directions to overcome the insufficient of the word co-occurrence problem. One is using the external resources to guide the model generation, another is aggregating several short texts into a long text, and the other is improving the model to satisfy the short text properties. For the first direction, Phan *et al.* (Phan *et al.*, 2008) proposed a framework that adopt the large external resources (such as Wiki and blog) to deal with the data sparsity problem. R.Z. Michal *et al.* proposed an author topic model (Rosen-Zvi *et al.*, 2004) which adopt the user information and make the model suitable for specific users. Jin *et al.* proposed the Dual-LDA model (Jin *et al.*, 2011), it use not only the short text corpus but also the related long text corpus to generate topics, respectively. The generation process use the long text to help the short text modeling. If the quality of the external long text or knowledge base is high, the generated topic quality will be improve. However, we cannot always obtain the related long text to guide short text and the related long text is very domain specific. So, using external resources is not suitable for the general short text dataset. In addition to adopt the long text, Hong *et al.* aggregate the tweets which shared the same words and get better results than the original tweet text (Hong & Davison, 2010).

For the model improvement, Wayne *et al.* use the mixture of unigrams model to model the tweets topics from whole corpus text (Zhao *et al.*, 2011). Their experimental results verify

that the mixture of unigram model can discover more coherent topics than LDA in the short text corpus. Further, Xiaohui Yan *et al.* proposed the Bi-term Topic Model (BTM) (Yan *et al.*, 2013; Cheng *et al.*, 2014) which directly model the word co-occurrence and use the corpus level bi-term to overcome the lack of the text information problem. A bi-term is a word pair containing a co-occur relation in this two words. The advantage is that BTM can model the general text without any domain specific external data. Comparing with the mixture of unigram, BTM is a special case of the mixture of unigram. They both model the corpus level topic but BTM generates two words (bi-term) every time the generation process. However, BTM discovers the word co-occurrence just by considering the bi-term frequency. The bi-term frequency will be failed to judge the word co-occurrence when the bi-term frequency is high but one of the frequency of two words in a bi-term is high and another is low.

3. The Word Co-occurrence Augmented Methods

Topic models learn topics base on the amount of the word co-occurrence in the documents. The word co-occurrence is a degree which describes how often the two words appear together. BTM, discovers topics from bi-terms in the whole corpus to overcome the lack of local word co-occurrence information. However, BTM will make the common words be performed excessively because BTM identifies the word co-occurrence information by the bi-term frequency in corpus-level. Thus, we propose a PMI- β priors methods on BTM. Our PMI- β priors method can adjust the co-occurrence score to prevent the common words problem. Next, we will describe the detail of our method of PMI- β priors.

We first describe the detail of BTM. First, we introduce the notation of “bi-term”. Bi-term is the word pair co-occurring in the short text. Any two distinct words in a document construct a bi-term. For example, a document with three terms will generate three bi-term (Yan *et al.*, 2013):

$$(t_1, t_2, t_3) \Rightarrow \{(t_1, t_2), (t_2, t_3), (t_1, t_3)\}. \quad (1)$$

Note that each bi-term is unordered. For a real case example, we have a document and the context is “I visit apple store”. Because “I” is a stop-word, we remove it. The remaining three terms “visit”, “apple” and “store” will generate three bi-terms “visit apple”, “apple store”, and “visit store”. We generate all possible bi-terms for each document and put all bi-terms in the bi-term set B.

Second, we describe the parameter estimation of the BTM. The aim of the parameter estimation of BTM is to estimate the topic assignment z , the corpus-topic posteriori distribution θ and the topic-word posteriori distribution ϕ . But the Gibbs sampling can integrate θ and ϕ due to use the conjugate priors. Thus, the only one parameter z should be

estimate. Clearly, we should assign a suitable topic for each bi-term. The Gibbs sampling equation shows below:

$$P(z = k | \mathbf{z}_{-b}, \mathbf{B}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \propto \theta \cdot \varphi, \quad (2)$$

where z is the topic assignment, k means the k th topic, \mathbf{B} is the bi-term set, $\boldsymbol{\alpha}$ is the corpus-topic prior distribution and $\boldsymbol{\beta}$ is the topic-word prior distribution. The θ and φ in Eq. (2) show following:

$$\theta = \frac{(n_{k,-b} + \alpha_k)}{\sum_{k=1}^K (n_{k,-b} + \alpha_k)}, \quad (3)$$

$$\varphi = \frac{(n_{k,-b}^{w_1} + \beta_k^{w_1})}{\sum_{t=1}^V (n_{k,-b}^{w_t} + \beta_k^{w_t})} \times \frac{(n_{k,-b}^{w_2} + \beta_k^{w_2})}{\sum_{t=1}^V (n_{k,-b}^{w_t} + \beta_k^{w_t})}, \quad (4)$$

where V is the number of unique words in the corpus, $n_{k,-b}$ is the statistical count for the document-topic distribution, and $n_{k,-b}^{w_i}$ is the statistical count for the document-topic distribution. When the frequency of bi-term is high the two terms in this bi-term tend to be put into the same topic. Otherwise, to overcome the lack of words in a single document BTM abandons the document-level directly. A topic in BTM contains several bi-term and a bi-term crosses many documents. BTM emphasizes that the co-occurrence information comes from all bi-terms in whole corpus.

However, just consider the frequency of bi-term in corpus-level will generate the topics which contain too many common words. To solve this problem, we consider the Pointwise Mutual Information (PMI) (Church & Hanks, 1990). Since the PMI score not only considers the co-occurrence frequency of the two words, but also normalizes by the single word frequency. Thus, we want to apply PMI score in the original BTM. A suitable way to apply PMI scores is modifying the priors in the BTM. The reason is that the priors modifying will not increase the complexity in the generation model and very intuitive. Clearly, there are two kinds of priors in BTM which are β -prior and β -priors. The β -prior is a corpus-topic bias without the data. While the β -priors are topic-word biases without the data. Applying the PMI score to the β -priors is the only one choice because we can adjust the degree of the word co-occurrence by modifying the distributions in the β -priors. For example, we assume that a topic contains three words “pen”, “apple” and “banana”. In the symmetric priors, we set $\langle 0.1, 0.1, 0.1 \rangle$ which means no bias of these three words, while we can apply $\langle 0.1, 0.5, 0.5 \rangle$ to enhance the word co-occurrence of “apple” and “banana”. Thus the topic will prefer to put the “apple” and “banana” together in the topic sampling step.

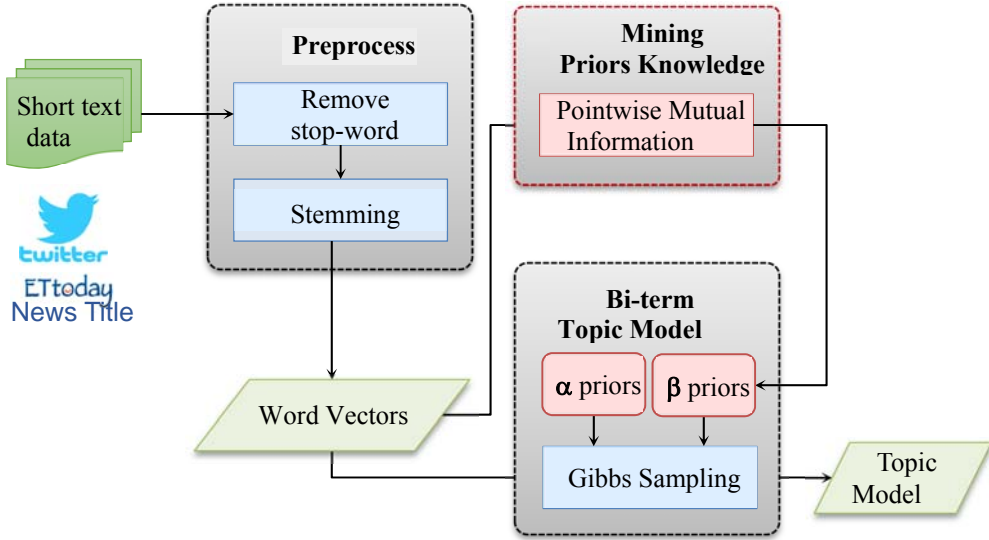


Figure 3. The PMI- β priors approach

Figure 3 shows our PMI- β -priors approach. After pre-processing, we first calculate the PMI score of each bi-term $\langle w_x, w_y \rangle$ as

$$PMI(w_x, w_y) = \log \frac{p(w_x, w_y)}{p(w_x)p(w_y)}, \quad (5)$$

Because the priors can view as an additional statistics count of the target probability, the value ordinarily should be greater than or equal to zero. Thus, we adjust the value of NPMI to $[0, 2]$ by adding one as:

$$NPMI(w_x, w_y) = \frac{PMI(w_x, w_y)}{-\log p(w_x, w_y)} + 1. \quad (6)$$

After getting the NPMI scores, we transform these scores to meet the β -priors. Let β_{SYM} is the original symmetric β -priors and the PMI β -priors, denote β_{PMI} , define as

$$\beta_{PMI}^{w_x, w_y} = \beta_{SYM} + 0.1 \times NPMI(w_x, w_y). \quad (7)$$

There is a constant value 0.1 in Eq. (7). This constant value 0.1 prevent the target probability being dominated by the priors. The partial of the word co-occurrence information should still be captured by the original model and the priors provide the additional information to enhance the word co-occurrence in the model. The following shows how we apply PMI- β -priors into the BTM. We apply the β_{PMI} of w_1 and w_2 in Eq. (6) and the new equation of shows below:

$$\varphi = \frac{(n_{k,-b}^{w_1} + \beta_{PMI}^{w_1, w_2})}{\sum_{t=1}^V (n_{k,-b}^{w_t} + \beta_k^{w_t})} \times \frac{(n_{k,-b}^{w_2} + \beta_{PMI}^{w_1, w_2})}{\sum_{t=1}^V (n_{k,-b}^{w_t} + \beta_k^{w_t})}. \quad (8)$$

Finally, we sample topic assignments by Gibbs sampling (Liu, 1994) approach.

4. Experiments

How to justly evaluate the quality of the topic model is still a problem. The reason is that the topic model is an unsupervised method. There are no prominent words or labels can directly assign to each topic. Thus, many researchers apply topic model in other applications, such as clustering, classification and information retrieval (Blei *et al.*, 2003; Yan *et al.*, 2013). In classification task, instead of using the original word vectors to identify the document categories, it use the reduced vectors which generating from the topic model. The topic model plays as a dimensional reduction role and the classification result shows how well the model to represent the original features. Topic model can also look as the document clustering approach by just considering a document assign to which topic(s). In this paper, we evaluate topic models by clustering and classification tasks. Otherwise, to make our experiment more robust, we adopt two different types of short text dataset - Twitter2011 and ETtoday Chinese news title. The properties of these two corpus are different. The text of ETtoday Chinese news title is very regular, while the text of Twitter2011 usually contains emotional words, simplified texts and some unformed words. For example, “haha” is the emotional word, and “agreeeee” is the unformed word.

Table 1 shows the statistics of short text datasets. The number of average words per document is not more than ten words. The number of documents in each class are shown in Figure 4. The property of both two dataset is skew. The skew dataset may cause the results that the fewer documents are dominated by the larger one. In summary, the challenges of these two datasets are not only the short text problem but also the unbalance category. The top-3 classes in the Twitter2011 dataset are “#jan25”, “#superbowl” and “#sotu”. And the top-3 classes in the ETtoday News Title dataset are “entertainment”, “physical” and “political”.

Table 1. The Statistics of Two Short Text Datasets

Property	Twitter2011	ETtoday News title
The number of documents	49,461	17,814
The number of domains	50	25
The number of distinct words	30,421	31,217
Avg. words per document	5.92	9.25

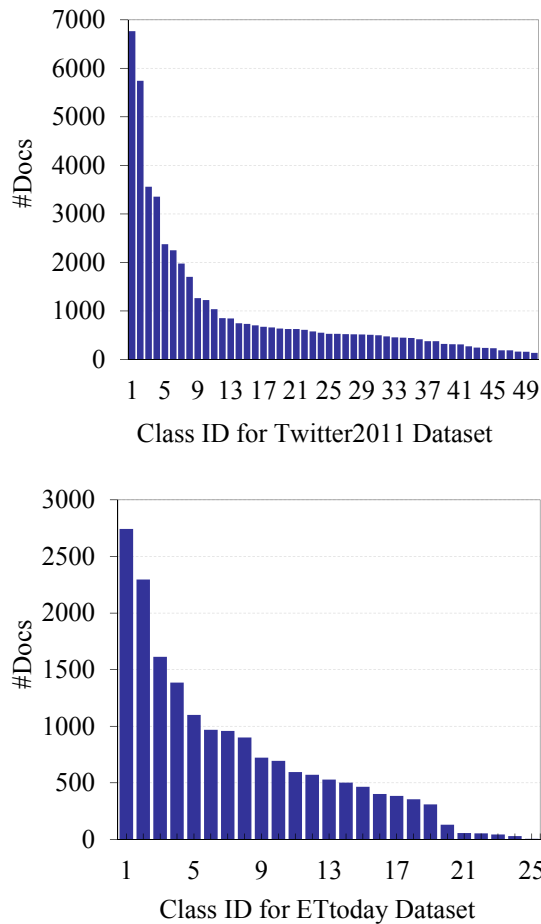


Figure 4. *The number of documents in each class*

4.1 Experimental Setup

All of the experiments were done on the Intel i7 3.4 GHz CPU and 16G memory PC. All of the pre-process and topic models were written by JAVA code. The parameters α priors and the base β priors of topic models are all set $\langle 0.1 \rangle$. The number of iterations in Gibbs sampling is set 1,000. To make our results more reliable, we run each experiments 10 times and average these scores.

For the clustering experiment, we first get the document-topic posteriori probability distribution ϕ and we use the highest probability topic $P(z|d)$ as the cluster assignment for each document in ϕ . For the classification experiment, we divide our dataset into five parts in which four parts for training and one for testing. After training the topic model, we fix the topic-word distribution ϕ and then we re-infer document-topic posteriori probability

distribution θ of all original short text documents. Instead of using the original word vectors to do the classification task, we take this re-inferred posteriori probability distribution θ as the reduced feature matrix. Finally we use this reduced feature matrix to classify the documents by LIBLINEAR¹.

We compare our methods with the previous topic models: 1) LDA, 2) Mixture of unigrams, and 3) BTM. In addition to the above three topic models, we also compare with our PCA- β priors methods. We use the principal component analysis (PCA) to discover the whole corpus principal component. Then, we transform the principal component to the topic-word prior distribution.

4.2 Evaluation Criteria

In this part, we list three criteria for the clustering experiment and one for classification. In the clustering experiment, let $\Omega = \{\omega_1, \omega_2, \dots, \omega_K\}$ is the output cluster labels, and $C = \{c_1, c_2, \dots, c_p\}$ is the gold standard labels of the documents. We first describe the three criteria for the clustering.

- **Purity**

Purity is a simple and transparent measure which perform the accuracy of all cluster assignments as the following equation:

$$\text{Purity}(\Omega, C) = \frac{\sum_k \max_j \|\omega_k \cap c_j\|}{N}, \quad (9)$$

where N is the total number of documents. Note that the high purity is easy to achieve when the number of clusters is large. In particular, purity is 1 if each document gets its own cluster.

- **Normalized Mutual Information (NMI)**

NMI score is based on the information theory. Let $I(\Omega, C)$ denotes the mutual information between the output cluster Ω and the gold standard cluster C . The mutual information of NMI is normalized by each entropy denoted $H(\Omega)$ and $H(C)$. This normalization can avoid the influence of the number of clusters. The equation of NMI shows following:

$$\text{NMI}(\Omega, C) = \frac{I(\Omega, C)}{[H(\Omega) + H(C)]/2}, \quad (10)$$

where $I(\Omega, C)$, $H(\Omega)$ and $H(C)$ denote:

¹ <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>

$$I(\Omega, C) = \sum_k \sum_j P(\varpi_k \cap c_j) \log \frac{P(\varpi_k \cap c_j)}{P(\varpi_k)P(c_j)}, \quad (11)$$

$$H(\Omega) = -\sum_k P(\varpi_k) \log P(\varpi_k). \quad (12)$$

- **Rand Index**

Rand Index (RI) (Rand, 1971) consider the clustering result as a pair-wise decision. More clearly, RI penalizes both true positive and true negative decisions during clustering. If two documents are both in the same class and the same cluster, or both in different classes and different clusters, this decision is correct. For other cases, the decision is false. The equation of RI shows following:

$$RI = \frac{TP + TN}{TP + FP + FN + TN}, \quad (13)$$

where TP , FP , FN , and TN are the true positive count, false positive count, false negative count and true negative count respectively. For the classification experiment, we adopt the accuracy as the measure. The definition of the accuracy is the same as the RI score in Eq. (13), but just change the cluster label to the classification label.

4.3 Experimental Results for the Twitter2011 Dataset

The Twitter2011 dataset was published in TREC 2011 microblog track². It contains approximately 16 million tweets sampled between January 23rd and February 8th, 2011. It is worth mentioning that there are some semantics tags, called hashtag, in some tweets. The hashtags had been given when the author wrote a tweet. Because these hashtags can identify the semantics of tweets, we use the hashtags as our ground truth for both clustering and classification experiments. However, there are about 10 percentages of all tweets contain hashtags and some hashtags are very rare. Also, there are contains multilingual tweets. To reduce the effect of noise in this dataset, we just extract the English tweets with top-50 frequent hashtags. After tweet extraction, we totally get the 49,461 tweets. Then, we remove the hashtags and stop-words from the context. Finally, we stem all the words in all tweets by the English stemming in the Snowball library.

Table 2 shows the clustering results on the Twitter2011 dataset, when we set the number of topic to 50. As expected, BTM is better than Mixture of unigram and LDA got the worst result when we adopt the symmetric priors $\langle 0.1 \rangle$. When apply the PMI- β priors, we get the

² <http://trec.nist.gov/data/tweets/>

better result than BTM with symmetric priors. Otherwise, our baseline method, PCA- β , is better than the original LDA because the PCA- β prior can make up the lack of the global word co-occurrence information in the original LDA.

Table 2. The Clustering Results on Twitter2011 dataset

Model	β priors	Purity	NMI	RI
LDA	<0.100>	0.4174	0.3217	0.9127
	PCA- β	0.4348	0.3325	0.9266
Mix	<0.100>	0.4217	0.3358	0.8687
	PCA- β	0.3748	0.3305	0.7550
BTM	<0.100>	0.4318	0.3429	0.9092
	PCA- β	0.4367	0.4000	0.8665
	PMI- β	0.4427	0.3927	0.9284

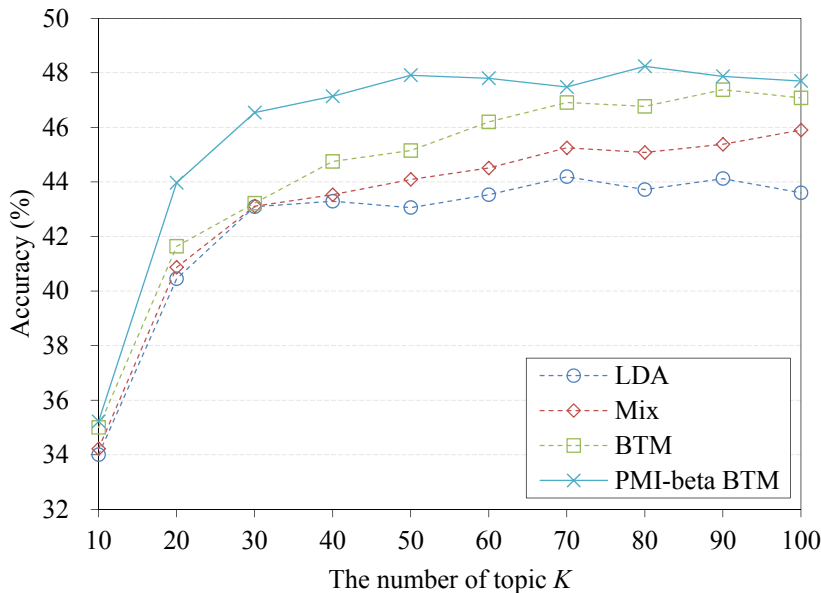


Figure 5. The Classification Results on Twitter2011 dataset

Figure 5 shows the classification results on the Twitter2011 dataset by using LIBLINEAR classifier. When apply the PMI- β priors, we get the better result than BTM with symmetric priors. Table 3 presents the top-10 topic words of the “job” topic in the Twitter2011 dataset for LDA, mixture of unigram, BTM and PMI- β -BTM respectively, when the number of topic is 70. The top-10 words are the 10 highest probability words of the topics. The bold words in this table are the words which highly correlated with the topic by the

human judgment. The topic words in the LDA and mixture of unigram models are almost non-correlated or low-correlated with the topic “job”, such as “jay” and “emote”. In BTM and PMI- β -BTM, the model capture the more high-correlated words, such as “engineer” and “management”.

Table 3. The top-10 topic words of the “job” topic in Twitter2011 dataset

	Top-10 Topic words
LDA	job , house, jay, steal, material, burglary, construct , park, pick, ur
Mix	job , robbery, material, construct , steal, warehouse, emote, feel, woman, does
BTM	job , management , engineer , media, social, open, sale , analyst, develop , senior
PMI- β -BTM	job , real, open, estate, management , market , company , sale , develop , engineer

4.4 Experimental Results for ETtoday News Title Dataset

The ETtoday News Title dataset is collected from the overview list of the ETtoday News website³ between January 1st and January 31, 2015. There are totally 25 predefined news labels in the dataset. These labels include some classical news category such as “society news”, “international news” and “political news”, and some special news category such as “animal and pets”, “3C” and “games”. In both the clustering and the classification experiments, we use these labels as the ground-truth. Because the Chinese text does not contain the break word, we must adopt the additional word breaker in the pre-process step. We adopt the jieba⁴, the Python Chinese word segmentation module, to segment all news title into several words.

Figure 6 shows the classification results on the ETtoday News Title dataset. The three original topic model LDA, mixture of unigram, and BTM perform the same order as the results of the Tweet2011 dataset. The PMI- β BTM is outperform all other methods. Our PMI- β -BTM is also suitable to model the regular short text. The top-10 topic words of the “baseball” topic of ETtoday news title dataset lists in the Table 4. Because these words are almost Chinese, we also attach the simple explanation in English. There are many non-related words in the LDA and mixture of unigram, such as “年終” (Year-end bonuses) and “不” (no). Especially, we compare the topic words in BTM with in PMI- β -BTM, the topic words in BTM contain some frequent but low-correlated words with the topic, such as “年” (means year) and “萬” (means ten thousand). While in the PMI- β -BTM, this noisy words do not appear. The reason is that the original BTM just consider the simple bi-term frequency and this bi-term frequency make some frequent words be extracted together with other words from the

³ <http://www.ettoday.net/news/news-list.htm>

⁴ <https://github.com/fxsjy/jieba>

document. Our PMI- β priors can decrease the probability of the common words by the word normalization effect in the PMI.

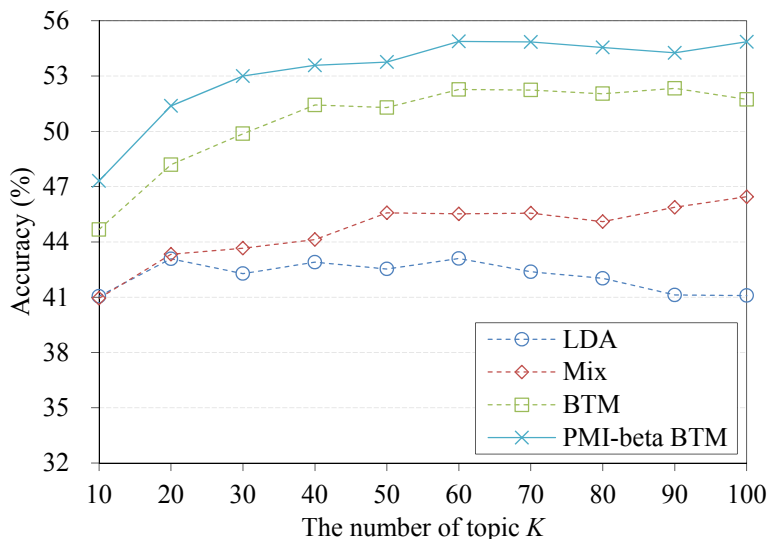


Figure 6. The Classification Results on ETtoday dataset

Table 4. The top-10 topic words of the “baseball” topic in ETtoday News Title dataset

	Top-10 Topic words
LDA	中職 (baseball game in Taiwan), 月 (month), 萬 (ten thousand), 年 (year), 大 (big), 元 (dollars), 吳誌揚 (a politician), 臺北 (Taipei), 臺灣 (Taiwan), 年終 (Year-end bonuses)
Mix	中職, 日 (day), 臺灣, 大, 英雄 (hero), 聯盟 (league baseball), 世界 (world), 棒球 (baseball), 不 (no), 挑戰 (challenge)
BTM	中職, 義大 (a baseball team), 兄弟 (a baseball team), MLB, 統一 (a baseball team), 年, 桃猿 (a baseball team), 萬, 獅 (a baseball team), 人 (human)
PMI- β -BTM	中職, MLB, 兄弟, 日職 (baseball game in Japan), 棒球, 桃猿, 先發 (Starting Pitcher), 總冠軍 (champion), 陳偉殷 (a Taiwanese professional baseball pitcher), 統一 (a baseball team)

5. Conclusions

In this paper, we propose a solution for topic model to enhance the amount of the word co-occurrence relation in the short text corpus. First, we find the BTM identifies the word co-occurrence by considering the bi-term frequency in the corpus-level. BTM will make the

common words be performed excessively because the frequency of bi-term comes from the whole corpus instead of a short document. We propose a PMI- β priors method to overcome this problem. The experimental results show our PMI- β -BTM get the best results in the regular short news title text.

Moreover, there are two advantages in our methods. We do not need any external data and the proposed two improvement of the word co-occurrence methods are both based on the original topic model and easy to extend. Bases on the original topic model means we did not modify the model itself, thus our methods can easily apply to some other existing BTM based models to overcome the short text problem without any modification. In the future, we can extend some other steps in PMI-priors to deal the further improvement, such as removing the redundant documents by clustering.

References

- Hofmann, T. (1999). Probabilistic latent semantic analysis. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, 289-296.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, 3, 993-1022.
- Divya, M., Thendral, K., & Chitrakala, S. (2013). A Survey on Topic Modeling. *International Journal of Recent Advances in Engineering & Technology (IJRAET)*, 1, 57-61.
- Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 262-272.
- Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A biterm topic model for short texts. In *Proceedings of the 22nd international conference on World Wide Web*, Rio de Janeiro, Brazil, 1445-1456.
- Nigam, K., McCallum, A. K., Thrun, S., & Mitchell, T. (2000). Text classification from labeled and unlabeled documents using EM. *Machine learning*, 39(2), 103-134.
- Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., *et al.* (2011). *Comparing twitter and traditional media using topic models*. In *Advances in Information Retrieval*. ed: Springer, 338-349.
- Cheng, X., Yan, X., Lan, Y., & Guo, J. (2014). BTM: Topic Modeling over Short Texts. *Knowledge and Data Engineering, IEEE Transactions on*, 26(12), 2928-2941.
- Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational linguistics*, 16(1), 22-29.
- Wallach, H. M., Mimno, D., & McCallum, A. (2009). Rethinking LDA: Why priors matter. In *Advances in Neural Information Processing Systems 22 (NIPS 2009)*.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes*, 25(2&3), 259-284.

- Wallach, H. M. (2006). Topic modeling: beyond bag-of-words. In *Proceedings of the 23rd international conference on Machine learning*, 977-984.
- Wang, X., McCallum, A., & Wei, X. (2007). Topical n-grams: Phrase and topic discovery, with an application to information retrieval. In *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, 697-702.
- Griffiths, T. L., Steyvers, M., Blei, D. M., & Tenenbaum, J. B. (2004). Integrating topics and syntax. In *Advances in neural information processing systems 17*, 537-544.
- Li, W. & McCallum, A. (2006). Pachinko allocation: DAG-structured mixture models of topic correlations. In *Proceedings of the 23rd international conference on Machine learning*, 577-584.
- Chen, Z. & Liu, B. (2014). Mining topics in documents: standing on the shoulders of big data. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, New York, USA, 1116-1125.
- Lin, C. X., Zhao, B., Mei, Q., & Han, J. (2010). PET: a statistical model for popular events tracking in social communities. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 929-938.
- Ramage, D., Dumais, S. T., & Liebling, D. J. (2010). Characterizing Microblogs with Topic Models. In *Fourth International AAAI Conference on Weblogs and Social Media*.
- Chen, J., Nairn, R., Nelson, L., Bernstein, M., & Chi, E. (2010). Short and tweet: experiments on recommending content from information streams. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1185-1194.
- Phelan, O., McCarthy, K., & Smyth, B. (2009). Using twitter to recommend real-time topical news. In *Proceedings of the third ACM conference on Recommender systems*, 385-388.
- Phan, X.-H., Nguyen, L.-M., & Horiguchi, S. (2008). Learning to classify short and sparse text & web with hidden topics from large-scale data collections. In *Proceedings of the 17th international conference on World Wide Web*, 91-100.
- Rosen-Zvi, M., Griffiths, T., Steyvers, M., & Smyth, P. (2004). The author-topic model for authors and documents. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, 487-494.
- Jin, O., Liu, N. N., Zhao, K., Yu, Y., & Yang, Q. (2011). Transferring topical knowledge from auxiliary long texts for short text clustering. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, 775-784.
- Hong, L. & Davison, B. D. (2010). Empirical study of topic modeling in twitter. In *Proceedings of the First Workshop on Social Media Analytics*, 80-88.
- Liu, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89(427), 958-966.
- Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336), 846-850.

- Ramage, D., Hall, D., Nallapati, R., & Manning, C. D. (2009). Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, 1*, 248-256.

節錄式語音文件摘要使用表示法學習技術

Extractive Spoken Document Summarization with Representation Learning Techniques

施凱文*、陳冠宇+、劉士弘+、王新民+、陳柏林*

Kai-Wun Shih, Kuan-Yu Chen, Shih-Hung Liu,

Hsin-Min Wang and Berlin Chen

摘要

大量多媒體內容的與日俱增促使自動語音文件摘要成為一項重要的研究議題。其中最為廣泛地被探究的是節錄式語音文件摘要(Extractive Spoken Document Summarization)，其目的是根據事先定義的摘要比例，從語音文件中選取一些重要的語句，用以代表原始語音文件的主旨或主題。另一方面，表示法學習(Representation Learning)是近期相當熱門的一個研究議題，多數的研究成果也證明了這項技術在許多自然語言處理(Natural Language Processing, NLP)的相關任務上，可以進一步地獲得優良的成效。有鑑於此，本論文主要探討使用詞表示法(Word Representations)及語句表示法(Sentence Representations)於節錄式中文廣播新聞語音文件摘要之應用。基於詞表示法及語句表示法，本論文提出三種新穎且有效的排序模型(Ranking Models)。除了文件中的文字資訊外，本論文更進一步地結合語音文件上的各式聲學特徵，如韻律特徵(Prosodic Features)等，期望可以獲得更好的摘要成效。本論文的語音文件摘要實驗語料

*國立臺灣師範大學資訊工程學系

Department of Computer Science & Information Engineering, National Taiwan Normal University

E-mail: {60247065S, berlin}@ntnu.edu.tw

+中央研究院資訊科學所

Institute of Information Science, Academia Sinica.

E-mail: {kychen, journey, whm}@iis.sinica.edu.tw

The author for correspondence is Berlin Chen.

是採用公視廣播新聞；實驗結果顯示，相較於其它現有的摘要方法，我們所發展的新穎式摘要方法能夠提供顯著的效能改善。

關鍵詞：語音文件、節錄式摘要、詞表示法、語句表示法、韻律特徵

Abstract

The rapidly increasing availability of multimedia associated with spoken documents on the Internet has prompted automatic spoken document summarization to be an important research subject. Thus far, the majority of existing work has focused on extractive spoken document summarization, which selects salient sentences from an original spoken document according to a target summarization ratio and concatenates them to form a summary concisely, in order to convey the most important theme of the document. On the other hand, there has been a surge of interest in developing representation learning techniques for a wide variety of natural language processing (NLP)-related tasks. However, to our knowledge, they are largely unexplored in the context of extractive spoken document summarization. With the above background, this study explores a novel use of both word and sentence representation techniques for extractive spoken document summarization. In addition, three variants of sentence ranking models building on top of such representation techniques are proposed. Furthermore, extra information cues like the prosodic features extracted from spoken documents, apart from the lexical features, are also employed for boosting the summarization performance. A series of experiments conducted on the MATBN broadcast news corpus indeed reveal the performance merits of our proposed summarization methods in relation to several state-of-the-art baselines.

Keywords: Spoken Document, Extractive Summarization, Word Representation, Sentence Representation, Prosodic Feature

1. 緒論

巨量資料充斥著現今的世界，在全球資訊網(World Wide Web)中已存在有數十億篇網頁，並且以指數的倍數持續成長著。為此，人們必須仰賴及時摘要各類資訊的自動化工具，以減緩資訊過載(Information Overload)的問題。這些迫切的需求促使了自動摘要(Automatic Summarization)技術的蓬勃發展(Luhn, 1958)。自動摘要技術可概分為節錄式(Extractive)摘要以及抽象式(Abstractive)摘要。前者主要是依據特定的摘要比例，從原始的文件中選取重要的語句子集(Sentence Subset)，透過該語句子集簡潔地表示原始文件的大致內容；而後者是在完全理解文件內容之後，重新撰寫產生摘要來代表原始文件的內容。雖然抽象式摘要是最為貼近人們日常撰寫摘要的形式，但其涉及深層的自然語言處

理能力(Mitra *et al.*, 1997)，較為困難許多；目前大多數的研究主要集中在節錄式摘要的自動產生(Jones, 1999)。除了傳統的文字文件外，多媒體文件亦迅速地在世界各地傳播，例如語音郵件、會議錄音、電視新聞以及課程演講等；因此，語音文件摘要(Spoken Document Summarization)自然地成為近年來的一項受關注的研究議題。本論文主要探討節錄式語音文件摘要，其目標是基於定量的摘要比例(Summarization Ratio)，從多媒體內容所對應的語音文件中選取能夠表示其內容主題資訊之語句，讓使用者可以迅速地理解多媒體內容的主要意涵。

本論文延續過去學者的經驗、成果以及貢獻，針對語音文件的特性及困難點，進一步地深入研究語音文件摘要方法與特徵使用，希望藉由自動語音文件摘要技術的改進，使自動摘要的結果能更適切地詮釋語音文件內容及主題。本論文提出兩個主要的研究貢獻：(1)基於表示法學習(Representation Learning)技術，本論文提出三種排序模型(即統計式模型、機率式模型以及圖論式模型)，嘗試將表示法學習技術運用於語音文件摘要任務之中。(2)除了利用文件中的文字資訊外，本論文進一步地結合語音文件中的各種韻律特徵，期望增進摘要系統的成效。

本論文的後續安排如下：第二節首先簡介當前主要的文件摘要方法。第三節介紹本論文所探究的表示法學習技術。第四節介紹本論文提出的三種排序模型於表示法學習之技術。第五節簡介語音文件的多種特徵。第六節介紹實驗語料及摘要評估之方法。第七節說明實驗結果及其分析。第八節為本論文之結論與未來展望。

2. 基礎文件摘要方法之簡介

2.1 以文件結構為基礎之摘要方法

摘要單位元(例如詞彙或語句)在文件中的位置資訊(Positional Information)可以作為文件尋找重要語句選取的判斷依據。一般而言，位於段落開頭的摘要單位元通常有較高的重要性及代表性。因此，前導(LEAD)摘要方法是根據摘要比例選取文件前 $M\%$ 的部分作為摘要(Hajime & Manabu, 2000)。這種摘要的方式簡單且直覺，但僅適用於特定的結構內容，並且文件需遵循某種編排方式。對於某些沒有特定結構編排的文件，或是缺乏文件結構的語音文件內容有可能會不適用。

2.2 以統計值為基礎之摘要方法

A. 向量空間模型(Vector Space Model, VSM)

向量空間模型已被廣泛地應用於資訊檢索中，用以估測使用者查詢(Query)與文件(Document)之間的相關程度(Salton & Lesk, 1968)。節錄式語音文件摘要任務可以被視為是一個資訊檢索的問題，重要語句的選取是以句子與語音文件內容的相關程度而定；亦即將文件內容視為查詢來檢索最相關的 m 個語句。因此，可以將文件表示成向量 \vec{D} ，文件中的每一語句表示為向量 \vec{S}_i ，透過餘弦相似度(Cosine Similarity)計算，就可估測兩者

之間的相似性程度。雖然直接地利用文件中的文字資訊已被證明在許多自然語言處理的相關問題中可以獲得一定的成效，但這樣的方法忽略了隱藏於文字中的語意資訊 (Semantic Information)。為了有效地運用這些語意資訊，最早於 2001 年開始有學者提出潛藏語意的概念應用於文件摘要的方法(Gong & Liu, 2001)。

B. 最大邊際關聯法(Maximum Margin Relevance, MMR)

最大邊際關聯法是於 1998 年被提出的自動摘要準則(Carbonell & Goldstein, 1998)。該準則是以遞迴的方式一次一句地挑選可能的摘要語句，其考量的重點不僅是希望所選取出來的摘要語句與文件的關聯性分數要高，更進一步地考慮了候選語句與已被選取的摘要語句之間的重複性分數要低。如此一來，摘要系統選取出的語句不僅可以代表文件中的重要主題，更可以充分的涵蓋文件中的各個面向。直到今日，最大邊際關聯法仍為節錄式摘要方法常使用的重要準則；它也常是一個主要的基礎系統，做為新摘要方法發展時效能比較與評估的依據。

2.3 以機率式模型為基礎之摘要方法

A. 單連語言模型(Unigram Language Model, ULM)

語言模型被廣泛地應用於語音辨識(Speech Recognition)與機器翻譯(Machine Translation)等方面，學者 Ponte 等人在 1998 年時將其運用於資訊檢索的問題中(Ponte & Croft, 1998)。由於我們可以將節錄式語音文件摘要任務視為一個資訊檢索的問題，即當給予一篇文件 D 時，希望對文件中的語句 S 依照機率值 $P(S|D)$ 進行排序；藉由貝式定理的推導，可得 (Chen *et al.*, 2009)：

$$P(S|D) = \frac{P(D|S)P(S)}{P(D)} \propto P(D|S) \quad (1)$$

其中 $P(D)$ 對於每一語句皆相同，故可忽略。而我們假設每一語句 S 的事前機率 $P(S)$ 為一個均勻分佈(Uniform Distribution)，因此 $P(S)$ 亦可忽略。值得一提的是，由於文件中的語句通常較為簡短，不容易建立一個準確的模型來完整地描述每一語句的內容涵意。為此，有研究學者陸續提出各式較為強健性的語言模型，例如關聯模型 (Relevance Model)(Lavrenko & Croft, 2001)等，期望可以改善此一問題。關聯模型的優點在於融入關聯文件中的資訊，藉此豐富語句模型使得更準確地表達語句的主題特性，以提升摘要的成效。

B. Okapi Best Match 25 (BM25)

Okapi BM25 是於 1994 年由學者 Robertson 等人所提出的權重計算公式，是現今資訊檢索模型中最著名的機率式檢索模型之一。其權重計算方式主要是將詞頻對文件長度作正規化，有效降低因文件長度不同而產生的檢索誤差(Robertson & Jones, 1976; Robertson &

Walker, 1994; Robertson *et al.*, 1996)。當利用該方法於文件摘要任務時，我們首先對文件的詞序列 $D = (w_1 w_2 \dots w_{|d|})$ 計算出每個詞 w_i 與語句 S 之間的相似性分數，接著將每個詞 w_i 對於語句 S 的相似性分數進行加權求和，進而得到文件 D 與語句 S 的相似性分數，公式如下：

$$BM25(D, S) = \sum_{w \in D} F(w, D) \cdot Sim(w, S) \cdot \log \frac{N}{1 + df_w} \quad (2)$$

$$F(w, D) = \frac{c(w, D) \cdot (k_2 + 1)}{c(w, D) + k_2} \quad (3)$$

$$Sim(w, S) = \frac{c(w, S) \cdot (k_1 + 1)}{c(w, S) + k_1 \cdot \left(1 - b + b \cdot \frac{|S|}{|avgS|}\right)} \quad (4)$$

其中 k_1 、 k_2 以及 b 均為自由參數，根據經驗設置，一般 $k_1 \in [1.2, 2.0]$ 、 $b = 0.75$ ； $c(w, S)$ 是詞 w 在語句 S 中出現的次數； $c(w, D)$ 是詞 w 在文件 D 中出現的次數；而 $|S|$ 是表示語句 S 的長度， $|avgS|$ 是在文件中所有語句的平均長度； N 是在集合中的文件總數； df_w 為在集合中文件包含詞 w 的篇數。

2.4 以圖論為基礎摘要之方法

A. 詞權重-逆向文件頻率(Term Weight-Inverse Document Frequency, TW-IDF)

詞權重-逆向文件頻率模型是由學者 Rousseau 與 Vazirgiannis 於 2013 年所提出(Rousseau & Vazirgiannis, 2013)。首先，此方法為每一篇文件建立一個有向圖(Directed Graph)，圖中的每一個頂點(Vertex)代表文件中的一個唯獨詞(Unique Word)。如果任兩個詞在文件中曾經相鄰出現，則此兩個頂點可以用每一個邊(Edge)相連，邊的方向表示這兩個詞出現時的先後次序。最後，統計有向圖中每一個頂點的內分支度(In-degree)個數，並與 BM25 模型相結合，即可求得文件與語句間的關聯程度。相較於大多數現存的摘要模型(例如 TF-IDF 與 BM25)，僅考慮每一個詞出現的頻率，詞權重-逆向文件頻率模型基於內分支度個數，重新賦予每一個詞一個權重，進一步地考慮了文字間在文件中的先後次序關係。

B. 馬可夫隨機漫步(Markov Random Walk, MRW)

馬可夫隨機漫步模型的概念是將文件視為一個網際網路，文件中的每一語句代表網路上的一個節點(Node)，而語句之間的相關程度則為節點間邊界(Edge)的權重(Wan & Yang, 2008)。馬可夫隨機漫步模型提出一套遞迴更新的演算法，利用節點間邊界的權重關係不斷地重複更新節點的重要性，最終獲得每一語句的重要性分數。更明確地，語句 S_i 的重要性分數 $SenScore(S_i)$ 是由相鄰的語句 S_j 分數的線性組合而得，我們可以將語句間邊界的權重關係以一個矩陣 $\tilde{M} = (\tilde{M}_{i,j})_{|D| \times |D|}$ 表示之，則馬可夫隨機漫步模型可以表示為：

$$\tilde{M}_{i,j} = \begin{cases} \frac{\text{sim}(S_i, S_j)}{\sum_{k=1}^{|D|} \text{sim}(S_i, S_k)} & , \text{if } \sum_{k=1}^{|D|} \text{sim}(S_i, S_k) \neq 0 \\ 0 & , \text{otherwise} \end{cases} \quad (5)$$

$$\text{SenScore}(S_i) = \mu \cdot \sum_{j \neq i} \text{SenScore}(S_j) \cdot \tilde{M}_{j,i} + \frac{(1 - \mu)}{|V|} \quad (6)$$

其中 $|D|$ 為文件 D 中語句的個數， $\text{sim}(\cdot, \cdot)$ 為相似度函數，用以計算兩個語句之間的相似程度。

3. 表示法學習(Representation Learning)

3.1 詞表示法(Word Representation)

當一種自然語言處理的問題要轉化為機器學習的問題，首先需要找到一種方法將這些語言符號數學化。傳統的自然語言處理中最直觀的方式是採用 **One-hot** 表示法，即每一個詞皆以一個 K 維(通常 K 為詞彙的大小)的向量表示之，而此向量中僅有某一個維度為 1，其餘為零。明顯地，此種表示法中任意兩個詞之間彼此互相獨立，意即我們無法計算出任兩個詞之間的相似程度。為了解決上述問題，學者 Hinton 首先於 1986 年提出了一種分散式表示法(Distributed Representation)模型(Hinton, 1986)，藉由訓練將每一個詞重新以一個較低維度的實數向量表示之，透過這個低維度的向量表示法，詞與詞之間的關係可以簡單地透過距離公式(如餘弦、歐式距離)來計算，並依此判斷詞與詞之間語意的相近程度。學者 Bengio 等人於 2003 年提出以前饋式類神經網路(Feed-Forward Neural Network, FFNN)來建立語言模型，在語言模型的建構過程中，每一個詞即會獲得一個低維度的實數向量表示法(Bengio *et al.*, 2003)。Google 於 2013 年開發出一套詞表示法工具 word2vec，當中包含連續型詞袋模型(Continuous Bag-of-Words, CBOW)(Mikolov *et al.*, 2013a)與跳躍式模型(Skip-Gram, SG)(Mikolov *et al.*, 2013b)。據我們所知，這些方法雖然已被使用來解決許多自然語言相關的問題，但卻鮮少被應用於語音文件摘要的任務之中。

A. 連續型詞袋模型(Continuous Bag-of-Words, CBOW)

連續型詞袋模型(CBOW)是由 Mikolov 所提出的兩個經典架構之一，該模型是設法直接地獲得每個詞的向量表示，而不是尋求學習一個統計語言模型(Mikolov *et al.*, 2013a)。CBOW 的架構類似於前饋式類神經網路(Feed-Forward Neural Network)，不同之處在於(1)CBOW 移除非線性隱藏層(Non-Linear Hidden Layer)。如此一來，大大的降低了類神經網路模型訓練時間過長的問題，實驗結果顯示，簡化的模型在許多應用中，依然保有優異的性能。(2)每個詞皆共享投影層(Projection Layer)，因此所有的詞皆會投影至相同的位置(即向量相加)，這樣的架構使得詞序列不會影響投影的結果。更明確地，CBOW 的訓練目標是在給定一個詞的上下文(Context)後，期望可以準確地預測該詞的出現，其圖形表示如圖 1(a)所示。該模型不同於傳統的詞袋模型，它是使用連續分散式表示法

(Continuous Distributed Representation)。形式上，給定一詞序列 $w_1 w_2 \dots w_T$ ，CBOW 的目標函數(Objective Function)是要最大化對數機率(Log-Probability)：

$$\sum_{t=1}^T \log P(w^t | w^{t-c}, \dots, w^{t-1}, w^{t+1}, \dots, w^{t+c}) \quad (7)$$

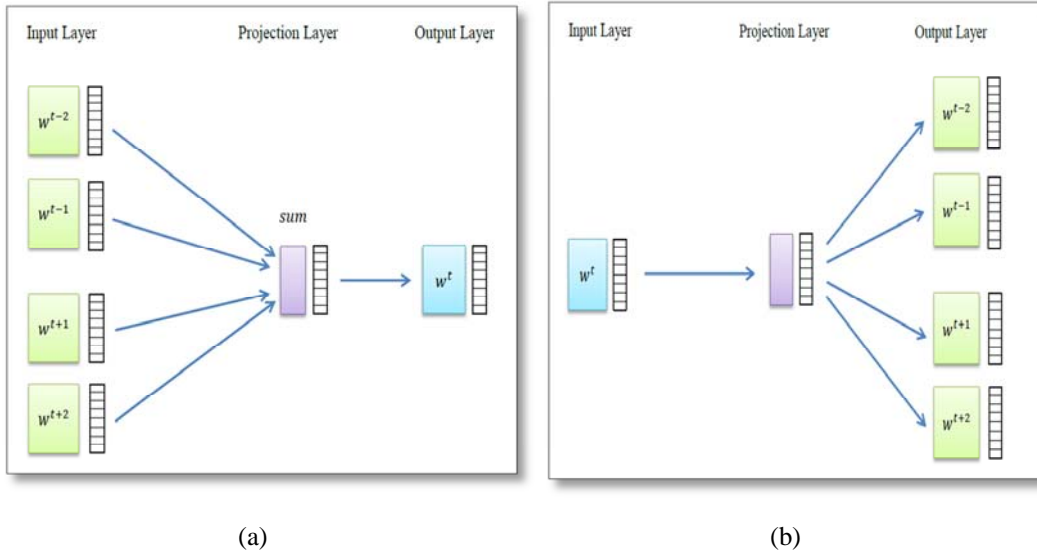


圖 1 (a). 連續型詞袋模型之示意圖

(b). 跳躍式模型之示意圖

其中 c 為中間詞 w_t 的上下文之窗口大小(Window Size)， T 代表訓練語料的長度，且

$$P(w^t | w^{t-c}, \dots, w^{t-1}, w^{t+1}, \dots, w^{t+c}) = \frac{\exp(v_{\bar{w}^t} \cdot v_{w^t})}{\sum_{i=1}^V \exp(v_{\bar{w}^t} \cdot v_{w_i})} \quad (8)$$

其中 v_{w^t} 為位置 t 詞 w 的詞表示法， V 是詞彙的大小， $v_{\bar{w}^t}$ 代表 w^t 的上下文詞表示法之加權 (Qiu *et al.*, 2014)。CBOW 的概念是透過一分佈式假設(Miller & Charles, 1991)，此假設指出具有類似語意的詞會經常出現於類似的上下文之中，因此建議找出可以很好地獲取上下文分佈的 w^t 詞表示法。

B. 跳躍式模型(Skip-Gram, SG)

跳躍式模型(SG)是由學者 Mikolov 等人於 2013 年時所提出的另一經典架構(Mikolov *et al.*, 2013b)，該模型同樣以簡化的前饋類神經網路來學習詞表示法。更明確地，跳躍式模型與連續型詞袋模型的模型訓練目標恰好相反，跳躍式模型是希望在給定一個詞 w 後，可以準確地預測其上下文中，詞出現的可能性。訓練的過程中，該模型是使用每一當前詞做為對數線性分類器(Log-Linear Classifier)的輸入，並預測此當前詞一定範圍內的前後的詞，其圖形表示如圖 1(b)所示。當給定一詞序列 $w_1 w_2 \dots w_T$ 後，SG 的目標函數是要最大化對數機率：

$$\sum_{t=1}^T \sum_{j=-c, j \neq 0}^c \log P(w^{t+j}|w^t) \quad (9)$$

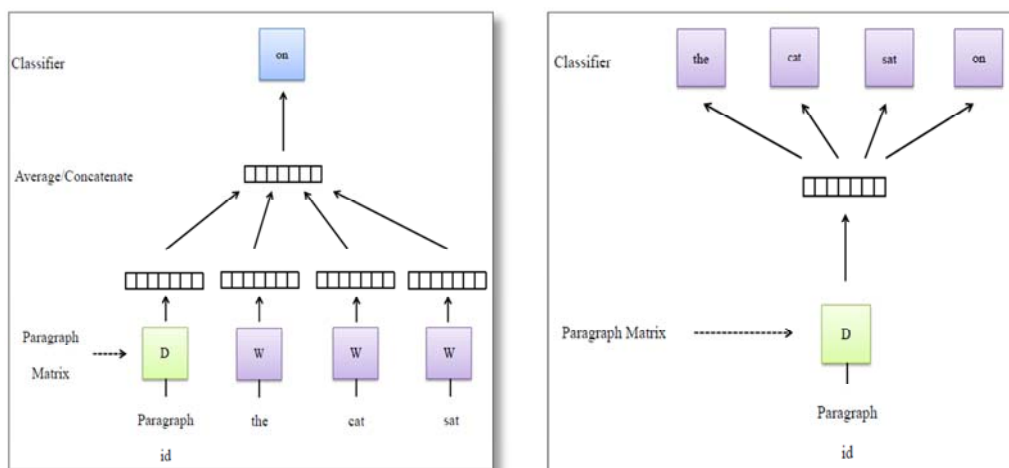
其中 c 為中間詞 w_t 的上下文之窗口大小(Window Size)，而條件機率(Conditional Probability)經由下式計算：

$$P(w^{t+j}|w^t) = \frac{\exp(v_{w^{t+j}} \cdot v_{w^t})}{\sum_{i=1}^V \exp(v_{w_i} \cdot v_{w^t})} \quad (10)$$

其中 w^{t+j} 與 w^t 分別為位置 $t+j$ 及 t 的詞表示法。在 CBOW 與 SG 的實作中皆引入階層軟式最大化法(Mikolov *et al.*, 2013b; Morin & Bengio, 2005)及負例採樣法(Mikolov *et al.*, 2013b; Mnih & Kavukcuoglu, 2013)，以增進訓練過程中參數估測的效能。

3.2 語句表示法(Sentence Representation)

雖然詞表示法已被廣泛使用，但許多自然語言處理的相關任務所需要的是語句的表示法。延續詞表示法的基本模型架構與精神，學者 Le 與 Mikolov 提出兩種學習語句表示法的模型，分別是分散式儲存模型與分散式詞袋模型(Le & Mikolov, 2014)。



(a)

圖 2 (a). 分散式儲存模型之示意圖

(b)

圖 2 (b). 分散式詞袋模型之示意圖

A. 分散式儲存模型(Distributed Memory Model of Paragraph Vector, PV-DM)

分散式儲存模型(PV-DM)類似於連續型詞袋模型。PV-DM 同樣以最大化目標中間詞輸出的機率為目標，其主要差異為：(1)訓練過程中於輸入層(Input Layer)引入一個段落編號(Paragraph ID)，亦即訓練語料中每一語句皆有一個唯一的段落編號。段落編號與一般的

詞相同，亦是先映射成一個向量，即段落向量(Paragraph Vector)。然而段落向量與詞向量的維度相同。在往後的計算中，將詞向量與段落向量串聯作為輸出層軟式最大化法(Soft-max)的輸入。在一個語句或是文件的訓練過程中，段落編號會保持不變，共享相同的段落向量，相當於每次在預測一個詞的機率時，皆利用了整個語句的語意。(2)在預測階段時，給予待預測的語句分配一個新的段落編號，保持詞向量與輸出層軟式最大化於訓練階段所得之參數，重新利用隨機梯度法訓練待預測語句，待收斂完畢後，即得到待預測語句的段落向量，其圖形表示如圖 2(a)所示。

B. 分散式詞袋模型(Distributed Bag-of-Words of Paragraph Vector, PV-DBOW)

分散式儲存模型(PV-DM)是採用詞向量與段落向量的平均或是串聯，進行預測一個詞。分散式詞袋模型(PV-DBOW)則是以段落向量作為輸入，從該向量對應的段落中隨機採樣詞序列作為輸出，該方法減少了輸入層的參數量。類似於跳躍式模型，其圖形表示法如圖 2(b)所示。該模型的概念簡單且僅需少量儲存空間(詞向量與輸出層軟式最大化法於訓練階段所得之參數)。

4. 運用表示法學習於語音文件摘要

4.1 餘弦相似度(Cosine Similarity)

由於向量空間模型簡單、直觀且有效，因此被廣泛地應用於各式自然語言處理的相關研究。藉助於詞表示法模型(例如 CBOW 與 SG)我們可以將文件或語句中所有詞所對應的詞表示法加總後取平均，作為該篇文件或語句的表示法：

$$v_D = \frac{\sum_{w \in D} v_w}{|D|}, \quad v_S = \frac{\sum_{w \in S} v_w}{|S|} \quad (11)$$

其中 v_w 為詞 w 的詞表示法， v_D 、 v_S 為代表文件 D 與語句 S 的表示法， $|D|$ 、 $|S|$ 為文件 D 及語句 S 長度。或是直接藉由語句表示法模型求得文件或語句的向量表示法：

$$v_D = PV_D, \quad v_S = PV_S \quad (12)$$

如此一來，文件 D 及其語句 S 皆有一固定長度的向量表示，其相關性就可藉由餘弦相似度計算而得：

$$Sim(S, D) = \frac{v_S \cdot v_D}{\|v_S\| \cdot \|v_D\|} \quad (13)$$

4.2 馬可夫隨機漫步(Markov Random Walk, MRW)

當結合各式詞與語句表示法模型於馬可夫隨機漫步模型中時，我們首先將文件與語句表示成一個固定維度的向量表示法。接著，我們使用餘弦相似度估測來計算兩兩語句間的相似程度，再透過馬可夫隨機漫步模型所提出的遞迴更新演算法，就可以求得每一語句

的重要性分數。最後，將語句依此分數遞減的方式排列後，根據事先定義的摘要比例來依序挑選語句並作為最後的輸出。

4.3 文件相似度量值(Document Likelihood Measure, DLM)

我們亦可以運用語言模型(LM)為基礎的方法於節錄式語音文件摘要，其實現的方式是計算文件被每一個語句模型生成的可能性 $P(D|S)$ ，並依此進行語句重要性之排序。利用詞表示法，我們首先定義一個以詞為基礎的語言模型。在給定一個詞 w_i 後，詞 w_j 出現機率為：

$$P(w_j|w_i) = \frac{\exp(v_{w_j} \cdot v_{w_i})}{\sum_{w_k \in V} \exp(v_{w_k} \cdot v_{w_i})} \quad (14)$$

接著，透過線性組合(Linear Combination)的方式，可以形成一個複合式的語句語言模型，而文件的生成機率就可以經由下式計算：

$$P(D|S) = \prod_{w_j \in D} \left[\lambda \cdot \sum_{w_i \in S} P(w_i|S) \cdot P(w_j|w_i) + (1 - \lambda) \cdot P(w_j|C) \right]^{c(w_j, D)} \quad (15)$$

其中 $P(w_i|S)$ 為一個權重係數，代表詞 w_i 出現在語句中的出現的可能性；並且，為了解決資料稀疏的問題，我們透過背景語言模型 $P(w_j|C)$ 對語句模型進行機率平滑化。另一方面，當使用語句表示法時，我們首先為每一個語句 S 建構出一個以語句表示法為基礎的語言模型，用以預測一個詞 w_j 發生的可能性：

$$P(w_j|S) = \frac{\exp(v_{w_j} \cdot v_S)}{\sum_{w_k \in V} \exp(v_{w_k} \cdot v_S)} \quad (16)$$

其中 v_S 是以 PV-DBOW 或 PV-DM 所求得的語句表示法。同樣地，文件的生成機率就可以經由下式計算：

$$P(D|S) = \prod_{w_j \in D} [\lambda \cdot P(w_j|S) + (1 - \lambda) \cdot P(w_j|C)]^{c(w_j, D)} \quad (17)$$

5. 語音文件之各種特徵簡介

文字文件內容除了提供文字訊息作為重要語句選取依據之外，語句中更包含文法(Grammar)、語意(Semantic)以及結構(Structure)等資訊，皆可視為重要的特徵。不同於文字文件，語音文件內容可能因辨識錯誤或語句邊界定義等問題，使得語句文法、語意以及結構等資訊相對較缺乏，但語音文件卻含有豐富的韻律特徵(Prosodic Features)，如語者在說話時發音的長短快慢、語氣的抑揚頓挫以及高低起伏等。因此若將這些語言學、聲韻學以及文件結構等資訊加以善用，相信有助於提升節錄式語音文件摘要的效能。

5.1 韻律特徵(Prosodic Features)

A. 音高(Pitch)

一般語者在敘述一件事情時，會以說話的高低起伏、抑揚頓挫來強調說話的內容以吸引聽者的注意，語者表達自身的感覺使得對方接受到強調的訊息，因此音高可視為一種語音中重要的資訊。

B. 能量(Energy)

能量可用來表示語者說話音量的大小，經常被視為一種可利用的重要資訊。一般語者在特別強調一件事情或是敘述重點時，會刻意地提高音量來表示強調關鍵字或是說話的內容以希望引起聽者的注意。

C. 音框長度(Duration)

類似於語句長度，語句越長所包含的資訊越多，而語句的音框長度代表語者說該語句的時間長度，因此說話時間越長的語句其包含的資訊亦越多。

D. 頻譜峰(Peak)與共振峰(Formant)

共振峰被定義為“聲譜中的頻譜峰”，差異在於母音(Vowel)有共振峰的結構，在母音發音較為清楚的音節(Syllable)，共振峰會較高。共振峰是用來描述聲學共振現象的一種概念，是決定語者特徵的主要因素。在有效頻寬範圍中會有約五個共振峰，從低頻率至高頻率依序排列為第一共振峰(F1)、第二共振峰(F2)、第三共振峰(F3)、第四共振峰(F4)以及第五共振峰(F5)，而通常以 F1、F2、F3 較為明顯，因此通常以這三個共振峰為代表。若語者在表達某語句較為字正腔圓，希望聽眾可以聽得清楚時，該語句可能為重要語句，共振峰整體來說會較高；若是語者所含糊帶過的語句則可能為非重要語句，其共振峰整體來說會較低。

5.2 詞彙特徵(Lexical Features)

A. 雙連語言模型分數(Bigram Language Model Score)

N 連語言模型(N -gram Language Model)是自然語言處理常用到的方法，其假設第 N 個詞的出現僅與前面 $N-1$ 個詞相關，模型參數則通常藉由最大化相似度估測(MLE)來求得。對一個語句的重要性估測是透過計算在語句中所出現的詞的條件機率之乘積，通常採用二連(Bigram)與三連(Trigram)語言模型。

B. 正規化雙連語言模型分數(Normalized Bigram Language Model Score)

為了避免在計算時因語句長度的影響，透過該語句長度將其雙連語言模型分數進行正規

化並做為另一項特徵。

C. 專有名詞(Named Entities)個數

根據專有名詞詞典(Lexicon)計算語句中的詞與專有名詞詞典重複的數量；其主要想法是含括愈多專有名詞的語句愈可能為重要語句。而專有名詞則包含公司名稱、地點、人名以及時間等。

D. 停用詞(Stop Words)個數

計算語句中所包含停用詞的數量，如中文詞的“了”、“的”等詞，以及英文詞如“a”、“the”等詞，即使出現的頻率很高，但通常不具有太多資訊，因此在檢索過程中經常被濾除，不列入搜尋的考慮範圍。

5.3 關聯特徵(Relevance Features)

通常為來自不同文件摘要模型所產生的摘要特徵分數，如以統計值為基礎的向量空間模型(Vector Space Model, VSM)、以圖論為基礎的馬可夫隨機漫步(Markov Random Walk, MRW)以及以機率生成模型為基礎的語言模型(Language Model, LM)等。

6. 實驗語料及評估方法

6.1 實驗語料

本論文實驗語料為公視新聞語料(Mandarin Chinese Broadcast News Corpus, MATBN)，由中央研究院資訊所與公共電視台合作錄製整理，其錄製內容為每天一個小時的公視晚間新聞深度報導(Wang *et al.*, 2005)。我們選取其中從 2001 年 11 月至 2002 年 8 月共 205 篇的新聞報導，並區分為發展集(185 篇)與測試集(20 篇)兩個部分。全部 205 篇語音文件長度約為 7.5 個小時。我們將語音文件進行人工切音處理，得到真正含有講話內容的音訊段落，再透過自動語音辨識系統進行轉寫，我們稱之為語音文件(Spoken Document, SD)，含有語音辨識錯誤與語句邊界偵測錯誤。此外，我們亦將此 205 篇語音文件透過人工聽寫的方式產生出沒有辨識錯誤的對應文字內容，我們稱之為文字文件(Text Document, TD)。每一篇文字文件皆有三位標記專家所提供的三份摘要結果，我們將此作為語音文件與文字文件的正確摘要答案。透過比較語音文件和文字文件的摘要效能，我們可以觀察語音辨識錯誤對於各種摘要方法的影響。本研究的背景語言模型訓練語料取材自 2001 至 2002 年的中央社新聞文字語料(Central News Agency, CNA)，並且以 SRI 語言模型工具訓練出經平滑化的單連語言模型。此外，本論文蒐集 2002 年中央通訊社的 101,268 篇同時期新聞文件作為詞表示法以及語句表示法的訓練語料以及虛擬關聯文件。我們設定摘要比例為 10%，其定義是摘要字數占整篇文件字數的比例，其詳細的統計資訊如表 1 所示。

表1. 廣播新聞文件之統計資訊

	訓練集	測試集
紀錄時段	2001/11/07-2002/08/22	2002/01/24-2002/08/20
文件個數	185	20
文件平均持續秒數	129.4	141.3
文件平均詞個數	326.0	290.3
文件平均語句個數	20.0	23.3
文件平均詞錯誤率	38.0%	39.4%

6.2 評估方法

本論文採用 ROUGE 作為文件摘要的評估方式。該方法是計算自動摘要結果與人工摘要之間的重疊單位元(Overlap Units)數目占人工摘要長度的比例。由於該方法是採用單位元比對的方式，不會產生語句邊界定義的問題，且適合用於多份人工摘要的評估。我們使用了較普遍的 ROUGE-1(Unigram)、ROUGE-2(Bigram)以及 ROUGE-L(Longest Common Subsequence, LCS)分數，其中 ROUGE-1 是評估自動摘要的訊息量，ROUGE-2 是評估自動摘要的流暢性，ROUGE-L 是最長共同字串。ROUGE- N 是自動摘要和人工摘要之間 N 連詞(N -gram)的召回率，人工標記的參考摘要為一集合 R ，故 ROUGE- N 計算公式如下(Lin, 2004)：

$$\text{ROUGE} - N = \frac{\sum_{sum \in R} \sum_{gram_N \in sum} \text{Count}_{match}(gram_N)}{\sum_{sum \in R} \sum_{gram_N \in sum} \text{Count}(gram_N)} \quad (18)$$

其中 sum 為人工摘要集合 R 中的任一個摘要， N 代表詞彙串之連續長度，而 $\text{Count}(gram_N)$ 是 N 連詞同時出現於自動摘要與人工摘要的最大數量。ROUGE-L 的計算方式與 ROUGE- N 相似，但前者僅考慮自動摘要與參考摘要的最長共同字串。

7. 實驗結果

7.1 基礎文件摘要之實驗結果

表 2 為測試集中的文字文件(TD)與語音文件(SD)在 ROUGE-1、ROUGE-2 以及 ROUGE-L 評估下的摘要結果；在此我們進行各式的基礎摘要方法的比較，包含前導方法(LEAD)、向量空間模型(VSM)、最大邊際關聯法(MMR)、潛藏語意分析(LSA)、單連語言模型(ULM)、關聯模型(RM)、Okapi Best Match 25(BM25)、詞權重-逆向文件頻率(TW-IDF)以及馬可夫隨機漫步(MRW)。首先在 TD 的實驗中，RM 的摘要效果是所有模型中最佳的，表示使用額外的關聯文件可以有效地彌補語句內容的不足，提高語句的估測能力。其次為 BM25，我們認為在文件摘要的問題中，詞彙的頻率(TF)、反文件頻率(IDF)以及文件長度的正規化(Normalized)是重要且不可或缺的特徵資訊。ULM 無論在 TD 或是 SD 上的摘要成效皆

優於圖論式模型 TW-IDF 與 MRW。TW-IDF 在計算詞頻(TF)時，多考慮了上下文(Context)的資訊，而 MRW 在計算重要語句時，除了使用其它語句的分數之外，亦考慮到語句彼此之間的相關度作為權重來調整，因此兩者效果皆會較僅考慮詞頻的 VSM 為佳。MMR 在進行語句選取時多考慮了冗餘資訊，因此摘要效果較 VSM 佳。

表 2. 基礎實驗於文字文件與語音文件之摘要結果

方法	文字文件 (TD)			語音文件 (SD)		
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
LEAD	0.312	0.196	0.278	0.254	0.117	0.220
VSM	0.347	0.228	0.290	0.343	0.189	0.288
MMR	0.365	0.242	0.316	0.360	0.206	0.309
LSA	0.362	0.233	0.316	0.345	0.201	0.301
ULM	0.411	0.299	0.362	0.364	0.218	0.313
RM	0.458	0.345	0.408	0.384	0.236	0.330
BM25	0.422	0.317	0.380	0.394	0.251	0.341
TW-IDF	0.374	0.260	0.317	0.322	0.164	0.270
MRW	0.415	0.296	0.357	0.339	0.194	0.289

LSA 在潛藏語意空間計算文件與語句的餘弦相似度，其結果亦顯示較 VSM 為佳。而 VSM 每個詞彙所構成的向量維度皆為獨立，因此無法得知出文件中詞彙之間的關聯性，使得進行文件相似度的比對時可能造成誤判的情況。

在 SD 的實驗中，BM25 反而超越 RM 成為所有模型中最佳的摘要方法，我們認為這可能是因為 RM 中所使用的語句模型受到語音辨識錯誤的影響，因此降低尋找有效的虛擬關聯文件(Pseudo Relevant Documents)的能力。此外，TW-IDF 與 MRW 的摘要效能皆較 LSA 及 MMR 差，我們認為亦是受到語音辨識錯誤的影響，因一個詞或是一個語句的重要性分數是來自鄰近其它詞或是語句的貢獻。而 LEAD 無論在 TD 或是 SD 上，相較於其它模型皆得到較差的效果，主要原因是 LEAD 僅適用於特殊文件結構，因此若摘要文件不具有某種特殊的結構，其摘要效能就會有所侷限。

7.2 詞表示法與語句表示法於節錄式語音文件摘要之實驗結果

在此我們利用目前兩種最先進的詞表示法—連續型詞袋模型(CBOW)和跳躍式模型(SG)，與最先進的兩種語句表示法—分散式儲存模型(PV-DM) 和分散式詞袋模型(PV-DBOW)之技術來從事語音文件摘要；實驗共分三組來進行，分別結合於餘弦相似度(Cosine Similarity)、馬可夫隨機漫步(MRW)以及文件相似度量值(DLM)的方法作為挑選摘要語句之依據。

表3. 詞表示法結合於餘弦相似度之摘要結果

方法	文字文件(TD)			語音文件(SD)		
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
CBOW	0.402	0.280	0.349	0.377	0.228	0.327
SG	0.401	0.265	0.347	0.361	0.214	0.312

首先，我們將詞表示法結合於餘弦相似度(Cosine Similarity)作為選取摘要語句的方法，其結果示於表 3。從實驗結果中觀察到，由於這兩種詞表示法各有著不同的模型結構與學習方式，因此在文字文件(TD)或是語音文件(SD)中，該兩種模型的摘要成效有稍微的差異。根據 TD 的結果顯示，CBOW 的摘要效能較 SG 佳，在 SD 中仍保持相同的情況。儘管該兩種詞表示法皆優於向量空間模型(VSM)與潛藏語意分析(LSA)，卻僅達到詞權重-逆向文件頻率(TW-IDF)差不多的水平，而且在 SD 的情況下的表現 SG 不及單連語言模型(ULM)(表 2)。

表4. 語句表示法結合於餘弦相似度之摘要結果

方法	文字文件(TD)			語音文件(SD)		
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
PV-DM	0.429	0.313	0.382	0.387	0.236	0.335
PV-DBOW	0.398	0.277	0.348	0.368	0.227	0.329

同樣地，我們將語句表示法結合於餘弦相似度作為選取摘要語句的方法，其結果示於表 4。在 TD 的結果中，PV-DM 與 PV-DBOW 該兩種語句表示法的摘要效果分別超越 CBOW 及 SG 詞表示法模型(表 3)。PV-DM 摘要成效較傳統的馬可夫隨機漫步(MRW)佳，但較 BM25 差。而在 SD 的結果中，兩種語句表示法的摘要成效比起詞表示法沒有太大的進步，我們認為語句表示法搭配餘弦相似度選取語句的方式亦受語音辨識的影響。

表5. 詞表示法結合於馬可夫隨機漫步之摘要結果

方法	文字文件(TD)			語音文件(SD)		
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
CBOW	0.436	0.310	0.384	0.393	0.246	0.346
SG	0.316	0.283	0.351	0.372	0.233	0.325

在第二組實驗中，我們將詞表示法結合馬可夫隨機漫步(MRW)以對語句進行選取，其結果呈現在表 5。從結果中可以觀察到，無論在 TD 或是 SD 上，相較於同樣以詞表示法的技術結合餘弦相似度的方法，使用該方法挑選語句的摘要成效皆優於以餘弦相似度的方式(表 3)。在 TD 實驗中，CBOW 摘要效能較 BM25 差，而 SG 未達到 MRW 的水平。在 SD 實驗中，仍然以 BM25 的摘要效果為佳。

表 6. 語句表示法結合於馬可夫隨機漫步之摘要結果

方法	文字文件(TD)			語音文件(SD)		
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
PV-DM	0.446	0.343	0.400	0.395	0.253	0.347
PV-DBOW	0.451	0.336	0.398	0.387	0.243	0.337

同樣地，我們以語句表示法結合馬可夫隨機漫步(MRW)對語句進行選取，其結果展示於表 6。從結果中發現到，無論在 TD 或是 SD 上，該方法的摘要成效，顯著地優越以詞、語句表示法結合於餘弦相似度(表 3 和 4)之選取摘要語句方法，亦超越以詞表示法結合於馬可夫隨機漫步的方式(表 5)。在 TD 實驗中，儘管該兩種詞表示法的摘要成效較 BM25 佳，但皆不及關聯模型(RM)。然而於 SD 實驗中，PV-DM 的摘要成效超越所有的傳統文件摘要模型。

表 7. 詞表示法結合於文件相似度量值之摘要結果

方法	文字文件 (TD)			語音文件 (SD)		
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
CBOW	0.444	0.329	0.386	0.372	0.221	0.314
SG	0.436	0.323	0.385	0.343	0.197	0.295

在最後一組實驗中，我們探討以詞表示法結合於文件相似度量值(DLM)對語句進行選取，其結果展示於表 7。我們將結果與同樣以詞表示法結合餘弦相似度(表 3)以及馬可夫隨機漫步的方法(表 5)進行比較。從 TD 實驗結果中可以觀察到，文件相似度量值充分地運用詞表示法於文件摘要，表現顯然較佳。我們亦注意到 SG 的摘要成效幾乎接近 CBOW。然而於 TD 與 SD 的實驗中，該兩種詞表示法皆仍不及 RM 的摘要成效。

表 8. 語句表示法結合於文件相似度量值之摘要結果

方法	文字文件(TD)			語音文件(SD)		
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
PV-DM	0.480	0.375	0.430	0.384	0.240	0.333
PV-DBOW	0.433	0.323	0.384	0.364	0.236	0.321

同樣地，我們以語句表示法於文件相似度量值對語句進行選取，其結果顯示在表 8。從 TD 的實驗結果中可以觀察到，PV-DM 的摘要效能顯著地優於表 2 中所有的傳統文件摘要模型，亦是所有表示法中具最佳摘要效能之模型。我們亦觀察到 PV-DBOW 與表 7 中的詞表示法 SG 有著相同的摘要成效。然而於 SD 中，該兩種語句表示法僅達到 RM 的水平，但皆仍不及 BM25。

7.3 利用聲學特徵結合支持向量機於文件摘要

本論文所使用的語音語料是經由人工切音，不會有語音邊界錯誤的問題，僅須考量語音辨識錯誤於文件摘要的影響，因此文字文件(TD)與語音文件(SD)兩者會有相同的語音邊界，而抽取出的韻律特徵亦會是一致。本論文總共使用 12 種不同的摘要特徵作為支持向量機(Support Vector Machine, SVM)的輸入，可概略分成三大類，分別為詞彙特徵(Lexical Features)、韻律特徵(Prosodic Features)以及關聯特徵(Relevance Features)，詳細的特徵資訊如表 9 所示。

表 9. 實驗採用之各式特徵

韻律特徵(Prosodic Features)	音高(Pitch):最大、最小、平均、差值 能量(Energy):最大、最小、平均、差值 音框長度(Duration):最大、最小、平均、差值 共振峰(Formant):最大、最小、平均、差值 頻譜峰值(Peak):最大、最小、平均、差值
詞彙特徵(Lexical Features)	專有名詞個數(Named Entity) 停用詞個數(Stop Word) 二連語言模型分數(Bigram) 正規化二連語言模型分數(Normalized Bigram)
關聯特徵(Relevance Features)	向量空間模型分數(VSM) 馬可夫隨機漫步分數(MRW) 語言模型分數(LM)

由表 10 中得到，無論在文字文件(TD)或是語音文件(SD)中，韻律特徵(Prosodic Features)相對於其它兩種特徵產生較為顯著的摘要效能，因此韻律特徵比起其它兩種特徵更能夠判斷摘要語句的重要資訊。在 TD 實驗中，詞彙特徵(Lexical Features)在這三種摘要特徵中的表現最差，其原因可能是該特徵描述的是表淺(Shallow)語句性質，包含專有名詞的數量、停用詞的數量以及語句的流暢性，沒有考慮語句的語意內容，因此單憑該特徵無法選取出較正確的摘要語句。此外，關聯特徵(Relevance Features)比起詞彙特徵有較好的摘要成效。在 SD 實驗中得到的結論，與 TD 的結論具一致性，但關聯特徵與韻律特徵之間效果差異較無 TD 來得顯著。

表 10. 單類特徵之摘要結果

	文字文件(TD)			語音文件(SD)		
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
韻律特徵	0.452	0.349	0.409	0.363	0.219	0.322
詞彙特徵	0.362	0.237	0.311	0.298	0.176	0.266
關聯特徵	0.389	0.254	0.332	0.355	0.200	0.300

我們進行使用所有摘要特徵於支持向量機器(Support Vector Machine, SVM)之實驗，其結果示於表 11。從實驗結果中可以發現，無論於 TD 或是 SD 中，經過各種面向的考量後，確實可以獲得較好的摘要成效。接著進行探討關聯特徵中使用其它模型分數對摘要效能的影響。因此我們將關聯特徵中的向量空間模型(VSM)、馬可夫隨機漫步(MRW)以及單連語言模型(ULM)的分數，以詞表示法模型摘要之分數作為替換，分別根據於表 3、5 和 7 中最佳的摘要表現，從各表中可以發現 CBOW 的摘要效果始終最佳。

表 11. 結合所有特徵之摘要結果

方法	文字文件(TD)			語音文件(SD)		
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
所有特徵	0.484	0.384	0.440	0.387	0.247	0.348

同樣地結合所有特徵一併做為支持向量機的輸入，其摘要效能如表 12 所示。從實驗結果中發現到，無論在 TD 或是 SD 中，以詞表示法模型作為關聯特徵，皆使得摘要成效非常顯著，尤其在 TD 中的實驗結果，產生最佳之摘要成效。

表 12. 以詞表示法模型摘要分數為關聯特徵之摘要結果

方法	文字文件(TD)			語音文件(SD)		
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
所有特徵	0.497	0.406	0.451	0.396	0.254	0.353

我們亦考慮語句表示法模型分數對摘要效能的影響。同樣將關聯特徵中的模型分數替換為語句表示法模型摘要之分數，分別根據於表 4、6 和 8 中最佳的摘要表現，從各表中的結果可觀察到 PV-DM 的摘要效果始終最佳；其摘要成效如表 13 所示。從 TD 的實驗結果中可以觀察到，使用語句表示法模型分數作為特徵之摘要成效較使用詞表示法來得差(表 12)。然而在 SD 中，結合以語句表示法模型分數作為關聯特徵可以達到最佳之摘要效果。

表 13. 以語句表示法模型摘要分數為關聯特徵之摘要結果

方法	文字文件(TD)			語音文件(SD)		
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
所有特徵	0.487	0.393	0.446	0.385	0.255	0.350

8. 結論與未來展望

過去在自動文件摘要的研究主要仍著重於文字文件摘要，直到 1990 年後期，由於影音多媒體技術的進步與成熟，才逐漸開始有語音文件摘要的研究。文件摘要可分為節錄式摘要與抽象式摘要，本論文旨在探討節錄式中文廣播新聞文件摘要方法。我們提出兩種詞表示法—連續型詞袋模型(CBOW)和跳躍式模型(SG)，以及兩種語句表示法—分散式儲

存模型(PV-DM)和分散式詞袋模型(PV-DBOW)於文件摘要的應用；透過表示法學習的技術能將詞之間、語句之間的關聯性表現出來，用以幫助選取語音文件中重要的摘要語句。經由一連串的實驗分析與討論，證明所提之方法的確可以較其它基礎實驗的摘要方法獲得更高的摘要效能。此外，我們除了利用文字文件的詞彙特徵及關聯特徵之外，亦利用語音訊號中之韻律特徵，希望能對摘要語句的選取提供更多有幫助的資訊。

未來，我們考慮其它先進的詞表示法，如全域向量(Global Vectors, GloVe)，以及希望可以運用詞性(Part of Speech, POS)資訊的詞性表示法(POS Representation)於語音節錄式文件摘要，並且將詞、語句以及詞性表示法結合於其它的語言模型之中，如關聯模型(Relevance Model)(Lavrenko & Croft, 2001)，以進一步地提升摘要成效。

致謝

本論文之研究承蒙教育部 - 國立臺灣師範大學邁向頂尖大學計畫(102J1A0800)與行政院科技部研究計畫(MOST 104-2221-E-003-018-MY3 和 MOST 103-2221-E-003-016-MY2)之經費支持，謹此致謝。

參考文獻

- Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research*, 3, 1137-1155.
- Carbonell, J., & Goldstein, J. (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval*, 335-336.
- Chen, Y.-T., Chen, B., & Wang, H.-M. (2009). A probabilistic generative framework for extractive broadcast news speech summarization. *IEEE Transactions on Audio, Speech and Language Processing*, 17(1), 95-106.
- Gong, Y., & Liu, X. (2001). Generic text summarization using relevance measure and latent semantic analysis. In *Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval*, 19-25.
- Hajime, M., & Manabu, O. (2000). A comparison of summarization methods based on task-based evaluation. In *Proceedings of the International Conference on Language Resources and Evaluation*, 633-639.
- Hinton, G. E. (1986). Learning distributed representations of concepts. In *Proceedings of the Annual Conference of the Cognitive Science Society*, 1-12.
- Jones, K. S. (1999). Automatic summarising: factors and directions. *Advances in Automatic Text Summarization*, 1-12.
- Lavrenko, V., & Croft, W. B. (2001). Relevance-based language models. In *Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval*, 120-127.

- Le, Q. V., & Mikolov, T. (2014). Distributed representations of sentences and documents. In *Proceedings of the International Conference on Machine Learning*.
- Lin, C. Y. (2004). ROUGE: a package for automatic evaluation of summaries. In *Proceedings of the Workshop on Text Summarization Branches Out*.
- Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2), 159-165.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a). Efficient estimation of word representations in vector space. In *Proceedings of the International Conference on Learning Representations*, 1-12.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. In *Proceedings of the International Conference on Learning Representations*, 1-9.
- Miller, G., & Charles, W. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6(1), 1-28.
- Mitra, M., Singhal, A., & Buckley, C. (1997). Automatic text summarization by paragraph extraction. In *Proceedings of the ACL/EACL Workshop on Intelligent Scalable Text Summarization*, 39-46.
- Mnih, A., & Kavukcuoglu, K. (2013). Learning word embeddings efficiently with noise-contrastive estimation. In *Proceedings of the Annual Conference on Neural Information Processing Systems*, 2265-2273.
- Morin, F., & Bengio, Y. (2005). Hierarchical probabilistic neural network language model. In *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*, 246-252.
- Ponte, J. M., & Croft, W. B. (1998). A language modeling approach to information retrieval. In *Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval*, 275-281.
- Qiu, L., Cao, Y., Nie, Z., & Rui, Y. (2014). Learning word representation considering proximity and ambiguity. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 1572-1578.
- Robertson, S. E., & Jones, K. S. (1976). Relevance weighting of search terms. *Journal of the American Society for Information Science*, 27(3), 129-146.
- Robertson, S. E., Walker, S., Jones, K. S., Hancock-Beaulieu, M., & Gatford, M. (1996). Okapi at TREC-4. In *Proceedings of the Fourth Text Retrieval Conference*, 73-97.
- Robertson, S. E., & Walker, S. (1994). Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval*, 232-241.
- Rousseau, F., & Vazirgiannis, M. (2013). Graph-of-word and TW-IDF: New approach to Ad hoc IR. In *Proceedings of the International Conference on Conference on Information, Knowledge Management*, 59-68.

- Salton, G., & Lesk, M. E. (1968). Computer evaluation of indexing and text processing. *Journal of the ACM*, 15(1), 8-36.
- Wan, X., & Yang, J. (2008). Multi-document summarization using cluster-based link analysis. In *Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval*, 299-306.
- Wang, H.-M., Chen, B., Kuo, J.-W., & Cheng, S.-S. (2005). MATBN: a Mandarin Chinese broadcast news corpus. *Journal of Computational Linguistics and Chinese Language Processing*, 10(2), 219-236.

調變頻譜分解技術於強健語音辨識之研究

Investigating Modulation Spectrum Factorization Techniques for Robust Speech Recognition

張庭豪*、洪孝宗*、陳冠宇+、王新民+、陳柏琳*

Ting-Hao Chang, Hsiao-Tsung Hung, Kuan-Yu Chen,

Hsin-Min Wang and Berlin Chen

摘要

自動語音辨識(Automatic Speech Recognition, ASR)系統常因環境變異而導致效能嚴重地受影響；所以長久以來語音強健(Robustness)技術的發展是一個極為重要且熱門的研究領域。本論文旨在探究語音強健性技術，希望能透過有效的語音特徵調變頻譜處理來求取較具強健性的語音特徵。為此，我們使用非負矩陣分解(Nonnegative Matrix Factorization, NMF)以及一些改進方法來正規化調變頻譜強度成分，藉以獲得較具強健性的語音特徵。本論文有下列幾項貢獻。首先，結合稀疏性的概念，期望能夠求取到具調變頻譜局部性的資訊以及重疊較少的 NMF 基底向量表示。其次，基於局部不變性的概念，希望發音內容相似的語句之調變頻譜強度成分，在 NMF 空間有越相近的向量表示以維持語句間的關聯程度。再者，在測試階段經由正規化 NMF 之編碼向量，更進一步提升語音特徵之強健性。最後，我們也結合上述三種 NMF 的改進方法。本論文的所有實驗皆於國際通用的標竿語料——Aurora-2 連續數字資料庫進行；實驗結果顯示相較於僅使用梅爾倒頻譜特徵之基礎實驗，我們所提出的改進方法皆

*國立臺灣師範大學資訊工程學系

Department of Computer Science & Information Engineering, National Taiwan Normal University

E-mail: {60247029S, 60047064S, berlin}@ntnu.edu.tw

+中央研究院資訊科學所

Institute of Information Science, Academia Sinica.

E-mail: {kychen, whm}@iis.sinica.edu.tw

The author for correspondence is Berlin Chen.

能顯著地降低語音辨識錯誤率。此外，我們也嘗試將所提出的改進方法與一些知名的特徵強健技術做比較和結合，以驗證這些改進方法之實用性。

關鍵詞：語音辨識、雜訊、強健性、調變頻譜、非負矩陣分解

Abstract

The performance of an automatic speech recognition (ASR) system often deteriorates sharply due to the interference from varying environmental noise. As such, the development of effective and efficient robustness techniques has long been a challenging research subject in the ASR community. In this article, we attempt to obtain noise-robust speech features through modulation spectrum processing of the original speech features. To this end, we explore the use of nonnegative matrix factorization (NMF) and its extensions on the magnitude modulation spectra of speech features so as to distill the most important and noise-resistant information cues that can benefit the ASR performance. The main contributions include three aspects: 1) we leverage the notion of sparseness to obtain more localized and parts-based representations of the magnitude modulation spectra with fewer basis vectors; 2) the prior knowledge of the similarities among training utterances is taken into account as an additional constraint during the NMF derivation; and 3) the resulting encoding vectors of NMF are further normalized so as to further enhance their robustness of representation. A series of experiments conducted on the Aurora-2 benchmark task demonstrate that our methods can deliver remarkable improvements over the baseline NMF method and achieve performance on par with or better than several widely-used robustness methods.

Keywords: Speech Recognition, Language Model, Concept Information, Model Adaptation

1. 研究動機

大多數的自動語音辨識系統，在不受雜訊干擾的理想實驗室環境下，皆可獲得良好的辨識效果；但是在真實的日常環境中，往往因為環境中諸多複雜因素的影響，造成訓練環境與測試環境存在不匹配問題，使得此系統之辨識精確率大幅度降低。造成環境不匹配問題的因素有語者變異、加成性背景雜訊、摺積性通道雜訊及其他語者發音的干擾等。本研究探討語音辨識之強健性技術，希望降低上述因素所帶來的負面影響，進而使語音辨識系統在實際應用時仍能保有一定的效能表現。

當前所發展出的各種語音強健技術大致可分為三種類型(Lin *et al.*, 2009; Chu *et al.*, 2011)：第一種類型為以聲學模型(Acoustic Model)為基礎之強健性技術(Model-Based Techniques)，此類方法大多是期望透過少量在測試環境所錄製的調適語料來對聲學模型

進行調整，使聲學模型可以近似於輸入含雜訊語音的機率分布參數，達到降低環境不匹配所造成影響的目的。第二類是以語音特徵為基礎之強健性技術(Feature-Based Techniques)。此類方法期望經過適當的正規化處理後，能使含雜訊語音與其原始乾淨語音趨於一致。最後第三類型為綜合式強健性技術，即同時在特徵處理和模型訓練兩階段做改善。

本論文將探討以語音特徵為基礎之強健性技術。其研究的議題主要圍繞在對何種空間正規化？以及在該空間應如何正規化？典型方法是將時間序列域(Temporal Domain)上的語音特徵視為是隨機變數(Random Variable)的樣本(Samples)，利用觀測到樣本去估測隨機變數之統計特性，進而對語音特徵時間序列做線性或非線性的轉換，使其在部分或整體之統計特性能經過正規化的處理。常見的方法有統計圖等化法(Histogram Equalization, HEQ)(Torre *et al.*, 2005)、倒頻譜平均值減去法(Cepstral Mean Subtraction, CMS)(Furui, 1981)以及倒頻譜平均數與變異數正規化法(Cepstral Mean and Variance Normalization, CMVN)(Vikki & Laurila, 1998)。上述方法所利用的統計資訊仍有所不足，無法觀察出明確的時序結構(Temporal Structure)改變。特徵參數時間序列之調變頻譜(Modulation Spectrum)為一有效描繪時空結構之媒介，相對於時間序列域之語音特徵正規化法的觀念而言，可能具有更廣泛的分析面向。例如，人們發出的聲音大多集中在調變頻譜的低頻處，因為發聲器官限制了語速。加成性的噪音則可能會反應在每個頻率，那些在高頻或與人聲不同的頻帶的資訊就可以被區分出來。近年來在調變頻譜域的語音強健性研究相當熱門，學者們致力於正規化特徵參數之時空結構，藉由強化語音特徵之調變頻譜來提升語音特徵的強健性。相關的技術包括了調變頻譜統計圖等化法(Spectral Histogram Equalization, SHE)(Sun *et al.*, 2007)、分頻式調變頻譜統計正規化法(Sub-Band Modulation Spectrum Compensation)(Huang *et al.*, 2009)與其它一系列資料導向(Data-Driven)之時間序列濾波器法(Xiao *et al.*, 2008; Hermansky & Morgan, 1994)等。

本論文旨在探究使用非負矩陣分解(Nonnegative Matrix Factorization, NMF)以及一些改進方法來正規化調變頻譜強度成分，以獲得較具強健性的語音特徵。首先，結合稀疏性的概念，期望能夠求取到具調變頻譜局部性的資訊以及重疊較少的 NMF 基底向量表示。其次，基於局部不變性的概念，希望發音內容相似的語句之調變頻譜強度成分，在 NMF 空間有越相近的向量表示以維持語句間的關聯程度。再者，在測試階段經由正規化 NMF 之編碼向量，更進一步提升語音特徵之強健性。最後，我們也結合上述三種 NMF 的改進方法。此外，也嘗試將我們所提出的改進方法與一些現有的特徵強健技術做比較和結合，以驗證這些改進方法之實用性。

2. 調變頻譜正規化法

2.1 調變頻譜之簡介

對於任一特定維度語音頻譜特徵所成的時間序列 $x[n]$ 而言，其調變頻譜定義如下：

$$X[k] = DFT(x[n]) = \sum_{t=0}^{N-1} x[t]e^{-j\frac{2\pi tk}{N}}, \quad 0 \leq k \leq \frac{N}{2} \quad (1)$$

其中， n 與 k 依序為音框索引與調變頻率索引， DFT 為離散傅立葉轉換(Discrete Fourier Transform, DFT)， $X[k]$ 代表語音特徵時間序列 $x[n]$ 的調變頻譜。由式(1)可看出調變頻譜可以被用來廣泛地分析語句中語音特徵隨時間變化的資訊。而 $X[k]$ 頻譜序列可視為一種對於原始語音訊號作降低取樣(Down-Sampling)後的調變訊號(由訊號取樣率轉至音框取樣率)，此序列即為所屬語音特徵時間序列之調變頻譜(Modulation Spectrum)。由式(1)可知，調變頻譜 $X[k]$ 之最高頻率與特徵序列 $x[n]$ 之取樣頻率(音框取樣率)有關。例如，在一般設定下，若音框取樣率為 100 Hz，則最高調變頻率為 50 Hz。

過去已有不少學者研究語音特徵之調變頻譜的特性，發現了調變頻譜中的低頻成分比高頻成分還要重要的特性(Kaneder *et al.*, 1997)。同時，調變頻譜之低頻成分(約 1Hz 至 16Hz)對於語音辨識正確率也有密切的關係，潛藏有重要的語意資訊。其中，最重要的是位於 4 Hz 附近，有學者指出，4 Hz 是人耳聽覺最為敏感之調變頻率(Hermansky, 1998)；另有學者也認為，4 Hz 為人類大腦皮層感知之重要調變頻率(Greenberg, 1997)。當語音訊號受到雜訊影響時，其語音特徵時間序列會受到影響而失真，及其調變頻譜也會跟著受到牽連。很多學者提出作用在調變頻譜的正規化法，以改善調變頻譜受到雜訊干擾的影響。因此，我們可將許多發展在語音特徵時間序列的正規化法應用在調變頻譜使其正規化；而正規化的對象是對其調變頻譜強度(Magnitude)成分 $|X[k]|$ 來進行處理，並保持其相位角不變 $\theta[k]=\angle X[k]$ 的部分。接著，經處理後被更新的強度成分會與原始相位成分結合，再藉由反傅立葉轉換(Inverse Discrete Fourier Transform, IDFT)來求得新的語音特徵時間序列。若調變頻譜的強度能夠被有效的正規化，便能夠有效解決雜訊產生的環境不匹配問題，使自動語音辨識系統在使用新的語音特徵的情況下能夠獲得較佳的辨識率。以下將會簡單回顧一些常見的調變頻譜正規化法。

2.2 調變頻譜平均正規化法(Spectral Mean Normalization, SMN)

假設當各種音素在一般環境中分布的比例接近一致時，每一維度語音特徵的調變頻譜之平均值應該為一個定值(Huang *et al.*, 2009)：

$$|\tilde{X}[k]| = |X[k]| - \mu_s + \mu_a \quad (2)$$

在式(2)中， $|X[k]|$ 為原始的調變頻譜強度成分， μ_s 為單一語句的調變頻譜強度成分之平均值， μ_a 為所有訓練語句的調變頻譜強度成分之平均值，而 $|\tilde{X}[k]|$ 便是更新過後的調變頻譜強度成分。

2.3 調變頻譜平均與變異數正規化法(Spectral Mean and Variance Normalization, SMVN)

除了要正規化調變頻譜強度成分之平均值外，也可同時正規化其標準差(Huang *et al.*, 2009)。假設特徵向量參數之平均值與變異數在一般環境中分布的比例接近一致時，我們

可以同時對其平均值和標準差來進行正規化：

$$|\tilde{X}[k]| = \frac{|X[k]| - \mu_s}{\sigma_s} \sigma_a + \mu_a \quad (3)$$

在式(3)中， μ_s 與 σ_s 為單一語句的調變頻譜強度成分之平均值與標準差； μ_a 與 σ_a 為所有訓練語句的調變頻譜強度成分之平均值與標準差， $|\tilde{X}[k]|$ 便是更新過後的調變頻譜強度成分。

2.4 調變頻譜統計圖等化法(Spectral Histogram Equalization, SHE)

利用非線性的轉換(Nonlinear Transformation)，不僅將調變頻譜強度成分之平均值與標準差(或變異數)作正規化，而是整體上使得訓練語句與測試語句的調變頻譜強度成分趨於擁有同一個機率分布函數，正規化全部階層的動差(Sun *et al.*, 2007)：

$$|\tilde{X}[k]| = F_{ref}^{-1}(F_X(|X[k]|)) \quad (4)$$

在式(4)中， $F_X(\cdot)$ 為單一語句某一特徵維度的調變頻譜強度之累積分布函數(Cumulative Distribution Function, CDF)， F_{ref} 則是利用所有訓練語句之調變頻譜強度所求得的對應之參考累積分布函數， $|\tilde{X}[k]|$ 便是更新過後的調變頻譜強度成分。

2.5 分頻段調變頻譜統計正規化法

此方法的概念是想要改進原始調變頻譜統計正規化法；原始調變頻譜統計正規化法是將全部調變頻帶的頻譜強度值視為是屬於同一隨機變數的樣本(Samples)，且將之一併進行正規化的動作。但是前面提到在語音辨識中，不同調變頻率的成分有不同的重要性，低頻成分是比高頻成分還要相對重要的，因為語言的重要資訊較集中於低頻成分。因此，有學者提出將調變頻帶分成許多子頻段，再分別對每一個子頻段的頻譜強度作上述所提的調變頻譜正規化的方法，而不是單純直接對整個全部調變頻帶做處理(Huang *et al.*, 2009)。因為要強調低調變頻率的重要性，所以在低頻部分的子頻段擁有較窄的頻寬，子頻段的數量也比較多，而高調變頻率便持有相反的特性。由於能更細緻地分析與處理低頻成分的資訊，過去的一些實驗數據顯示出將調變頻率分頻段來正規化的做法，能比全頻帶正規化的方式獲得較好的效能。

3. 三種新穎NMF改進方法用於調變頻譜分解

3.1 傳統非負矩陣分解法(NMF)

在很多領域中如何尋找重要的潛藏資訊成分是個重要的議題，而基於非負矩陣分解法(Nonnegative Matrix Factorization, NMF)(Lee & Seung, 1999)的技術可以被用於處理此議題。顧名思義，此方法就是將非負的原始資料所成的矩陣進行分解，表示成兩個也是非負的矩陣乘積，接著利用線性組合的特性來表示原始資料中各個樣本之目的。而其它常見的線性表示法有主成分分析(Principal Component Analysis, PCA)與獨立成分分析(Independent Component Analysis, ICA)。非負矩陣分解法與這兩種線性表示法之差異就是

能夠提供非負的基底向量(Nonnegative Basis Vectors)，且也能夠擁有保證由基底向量組合而成之資料也為非負的特性。非負矩陣分解法的另一個重要特性是想要學習以部分為基礎(Parts-Based)之線性表示法來表示原始的資料，且此線性表示法是一個加法的組合模式。這種以部分為基礎的概念方法擁有直觀的性質，而且對於一個特定任務來說，在與其它分解方法相比下可以得到比較高的解釋性。過去有學者應用非負矩陣分解法在影像處理的領域，例如人臉影樣可以用為五官等局部影像做為非負基底向量經由線性組合(線性編碼)而產生。若是使用上述所提到的，例如 PCA，在分解舉證產生基底向量的過程中可能會產生負值，這些負值在影像處理當中會難以解釋。而在語音領域方面，語音的特徵值有正有負，所以較難以直接地使用非負矩陣分解法；直到近期有學者將非負矩陣分解法用在分析調變頻譜強度以擷取重要語音特徵(Chu *et al.*, 2011)，而可以得到了不錯的強健性效果。NMF 的數學式表示如下：

$$V \approx WH = \sum_k W_{ik} H_{kj} \quad (5)$$

其中 $V \in R^{I \times J}$ 為一個非負矩陣，而兩個被分解出來的非負矩陣分別為 $W \in R^{I \times K}$ 和 $H \in R^{K \times J}$ ，如圖 1 所示。

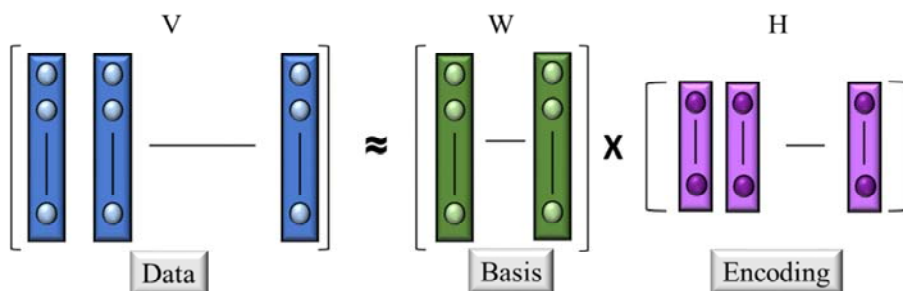


圖 1. 非負矩陣分解法(NMF)示意圖

其中矩陣 W 所包的 K 行即為基底向量，矩陣 H 中的每一行則通常被稱為編碼向量(Encoding)，有著權重的概念，與基底向量進行線性組合去近似資料矩陣 V 。 I 是每筆資料向量的維度大小； J 為所有資料向量的個數； K 為基底向量的數量。參數 K 是可以自行決定的，通常會選擇小於 I 與 J ，但還是會有選擇的限制：

$$(I + J) \times K < I \times J \quad (6)$$

式(6)是學者過去所提出更確切的基底向量個數選擇限制關係式。在非負矩陣分解法的方法中，有著資料壓縮的概念，若是 K 的數目選擇得越少，代表壓縮的比率越高。因為我們對資料進行了壓縮的動作，所以壓縮後的資料跟原始的資料來比較必定會有一些資料是在壓縮過程中被遺失了。我們希望遺失的部分資料越少越好，所以可以定義減損函數(Loss Function)來測量資料前後的相似度。測量由兩個因子矩陣 W 與 H 所重建的訊號 A 與原始訊號 V 之間的距離，對分解結果與原始資料的近似程度作量化(Quantification)。

非負矩陣分解法常見的減損函數為歐氏距離(Euclidian Distance 或 Frobenius Norm)：

$$D_F(V||WH) = \|V - WH\|_F^2 = \sum_{i,j} (V_{ij} - (WH)_{ij})^2 \quad (7)$$

$D_F(V||WH)$ 是藉由歐氏距離所提出的減損函數。當重建訊號 Λ 與原始信號 V 相等時，則 $D_F(V||WH) = 0$ 。另一個減損函數則是基於 KL 散度(Kullback-Leibler Divergence)：

$$D_{KL}(V||WH) = \sum_{i,j} \left(V_{ij} \ln \frac{V_{ij}}{(WH)_{ij}} - V_{ij} + (WH)_{ij} \right) \quad (8)$$

當原始信號 V 與重建訊號 Λ 相等時， $D_{KL}(V||\Lambda) = 0$ 。因為 KL 散度不具對稱性(Symmetric)，因此減損函數值不能稱為兩個訊號之間的距離值(Distance)，而是兩訊號之間的差異值(Divergence)，而 KL 散度也稱為相對熵(Relative Entropy)。

由於要將資料矩陣 V 分解成 W 與 H ，而使誤差最小化。所以使用迭代更新規則將 W 與 H 更新去求得局部最小值(Local Minimum)。最起初提出的方法是使用梯度下降演算法(Gradient Descent Algorithm)與加法迭代(Iteration)規則。後來又有學者提出乘法迭代規則；乘性迭代規則能夠直接地賦予非負矩陣分解法之非負限制的特性。以下是乘法迭代更新規則(Lee & Seung, 2000)：

Euclidian Distance 的乘法更新規則：

$$\begin{aligned} H_{kj} &\leftarrow H_{kj} \frac{(W^T V)_{kj}}{(W^T W H)_{kj}} \\ W_{ik} &\leftarrow W_{ik} \frac{(V H^T)_{ik}}{(W H H^T)_{ik}} \end{aligned} \quad (9)$$

Kullback-Leibler Divergence 的乘法更新規則：

$$\begin{aligned} H_{kj} &\leftarrow H_{kj} \frac{\sum_i W_{ik} V_{ij} / (WH)_{ij}}{\sum_i W_{ik}} \\ W_{ik} &\leftarrow W_{ik} \frac{\sum_j H_{kj} V_{ij} / (WH)_{ij}}{\sum_j H_{kj}} \end{aligned} \quad (10)$$

3.2 非平滑非負矩陣分解法(NSNMF)

非平滑非負矩陣分解法(Pascual-Montano *et al.*, 2006)直接修改傳統非負矩陣分解法的模型，利用模型的乘法性質，達到矩陣全面的稀疏，以能擷取更局部的資訊(如圖 2 所示意)。非負矩陣分解法將資料矩陣分成兩個矩陣相乘，也就是基底矩陣乘以編碼矩陣。若在一個矩陣中，其元素是非稀疏或平滑的，為了要補償最後兩個矩陣相乘之後能盡可能地近似原始資料矩陣，這將會迫使另一個矩陣面臨稀疏或非平滑的情況。非平滑非負矩陣分解法可以定義如下：

$$V = WSH \quad (11)$$

在式(11)中，矩陣 $V \in R^{I \times J}$ 為資料矩陣；矩陣 $W \in R^{I \times K}$ 為基底矩陣；矩陣 $H \in R^{K \times J}$ 為編碼矩陣；而矩陣 $S \in R^{K \times K}$ 稱為平滑矩陣，其定義如下：

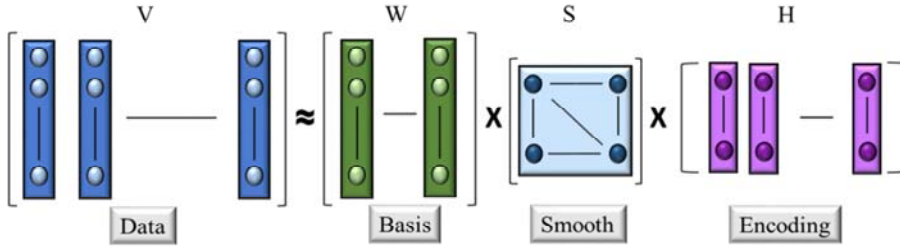


圖2. 非平滑非負矩陣分解法(NSNMF)示意圖

$$S = (1 - \theta)I + \frac{\theta}{K}11^T \quad (12)$$

式(12)中 1 是一個元素都是 1 的向量， I 是單位矩陣，以及 θ 是一個用來控制整體稀疏程度的參數，此參數 θ 滿足 $0 \leq \theta \leq 1$ 的範圍中。對平滑矩陣 S 可以解釋為：假設 X 為一個正的非零值向量，而 $Y=SX$ 為轉換後的向量。如果 $\theta=0$ ， $Y=X$ ，意謂著向量 X 中沒有平滑發生；如果 $\theta=1$ ，向量 Y 中所有的元素會變成一致的數值，此數值會等於向量 X 所有元素的平均，這就是最平滑的向量。由上述可知參數 θ 用來控制平滑矩陣 S 的平滑程度。由於模型的乘法性質，平滑矩陣 S 中若有強烈的平滑情況，將會迫使在基底向量與編碼向量中造成強烈的稀疏，因此也可以說參數 θ 是用來控制整個非負矩陣分解法模型的稀疏程度。特別的是，當參數 $\theta=0$ 時，平滑矩陣 S 會等同於一個單位矩陣 I ，此時模型會回歸到傳統的非負矩陣分解法的模型。在此，我們將更進一步地說明整個非平滑非負矩陣分解法的流程與乘法更新規則。首先，式(11)中非平滑非負矩陣分解法的模型可以等價地寫成：

$$V = (WS)H = W(SH) \quad (13)$$

用括號來表示平滑矩陣 S 是先與哪個矩陣做相乘。若是平滑矩陣 S 先與基底矩陣 W 做相乘，代表說基底矩陣 W 會變得平滑，這將會迫使編碼矩陣 H 變得稀疏；同樣地，若是平滑矩陣 S 先與編碼矩陣 H 做相乘，代表說編碼矩陣 H 會變得平滑，這將會迫使基底矩陣 W 變得稀疏。在非負矩陣分解與更新過程中，上述兩種情況將都會發生的，所以基底矩陣 W 與編碼矩陣 H 都會被強制變為具稀疏性。相較於傳統非負矩陣分解法，非平滑非負矩陣分解法的乘法迭代更新規則為：在更新編碼矩陣 H 時，將 W 換成 (WS) ；更新基底矩陣 W 時，將 H 換成 (SH) 。

Euclidian Distance 的乘法更新規則：

$$\begin{aligned} H_{kj} &\leftarrow H_{kj} \frac{((WS)^T V)_{kj}}{((WS)^T (WS)H)_{kj}} \\ W_{ik} &\leftarrow W_{ik} \frac{(V(SH)^T)_{ik}}{(W(SH)(SH)^T)_{ik}} \end{aligned} \quad (14)$$

Kullback-Leibler Divergence 的乘法更新規則：

$$\begin{aligned} H_{kj} &\leftarrow H_{kj} \frac{\sum_i (WS)_{ik} V_{ij} / ((WS)H)_{ij}}{\sum_i (WS)_{ik}} \\ W_{ik} &\leftarrow W_{ik} \frac{\sum_j (SH)_{kj} V_{ij} / (W(SH))_{ij}}{\sum_j (SH)_{kj}} \end{aligned} \quad (15)$$

而其它部分的演算法流程同傳統非負矩陣分解法。

3.3 基於圖正則化非負矩陣分解法(GNMF)

基於圖正則化非負矩陣分解法(Graph Regularized Non-negative Matrix Factorization, GNMF)(Cai *et al.*, 2011)的主要目的在於保留資料的局部不變性(Locally Invariant)(Hadsell *et al.*, 2006)，意指原本相鄰的資料向量經過降維或投影後仍然維持相鄰近。資料向量間的遠近關係，或幾何結構資訊可以用一權重矩陣 \mathbf{E} 表示，其維度是等於資料向量數量所形成的方陣。最後將權重矩陣 \mathbf{E} 納入減損函式中，做為編碼矩陣的正則項(Regularization Term)。

令 $\mathbf{h}_j = [h_{j1}, \dots, h_{jk}]^T$ 為編碼矩陣 \mathbf{H} 的第 j 行， \mathbf{h}_j 可被視為是第 v_j 個資料向量相對於新的基底矩陣 \mathbf{W} 之新表示。在此我們討論較常見的歐式距離：

$$d(\mathbf{h}_j, \mathbf{h}_l) = \|\mathbf{h}_j - \mathbf{h}_l\|^2 \quad (16)$$

此距離用來測量相對於新的基底矩陣 \mathbf{W} ，而兩個資料向量 \mathbf{h}_j 與 \mathbf{h}_l 在低維度空間中表示之間的差異(Dissimilarity)，距離函式值越大代表此兩個資料向量 \mathbf{h}_j 與 \mathbf{h}_l 彼此差異越大。

$$\begin{aligned} R_1 &= \frac{1}{2} \sum_{j,l=1}^v \|\mathbf{h}_j - \mathbf{h}_l\|^2 E_{jl} \\ &= \sum_{j=1}^v \mathbf{h}_j^T \mathbf{h}_j D_{jj} - \sum_{j,l=1}^v \mathbf{h}_j^T \mathbf{h}_l E_{jl} \\ &= Tr(\mathbf{H}^T \mathbf{D} \mathbf{H}) - Tr(\mathbf{H}^T \mathbf{E} \mathbf{H}) \\ &= Tr(\mathbf{H}^T \mathbf{L} \mathbf{H}) \end{aligned} \quad (17)$$

其中 R_1 是編碼矩陣的正則項， $Tr(\cdot)$ 為矩陣的跡數(Trace)， $D_{jj} = \sum_l E_{jl}$ ， $\mathbf{L} = \mathbf{D} - \mathbf{E}$ ， \mathbf{L} 稱作圖拉普拉斯算子(Graph Laplacian)。在此希望使 R_1 最小化，達到保留資料局部不變性的

目的。將上述所求出的 R_1 當作懲罰項，加入到傳統 NMF 之歐式距離減損函式中可以得到新的減損函數：

$$O_{Euclidean} = \|V - WH\|^2 + \lambda Tr(HLH^T) \quad (18)$$

同樣地，可利用梯度下降演算法去求出基於圖正則化非負矩陣分解法的乘法更新規則：

$$H_{kj} \leftarrow H_{kj} \frac{(W^T V + \lambda HE)_{kj}}{(W^T W H + \lambda H D)_{kj}}$$

$$W_{ik} \leftarrow W_{ik} \frac{(V H^T)_{ik}}{(W H H^T)_{ik}} \quad (19)$$

其中 $\lambda \geq 0$ ，為正則化參數，去控制新的表示之平滑性。

有別於傳統 NMF 方法僅在歐氏空間中求解，GNMF 方法可以視不同應用問題而設計合適的權重矩陣 E 。在語音辨識的任務中，聲學模型通常被建立在音素層次，而且主宰語音辨識的表現。因此，本論文也提出利用音素錯誤率建立權重矩陣 E ，詳細的描述與實驗稍後將被呈現在第 4.3 節。

3.4 非負編碼矩陣統計圖等化法(HNMF)

傳統的 NMF 方法將訓練資料分解成非負基底矩陣 W_{clean} 和編碼矩陣 H_{clean} 兩部分，在測試階段時只保留基底矩陣，而丟棄了編碼矩陣的資訊。在應用中，受噪音干擾的語音可能會得到與乾淨語料不相似的編碼向量，此時我們不能確定這樣的資料表示是否已經排除大部分雜訊？再者，即便是乾淨語料中也存在著許多變異性。為了克服上述問題，本論文提出利用統計圖等化法將編碼矩陣做正規化處理。在訓練階段時，我們利用統計圖等化法(HEQ)將乾淨訓練語料的編碼矩陣 H_{clean} 的資訊儲存建表，統計編碼矩陣 H_{clean} 的參考分布，如圖 3。而在測試階段時求出編碼向量 h ，再將 h 每一個元素執行統計圖等化法之查表的動作，試圖將含有雜訊的 h 還原回到對應的乾淨的編碼向量。以能夠去對應由乾淨訓練語料所估測出來的參考分布，使語句的編碼向量在訓練環境與測試環境之機率分布一致，如圖 4 所示。我們認為乾淨的基底向量矩陣乘上正規化後的編碼矩陣應較能夠還原回乾淨的語音特徵。

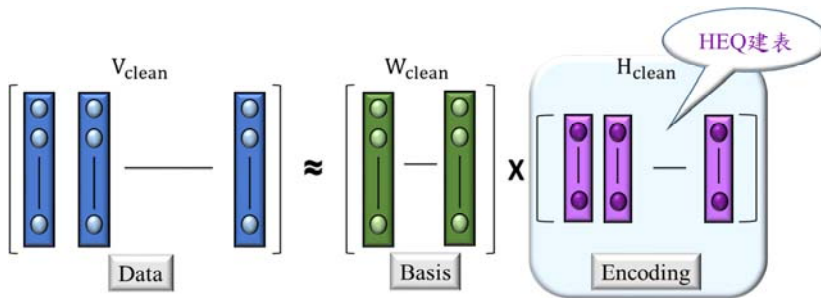


圖3. 非負編碼矩陣統計圖等化法(HNMF)訓練階段示意圖

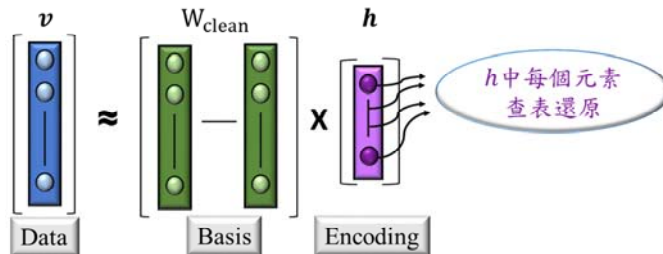


圖4. 非負編碼矩陣統計圖等化法(HNMF)還原示意圖

4. 實驗結果與分析

4.1 實驗語料庫

本論文實驗所採用的語料庫是 Aurora-2，它是由歐洲電信標準協會 (European Telecommunications Standards Institute, ETSI) 所發行的語料庫 (Hirsch & Pearce, 2000)，以美國成年人的聲音作為錄音來源，內容是連續的英文數字由 0 (Zero) 到 9 (Nine) 跟 Oh 等發音字詞。語料庫內有乾淨及含有雜訊的語音，雜訊中有八種不同的加成性雜訊與兩種不同的通道效應，而通道效應是使用國際電信聯合會 (ITU) 標準中的 G.712 和 MIRS。根據不同的雜訊干擾，分成三個測試集：Set A、Set B 及 Set C。Set A 的語音分別含有地下鐵 (Subway)、人聲 (Babble)、汽車 (Car) 和展覽會館 (Exhibition) 等四種加成性雜訊與 G.712 通道效應；Set B 的語音則分別含有餐廳 (Restaurant)、街道 (Street)、機場 (Airport) 和火車站 (Train Station) 等四種加成性雜訊與 G.712 的通道效應；Set C 分別加入了地下鐵 (Subway) 與街道 (Street) 兩種雜訊與 MIRS 通道效應。而其中的訊噪比 (SNR) 則有七種，為 Clean、20dB、15dB、10dB、5dB、0dB 和 -5dB，並且提供二種訓練模式：乾淨情境訓練模式 (Clean-Condition Training) 與複合情境訓練模式 (Multi-Condition Training)。本研究的基礎實驗皆使用乾淨情境訓練模式，故在聲學模型訓練時並沒有使用到任何加成性雜訊的資訊或內涵。

4.2 實驗設定

在本論文中的基礎實驗是採用梅爾倒頻譜係數 (Mel-frequency Cepstral Coefficients, MFCC) 做為語音特徵參數，取樣頻率 (Sampling Rate) 為 8,000Hz，預強調 (Pre-emphasis) 參數設為 0.97；使用的窗函數為漢明窗 (Hamming Window)，音框長度 (Frame Length) 是 25 毫秒，音框間距 (Frame Shift) 為 10 毫秒。每一個音框的語音特徵是使用 13 維梅爾倒頻譜係數 (第 1 維至第 12 維還有第 0 維)，加上其一階差量和二階差量，共 39 維之特徵參數。本論在對語音特徵進行強健性 (正規化) 處理時，只針對 13 維的靜態特徵參數進行處理，待處理完成後才額外將語音特徵的一階差量和二階差量加入形成最後每一個音框的語音特徵。

4.3 辨識效能評估方式

辨識效能的評估方式是依照美國國家標準與科技局(National Institute of Standards and Technology, NIST)所訂立的評估標準，進行每一句測試語句之正確轉寫詞串與語音辨識詞串的比較。評估方式是以詞正確率(Word Accuracy Rate)為主，計算正確轉寫詞串與語音辨識詞串彼此間的詞取代個數(Substitutions)、詞插入個數(Insertions)和詞刪除個數(Deletions)：

$$\text{詞正確率(\%)} = \frac{\text{詞正確辨識個數} - \text{詞插入個數}}{\text{輸入詞總數}} \times 100\% \quad (20)$$

最後在評估整體語音辨識效能時，我們參照國際學者之設定，對測試語句在每一種噪音的訊噪比的詞正確率結果做加總與取平均的動作(去掉極端的訊噪比 Clean 跟-5，只計算範圍 20dB 到 0dB 中的平均詞正確率)；本論文以下的全部實驗皆是利用平均詞正確率來評估語音辨識的效能。

4.4 非平滑非負矩陣分解法(NSNMF)之實驗結果

由 3.2 節的敘述可知 NSNMF 因為乘法的性質，若是平滑程度高的矩陣S與W或H矩陣其中一個相乘，為了要補償能儘可能的近似重建原始資料，會迫使另一個矩陣達到稀疏的效果。如此經遞迴地使用乘法更新規則，最終可達到稀疏化矩陣W與H的效果。實驗數據如表 1 所示；由實驗結果可見，隨著 θ 的增加語音辨識之詞正確率的確能夠逐漸地被提高。雖然在 $\theta = 0.3$ 以前較無明顯效果，可能因為迫使矩陣稀疏的程度並不高；而 $\theta = 1$ 時，表示迫使矩陣稀疏程度最高，而數據顯示的確能夠表現得最好。

表1. 非平滑之非負矩陣分解法(NSNMF)在使用不同 θ 值下之詞正確率(%)

	Set A	Set B	Set C	Average
NMF	67.09	70.98	68.22	68.87
$\theta = 0.1$	67.54	71.62	66.01	68.87
$\theta = 0.2$	66.91	71.35	64.72	68.25
$\theta = 0.3$	66.89	71.63	67.85	68.98
$\theta = 0.4$	70.19	73.64	68.72	71.28
$\theta = 0.5$	69.46	73.60	66.22	70.47
$\theta = 0.6$	70.82	74.18	71.20	72.24
$\theta = 0.7$	72.07	75.29	69.73	72.89
$\theta = 0.8$	72.99	76.25	72.75	74.25
$\theta = 0.9$	74.12	76.98	74.67	75.37
$\theta = 1$	77.05	79.75	77.47	78.21

4.5 基於圖正則化非負矩陣分解法(GNMF)之實驗結果

在探討 GNMF 的效能時，我們首先在求取權重矩陣 E 時使用 0-1 權重的方式；為此，我們先算出所有訓練語句(8,440 句)彼此間的關聯度。關聯度的估測是透過計算兩兩訓練語句彼此間的音素錯誤率(Phone Error Rate, PER)而得；我們會先求得每一句訓練語句經人工轉寫(Transcription)之音素序列(Phone Sequence)。本論文音素錯誤率的算法是使用編輯距離(Edit Distance)的算法，計算每一句訓練語句(當成目標語句)的音素序列與其它訓練語句的音素序列彼此間的音素取代個數、音素插入個數、音素刪除個數，並依下式計算音素錯誤率：

$$\text{音素錯誤率(\%)} = \frac{\text{音素取代數} + \text{音素插入數} - \text{音素刪除數}}{\text{目標訓練語句音素總數}} \times 100\% \quad (21)$$

所以最後求得權重矩陣 E 是個維度大小為 8,400*8,400 的矩陣，其中每個元素都紀錄著每一句訓練語句與其它語句彼此間的音素錯誤率，對角線上的為一個位置為某一句訓練語句自己本身所以差異是 0。值得一提的是，當使用編輯距離算差異度時， E_{ji} 與 E_{ij} 的值可能不會一樣，是因為不同的目標語句(任兩句訓練語句，彼此的音素序列長度可能會是不同的)的假定，所以權重矩陣 E 是不對稱的。但我們認為兩個訓練語句間彼此的關聯度應該是對稱的(一樣的)；因此我們採折衷方式，將 E_{ji} 與 E_{ij} 的值都改為兩者的相加取平均，使權重矩陣 E 變成一個對稱矩陣。再者，我們設了一個門檻值(Threshold) α ：

$$\begin{cases} E_{ji} \leq \alpha, & E_{ji} = 1 \\ E_{ji} > \alpha, & E_{ji} = 0 \end{cases} \quad (22)$$

當 E_{ji} 或 E_{ij} 大於門檻值時，代表說這兩句訓練語句彼此間的音素錯誤率較大(訓練語句差異大)，因此關聯度設為 0，希望兩個訓練語句彼此間沒有關聯；而當 E_{ji} 或 E_{ij} 小於等於門檻值時代表這兩句訓練語句彼此間的音素錯誤率較小，應有較大的關聯性，因此設為 1。因為設定了一個門檻值 α ，若門檻值 α 設定的較嚴格(值較小時)，權重矩陣 E 就會顯得零值越多而越稀疏化。另外，在本論文中對於式(18)中的 λ 值設定為 100。GNMF 的實驗數據如表 2 所示。當 α 值越高時，代表門檻越寬鬆，權重矩陣 E 中非零值的元素也會越多；

表2. GNMF 使用不同門檻值的之詞正確率(%)

	Set A	Set B	Set C	Average
NMF	67.09	70.98	68.22	68.87
$\alpha = 0.3$	68.00	72.09	67.92	69.62
$\alpha = 0.5$	67.86	72.22	68.02	69.64
$\alpha = 0.7$	67.07	71.18	66.77	68.65
$\alpha = 0.8$	67.97	72.48	67.79	69.74
$\alpha = 0.9$	68.49	72.64	68.00	70.05

數據顯示語音辨識的詞會隨 α 值變大而逐漸提高。可能是因為若 α 的值越低，門檻越嚴格，權重矩陣 E 中非零值的元素就越少，導致任兩語句間的關聯度在求取矩陣 W 與 H 時較不會被強調。

表3. GNMF-a 使用權重矩陣全域給值之詞正確率(%)

	Set A	Set B	Set C	Average
NMF	67.09	70.98	68.22	68.87
$\alpha = 0.9$	68.49	72.64	68.00	70.05
GNMF-a	70.63	74.27	70.78	72.12

基於此觀察，本論文嘗試改成讓權重矩陣 E 的每一個元素都能擁有適當的權重值，而不使用0-1權重(我們認為只設定一個門檻值就將權重值一分為二的作法可能會較粗糙一些)；我們利用式(23)將 E 中的每個元素之音素錯誤率做轉換權重的動作，可將權重值限制在0到1之間：

$$E_{jl} = \frac{1}{1 + \text{PER}_{j,l}} \quad (23)$$

如此做法，可以將各個訓練語句彼此間的關聯程度做比較精細的描述，而不是只有0或1的權重值。例如：音素錯誤率0%的轉換後會變成1；音素錯誤率40%的轉換後會變成0.714；音素錯誤率100%的轉換後會變成0.5；音素錯誤率160%的轉換後會變成0.385。讓越低的音素錯誤率能夠有越高的權重值。特別的是，權重矩陣 E 之對每一個角線位置的值，也就是代表某一語句本身的關聯程度；因其音素錯誤率為0%，所以關聯程度會為1。相關的數據如表3所示；當改成使用全域都有值之權重矩陣 E （對應方法簡稱為GNMF-a）會比使用預設的一個門檻值之權重矩陣 E 的效果會來的好一些，使詞正確率提高了2.07%；同時也比傳統非負矩陣分解法(NMF)提高了3.25%的詞正確率。

值得注意的是，在上述實驗中使用基於音素錯誤率求取權重矩陣 E 的方式，在此方式中語音特徵的所有維度的調變頻譜強度是使用相同的權重矩陣 E 。另一方面，我們也嘗試基於語句每一維度的調變頻譜強度，個別地利用歐式距離來計算的語句間的關聯程度，並且也使用類似式(23)的轉換式，求出不同維度的權重矩陣 E ，此方式簡稱為GNMF-eu。實驗結果如表4所示，基於歐式距離使得不同維度對應著不同的權重矩陣去進行NMF中矩陣 W 與 H 求取，最後反應在語音辨識效能上似乎沒有較使用音素錯誤率的方式來的好。

表4. GNMF-eu 之詞正確率(%)

	Set A	Set B	Set C	Average
NMF	67.09	70.98	68.22	68.87
GNMF-a	70.63	74.27	70.78	72.12
GNMF-eu	69.34	72.26	69.44	70.53

4.6 非負編碼矩陣統計圖等化法(HNMF)之實驗結果

如第三節所提及，我們進一步保留在訓練階段獲得之乾淨訓練語料編碼矩陣 H 的累積分布函數(CDF)資訊，將其儲存建表以供測試階段使用統計圖等化法來正規化每一測試語句的編碼向量。其數據如表 5 所示，此方法(HNMF)在基底個數等於 5 時，可以達到優於非平滑非負矩陣分解法(NSNMF)的效能。但可發現若基底個數若持續增加時，似乎就無法與 NSNMF 競爭，且效能提昇的程度不明顯。

表 5. HNMF 之不同基底個數之詞正確率(%)

	Set A	Set B	Set C	Average
$K=5$	77.65	80.16	77.25	78.57
$K=10$	69.55	74.32	67.77	71.10
$K=15$	67.73	72.60	65.26	69.18
$K=20$	66.71	71.74	63.56	68.09
$K=30$	67.43	72.66	64.05	68.84

4.7 三種非負矩陣分解法改進方法之結合

接著我們結合非平滑非負矩陣分解法(NSNMF)以及基於圖正則化非負矩陣分解法(GNMF)，稱為 NSGNMF(以下所結合之 GNMF 皆使用 GNMF-a)，實驗數據如表 6 所示。雖然非平滑非負矩陣分解法原本就有 78.21% 之不錯的詞正確率，不過加上基於圖正則化非負矩陣分解法利用訓練語句間的相關聯度之概念，能夠有 1.24% 的正確率提升。最後我們再對編碼矩陣做 HEQ 正規化處理來提升語音辨識效能(表示成 NSHGNMF)；不過再結合 HNMF 之後效果並沒有預料中顯著，只有些許的詞正確率提昇。在表 6 中也列出兩種常見的調變頻譜正規化法(SHE 與 PCA)來作為比較比較(Kao *et al.*, 2014)。SHE 是利用

表 6. NMF 改良方法結合與之詞正確率(%)比較

	Set A	Set B	Set C	Average
NMF	67.09	70.98	68.22	68.87
GNMF	70.63	74.27	70.78	72.12
NSNMF	77.05	79.75	77.47	78.21
HNMF	77.65	80.16	77.25	78.57
NSGNMF	78.22	80.92	78.95	79.45
NSHGNMF	78.28	80.96	78.98	79.49
SHE	74.82	77.44	76.47	76.20
PCA	70.90	73.34	71.39	71.97

HEQ 將調變頻譜強度成分的平均值與標準差正規化，並同時正規化其它階層的動差使訓練語句與測試語句的調變頻譜強度的機率分布趨於一致。PCA 則是對所有訓練語句的調變頻譜強度成分求取共變異數，接著利用前 r 個特徵值(Eigenvalues)去找其對應的 r 個特徵向量(Eigenvectors)以當作調變頻譜強度成分的 PCA 子空間之基底，使測試語句的調變頻譜強度成分能夠投影到 PCA 子空間以達到正規化的目的。

表7. 結合CMVN 與NMF 之詞正確率(%)

	Set A	Set B	Set C	Average
CMVN	75.93	76.76	76.82	76.44
CMVN+NSNMF	83.56	85.51	83.27	84.28
CMVN+GNMF	83.58	84.78	82.36	83.81
CMVN+HNMF	82.88	84.84	82.37	83.56
CMVN+NSGNMF	83.94	85.76	83.61	84.61
CMVN+NSHGNMF	83.98	85.85	83.71	84.67

表8. 結合HEQ 與NMF 之詞正確率(%)

	Set A	Set B	Set C	Average
HEQ	80.03	82.05	80.10	80.85
HEQ+NSNMF	83.84	85.88	83.70	84.63
HEQ+GNMF	83.71	84.76	82.53	83.89
HEQ+HNMF	82.89	85.52	83.59	84.08
HEQ+NSGNMF	84.02	85.89	83.79	84.72
HEQ+NSHGNMF	84.05	85.93	83.82	84.76

表9. 結合AFE 與NMF 之詞正確率(%)

	Set A	Set B	Set C	Average
AFE	87.68	87.10	86.29	87.17
AFE+NSNMF	87.74	87.65	86.32	87.42
AFE+GNMF	87.45	87.72	86.23	87.31
AFE+HNMF	87.81	87.22	86.36	87.28
AFE+NSGNMF	87.85	87.66	86.54	87.51
AFE+NSHGNMF	87.82	87.70	86.55	87.52

4.8 結合不同時間序列正規化法之結果

最後我們探討額外結合不同時間序列正規化法(CMVN 與 HEQ)與本論文所提出三種調變頻譜非負矩陣分解法的實驗結果，如表 7 與 8 所示。本論文所提出三種調變頻譜非負矩陣分解法都能與先經過不同時間序列正規化法處理過後的語音特徵相結合使用而得到效能提昇。值得注意的是，CMVN 與 HEQ 皆是在語句的音框層面(Frame Level)對每個音框分別作正規化，而 NMF 的方法是在整體語句層次(Utterance Level)正規化，因分別處理不同的面向，所以在結合後有加成性的效果。效果提升最顯著的是 CMVN 與 NSHG NMF 的結合；其次是與 HEQ 結合的 NSHG NMF，皆能有不錯的進步。我們也將所提出方法與與進階前端標準(Advanced Front-End Standard, AFE)處理過語音特徵(Macho *et al.*, 2002)做結合，其結果如表 9 所示。AFE 是近年來歐洲電信標準協會(ETSD)所推出的特徵向量擷取方法，是一個著名且成效非常好的常見基礎系統設置，在多種任務上被證實能顯著地提升語音辨識系統在雜訊環境中的效能。當 AFE 與 NSHG NMF 做結合時只能有些微提升；我們猜測可能是因為 AFE 本身已具備有很完善的語音特徵正規化處理程序，若再加 NSHG NMF 時，語音特徵可能會被過度地正規化而導致語音辨識效能無法被顯著提升。

5. 結論

本論文探討了非負矩陣分解法的三種改進方法並將之運用在語音特徵的調變頻譜正規化上；希望藉此能夠擷取出更強健性的調變頻譜基底向量，而達到增進語音強健性的目的。第一種是非平滑非負矩陣分解法(NSNMF)，利用添加了一個平滑矩陣 S ，變更傳統非負矩陣分解法的模型；利用模型乘法的性質，使一個矩陣平滑，進而迫使另一個矩陣達到稀疏的效果。第二種是基於圖正則化非負矩陣分解法(GNMF)，在減損函式中增加了一個額外的正則項。利用幾何結構與局部不變性的特性，求得訓練語句間的關聯程度並創造一個權重矩陣以供使用，使經正規化的語音特徵能夠增加鑑別力。第三種是非負編碼矩陣統計圖等化法(HNMF)，希望能夠利用在訓練階段時可獲得的編碼矩陣，利用統計圖等化法將其累積分布函數資訊建表儲存，希望在測試階段時能藉此將含雜訊語句的編碼向量進一步正規化。

當將此三種非負矩陣分解法之改進方式運用在 Aurora-2 上時，皆能使語音辨識效能有所進步。整體上來說，NSNMF 使用的矩陣稀疏性而有較顯著且一致的效能提升；GNMF 雖沒有帶來大幅度的效能提升，但是其所利用語句之間的關聯程度資訊也能對語音特徵正規化有所幫助，像是與 NSNMF 結合，也能稍微提昇精確率；另外，HNMF 在少許基底個數時能提供不錯的語音辨識效能提升。

致謝

本論文之研究承蒙教育部 - 國立臺灣師範大學邁向頂尖大學計畫(102J1A0800)與行政院科技部研究計畫(MOST 104-2221-E-003-018-MY3 和 MOST 103-2221-E-003-016-MY2)之經費支持，謹此致謝。

參考文獻

- Cai, D., He, X., Han, J., & Huang, T. S. (2011). Graph regularized nonnegative matrix factorization for data representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8), 1548-1560.
- Chu, W.-Y., Hung, J.-W., & Chen, B. (2011). Modulation spectrum factorization for robust speech recognition. In *Proceedings of the APSIPA Annual Summit and Conference*, 18-21.
- Furui, S. (1981). Cepstral analysis techniques for automatic speaker verification. *IEEE Transactions on Acoustic, Speech and Signal Processing*, 29(2), 254-272.
- Greenberg, S. (1997). On the origins of speech intelligibility in the real world. In *Proceedings of the ESCA-NATO Tutorial and Research Workshop on Robust Speech Recognition for Unknown Communication Channels*.
- Hadsell, R., Chopra, S., & LeCun Y. (2006). Dimensionality reduction by learning an invariant mapping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1735-1742.
- Hermansky, H., & Morgan, N. (1994). RASTA processing of speech. *IEEE Transactions on Speech and Audio Processing*, 2(4), 578-589.
- Hermansky, H. (1998). Should Recognizers Have Ears? *Speech Communication*, 25(1-3), 3-27.
- Hirsch, H. G., & Pearce, D. (2000). The AURORA experimental framework for the performance evaluations of speech recognition systems under noisy conditions. In *Proceedings of the ISCA ITRW ASR*.
- Huang, S.-Y., Tu, W.-H., & Hung, J.-W. (2009). A study of sub-band modulation spectrum compensation for robust speech recognition. In *Proceedings of the ROCLING XXI: Conference on Computational Linguistics and Speech Processing*.
- Kanedera, N., Arai, T., Hermansky, H., & Pavel, M. (1997). On the importance of various modulation frequencies for speech recognition. In *Proceedings of the European Conference on Speech Communication and Technology*.
- Kao, Y.-C., Wang, Y.-T., & Chen, B. (2014). Effective modulation spectrum factorization for robust speech recognition. In *Proceedings of the Annual Conference of the International Speech Communication Association*, 2724-2728.
- Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401, 788-791.
- Lee, D. D., & Seung, H. S. (2000). Algorithms for Non-negative Matrix Factorization. In *Proceedings of the Annual Conference on Neural Information Processing Systems*, 556-562.
- Lin, S.-H., Chen, B., & Yeh, Y.-M. (2009). Exploring the use of speech features and their corresponding distribution characteristics for robust speech recognition. *IEEE Transactions on Audio, Speech and Language Processing*, 17(1), 84-94.

- Macho, D., Mauuary, L., Noé, B., Cheng, Y. M., Ealey, D., Juvet, D., Kelleher, H., Pearce, D., & Saadoun, F. (2002). Evaluation of a noise-robust DSR front-end on Aurora databases. In *Proceedings of the Annual Conference of the International Speech Communication Association*.
- Pascual-Montano, A., Carazo, J. M., Kochi, K., Lehmann, D., & Pascual-Marqui, R. D. (2006). Nonsmooth nonnegative matrix factorization (nsNMF). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3), 403-415.
- Sun, L.-C., Hsu, C.-W., & Lee, L.-S. (2007). Modulation Spectrum Equalization for robust Speech Recognition. In *Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding*.
- Torre, A. D. L., Peinado, A. M. J., Segura, C., Perez-Cordoba, J. L., Benitez, M. C., & Rubio, A. J. (2005). Histogram equalization of speech representation for robust speech recognition. *IEEE Transactions on Speech and Audio Processing*, 13(3), 355-366.
- Vikki, A., & Laurila, K. (1998). Segmental feature vector normalization for noise robust speech recognition. *Speech Communication*, 25, 133-147.
- Xiao, X., Chng, E. S., & Li, H. (2008). Normalization of the speech modulation spectra for robust speech recognition. *IEEE Transactions on Speech and Audio Processing*, 16(8), 1662-1674.

透過語音特徵建構基於堆疊稀疏自編碼器演算法之
婚姻治療中夫妻互動行為量表自動化評分系統

**Automating Behavior Coding for Distressed Couples
Interactions Based on Stacked Sparse Autoencoder
Framework using Speech-acoustic Features**

陳柏軒*、李祈均*

Po-Hsuan Chen and Chi-Chun Lee

摘要

在過去人類行為分析是透過傳統人為觀察方式來記錄。像婚姻治療方面，評分者利用觀看錄影的方式來對一整段夫妻對話中所展現的行為作評分。藉由這樣取得各種行為表達程度的量化，針對此量化分數來更進一步研究夫妻婚姻治療成效，但這種做法非常耗時且會因為評分者的各種主觀因素影響最後的準確性。如果能透過機器學習的方式來自動化處理辨識，將會節省非常多的人工時間和提升客觀性。深度學習(Deep Learning)在目前機器學習上是很熱門的話題。本論文提出以堆疊稀疏自編碼器(Stacked Sparse Autoencoder, SSAE)方式對聲音訊號特徵進行降維，並找出相對關鍵的高階特徵，最後再利用邏輯迴歸分析(Logistic Regression, LR)來辨識。此方法的整體準確率為 75%(丈夫行為平均辨識準確率為 74.9%、太太為 75%)。相對於過去研究的 74.1% (丈夫行為平均準確率 75%，太太為 73.2%) (Black *et al.*, 2013)，提升 0.9%。我們提出的方法在使用更低維度的聲音特徵值中可有效的提升行為辨識準確率。

關鍵詞：深度學習，堆疊自編碼器，婚姻治療，人類行為分析，情緒分析

*國立清華大學電機工程學系

Department of Electrical Engineering, National Tsing Hua University

E-mail: theusa20@gmail.com; ccleee@ee.nthu.edu.tw

Abstract

Traditional way of conducting analyses of human behaviors is through manual observation. For example in couple therapy studies, human raters observe sessions of interaction between distressed couples and manually annotate the behaviors of each spouse using established coding manuals. Clinicians then analyze these annotated behaviors to understand the effectiveness of treatment that each couple receives. However, this manual observation approach is very time consuming, and the subjective nature of the annotation process can result in unreliable annotation. Our work aims at using machine learning approach to automate this process, and by using signal processing technique, we can bring in quantitative evidence of human behavior. Deep learning is the current state-of-art machine learning technique. This paper proposes to use stacked sparse autoencoder (SSAE) to reduce the dimensionality of the acoustic-prosodic features used in order to identify the key higher-level features. Finally, we use logistic regression (LR) to perform classification on recognition of *high* and *low* rating of six different codes. The method achieves an overall accuracy of 75% over 6 codes (husband's average accuracy of 74.9%, wife's average accuracy of 75%), compared to the previously-published study of 74.1% (husband's average accuracy of 75%, wife's average accuracy of 73.2%) (Black *et al.*, 2013), a total improvement of 0.9%. Our proposed method achieves a higher classification rate by using much fewer number of features (10 times less than the previous work (Black *et al.*, 2013)).

Keywords: Deep Learning, Stacked Autoencoders, Couple Therapy, Human Behavior Analysis, Emotion Recognition

1. 緒論

人與人之間交談互動，常透過語言傳達彼此的想法，並在這交談過程中得知雙方的行為反應。利用人為觀察來分析雙方行為反應，這部分最早常應用在心理學和精神學方面 (O'Brian *et al.*, 1994)。人為行為觀察相當的成功研究在親密關係 (Karney & Bradbury, 1995) (Gonzaga *et al.*, 2007)，即夫妻的行為是影響親密關係程度的因素之一。然而用於人為觀察行為的方式存在一些困難，一方面太消耗時間，另一面也浪費成本。

如果能透過電腦工程的方式來取代人為觀察將大大提升效率，透過低層描述映射高層描述來預測人類行為 (Schuller *et al.*, 2007)，這項研究領域是正在不斷發展的一部分。人類行為信號處理(Behavioral Signal Processing, BSP)目的在幫助連接信號科學和行為處理的方法，建立在傳統的信號處理研究，如語音識別，面手部追蹤等等。相關顯著 BSP 研究已發產於以人為中心的提取音頻，視頻信號，來分析實際上人類行為或是情感方面 (Burkhardt *et al.*, 2009; Devillers & Campbell, 2011)。

婚姻治療中夫妻互動行為量表自動化評分系統

本論文利用 BSP 的基本思路應用在婚姻治療資料庫上面 (Christensen *et al.*, 2004)，婚姻治療資料庫會詳細說明在第二章。這個資料庫紀錄了夫妻在一段對話中談述了他們所選擇婚姻中的問題。評分者在根據他們一段話的種種行為根據不同行為量表進行評分(幽默行為、悲傷行為等等)。

延續上篇論文的研究內容來自動化分析夫妻一段對話的行為分數(Black *et al.*, 2013)，一段語音經過預處理，之後作聲音特徵擷取(acoustic feature extraction)，再使用機器學習來作分類辨識，得到最後的準確率。其中，特徵擷取和機器學習的算法都會影響最後的準確率，思考如何改進這些影響因素，對整體準確率的提升是一大重要的課題，也是我們提出這篇論文的因素之一。

在特徵擷取方面，我們沿用三種低階語音特徵(Low Level Descriptors, LLDs)，語韻(prosodic) LLDs、頻譜(spectrum) LLDs 和音質(voice quality) LLDs。切割三種說話者說話區間(speaker domain)，丈夫說話區間、太太說話區間、和不分人說話區間。再來對應各區間提取 20% 語句，經過 7 種統計函數(functionals)，產生 2940 種特徵值。最後我們利用非監督深度學習的做法來降維找出相對關鍵的主要特徵值表現。

深度學習在機器學習領域裡面是最近熱門的話題 (Hinton, 2006)。深度學習可看成是一種資訊的表達方式，利用多層神經網絡，第一層輸入的數據學習之後，產生新的組合輸出，輸出值為第二層的輸入值，再經由學習產生新的輸出值，依此類推重覆把每層的資訊堆疊下去，透過這樣多層學習，可以得到對一個目標值好的特徵表示，相對準確率就能有所提升。至今存在多種深度學習框架如深度神經網路(DNN)、深度信念網路(DBN)和卷積神經網路(CNN)已被應用在語音 (Hinton *et al.*, 2012)、影像辨識 (Smirnov *et al.*, 2014)和手寫識別 (Perwey & Chaturvedi, 2011)等等。

我們利用深度學習中的堆疊稀疏自編碼器(stacked sparse autoencoder, SSAE)，降低特徵值維度，提升特徵值整體相關性，最後利用簡單 LR 辨識行為分數高低。此初期研究結果顯示整體行為平均準確率 75%較之前研究使用 40479 維特徵值結合支持向量器(support vector machine) (Black *et al.*, 2013)提升了 0.9%。

以下簡述各章節的內容。第二章介紹本篇論文所使用的資料庫(database)，第三章介紹我們使用的 SSAE 架構和其演算法，第四章介紹我們提出的系統架構和研究結果，第五章節為結論。

2. 婚姻治療資料庫

為了測試我們提出方法的準確率，我們使用和之前論文相同的婚姻治療資料庫(couple therapy database)。以下簡單的介紹的資料庫相關內容：此資料庫的收集是基於研究綜合行為夫婦治療(integrative behavioral couple therapy, IBCT)成效 (Christensen *et al.*, 1995)。資料內容針對 134 對夫妻，每對都長期患有婚姻的問題，如夫妻相處不融洽或是爭執。

治療內容為每對夫妻接受為期一年的治療，研究團隊再讓每對夫妻由太太和丈夫各別選擇一個目前存在嚴重婚姻問題的題目來作為一段 10 分鐘對話，對話中沒有治療師和

研究團隊。透過這 10 分鐘的對話讓夫妻彼此了解雙方之間的問題並且試圖解決當前問題。

每對夫妻皆會進行三個不同階段的對話，治療前、治療中和治療兩年後。透過這三個時間點對話，再經由多位有專業背景的評分者經由兩個行為評分量表，基於社交互動行為評分系統(Social Support Interaction Rating System, SSIRS) (Jones & Christensen, 1998) 和基於夫妻互動行為評分系統(Couples Interaction Rating System, CIRS) (Heavey *et al.*, 2002)進行評分，依據評分結果來了解治療的成效。SSIRS 主要包含 19 種行為準則在四個社交互動分類裡，情感(affectivity)、屈從服從(dominance/submission)、互動表現行為(feature of interaction)和主題評價(topic definition)來作為評分的內容，CIRS 主要包含 13 種行為準則關於夫妻互動問題解決方面，如表 1。

表 1. 32 種人類行為準則包含在兩種行為量表 SSIRS 和 CIRS

Manual	Codes
<p style="text-align: center;">SSIRS (Social Support Interaction Rating System)</p>	<p>Global positive affect、global negative affect use of humor、sadness、anger/frustration、 belligerence/domineering、contempt/disgust、 tension/anxiety、defensiveness、affection、 satisfaction、solicits partner suggestions、 instrumental support offered、emotional support offered、submissive or dominant、topic a relationship issue、topic a personal issue、 discussion about husband、discussion about wife</p>
<p style="text-align: center;">CIRS (Couples Interaction Rating System)</p>	<p>Acceptance of other、blame、responsibility for self、solicits partner perspective、states external origins、discussion、clearly defines problem、 offers solutions、negotiates、make agreements、pressures for change、 withdraws、avoidance</p>

總共 32 個行為準則，每個行為評分區間為 1 到 9 分。同一對話中，丈夫與妻子會各別被評分。1 為對這項行為所表現的程度最低，9 為對這項行為所表現的程度最高。評分者為 3 到 4 個，透過觀察夫妻 10 分鐘的影片來各別對 32 個行為進行評分。最後總共有 569 個 10 分鐘的會話，117 對夫妻在這個婚姻治療庫裡。

本篇論文延續上一篇論文所使用的 6 種行為來下去作分析，包含認同對方(Acceptance of other)、責備行為(Blame)、夫妻之間正面的互動(Global positive affect)、夫妻之間負面的互動(Global positive affect)、悲傷行為(sadness)、幽默表現行為(humor)，如

表 2。之所以會選擇這 6 種行為，因為和其他 26 種行為評分比起來，這 6 種有較高的評分者認同度(Agreement)，認同度的計算方式為個別評分者的分數和其他評分者評分的平均分數取相關係數(correlation)。其餘行為的認同度介於 0.4 和 0.7 之間，第五章節會比較這 6 種行為預測準確率。

表 2. 對於 6 種行為準則的認同度(agreement)

Code	Agreement
Acceptance of other (acc)	0.751
Blame (bla)	0.788
Global positive affect (pos)	0.740
Global negative affect (neg)	0.798
Sadness (sad)	0.722
Use of humor (hum)	0.755

3. 研究方法

在本節，我們首先簡單的介紹自編碼器(Autoencoder)和堆疊稀疏自編碼(Stacked Sparse Autoencoder, SSAE)基本架構以及本篇論文用到的演算法。

3.1 自編碼器(Autoencoder)

深度學習中自編碼器利用非監督學習方式 (Rubanov, 2000)，目標從高維度的輸入特徵值學習到更具代表性的特徵值，最後透過解碼讓輸出值等於輸入值，基本的自編碼器架構如圖 1。

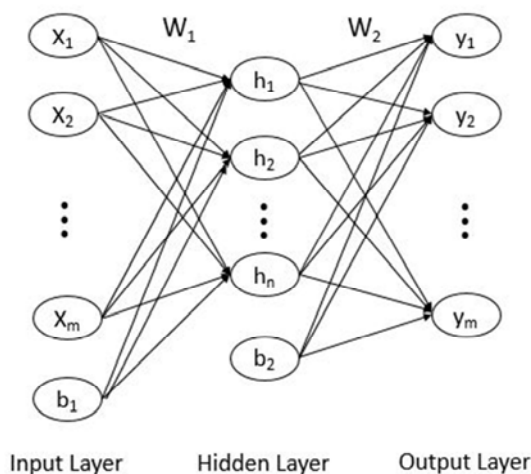


圖 1. 自編碼器

從圖 1，輸入值 $x_i, i = 1, 2, \dots, m, x \in R^m$ ，隱藏層(hidden layer)中的 $h_i, i = 1, 2, \dots, n, h \in R^n$ ，權重矩陣(weight matrix) $W_1 \in R^{n \times m}$ ，偏移向量(bias vector) $b_1 \in R^n$ 。由這些因子(factor)構成激活函數(activation function)，如式(1)。

$$h(x) = f(W_1x + b_1) \quad (1)$$

其中 $f(x) = 1/(1 + \exp(-z))$ 為 sigmoid function。輸出值 $y_i, i = 1, 2, \dots, m, y \in R^m$ ，權重矩陣 $W_2 \in R^{m \times n}$ ，偏移向量 $b_2 \in R^m$ ，自編碼器輸出為式(2)：

$$y = f(W_2h(x) + b_2) \quad (2)$$

為了要求得權重矩陣 W_1 和 W_2 ，偏移向量 b_1 和 b_2 ，假設一個樣本集為 $\{(x_1, y_1), (x_2, y_2) \dots (x_m, y_m)\}$ ，有 m 組樣本， x_i 為樣本輸入特徵值， y_i 為對應標籤值，利用代價函數(cost function)，如式(3)。

$$J(W, b) = \left[\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \|h(x^{(i)}) - y^{(i)}\|^2 \right) \right] + \frac{\lambda}{2} \sum_{l=1}^n \sum_{i=1}^{p_l} \sum_{j=1}^{p_{l+1}} (w_{i,j}^{(l)})^2 \quad (3)$$

式(3)中第一項為均方差項(sum-of-squares error term)，第二項為規則項(regularization term)，其中 λ 為權重衰減參數(weight decay parameter)， n 為自編碼器層數， p_l 為第 l 層節點數，這項是為了避免訓練過程發生過擬合(overfitting)，之後我們利用反向傳導(back-propagation)演算法和 L-BFGS 優化算法 (Andrew & Gao, 2007)，重複疊代減小 $J(W, b)$ 值，最後得到 W 和 b 。

而為了讓輸入特徵值更有效的歸類群集並且不同特徵之間的區隔明顯， $J(W, b)$ 加入稀疏項(sparsity term)如式(4)，取名為稀疏編碼器(sparse autoencoder) (Obst, 2014)。

$$J_s(W, b) = J(W, b) + \beta \sum_{j=1}^q KL(\rho || \hat{p}_j) \quad (4)$$

其中 $KL = \rho \log \frac{\rho}{\hat{p}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{p}_j}$ ， ρ 為稀疏參數(sparsity parameter)， $\hat{p}_j = \frac{1}{m} \sum_{i=1}^m h_j(x_i)$ ， β 為控制稀疏項(sparsity term)的參數， q 為隱藏層的節點數。

3.2 堆疊稀疏自編碼器(Stacked Sparse Autoencoder)

由多個稀疏自編碼器逐層訓練後，堆疊組成的架構為堆疊稀疏自編碼器(Stacked Sparse Autoencoder)，如圖 2，每一層的編碼後輸出為下一層的輸入。從圖 2 可看出，輸入層(Input layer)經由第一個稀疏自編碼器訓練完之後得到第一隱藏層(Hidden layer1)的 n 個節點，由這 n 個節點在經過第二個稀疏自編碼器訓練得到第二隱藏層(Hidden layer2)的 p 個節點，每層的隱藏層節點可視為由上一層產生新的一組特徵，透過這樣逐層訓練可以訓練更多層。

我們實驗採用堆疊稀疏自編碼器(Stacked Sparse Autoencoder, SSAE)，希望透過 SSAE 得到好的特徵表示方式，最後經由分類器產生更好的準確率。

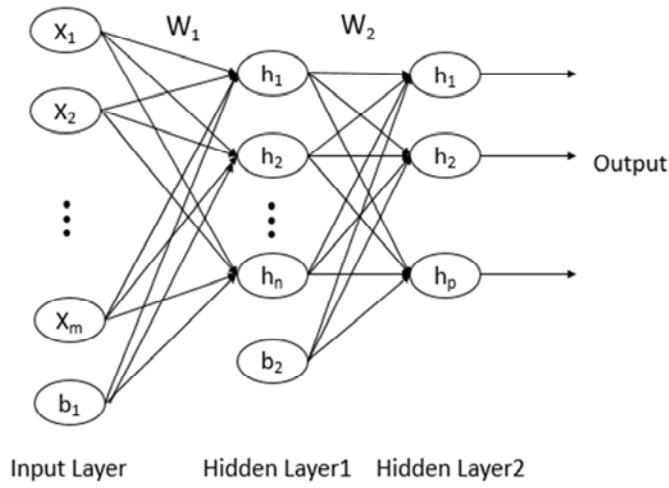


圖 2. 堆疊稀疏自編碼器

3.3 實驗架構

我們使用 3 層隱藏層的 SSAE 作為非監督學習的架構，來從低層級特徵(low level feature) 訓練成高層級特徵(high level feature)，然後用 LR 來監督學習作辨識，本實驗第一層稀疏自編碼器架構如圖 3。

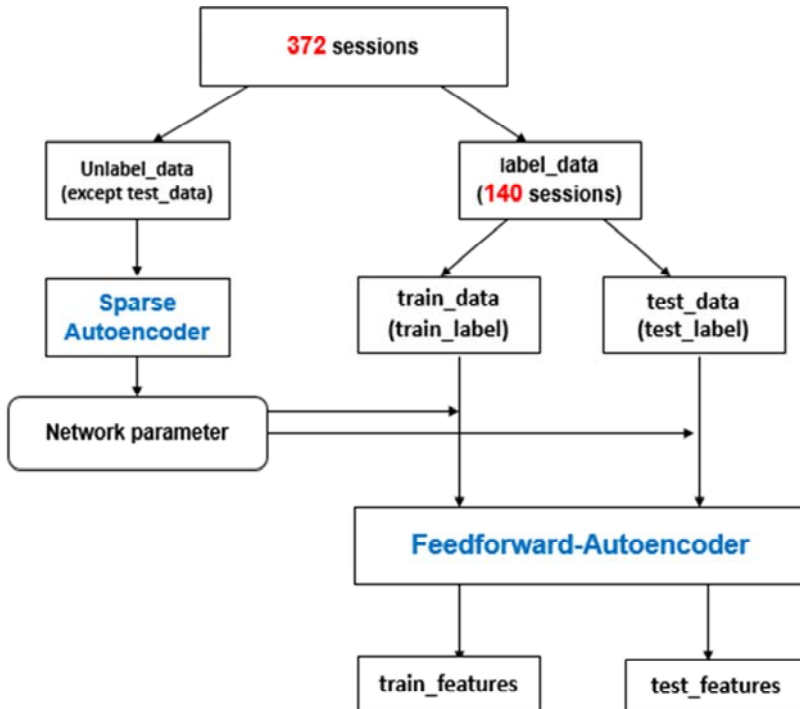


圖 3. 實驗架構

一段語音經過預處理，降低雜訊影響，才不會影響之後的特徵擷取，而這部分預處理在上篇論文已經被處理過了 (Black *et al.*, 2013)。本篇論文改變特徵擷取方法，這部分下一章會介紹。如圖 3，正規化後的特徵值，一種行為包含 372 筆 10 分鐘會話(session)，分為有標籤數據(labeled data)和沒有標籤數據(unlabeled data)，沒有標籤數據利用稀疏自編碼器來訓練網絡參數，訓練好後再把 140 筆有標籤數據分為訓練資料和測試資料，輸入自訓練好的網絡參數，產生新的一組特徵。新的一組特徵為下一層輸入值，重複利用圖 3 架構可以產生更多層。我們希望新的特徵值對於行為分數將有更好的表示，下面章節會證明之。

4. 實驗設計和結果

4.1 特徵值

如圖 4，利用原本 LLDs，在三種對話區間裡(speaker domain)，丈夫時間區間(husband、H)、太太時間區間(wife、W)和不分人時間區間(full、F)所說句子，切割成以 20% 句子為一個時間區間，切割完後合成一個行向量，行向量的特徵值，再經由如表 3 所列的 7 種 functionals 處理過後，產生最後 2940 個特徵值。在輸入 SSAE 以前，我們把這些特徵值正規化在 0 和 1 的區間。詳細的特徵值內容可參考 (Black *et al.*, 2013)。

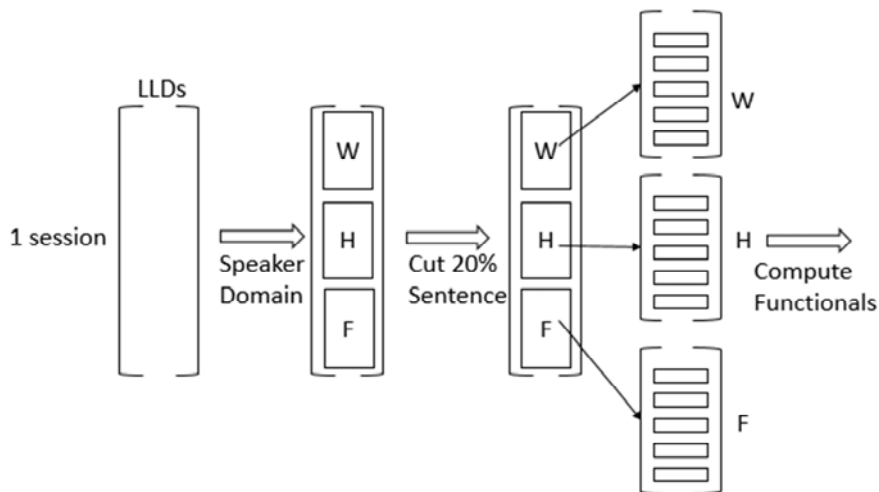


圖 4. 實驗特徵提取架構

表 3. 28 種特徵值和 7 種 functionals

LLDs	Functionals
1. MFCC[0-14]	1. Mean
2. MFB[0-7]	2. Median
3. F0normlog	3. Standard deviation
4. VAD(speech/no speech)	4. Skewness
5. Intensity	5. Kurtosis
6. Jitter	6. Max position
7. Jitter of Jitter	7. Min position
8. Shimmer	

4.2 資料

由原本資料庫 569 筆對話、117 對夫妻，經由上篇論文預處理過後(Black *et al.*, 2013)，產生最後的 372 筆對話、104 對夫妻。在 372 筆對話裡面丈夫和太太都會被評分到，對應在 6 種行為準則，我們選擇前 20% 的分數和後 20% 的分數的對話當作實驗的辨識，共 140 筆對話；兩種標籤值 0 和 1，1 為對應到高分，0 為對應到低分。而在這些取出來被預測的對話裡，夫妻數介於 68 到 77 對，利用這些行為對應到夫妻對數來作交叉驗證，1 對夫妻作驗證，其餘對數作訓練，重複循環 6 種行為對應到的夫妻對數來作驗證。

4.3 實驗設定

在這實驗裡，我們用 SSAE 來作為非監督學習，LR 來監督學習預測，留一對夫妻法則 (leave-one-couple-out) 的方式來作交叉驗證。一開始先用貪婪訓練算法 (greedy layerwise) 逐層預學習 (pre-training)，訓練完參數初始值輸入至 SSAE，SSAE 有五個因子會影響最後的表現，分別是隱藏層節點 (hidden units)、計算損失函數 (cost function) 的疊代次數和三個超參數 (hyper-parameters) 為 λ 、 ρ 、 β ， λ 為權重衰減參數 (weight decay parameter)， ρ 為稀疏參數 (sparsity parameter)， β 為控制稀疏項 (sparsity term) 的參數，這些參數在第三章有介紹過。我們先用 1 層隱藏層來測試準確率，如表 4。透過改變不同的隱藏層節點數，根據準確率來決定我們下一層所使用的隱藏層節點數。

如表 4 可得知，隱藏層數目為 300 的時候，丈夫和太太被評分的 6 種行為平均準確率為最高，使用的疊代次數為 15 次， $\rho = 0.1$ ， $\lambda = 0.002$ ， $\beta = 2$ 。

接下來測試二層隱藏層的準確率，第一層隱藏數已經決定好了，我們測試的第二層隱藏層節點數，如表 5。從表中得知，第二層隱藏層的節點數為 200 的時候，準確率為最高，使用的疊代次數為 15 次， $\rho = 0.1$ ， $\lambda = 0.0001$ ， $\beta = 1$ 。

表 4. 1st hidden unit 分析丈夫和太太對應到 6 種行為的準確率，粗體字為較高的準確

1 st hidden unit	Rated Spouse	Acc (%)	Bla (%)	Pos (%)	Neg (%)	Sad (%)	Hum (%)	Avg (%)
100	Husband	67.9	76.4	65.7	78.6	52.9	61.4	67.2
	Wife	70	73.6	65	74.3	58.6	59.3	66.8
200	Husband	72.9	76.4	71.4	82.1	57.1	67.1	71.2
	Wife	71.4	82.9	65.7	77.1	64.9	57.9	70
300	Husband	77.1	77.9	72.1	82.9	58.6	67.1	72.6
	Wife	75.7	82.1	71.4	78.6	58.6	63.6	71.7
500	Husband	70	78.6	68.6	82.9	55	62.1	69.5
	Wife	74.3	82.1	69.3	80.7	58.6	62.9	71.3
1000	Husband	75	77.9	69.3	84.3	58.6	65.7	71.8
	Wife	72.1	79.3	69.3	80	53.6	62.9	69.5
Previous method(Black et al., 2013)	Husband	78.6	72.9	72.1	84.3	60	71.4	73.2
	Wife	77.9	84.3	74.3	80	66.4	67.1	75

表 5. 2nd hidden unit 分析丈夫和太太對應到 6 種 code 的準確率，粗體字為較高的準確率

1 st Hidden Layer	2 nd Hidden Layer	Rated Spouse	Acc (%)	Bla (%)	Pos (%)	Neg (%)	Sad (%)	Hum (%)	Avg (%)
300	100	Husband	75	78.6	68.6	83.6	57.9	67.9	71.9
		Wife	71.4	80.7	72.9	77.1	58.6	62.9	70.6
	200	Husband	77.1	77.1	71.4	83.6	57.9	69.3	72.7
		Wife	72.1	82.1	72	77.1	62.1	65.7	71.9
	300	Husband	73.6	76.4	72.1	84.3	58.6	67.1	72
		Wife	72.9	80.7	71.4	76.4	55	70	71.3
Previous method (Black et al., 2013)	Husband	78.6	72.9	72.1	84.3	60	71.4	73.2	
	Wife	77.9	84.3	74.3	80	66.4	67.1	75	

最後一層的節點數我們設為 150，如表 6，得到最後的準確率。使用的疊代次數為 20 次， $\rho = 0.1$ ， $\lambda = 0.0001$ ， $\beta = 1$ 。

表 6. 3rd hidden unit 分析丈夫和太太對應到 6 種 code 的準確率和之前研究準確率比較

1 st Hidden Layer	2 nd Hidden Layer	3 rd Hidden Layer	Rated Spouse	Acc (%)	Bla (%)	Pos (%)	Neg (%)	Sad (%)	Hum (%)	Avg (%)
300	200	150	Husband	80	78.6	73.6	84.3	59.3	73.6	74.9
			Wife	80	83.6	72.9	81.4	65	67.9	75
Previous method (Black <i>et al.</i> , 2013)			Husband	78.6	72.9	72.1	84.3	60	71.4	73.2
			Wife	77.9	84.3	74.3	80	66.4	67.1	75

4.4 實驗結果比較

我們所使用三種不同層的稀疏自編碼器和之前的論文整體平均準確率結果如表 7。

表 7. 整體平均正確率對於四種不同方法

Method		Avg(%)
Previous (Black <i>et al.</i> , 2013)		74.1
SSAE	One Layer	72.2
	Two Layers	72.3
	Three Layers	75.0

由表 7 中可得知 3 層的 SSAE 較之前研究提高 0.9%。之前研究使用 40479 個特徵值來作預測，而我們使用 2940 個特徵值，理論上看來較多的特徵值相對於準確率會較高，但透過深度學習的方式，降低數據的維度，找出相對關鍵的特徵，對於準確率的提升是有幫助的，從表 7 中看來雖然訓練 1 層和 2 層準確率表現沒有比較好，在使用 3 層之後就有好的表現，此論點由此可證。

5. 結論

現今存在越來越多資料庫，如何快速且準確預測資料，是近來研究的熱門議題。在這篇論文中，我們提出堆疊稀疏自編碼器改變特徵提取的方法和以為主體架構，來比較和之前研究的準確率，目的在藉由降低特徵數量，提升訊息的含量的方式，找到相對關鍵的特徵，來達到更好的準確率並減少訓練時間。最後結果也證明了利用非監督學習來訓練出新的一組特徵值，經由監督學習作分類，準確率較之前研究來的好，提出新的方法整體平均為 75% 高於舊的研究 74.1%，提升 0.9%。

本研究由於是對男女句子取出特徵值，再利用深度學習和機器學習去分析，準確率能有所提升，但是上升並不明顯還有改進空間。在未來，不管是透過改變非監督學習的演算法，或者是改變特徵擷取，只要產生出好的特徵值對資料的表示，準確率勢必會有更大的突破，把訓練好的模型套用在其他資料庫來講，也大大縮短人為預測所消耗的時間和成本。

參考文獻

- Andrew, G., & Gao, J. (2007). Scalable training of L 1-regularized log-linear Models. In *Proceedings of the 24th international conference on Machine learning. ACM*, 33-40.
- Black, M., Katsamanis, A., Baucom, B., Lee, C., Lammert, A., Christensen, A., Georgiou, P., & Narayanan, S. (2013). Toward automating a human behavioral coding system for married couples' interactions using speech acoustic features. *Speech Communication*, 55(1), 1-21.
- Burkhardt, F., Polzehl, T., Stegmann, J., Metze, F., & Huber, R. (2009). Detecting real life anger. In *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 4761-4764.
- Christensen, A., Atkins, D.C., Yi, J., Baucom, D.H., & George, W.H. (2004). Couple and individual adjustment for 2 years following a randomized clinical trial comparing traditional versus integrative behavioral couple therapy. *J. Consult. Clin. Psychol.*, 72, 176-191.
- Christensen, A., Jacobson, N.S., & Babcock, J.C. (1995). Integrative behavioral couple therapy. In: Jacobsen, N.S., Gurman, A.S. (Eds.), *Clinical Handbook of Marital Therapy, second ed. Guilford Press, New York*, 31-64.
- Devillers, L., & Campbell, N. (2011). Special issue of computer speech and language on affective speech in real-life interactions. *Comput. Speech Lang.*, 25, 1-3.
- Gonzaga, G.C., Campos, B., & Bradbury, T. (2007). Similarity, convergence, and relationship satisfaction in dating and married couples. *J. Personal. Soc. Psychol.*, 93, 34-48.
- Heavey, C., Gill, D., & Christensen, A. (2002). Couples interaction rating system 2 (CIRS2)., *University of California, Los Angeles. Los Angeles, CA, USA*.
- Hinton, G. (2006). Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(5786), 504-507.
- Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A., Jaitly, N., et al. (2012). Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Process. Mag.*, 29(6), 82-97.
- Jones, J., & Christensen, A. (1998). Couples interaction study: Social support interaction rating system. *University of California, Los Angeles. Los Angeles, CA, USA*.
- Karney, B.R., & Bradbury, T.N. (1995). The longitudinal course of marital quality and stability: A review of theory, methods, and research. *Psychol. Bull.*, 118, 3-34.

- O'Brian, M., John, R.S., Margolin, G., & Erel, O. (1994). Reliability and diagnostic efficacy of parent's reports regarding children's exposure to marital aggression. *Violence and Victims*, 9(1), 45-62.
- Obst, O. (2014). Distributed machine learning and sparse representations. *Neurocomputing*, 124, 1.
- Perwej, Y., & Chaturvedi, A. (2011). Machine recognition of Hand written Characters using neural networks. *International Journal of Computer Applications*, 14(2), 6-9.
- Rubanov, N. (2000). The layer-wise method and the backpropagation hybrid approach to learning a feedforward neural network. *IEEE Trans. Neural Netw.*, 11(2), 295-305.
- Schuller, B., Batliner, A., Seppi, D., Steidl, S., Vogt, T., Wagner, J., et al. (2007). The relevance of feature type for automatic classification of emotional user states: Low level descriptors and functionals. In *Proc. Interspeech*, Antwerp, Belgium, 2253-2256.
- Smirnov, E., Timoshenko, D., & Andrianov, S. (2014). Comparison of Regularization Methods for ImageNet Classification with Deep Convolutional Neural Networks. *AASRI Procedia*, 6, 89-94.

The individuals listed below are reviewers of this journal during the year of 2015. The IJCLCLP Editorial Board extends its gratitude to these volunteers for their important contributions to this publication, to our association, and to the profession.

Guo-Wei Bian	Tan Lee
Jing-Shin Chang	Bor-Shen Lin
Tao-Hsing Chang	Shu-Yen Lin
Yu-Yun Chang	Chao-Hong Liu
Yi-Hsiang Chao	Wei-Yun Ma
Chien Chin Chen	Wei-Ho Tsai
Yeou-Jiunn Chen	Yu Tsao
Pu-Jen Cheng	Chin-Chin Tseng
Tai-Shih Chi	Hsu Wang
Chih-Yi Chiu	Jenq-Haur Wang
Hong-Jie Dai	Jia-Ching Wang
Wei-Tyng Hong	Chih-Hsuan Wei
Jen-Wei Huang	Jiun-Shiung Wu
Jeih-Weih Hung	Cheng-Zen Yang
Chih-Chung Kuo	Jui-Feng Yeh
Wen-Hsing Lai	Ming-Shing Yu
Chi-Chun Lee	Yue Zhang
Hong-Yi Lee	

2015 Index
International Journal of Computational Linguistics &
Chinese Language Processing
Vol. 20

IJCLCLP 2015 Index-1

This index covers all technical items---papers, correspondence, reviews, etc.---that appeared in this periodical during 2015.

The Author Index contains the primary entry for each item, listed under the first author's name. The primary entry includes the coauthors' names, the title of paper or other item, and its location, specified by the publication volume, number, and inclusive pages. The Subject Index contains entries describing the item under all appropriate subject headings, plus the first author's name, the publication volume, number, and inclusive pages.

AUTHOR INDEX

C

Chang, Tao-Hsing

Yao-Ting Sung, and Jia-Fei Hong.
Automatically Detecting Syntactic Errors in
Sentences Writing by Learners of Chinese as a
Foreign Language; 20(1): 49-64

Chang, Ting-Hao

Hsiao-Tsung Hung, Kuan-Yu Chen, Hsin-Min
Wang and Berlin Chen. Investigating
Modulation Spectrum Factorization
Techniques for Robust Speech Recognition;
20(2): 87-106

Chen, Berlin

see Shih, Kai-Wun, 20(2): 65-86
see Chang, Ting-Hao, 20(2): 87-106

Chen, Guan-Bin

and Hung-Yu Kao. Word Co-occurrence
Augmented Topic Model in Short Text; 20(2):
45-64

Chen, Howard Hao-Jan

see Tung, Tzu-Yun, 20(1): 79-96

Chen, Kuan-Yu

see Shih, Kai-Wun, 20(2): 65-86
see Chang, Ting-Hao, 20(2): 87-106

Chen, Po-Hsuan

and Chi-Chun Lee. Automating Behavior
Coding for Distressed Couples Interactions
Based on Stacked Sparse Autoencoder
Framework using Speech-acoustic Features;
20(2): 107-120

Cheng, Xueqi

see Xiong, Jinhua, 20(1): 1-22

Chu, Wei-Cheng

see Lin, Chuan-Jie, 20(1): 23-48

H

Hong, Jia-Fei

see Chang, Tao-Hsing, 20(1): 49-64

Hou, Jianpeng

see Xiong, Jinhua, 20(1): 1-22

Hsu, Kuang-Yi

see Lin, Yi-Chung, 20(2): 1-26

Huang, Chien-Tsung

Yi-Chung Lin and Keh-Yih Su. Explanation
Generation for a Math Word Problem Solver;
20(2): 27-44
see Lin, Yi-Chung, 20(2): 1-26

Hung, Hsiao-Tsung

see Chang, Ting-Hao, 20(2): 87-106

K

Kao, Hung-Yu

see Chen, Guan-Bin, 20(2): 45-64

Ku, Lun-Wei

see Lin, Yi-Chung, 20(2): 1-26

L

Lee, Chi-Chun

see Chen, Po-Hsuan, 20(2): 107-120

Liang, Chao-Chun

see Lin, Yi-Chung, 20(2): 1-26

Liau, Churn-Jung

see Lin, Yi-Chung, 20(2): 1-26

Lin, Chuan-Jie

and Wei-Cheng Chu. A Study on Chinese
Spelling Check Using Confusion Sets and
N-gram Statistics; 20(1): 23-48

Lin, Yi-Chung

Chao-Chun Liang, Kuang-Yi Hsu, Chien-Tsung
Huang, Shen-Yun Miao, Wei-Yun Ma,
Lun-Wei Ku, Churn-Jung Liau and Keh-Yih
Su. Designing a Tag-Based Statistical Math
Word Problem Solver with Reasoning and
Explanation; 20(2): 1-26
see Huang, Chien-Tsung, 20(2): 27-44

Liu, Shih-Hung

see Shih, Kai-Wun, 20(2): 65-86

M

Ma, Wei-Yun

see Lin, Yi-Chung, 20(2): 1-26

Miao, Shen-Yun

see Lin, Yi-Chung, 20(2): 1-26

Mochizuki, Keiko

Hiroshi Sano, Ya-Ming Shen, and Chia-Hou Wu.
Cross-Linguistic Error Types of Misused
Chinese Based on Learners' Corpora; 20(1):
97-114

S

Sano, Hiroshi

see Mochizuki, Keiko, 20(1): 97-114

Shen, Ya-Ming

see Mochizuki, Keiko, 20(1): 97-114

Shih, Kai-Wun

Kuan-Yu Chen, Shih-Hung Liu, Hsin-Min Wang and Berlin Chen. Extractive Spoken Document Summarization with Representation Learning Techniques; 20(2): 65-86

Su, Keh-Yih

see Lin, Yi-Chung, 20(2): 1-26

see Huang, Chien-Tsung, 20(2): 27-44

Sung, Yao-Ting

see Chang, Tao-Hsing, 20(1): 49-64

T

Tung, Tzu-Yun

Howard Hao-Jan Chen, and Hui-Mei Yang. The Error Analysis of “Le” Based on “Chinese Learner Written Corpus”; 20(1): 79-96

W

Wang, Hsin-Min

see Shih, Kai-Wun, 20(2): 65-86

see Chang, Ting-Hao, 20(2): 87-106

Wu, Chia-Hou

see Mochizuki, Keiko, 20(1): 97-114

X

Xiong, Jinhua

Qiao Zhang, Shuiyuan Zhang, Jianpeng Hou, and Xueqi Cheng. HANSpeller: A Unified Framework for Chinese Spelling Correction; 20(1): 1-22

Y

Yang, Hui-Mei

see Tung, Tzu-Yun, 20(1): 79-96

Yeh, Chan-Kun

see Yeh, Jui-Feng, 20(1): 65-78

Yeh, Jui-Feng

and Chan-Kun Yeh. Automatic Classification of the “De” Word Usage for Chinese as a Foreign Language; 20(1): 65-78

Z

Zhang, Qiao

see Xiong, Jinhua, 20(1): 1-22

Zhang, Shuiyuan

see Xiong, Jinhua, 20(1): 1-22

SUBJECT INDEX

A

Annotation System

Cross-Linguistic Error Types of Misused Chinese Based on Learners’ Corpora; Mochizuki, K., 20(1): 97-114

C

Chinese Grammar

Automatically Detecting Syntactic Errors in Sentences Writing by Learners of Chinese as a Foreign Language; Chang, T.-H., 20(1): 49-64

Chinese Learner Written Corpus

The Error Analysis of “Le” based on “Chinese Learner Written Corpus”; Tung, T.-Y., 20(1): 79-96

Chinese Spelling Check

A Study on Chinese Spelling Check Using Confusion Sets and N-gram Statistics; Lin, C.-J., 20(1): 23-48

Chinese Spelling Correction

HANSpeller: A Unified Framework for Chinese Spelling Correction; Xiong, J., 20(1): 1-22

Chinese Teaching

The Error Analysis of “Le” based on “Chinese Learner Written Corpus”; Tung, T.-Y., 20(1): 79-96

Chinese Written Corpus

Automatically Detecting Syntactic Errors in Sentences Writing by Learners of Chinese as a Foreign Language; Chang, T.-H., 20(1): 49-64

Classifier

Automatic Classification of the “De” Word Usage for Chinese as a Foreign Language; Yeh, J.-F., 20(1): 65-78

Concept Information

Investigating Modulation Spectrum Factorization Techniques for Robust Speech Recognition; Chang, T.-H., 20(2): 87-106

Confusion Set Expansion

A Study on Chinese Spelling Check Using Confusion Sets and N-gram Statistics; Lin, C.-J., 20(1): 23-48

Couple Therapy

Automating Behavior Coding for Distressed Couples Interactions Based on Stacked Sparse Autoencoder Framework using Speech-acoustic Features; Chen, P.-H., 20(2): 107-120

D

Decision-making

HANSpeller: A Unified Framework for Chinese Spelling Correction; Xiong, J., 20(1): 1-22

Deep Learning

Automating Behavior Coding for Distressed Couples Interactions Based on Stacked Sparse Autoencoder Framework using Speech-acoustic Features; Chen, P.-H., 20(2): 107-120

Document Classification

Word Co-occurrence Augmented Topic Model in Short Text; Chen, G.-B., 20(2): 45-64

Document Clustering

Word Co-occurrence Augmented Topic Model in Short Text; Chen, G.-B., 20(2): 45-64

E**Emotion Expression**

Automating Behavior Coding for Distressed Couples Interactions Based on Stacked Sparse Autoencoder Framework using Speech-acoustic Features; Chen, P.-H., 20(2): 107-120

Error Analysis

The Error Analysis of “Le”ased on “Chinese Learner Written Corpus”; Tung, T.-Y., 20(1): 79-96

Cross-Linguistic Error Types of Misused Chinese Based on Learners’ Corpora; Mochizuki, K., 20(1): 97-114

Explanation Generation

Explanation Generation for a Math Word Problem Solver; Huang, C.-T., 20(2): 27-44

Extractive Summarization

Extractive Spoken Document Summarization with Representation Learning Techniques; Shih, K.-W., 20(2): 65-86

G**Google Ngram Scoring Function**

A Study on Chinese Spelling Check Using Confusion Sets and N-gram Statistics; Lin, C.-J., 20(1): 23-48

H**HMM**

HANSpeller: A Unified Framework for Chinese Spelling Correction; Xiong, J., 20(1): 1-22

Human Behavior Analysis

Automating Behavior Coding for Distressed Couples Interactions Based on Stacked Sparse Autoencoder Framework using Speech-acoustic Features; Chen, P.-H., 20(2): 107-120

I**Interference of Mother Tongues**

Cross-Linguistic Error Types of Misused Chinese Based on Learners’ Corpora; Mochizuki, K., 20(1): 97-114

L**Language Model**

Investigating Modulation Spectrum Factorization Techniques for Robust Speech Recognition; Chang, T.-H., 20(2): 87-106

“le”

The Error Analysis of “Le”ased on “Chinese Learner Written Corpus”; Tung, T.-Y., 20(1): 79-96

Learner’s Corpus

Cross-Linguistic Error Types of Misused Chinese Based on Learners’ Corpora; Mochizuki, K., 20(1): 97-114

M**Machine Reading**

Designing a Tag-Based Statistical Math Word Problem Solver with Reasoning and Explanation; Lin, Y.-C., 20(2): 1-26

Explanation Generation for a Math Word Problem Solver; Huang, C.-T., 20(2): 27-44

Math Word Problem Explanation

Explanation Generation for a Math Word Problem Solver; Huang, C.-T., 20(2): 27-44

Math Word Problem Solver

Designing a Tag-Based Statistical Math Word Problem Solver with Reasoning and Explanation; Lin, Y.-C., 20(2): 1-26

Model Adaptation

Investigating Modulation Spectrum Factorization Techniques for Robust Speech Recognition; Chang, T.-H., 20(2): 87-106

N**Natural Language Processing**

Automatic Classification of the “De” Word Usage for Chinese as a Foreign Language; Yeh, J.-F., 20(1): 65-78

Natural Language Understanding

Designing a Tag-Based Statistical Math Word Problem Solver with Reasoning and Explanation; Lin, Y.-C., 20(2): 1-26

O**Online Dictionary of Misused Chinese based on Learners’ Corpora**

Cross-Linguistic Error Types of Misused Chinese Based on Learners’ Corpora; Mochizuki, K., 20(1): 97-114

P**Prosodic Feature**

Extractive Spoken Document Summarization with Representation Learning Techniques; Shih, K.-W., 20(2): 65-86

R

Ranker-Base Model

HANSpeller: A Unified Framework for Chinese Spelling Correction; Xiong, J., 20(1): 1-22

Rule Induction

Automatic Classification of the “De” Word Usage for Chinese as a Foreign Language; Yeh, J.-F., 20(1): 65-78

Rule-based Model

HANSpeller: A Unified Framework for Chinese Spelling Correction; Xiong, J., 20(1): 1-22

S

Secondary Language Learning

Automatic Classification of the “De” Word Usage for Chinese as a Foreign Language; Yeh, J.-F., 20(1): 65-78

Sentence Representation

Extractive Spoken Document Summarization with Representation Learning Techniques; Shih, K.-W., 20(2): 65-86

Short Text

Word Co-occurrence Augmented Topic Model in Short Text; Chen, G.-B., 20(2): 45-64

Speech Recognition

Investigating Modulation Spectrum Factorization Techniques for Robust Speech Recognition; Chang, T.-H., 20(2): 87-106

Spoken Document

Extractive Spoken Document Summarization with Representation Learning Techniques; Shih, K.-W., 20(2): 65-86

Stacked Autoencoders

Automating Behavior Coding for Distressed Couples Interactions Based on Stacked Sparse Autoencoder Framework using Speech-acoustic Features; Chen, P.-H., 20(2): 107-120

Syntactic Errors

Automatically Detecting Syntactic Errors in Sentences Writing by Learners of Chinese as a Foreign Language; Chang, T.-H., 20(1): 49-64

T

Topic Model

Word Co-occurrence Augmented Topic Model in Short Text; Chen, G.-B., 20(2): 45-64

W

Word Representation

Extractive Spoken Document Summarization with Representation Learning Techniques; Shih, K.-W., 20(2): 65-86

Word Usage

Automatic Classification of the “De” Word Usage for Chinese as a Foreign Language; Yeh, J.-F., 20(1): 65-78

The Association for Computational Linguistics and Chinese Language Processing

(new members are welcomed)

Aims :

1. To conduct research in computational linguistics.
2. To promote the utilization and development of computational linguistics.
3. To encourage research in and development of the field of Chinese computational linguistics both domestically and internationally.
4. To maintain contact with international groups who have similar goals and to cultivate academic exchange.

Activities :

1. Holding the Republic of China Computational Linguistics Conference (ROCLING) annually.
2. Facilitating and promoting academic research, seminars, training, discussions, comparative evaluations and other activities related to computational linguistics.
3. Collecting information and materials on recent developments in the field of computational linguistics, domestically and internationally.
4. Publishing pertinent journals, proceedings and newsletters.
5. Setting of the Chinese-language technical terminology and symbols related to computational linguistics.
6. Maintaining contact with international computational linguistics academic organizations.
7. Dealing with various other matters related to the development of computational linguistics.

To Register :

Please send application to:

The Association for Computational Linguistics and Chinese Language Processing
Institute of Information Science, Academia Sinica
128, Sec. 2, Academy Rd., Nankang, Taipei 11529, Taiwan, R.O.C.

payment : Credit cards(please fill in the order form), cheque, or money orders.

Annual Fees :

regular/overseas member : NT\$ 1,000 (US\$50.-)

group membership : NT\$20,000 (US\$1,000.-)

life member : ten times the annual fee for regular/ group/ overseas members

Contact :

Address : The Association for Computational Linguistics and Chinese Language Processing
Institute of Information Science, Academia Sinica
128, Sec. 2, Academy Rd., Nankang, Taipei 11529, Taiwan, R.O.C.

Tel. : 886-2-2788-3799 ext. 1502 Fax : 886-2-2788-1638

E-mail: acclcp@hp.iis.sinica.edu.tw Web Site: <http://www.acclcp.org.tw>

Please address all correspondence to Miss Qi Huang, or Miss Abby Ho

The Association for Computational Linguistics and Chinese Language Processing

Membership Application Form

Member ID# : _____

Name : _____ Date of Birth : _____

Country of Residence : _____ Province/State : _____

Passport No. : _____ Sex: _____

Education(highest degree obtained) : _____

Work Experience : _____

Present Occupation : _____

Address : _____

Email Add : _____

Tel. No : _____ Fax No : _____

Membership Category : Regular Member Life Member

Date : ____/____/____ (Y-M-D)

Applicant's Signature :

Remarks : Please indicated clearly in which membership category you wish to register,
according to the following scale of annual membership dues :

Regular Member : US\$ 50.- (NT\$ 1,000)

Life Member : US\$500.- (NT\$10,000)

Please feel free to make copies of this application for others to use.

Committee Assessment :

中華民國計算語言學學會

宗旨：

- (一) 從事計算語言學之研究
- (二) 推行計算語言學之應用與發展
- (三) 促進國內外中文計算語言學之研究與發展
- (四) 聯繫國際有關組織並推動學術交流

活動項目：

- (一) 定期舉辦中華民國計算語言學學術會議 (Rocling)
- (二) 舉行有關計算語言學之學術研究講習、訓練、討論、觀摩等活動項目
- (三) 收集國內外有關計算語言學知識之圖書及最新發展之資料
- (四) 發行有關之學術刊物，論文集及通訊
- (五) 研定有關計算語言學專用名稱術語及符號
- (六) 與國際計算語言學學術機構聯繫交流
- (七) 其他有關計算語言發展事項

報名方式：

1. 入會申請書：請至本會網頁下載入會申請表，填妥後郵寄或E-mail至本會
2. 繳交會費：劃撥：帳號：19166251，戶名：中華民國計算語言學學會
信用卡：請至本會網頁下載信用卡付款單

年費：

- 終身會員： 10,000.- (US\$ 500.-)
- 個人會員： 1,000.- (US\$ 50.-)
- 學生會員： 500.- (限國內學生)
- 團體會員： 20,000.- (US\$ 1,000.-)

連絡處：

地址：台北市115南港區研究院路二段128號 中研院資訊所(轉)
電話：(02) 2788-3799 ext.1502 傳真：(02) 2788-1638
E-mail：aclclp@hp.iis.sinica.edu.tw 網址：<http://www.aclclp.org.tw>
連絡人：黃琪 小姐、何婉如 小姐

中華民國計算語言學學會 個人會員入會申請書

會員類別	<input type="checkbox"/> 終身 <input type="checkbox"/> 個人 <input type="checkbox"/> 學生	會員編號	(由本會填寫)	
姓名		性別	出生日期	年 月 日
			身分證號碼	
現職		學歷		
通訊地址	□□□			
戶籍地址	□□□			
電話		E-Mail		
申請人：			(簽章)	
中華民國 年 月 日				

審查結果：

1. 年費：

- 終身會員： 10,000.-
- 個人會員： 1,000.-
- 學生會員： 500.- (限國內學生)
- 團體會員： 20,000.-

2. 連絡處：

地址：台北市南港區研究院路二段128號 中研院資訊所(轉)
 電話：(02) 2788-3799 ext.1502 傳真：(02) 2788-1638
 E-mail：acclcp@hp.iis.sinica.edu.tw 網址：<http://www.acclcp.org.tw>
 連絡人：黃琪 小姐、何婉如 小姐

3. 本表可自行影印

The Association for Computational Linguistics and Chinese Language Processing (ACLCLP) PAYMENT FORM

Name: _____(Please print) Date: _____

Please debit my credit card as follows: US\$ _____

VISA CARD MASTER CARD JCB CARD Issue Bank: _____

Card No.: _____ - _____ - _____ - _____ Exp. Date: _____(M/Y)

3-digit code: _____ (on the back card, inside the signature area, the last three digits)

CARD HOLDER SIGNATURE: _____

Phone No.: _____ E-mail: _____

Address: _____

PAYMENT FOR

US\$ _____ Computational Linguistics & Chinese Languages Processing (IJCLCLP)

Quantity Wanted: _____

US\$ _____ Journal of Information Science and Engineering (JISE)

Quantity Wanted: _____

US\$ _____ Publications: _____

US\$ _____ Text Corpora: _____

US\$ _____ Speech Corpora: _____

US\$ _____ Others: _____

US\$ _____ Membership Fees Life Membership New Membership Renew

US\$ _____ = Total

Fax 886-2-2788-1638 or Mail this form to:

ACLCLP

% IIS, Academia Sinica

Rm502, No.128, Sec.2, Academia Rd., Nankang, Taipei 115, Taiwan

E-mail: aclclp@hp.iis.sinica.edu.tw

Website: <http://www.aclclp.org.tw>

中華民國計算語言學學會 信用卡付款單

姓名：_____ (請以正楷書寫) 日期：_____

卡別： VISA CARD MASTER CARD JCB CARD 發卡銀行：_____

信用卡號：_____ - _____ - _____ - _____ 有效日期：_____ (m/y)

卡片後三碼：_____ (卡片背面簽名欄上數字後三碼)

持卡人簽名：_____ (簽名方式請與信用卡背面相同)

通訊地址：_____

聯絡電話：_____ E-mail：_____

備註：為順利取得信用卡授權，請提供與發卡銀行相同之聯絡資料。

付款內容及金額：

NT\$ _____ 中文計算語言學期刊(IJCLCLP) _____

NT\$ _____ Journal of Information Science and Engineering (JISE)

NT\$ _____ 中研院詞庫小組技術報告 _____

NT\$ _____ 文字語料庫 _____

NT\$ _____ 語音資料庫 _____

NT\$ _____ 光華雜誌語料庫1976~2010

NT\$ _____ 中文資訊檢索標竿測試集/文件集

NT\$ _____ 會員年費： 續會 新會員 終身會員

NT\$ _____ 其他：_____

NT\$ _____ = 合計

填妥後請傳真至 02-27881638 或郵寄至：

11529台北市南港區研究院路2段128號中研院資訊所(轉)中華民國計算語言學學會 收

E-mail: aclclp@hp.iis.sinica.edu.tw

Website: <http://www.aclclp.org.tw>

Publications of the Association for Computational Linguistics and Chinese Language Processing

	<u>Surface</u>	<u>AIR</u> <u>(US&EURP)</u>	<u>AIR</u> <u>(ASIA)</u>	<u>VOLUME</u>	<u>AMOUNT</u>
1. no.92-01, no. 92-04(合訂本) ICG 中的論旨角色與 A Conceptual Structure for Parsing Mandarin -- Its Frame and General Applications--	US\$ 9	US\$ 19	US\$15	_____	_____
2. no.92-02 V-N 複合名詞討論篇 & 92-03 V-R 複合動詞討論篇	12	21	17	_____	_____
3. no.93-01 新聞語料庫字頻統計表	8	13	11	_____	_____
4. no.93-02 新聞語料庫詞頻統計表	18	30	24	_____	_____
5. no.93-03 新聞常用動詞詞頻與分類	10	15	13	_____	_____
6. no.93-05 中文詞類分析	10	15	13	_____	_____
7. no.93-06 現代漢語中的法相詞	5	10	8	_____	_____
8. no.94-01 中文書面語頻率詞典 (新聞語料詞頻統計)	18	30	24	_____	_____
9. no.94-02 古漢語字頻表	11	16	14	_____	_____
10. no.95-01 注音檢索現代漢語字頻表	8	13	10	_____	_____
11. no.95-02/98-04 中央研究院平衡語料庫的內容與說明	3	8	6	_____	_____
12. no.95-03 訊息為本的格位語法與其剖析方法	3	8	6	_____	_____
13. no.96-01 「搜」文解字—中文詞界研究與資訊用分詞標準	8	13	11	_____	_____
14. no.97-01 古漢語詞頻表 (甲)	19	31	25	_____	_____
15. no.97-02 論語詞頻表	9	14	12	_____	_____
16. no.98-01 詞頻詞典	18	30	26	_____	_____
17. no.98-02 Accumulated Word Frequency in CKIP Corpus	15	25	21	_____	_____
18. no.98-03 自然語言處理及計算語言學相關術語中英對譯表	4	9	7	_____	_____
19. no.02-01 現代漢語口語對話語料庫標註系統說明	8	13	11	_____	_____
20. Computational Linguistics & Chinese Languages Processing (One year) (Back issues of <i>IJCLCLP</i> : US\$ 20 per copy)	---	100	100	_____	_____
21. Readings in Chinese Language Processing	25	25	21	_____	_____
TOTAL				_____	_____

10% member discount: _____ **Total Due:** _____

• **OVERSEAS USE ONLY**

- PAYMENT : Credit Card (Preferred)
 Money Order or Check payable to "The Association for Computation Linguistics and Chinese Language Processing " or “中華民國計算語言學學會”

• E-mail : aclclp@hp.iis.sinica.edu.tw

Name (please print): _____ Signature: _____

Fax: _____ E-mail: _____

Address : _____

中華民國計算語言學學會 相關出版品價格表及訂購單

編號	書目	會員	非會員	冊數	金額
1.	no.92-01, no. 92-04 (合訂本) ICG 中的論旨角色 與 A conceptual Structure for Parsing Mandarin--its Frame and General Applications--	NT\$ 80	NT\$ 100	_____	_____
2.	no.92-02, no. 92-03 (合訂本) V-N 複合名詞討論篇 與 V-R 複合動詞討論篇	120	150	_____	_____
3.	no.93-01 新聞語料庫字頻統計表	120	130	_____	_____
4.	no.93-02 新聞語料庫詞頻統計表	360	400	_____	_____
5.	no.93-03 新聞常用動詞詞頻與分類	180	200	_____	_____
6.	no.93-05 中文詞類分析	185	205	_____	_____
7.	no.93-06 現代漢語中的法相詞	40	50	_____	_____
8.	no.94-01 中文書面語頻率詞典 (新聞語料詞頻統計)	380	450	_____	_____
9.	no.94-02 古漢語字頻表	180	200	_____	_____
10.	no.95-01 注音檢索現代漢語字頻表	75	85	_____	_____
11.	no.95-02/98-04 中央研究院平衡語料庫的內容與說明	75	85	_____	_____
12.	no.95-03 訊息為本的格位語法與其剖析方法	75	80	_____	_____
13.	no.96-01 「搜」文解字—中文詞界研究與資訊用分詞標準	110	120	_____	_____
14.	no.97-01 古漢語詞頻表 (甲)	400	450	_____	_____
15.	no.97-02 論語詞頻表	90	100	_____	_____
16.	no.98-01 詞頻詞典	395	440	_____	_____
17.	no.98-02 Accumulated Word Frequency in CKIP Corpus	340	380	_____	_____
18.	no.98-03 自然語言處理及計算語言學相關術語中英對譯表	90	100	_____	_____
19.	no.02-01 現代漢語口語對話語料庫標註系統說明	75	85	_____	_____
20.	論文集 COLING 2002 紙本	100	200	_____	_____
21.	論文集 COLING 2002 光碟片	300	400	_____	_____
22.	論文集 COLING 2002 Workshop 光碟片	300	400	_____	_____
23.	論文集 ISCSLP 2002 光碟片	300	400	_____	_____
24.	交談系統暨語境分析研討會講義 (中華民國計算語言學學會1997第四季學術活動)	130	150	_____	_____
25.	中文計算語言學期刊 (一年四期) 年份: _____ (過期期刊每本售價500元)	---	2,500	_____	_____
26.	Readings of Chinese Language Processing	675	675	_____	_____
27.	剖析策略與機器翻譯 1990	150	165	_____	_____
			合 計	_____	_____

※ 此價格表僅限國內 (台灣地區) 使用

劃撥帳戶：中華民國計算語言學學會 劃撥帳號：19166251

聯絡電話：(02) 2788-3799 轉1502

聯絡人：黃琪 小姐、何婉如 小姐 E-mail: acclcp@hp.iis.sinica.edu.tw

訂購者：_____ 收據抬頭：_____

地 址：_____

電 話：_____ E-mail: _____

Information for Authors

International Journal of Computational Linguistics and Chinese Language Processing (IJCLCLP) invites submission of original research papers in the area of computational linguistics and speech/text processing of natural language. All papers must be written in English or Chinese. Manuscripts submitted must be previously unpublished and cannot be under consideration elsewhere. Submissions should report significant new research results in computational linguistics, speech and language processing or new system implementation involving significant theoretical and/or technological innovation. The submitted papers are divided into the categories of regular papers, short paper, and survey papers. Regular papers are expected to explore a research topic in full details. Short papers can focus on a smaller research issue. And survey papers should cover emerging research trends and have a tutorial or review nature of sufficiently large interest to the Journal audience. There is no strict length limitation on the regular and survey papers. But it is suggested that the manuscript should not exceed 40 double-spaced A4 pages. In contrast, short papers are restricted to no more than 20 double-spaced A4 pages. All contributions will be anonymously reviewed by at least two reviewers.

Copyright : It is the author's responsibility to obtain written permission from both author and publisher to reproduce material which has appeared in another publication. Copies of this permission must also be enclosed with the manuscript. It is the policy of the CLCLP society to own the copyright to all its publications in order to facilitate the appropriate reuse and sharing of their academic content. A signed copy of the IJCLCLP copyright form, which transfers copyright from the authors (or their employers, if they hold the copyright) to the CLCLP society, will be required before the manuscript can be accepted for publication. The papers published by IJCLCLP will be also accessed online via the IJCLCLP official website and the contracted electronic database services.

Style for Manuscripts: The paper should conform to the following instructions.

1. Typescript: Manuscript should be typed double-spaced on standard A4 (or letter-size) white paper using size of 11 points or larger.

2. Title and Author: The first page of the manuscript should consist of the title, the authors' names and institutional affiliations, the abstract, and the corresponding author's address, telephone and fax numbers, and e-mail address. The title of the paper should use normal capitalization. Capitalize only the first words and such other words as the orthography of the language requires beginning with a capital letter. The author's name should appear below the title.

3. Abstracts and keywords: An informative abstract of not more than 250 words, together with 4 to 6 keywords is required. The abstract should not only indicate the scope of the paper but should also summarize the author's conclusions.

4. Headings: Headings for sections should be numbered in Arabic numerals (i.e. 1.,2....) and start from the left-hand margin. Headings for subsections should also be numbered in Arabic numerals (i.e. 1.1. 1.2...).

5. Footnotes: The footnote reference number should be kept to a minimum and indicated in the text with superscript numbers. Footnotes may appear at the end of manuscript

6. Equations and Mathematical Formulas: All equations and mathematical formulas should be typewritten or written clearly in ink. Equations should be numbered serially on the right-hand side by Arabic numerals in parentheses.

7. References: All the citations and references should follow the APA format. The basic form for a reference looks like

Authora, A. A., Authorb, B. B., & Authorc, C. C. (Year). Title of article. *Title of Periodical*, volume number(issue number), pages.

Here shows an example.

Scruton, R. (1996). The eclipse of listening. *The New Criterion*, 15(30), 5-13.

The basic form for a citation looks like (Authora, Authorb, and Authorc, Year). Here shows an example. (Scruton, 1996).

Please visit the following websites for details.

(1) APA Formatting and Style Guide (<http://owl.english.purdue.edu/owl/resource/560/01/>)

(2) APA Style (<http://www.apastyle.org/>)

No page charges are levied on authors or their institutions.

Final Manuscripts Submission: If a manuscript is accepted for publication, the author will be asked to supply final manuscript in MS Word or PDF files to clp@hp.iis.sinica.edu.tw

Online Submission: <http://www.aclclp.org.tw/journal/submit.php>

Please visit the IJCLCLP Web page at <http://www.aclclp.org.tw/journal/index.php>

