

Automatic Semantic Classification for Chinese Unknown Compound Nouns

Keh-Jiann Chen & Chao-jan Chen

Institute of Information Science, Academia Sinica, Taipei

Abstract

The paper describes a similarity-based model to present the morphological rules for Chinese compound nouns. This representation model serves functions of 1) as the morphological rules of the compounds, 2) as a mean to evaluate the properness of a compound construction, and 3) as a mean to disambiguate the semantic ambiguity of the morphological head of a compound noun. An automatic semantic classification system for Chinese unknown compounds is thus implemented based on the model. Experiments and error analyses are also presented.

1. Introduction

The occurrences of unknown words cause difficulties in natural language processing. The word set of a natural language is open-ended. There is no way of collecting every words of a language, since new words will be created for expressing new concepts, new inventions. Therefore how to identify new words in a text will be the most challenging task for natural language processing. It is especially true for Chinese. Each Chinese morpheme (usually a single character) carries meanings and most are polysemous. New words are easily constructed by combining morphemes and their meanings are the semantic composition of morpheme components. Of course there are exceptions of semantically non-compositional compounds. In Chinese text, there is no blank to mark word boundaries and no inflectional markers nor capitalization markers to denote the syntactic or semantic types of new words. Hence the unknown word identification for Chinese became one of the most difficult and demanding research topic.

The syntactic and semantic categories of unknown words in principle can be determined by their content and contextual information. However many difficult problems have to be solved. First of all it is not possible to find a uniform representational schema and categorization algorithm to handle different types of unknown words, since each type of unknown words has very much different morpho-syntactic structures. Second, the clues for

identifying different type of unknown words are also different. For instance, identification of names of Chinese is very much relied on the surnames, which is a limited set of characters. The statistical methods are commonly used for identifying proper names (Chang et al. 1994, Sun et al. 1994). The identification of general compounds is more relied on the morphemes and the semantic relations between morphemes. There are co-occurrence restrictions between morphemes of compounds, but their relations are irregular and mostly due to common sense knowledge. The third difficulty is the problems of ambiguities, such as structure ambiguities, syntactic ambiguities and semantic ambiguities. For instances, usually a morpheme character/word has multiple meaning and syntactic categories. Therefore the ambiguity resolution became one of the major tasks.

Compound nouns are the most frequently occurred unknown words in Chinese text. According to an inspection on the Sinica corpus (Chen etc. 1996), 3.51% of the word tokens in the corpus are unknown, i.e. they are not listed in the CKIP lexicon, which contains about 80,000 entries. Among them, about 51% of the word types are compound nouns, 34% are compound verbs and 15% are proper names. In this paper we focus our attention on the identification of the compound nouns. We propose a representation model, which will be facilitated to identify, to disambiguate and to evaluate the structure of a compound noun. In fact this model can be extended to handle compound verbs also.

1.1 General properties of compounds and their identification strategy

The semantic category and syntactic category are closely related. For coarse-grained analysis, syntactic categorization and semantic categorization are close related. For instances, nouns denote entities; active verbs denote events and stative verbs denote states. For fine-grained analysis, syntactic and semantic classifications take different classification criterion. In our model the coarse-grained analysis is processed first. The syntactic categories of an unknown

word are predicted first and the possible semantic categories will be identified according to its top ranked syntactic categories. Different syntactic categories require different representational models and different fine-grained semantic classification methods.

The presupposition of automatic semantic classification for compounds is that the meaning of a compound is the semantic composition of its morphemic components and the head morpheme determines the major semantic class of this compound. There are many polysyllabic words of which the property of semantic composition does not hold, for instances the transliteration words, those words should be listed in the lexicon. Since for the majority of compounds the presupposition hold, the design of our semantic classification algorithm will be based upon this presupposition. Therefore the process of identifying semantic class of a compound boils down to find and to determine the semantic class of its head morpheme. However ambiguous morphological structures cause the difficulties in finding head morpheme. For instances, the compound in 1a) has two possible morphological structures, but only 1b) is the right interpretation.

- 1a) 美國人 ‘American’
- b) 美國 人 ‘America’ ‘people’ ,
- c) 美 國人 ‘beautiful’ ‘country-man’

Once the morphological head is determined, the semantic resolution for the head morpheme is the next difficulty to be solved. About 51.5% of the 200 most productive morphemes are polysemous and according to the Collocation Dictionary of Noun and Measure Words (CDNM), in average each ambiguous morpheme carries 3.5 different senses (Huang et al. 1997).

2. Representation Models

Compounds are very productive types of unknown words. Nominal and verbal compounds are easily coined by combining two/many words/characters. Since there are more than 5000 commonly used Chinese characters and each with idiosyncratic

syntactic behaviors, it is hard to derive a set of morphological rules to generate the set of Chinese compounds without over-generation or under-generation. The set of general compounds is an open-class. The strategy for automatic identification will be relied not only on the morpho-syntactic structures but also morpho-semantic relations. In general, certain interpretable semantic relationships between morphemic must be held. However there is no absolute means to judge whether the semantic relations between morphemic components are acceptable, i.e. the acceptability of such type of compounds is not simply ‘yes’ or ‘no’. The degree of properness of a compound should depend on the logical relation between morphemic components and their logicalness should be judged by common sense knowledge. It is almost impossible to implement a system with common sense knowledge. Chen & Chen (1998) proposed an example-based measurement to evaluate the properness of a newly coined compound instead. They postulate that for a newly coined compound, if the semantic relation of its morphemic components is similar to the existing compounds, then it is more likely that this newly coined compound is proper.

2.1 Example-based similarity measure

Supposed that a compound has the structure of XY where X and Y are morphemes and supposed without loss of generality Y is the head. For instance, 習字機 ‘learn-word-machine’ is a noun compound and the head morpheme Y is 機 ‘machine’ and the modifier X is 習字 ‘learn-word’. In fact the morpheme 機 has four different meanings. They are ‘machine’, ‘airplane’, ‘secret’ and ‘opportunity’. How do computers judge which one is the right meaning and how is the compound construction well-formed or logically meaningful? First of all, the examples with the head morpheme 機 are extracted from corpora and dictionaries. The examples are classified according to their meaning as shown in the Table 1.

Senses	semantic category	examples
(1) machine <機器>	Bo011	沖洗機 照相機 攝影機 錄音機 處理機
(2) airplane <飛機>	Bo223	飛機 軍機 客機 專機 轟炸機 直昇機
(3) opportunity <機會>	Ca042	危機 時機 契機 轉機 良機 先機
(4) secret <機密>	Da011	商機 先機 玄機 禪機 天機

Table 1. Four senses of the morpheme ‘機’ and their respective samples

The meaning of 習字機 is then determined by comparing the similarity between it and each class of examples. The meaning of the input unknown word will be assigned with the meaning of the class with the most similar morpho-semantic structures with this unknown word. The similarity measure is based on the following formula.

Supposed that each class of examples forms the following semantic relation rules. The rules show the possible semantic relations between prefix and suffix Y and their weight in term of the frequency distribution of each semantic category of the prefixes in the class.

Rules: Sem1 + Y Freq1
 Sem2 + Y Freq2
 ⋮ ⋮
 Semk + Y Freqk

(freqi: the number of the words of the form Semi + Y)

Take suffix 機 with meaning of 'machine' as example. For the morpheme 機 'machine', the extracted compounds of the form X+機 'machine' and the semantic categories of the modifiers are shown in Table 2 and the morphological rule derived from them is in Table 3. The semantic types and their hierarchical structure are adopted from the Chilin (Mei et al. 1984). The similarity is measured between the semantic class of the prefix X of the unknown compound and the prefix semantic types shown in the rule. One of the measurements proposed is:

$SIMILAR (Sem, Rule) = \{ \sum_{i=1,k} Information-Load(Sem \cap Semi) * Freq_i \} / Max-value$
 Where Sem is the semantic class of X. Max-value

X-機 Semi freqi

Hh031	3	Sem(習字) = Hg111
Ae173	1	Information-Load(Hg111 \cap Hh031)=Information-Load(H)=2.231
He032	2	Information-Load(Hg111 \cap He032)=Information-Load(H)=2.231
Eb342	2	Information-Load(Hg111 \cap Hg191)=Information-Load(Hg)=5.91
Ae162	1	$\sum_{i=1,k} Information-Load(Hg111 \cap Semi) * Freq_i$
Hg191	1	= 2.231*3 + 2.231* 2 + 5.912 * 1 +
:	:	= 104.632
:	:	Max-Value1 = $\sum_i Information-Load(Hg031 \cap Semi) * Freq_i$
:	:	= 155.164
:	:	SIMILAR= (104.632 / 155.164) = 0.6743

Table 3. The derived morphological rule for the morpheme 'machine' and the similarity measure of '習字機' as a noun compound which denotes a kind of machine.

is the maximal value of $\{ \sum Information Load(S \cap Semi) * Freq_i \}$ for all semantic classes S. The max-value normalizes the SIMILAR value to 0~1. $S \cap Semi$ denotes the least common ancestor of S and Semi. For instance, (Hh03 \cap Hb06) = H. The Information-Load(S) of a semantic class S is defined as Entropy(semantic system) – Entropy(S). Simply speaking it is the amount of reduced entropy after S is seen. Entropy(S) = $\sum_{i=1,k} -P(Semi|S) * Log P(Semi|S)$, where {Sem1, Sem2,...,Semk} is the set of the bottom level semantic classes contained in S.

沖洗(機) Fa211 Hh031	投影(機) Bg033
照相(機) Hh031	抽水(機) Ih033
攝影(機) Ae173 Hh031	石版(機) Bo171
錄音(機) Ae173 Hh032	升降(機) If083 Ih063
處理(機) Hc013 He032	吹風(機) Hi141 Hj345
販賣(機) He032	研磨(機) Fa221
自動(機) Eb342	接收(機) Hc071
電動(機) Eb342	放大(機) Ih031
翻譯(機) Ae162 Hg162	冷氣(機) Bg051
傳真(機) Dk162 Hg191	烘乾(機) Ih042
幻燈(機) Bo041	繪圖(機) Hg192
紡織(機) Hd121	升高(機) Ih061
時光(機) Ca032 Ca041	跑步(機) Fb012
發動(機) Hc122	開票(機) Hc231
果汁(機) Br121	掃除(機) Fa212 Jd102

Table 2. The semantic categories of modifiers of the compounds of X-“機 machine”

Take the word 習字機 'learning-word-machine' as example. In the table 3, the results show the estimated similarity between the

compound 習字機 and the extracted examples. The similarity value is also considered as the logical properness value of this compound. In this case is 0.67, which can be interpreted as that we have 67% of confidence to say that 習字機 ‘learning-word-machine’ is a well-formed compound.

The above representation model serves many functions. First of all it serves as the morphological rules of the compounds. Second it serves as a mean to implement the evaluation function. Third it serves as a mean to disambiguate the semantic ambiguity of the morphological head of a compound noun. For instance, there are four different 機. Each denotes ‘machine’, ‘airplane’, ‘opportunity’ and ‘secret’ and they are considered as four different morphemes. The example shows that ‘習字機’ denotes a machine not other senses, since the evaluation score for matching the rules of ‘機machine’ has the highest evaluation score among them.

The above discussion shows the basic concept and the base-line model of the example-based model. The above similarity measure is called over-all-similarity measure, since it takes the equal weight on the similarity values of the input compound with every member in the class. Another similarity measure is called maximal-similarity, which is defined as follows. It takes the maximal value of the similarity between input compound and every member in the class as the output.

$$\text{SIM2}(\text{Word}, \text{Rule}) = \text{Max}_{i=1,k} \{ (\text{Information Load}(\text{Sem} \cap \text{Semi})) / \text{Max-value}_2 \}$$

Both similarity measures are reasonable and have their own advantages. The experiment results showed that the combination of these two measures achieved the best performance on the weights of $w_1=0.3$ and $w_2=0.7$ (Chen & Chen 1998), i.e. $\text{SIM} = \text{SIM1} * w_1 + \text{SIM2} * w_2$, where $w_1+w_2 = 1$. We adopt this measure in our experiments.

It also showed a strong co-relation between the similarity scores and the human evaluation scores on the properness of testing compounds. The human considered bad compounds showed also low similarity scores by computers.

3. System Implementation

3.1 Knowledge sources

To categorize unknown words, the computer system has to equip with the linguistic and semantic knowledge about words, morphemes, and word formation rules. The knowledge is facilitated to identify words, to categorize their semantic and syntactic classes, and to evaluate the properness of

word formation and the confidence level of categorization. In our experiments, the available knowledge sources include:

- 1) CKIP lexicon: an 80,000 entry Chinese lexicon with syntactic categories for each entry (CKIP 1993).
- 2) Chilin: a thesaurus of synonym classes, which contains about 70,000 words distributed under 1428 semantic classes (Mei 1984).
- 3) Sinica Corpus: a 5 million word balanced Chinese corpus with word segmented and part-of-speech tagged (Chen 1996).
- 4) the Collocation Dictionary of Noun and Measure Words (CDNM) : The CDNM lists collocating measure words for nouns. The nouns in this dictionary are arranged by their ending morpheme, i.e. head morpheme. There are 1910 noun ending morphemes and 12,352 example nouns grouped according to their different senses.

Each knowledge source provides partial data for representing morphological rules, which includes lists of sample compounds, high frequency morphemes and their syntactic and semantic information. Unknown words and their frequencies can be extracted from the Sinica corpus. The extracted unknown words produce the testing data and the morpheme-category association-strength which are used in the algorithm for the syntactic category prediction for unknown words (Chen et al. 1997). The CKIP dictionary provides the syntactic categories for morphemes and words. The Chilin provides the semantic categories for morpheme and words. The CDNM provides the set of high frequency noun morphemes and the example compounds grouped according to each difference sense. The semantic categories for each sense is extracted from the Chilin and disambiguated manually.

The sample compounds for each sense-differentiated morpheme extracted from CDNM form the base samples for the morphological rules. Additional samples are supplemented from the Chilin.

3.2 The algorithm for morphological analysis

The process of morphological analysis for compound words is very similar to Chinese word segmentation process. It requires dictionary look-up for matching morphemes and resolution methods for the inherent ambiguous segmenta-

tions, such as the examples in 1). However conventional word segmentation algorithms cannot apply for the morphological analysis without modification, since the morpho-syntactic behavior is different from syntactic behavior. Since the structure of the Chinese compound nouns is head final and the most productive morphemes are monosyllabic, there is a simple and effective algorithm, which agrees with these facts. This algorithm segments input compounds from left to right by the longest matching criterion (Chen & Liu 1992). It is clear that the left to right longest matching algorithm prefers shorter head and longer modifier structures.

3.3 Semantic categories of morphemes

The semantic categories of morphemes are followed from the thesaurus Chilin. This thesaurus is a lattice structure of concept taxonomy. Morphemes/words may have multiple classification due to either ambiguous classification or inherent semantic ambiguities. For the ambiguous semantic categories of a morpheme, the lower ranking semantic categories will be eliminated and leave the higher-ranking semantic categories to compete during the identification process. For instances, in the table 2 only the major categories of each example are shown. Since the majority of morphemes are unambiguous, they will compensate the uncertainty caused by the semantically ambiguous morphemes. The rank of a semantic category of a morpheme depends on the occurring order of this morpheme in its synonym group, since the arrangement of the Chilin entries is by this natural. In addition, due to limit coverage of Chilin, many of the morphemes are not listed. For the unlisted morphemes, we recursively apply the current algorithm to predict their semantic categories.

4. Semantic Classification and Ambiguity Resolution for Compound Nouns

The demand of a semantic classification system for compound nouns was first raised while the task of semantic tagging for Chinese corpus was tried. The Sinica corpus is a 5 million-word Chinese corpus with part-of-speech tagging. In this corpus there are 47,777 word types tagged with common nouns and only 12,536 of them are listed in the Chilin. They count only 26.23%. In other words the semantic categories for most of the common nouns are unknown. They will be the target for automatic semantic classification.

4.1 Derivation of morphological rules

A list of most productive morphemes are first generated from the unknown words extracted from the Sinica corpus. The morphological rules of the set of the most productive head morphemes are derived from their examples. Both the CDMN and Chilin provide some examples. So far there are 1910 head morphemes for compound nouns with examples in the system and increasing. They are all monosyllabic morphemes. For the top 200 most productive morphemes, among them 51.5% are polysemous and in average each has 3.5 different meanings. The coverage of the current 1910 morphemes is about 71% of the unknown noun compounds of the testing data. The rest 29% uncovered noun morphemes are either polysyllabic morphemes or the low frequency morphemes.

4.2 Semantic classification algorithm

The unknown compound nouns extracted from the Sinica corpus were classified according to the morphological representation by the similarity-based algorithm. The problems of semantic ambiguities and out-of-coverage morphemes were two major difficulties to be solved during the classification stage. The complete semantic classification algorithm is as follows:

- 1) For each input noun compound, apply morphological analysis algorithm to derive the morphemic components of the input compound.
- 2) Determine the head morpheme and modifiers. The default head morpheme is the last morpheme of a compound.
- 3) Get the syntactic and semantic categories of the modifiers. If a modifier is also an unknown word, then apply this algorithm recursively to identify its semantic category.
- 4) For the head morpheme with the representational rules, apply similarity measure for each possible semantic class and output the semantic class with the highest similarity value.
- 5) If the head morpheme is not covered by the morphological rules, search its semantic class from the Chilin. If its semantic class is not list in the Chilin, then no answer can be found. If it is polysemous, then the top ranked classes will be the output.

In the step 1, the algorithm resolves the possible ambiguities of the morphological structures of the input compound. In the step 3, the semantic categories of the modifier are determined. There are some complications. The first complication is that the modifier has multiple semantic cate-

gories. In our current process, the categories of lower ranking order will be eliminated. The remaining categories will be processed independently. One of the semantic categories of the modifier pairing with one of the rule of the head morpheme with the category will achieve the maximal similarity value. The step 4 thus achieves the resolution of both semantic ambiguities of the head and the modifier. However only the category of the head is our target of resolution. The second complication is that the modifier is also unknown. If it is a not listed in the Chilin, there is no way of knowing its semantic categories by the current available resources. At the step 4, the prediction of semantic category of the input compound will depend solely on the information about its head morpheme. If the head morpheme is unambiguous then output the category of the head morpheme as the prediction. Otherwise, output the semantic category of the top rank sense of the head morpheme. The step 5 handles the cases of exceptions, i.e. no representational rule for head morphemes.

4.3 Experimental results

The system classifies the set of unknown common nouns extracted from the Sinica corpus. We randomly picked two hundred samples from the output for the performance evaluation by examining the semantic classification manually. The correction

rate for semantic classification is 84% and 81% for the first hundred samples and the second hundred samples respectively. We further classify the errors into different types. The first type is caused by the selection error while disambiguating the polysemous head morphemes. The second type is caused by the fact that the meanings of some compounds are not semantic composition of the meanings of their morphological components. The third type errors are caused by the fact that a few compounds are conjunctive structures not assumed head-modifier structure by the system. The fourth type errors are caused by the head-initial constructions. Other than the classification errors, there exist 10 unidentifiable compounds, 4 and 6 in each set, for their head morphemes are not listed in the system nor in the Chilin. Among the 190 identifiable head morphemes, 142 of them are covered by the morphological rules encoded in the system and 80 of them have multiple semantic categories. The semantic categories of remaining 48 head morphemes were found from the Chilin. If the type 1 selection errors are all caused by the 80 morphemes with multiple semantic categories, then the correction rate of semantic disambiguation by our similarity-based measure is $(80-15)/80 = 81\%$.

Testing data 1	Testing data 2
Total :100	Total: 100
error:12	error:13
type1(semantic selection error): 8	type1(semantic selection error): 7
type2(non-compositional): 2	type3(non-compositional): 5
type3(conjunction): 1	type3(conjunction): 0
type4(head-initial): 1	type4(head-initial): 1
unidentified: 4	unidentified: 6

Table 5. The performance evaluations of the semantic classification algorithm

5. Further Remarks and Conclusions

In general if an unknown word was extracted from corpora, both of its syntactic and semantic categories are not known. The syntactic categories will be predicted first according to its prefix-category and suffix-category associations as mentioned in (Chen et al. 1997). According to the top ranked syntactic predictions, each respective semantic representational rules or models will be applied to produce the morpho-semantic plausibility of the unknown word of its respective syntactic categorization. For instance if the predicted syntactic

categories are either a common noun or a verb, the algorithm present in this paper will be carried out to classify its semantic category and produce its plausibility value for the noun category. Similar process should deal with the case of verbs and produce the plausibility of being a verb. The final syntactic and semantic prediction will be based on their plausibility values and its contextual environments (Bai et al. 1998, Ide 1998). The advantages of the current representational model are:

- 1) It is declarative. New examples and new mor-

phemes can be added into the system without changing the processing algorithm, but the performance of the system might be increased due to the increment of the knowledge.

- 2) The representational model not only provides the semantic classification of the unknown words but also gives the value of the plausibility of a compound construction. This value could be utilized to resolve the ambiguous matching between competing compound rules.
- 3) The representational model can be extended for presenting compound verbs.
- 4) It acts as one of the major building block of a self-learning system for linguistic and world knowledge acquisition on the Internet environment.

The classification errors are caused by a) some of the testing examples have no semantic composition property, b) some semantic classifications are too much fine-grained. There is no clear cut difference between some classes, even human judge cannot make a right classification, c) there are not enough samples that causes the similarity-based model does not work on the suffixes with few or no sample data. The above classification errors can be resolved by collecting the new words, which are semantically non-compositional, into the lexicon and by adding new examples for each morpheme.

Current semantic categorization system only roughly classifies the unknown compound nouns according to their semantic heads. In the future deeper analysis on the semantic relations between modifier and head should also be carried out.

6. References

- Bai, M.H., C.J. Chen & K.J. Chen, 1998, "POS-tagging for Chinese Unknown Words by Contextual Rules" *Proceedings of ROCLING*, pp.47-62.
- Chang, J. S., S.D. Chen, S. J. Ker, Y. Chen, & J. Liu, 1994 "A Multiple-Corpus Approach to Recognition of Proper Names in Chinese Texts", *Computer Processing of Chinese and Oriental Languages*, Vol. 8, No. 1, pp. 75-85.
- Chen, C. J., M. H. Bai, K. J. Chen, 1997, "Category Guessing for Chinese Unknown Words." *Proceedings of the Natural Language Processing Pacific Rim Symposium 1997*, pp. 35-40. NLPRS '97 Thailand.
- Chen, K. J., C.J. Chen, 1998, "A Corpus-based Study on Computational Morphology for Mandarin Chinese", in *Quantitative and Computational Studies on the Chinese Language* Eds by Benjamin K. Tsou, City Univ. of Hong Kong, pp283-306.
- Chen, Keh-Jiann, Ming-Hong Bai, 1997, "Unknown Word Detection for Chinese by a Corpus-based Learning Method." *Proceedings of the 10th Research on Computational Linguistics International Conference*, pp159-174.
- Chen, K.J. & S.H. Liu, 1992, "Word Identification for Mandarin Chinese Sentences," *Proceedings of 14th Coling*, pp. 101-107.
- Chien, Lee-feng, 1999, "PAT-tree-based Adaptive Keyphrase Extraction for Intelligent Chinese Information Retrieval," *Information Processing and Management*, Vol. 35, pp. 501-521.
- Fung P., 1998, "Extracting Key Terms from Chinese and Japanese Texts," *Computer Processing of Oriental Languages*, Vol. 12, #1, pp 99-122.
- Huang, C. R. Et al., 1995, "The Introduction of Sinica Corpus," *Proceedings of ROCLING VIII*, pp. 81-89.
- Huang, Chu-Ren, Keh-Jiann Chen, and Ching-hsiung Lai, 1997, *Mandarin Daily Classification Dictionary*, Taipei: Mandarin Daily Press.
- Ide, Nancy & Jean Veronis, 1998, "Special Issue on Word Sense Disambiguation", *Computational Linguistics*, Vol. 24, # 1.
- Lee, J.C. , Y.S. Lee and H.H. Chen, 1994, "Identification of Personal Names in Chinese Texts." *Proceedings of 7th ROC Computational Linguistics Conference*.
- Lin, M. Y., T. H. Chiang, & K. Y. Su, 1993, "A Preliminary Study on Unknown Word Problem in Chinese Word Segmentation" *Proceedings of Rocling VI*, pp 119-137.
- Mei, Gia-Chu etc., 1984 *同義詞詞林*. (Chilin - thesaurus of Chinese words). Hong Kong, 商務印書館香港分館.
- McDonald D., 1996, "Internal and External Evidence in the Identification and Semantic Categorization of Proper Names", in *Corpus Processing for Lexical Acquisition*, J. Pustejovsky and B. Boguraev Eds, MIT Press 1996.
- Sun, M. S., C.N. Huang, H.Y. Gao, & Jie Fang, 1994, "Identifying Chinese Names in Unrestricted Texts", *Communication of COLIPS*, Vol.4 No. 2. 113-122.