

基於特徵為本及使用 SVM 的文本對蘊涵關係的自動推論方法

Textual Entailment Recognition Using Textual Features and SVM

張道行 Tao-Hsing Chang

許曜麒 Yao-Chi Hsu

張中維 Chung-Wei Chang

許堯銓 Yao-Chuan Hsu

國立高雄應用科技大學資訊工程系

Department of Computer Science and Information Engineering

National Kaohsiung University of Applied Sciences

changth@kuas.edu.tw

陳學志 Hsueh-Chih Chen

國立臺灣師範大學教育心理與輔導學系

Department of Educational Psychology and Counseling

National Taiwan Normal University

chcjyh@ntnu.edu.tw

摘要

這篇論文的目的是提出一個能判斷文本對蘊涵關係的系統。本系統主要使用 7 項由觀察資料所得之特徵作為輸入項，並以 SVM 作為預測模型。以 NTCIR-10 中 RITE-2 提供的資料評估本文所提方法，整體的 Macro F1-measure 為 46.35%，較先前研究所提出的方法為佳。而該方法使用的特徵數量少、模型運算與實作簡單，可以在實際應用上有更好的效果。

Abstract

The aim of this paper is to propose a system, which can automatically infer entailment relations of textual pairs. SVM is utilized as a prediction model of the system and seven features of textual pairs are employed to be input of the prediction model. The performance of this system is evaluated by dataset in CT-MC task held by RITE-2 of NTCIR. Macro-F1 of the proposed method is 46.35%.

關鍵詞：蘊涵關係，自動推論，支援向量機

Keywords: Textual Entailment, Automatic Inference, Support Vector Machine.

一、緒論

文本對蘊涵關係判讀是自然語言處理領域一個有趣且有意義的問題，其可能的應用也十

分廣泛。例如矛盾關係分析可以用來自動分析不同知識來源所提供資訊的一致性與正確性，這在網路的知識提供者如 **wiki** 以及利用網路資源進行數位學習等應用都是非常重要的功能，因為這些應用必須確保知識來源的正確性。

相關問題的許多研究已經被提出，為了評估這些方法的效能差異，有許多的評估基準資料被釋出，也舉辦了方法效能評比。在 **NTCIR-10** 的 **RITE-2** 中，各個接受評比的方法必須在給定一個由兩個句子所組成的文本對時，能夠分辨文本對是屬於雙向推論關係、正向推論關係、矛盾關係抑或獨立關係。以下列句子 **S1** 至 **S5** 所組成的各個文本對舉例說明。

S1：韭菜原產於中國，是常見的蔬菜之一。

S2：韭菜原產於中國。

S3：韭菜原產於日本。

S4：水菜原產於日本。

S5：水菜原產地為日本。

對文本對(**S1, S2**)而言，句子 **S2** 中的所有資訊內容可以由 **S1** 推論，但 **S1** 中有一句內容「常見的蔬菜之一」是不能由 **S2** 推斷出來的，因此這文句對稱為具有正向關係。以文本對(**S2, S3**)來說，「韭菜」的「原產地」應該只有一處，但此二句描述韭菜的原產地是不同的，因此是矛盾的資訊。在推論過程產生矛盾的文本對稱為具有矛盾關係。而考慮文本對(**S3, S4**)，雖然 **S4** 與 **S3** 相似，但與文本對(**S2, S3**)不同，「韭菜」與「水菜」的原產地沒有互斥關係，因此這兩句是描述不同的事件而不是同一事件的描述不同。這樣的情況稱該文本對具有獨立關係。又考慮文本對(**S4, S5**)，兩句表達相同的知識內容，由句子 **S4** 可以推論出 **S5**，而由句子 **S5** 也可推論出 **S4**，因此稱此文本對具有雙向推論關係。

這篇論文的目的是提出一個能判斷文本對蘊涵關係的系統。本系統主要使用 7 項由觀察資料所得之特徵作為輸入項，並以 **SVM** 作為預測模型。這 7 項特徵是藉由觀察資料所歸納，均具有合理的推論解釋以及數值定義。因此本文將分析比較在相同的預測模型時，不同的選取特徵造成預測文本對蘊涵關係效能上的差異。另外也比較在同樣 7 項特徵所產生的訓練資料下，哪一種預測模型在這個問題上會有較佳的表現。

這篇論文其餘部分組織如下。第二節探討相關研究，包括英文的文句蘊涵研究以及近年來國際評比中表現較佳的方法，並說明與我們提出的方法之間的關係。第三節介紹本文提出的 7 項特徵所代表的意義及定義值的計算方法。另外也說明將用來比較的預測模型，第四節是比較本文所提方法與其他方法，並以 **NTCIR RITE-2** 提供的訓練與測試資料為依據。最後討論本研究的侷限以及未竟之處，探討未來可行的研究方向。

二、文獻回顧

關於英語文本的蘊涵推論已經有很多相關研究。[1]運用字面上的相似度來判斷蘊涵關係。如果是相同的詞彙，此方法確實可以精準的判斷出蘊涵關係，但是這樣的方法在處理同義詞的文本時容易錯判。[2]提出的「淺層語意特徵」(**Shallow Semantic Features**)的篩選法則可以解決這個問題。[2]使用 **WordNet** 作為背景知識，解釋不同的詞彙是相同或相反的詞意。例如「兇手」、「受害者」與「謀殺」相關，但「兇手」是「謀殺」的衍生字，而「受害者」與「兇手」是反義字。

除了以詞彙和語意特徵的相似程度推論蘊涵關係之外，還有使用剖析樹(parsing tree)來分析句法結構以推論蘊涵關係的方法[3-5]。這些方法都是先使用 parser 把文句以樹狀結構表示，而 [3]運用 linear distance 與 tree edit distance 等方法計算文句差異。[4]和[5]則將兩個句子的樹狀結構在經過數次的插入、刪除、代換後將樹狀圖調整成相同的圖，而其過程中插入、刪除、代換的次數稱為樹距(tree distance)，可用來當作樹狀圖之間的差異標準。這些研究利用這個差異性來判斷文句的蘊涵關係。

2011 年由 NTCIR-9 所舉辦 RITE 文句蘊涵推論的任務中，[6]採用包括句子長度、內文關鍵字重複率、關鍵字重複的數量與詞性等等淺層特徵，藉此分辨出兩個句子之間的差異來判斷文本蘊涵關係。實驗結果證明，只使用淺層特徵推論蘊涵關係也有良好的效果。[7]則提出一種基於語法分析的方法。首先，他們使用 stanford parser[8]分析文句的語法樹，並標示出主要的動詞與名詞。接著在分析不同類型的主要動詞與名詞後歸納出幾種主要特徵，最後使用這些特徵來計算文句之間的句法相似度。實驗結果證明[7]的效能較只使用淺層特徵推論蘊涵關係有更好的表現。2013 年所舉辦的 RITE2 中，效能最佳的 IASL[10]提出以二元關係分類的概念。[10]認為兩個句子關係有三種：句子間是否衝突、第一句是否推論第二句、第二句是否推論第一句等。由兩個句子呈現的這三種關係結果判斷兩個句子屬於哪種蘊涵關係。而各種二元關係的依據則建立在各項特徵上。

上述的特徵與預測文本蘊涵關係都會需要一個整合個特徵的分類模型[11]。支持向量機(SVM)是最普遍的分類模型，例如[3]以經常使用的特徵包括文字、剖析樹、情緒正反意、名詞縮寫等，將文句轉換成特徵向量，並使用特徵向量推論出蘊涵關係。另一種常用的分類模型則是決策樹。決策樹可由專家建構或是藉由機器學習的方式產生，ID3 可以透過得到的資訊來最佳化樹的結構。根據先前的研究，本文將嘗試利用文本對的詞彙、語意及語法特徵配合 SVM 預測模型推論文本對的蘊涵關係。

在 RITE-2 中，支持向量機(SVM)也是最常使用的分類模型[12][18-20]。而各研究的主要差異就在輸入特徵的選擇上。例如[12]強調以多達 20 種的特徵輸入 SVM 進行判斷；[18]則提到以關鍵字的匹配及數量、剖析樹詞類分析、否定詞、同義詞作為特徵；[19]則使用時間與數字的表示以及否定；[20]則提到句法分析、專有名詞辨認、近義詞、常用詞的數量、文句長度、否定詞、反義詞的使用。雖然某些特徵同時被不同方法所使用，但是系統詮釋與訓練其特徵的方式仍有不同。本文將發展一些特徵並同樣以 SVM 為預測模型，以便比較先前研究和本文所提方法的特徵在預測蘊涵關係上的效能差異。

三、方法

本文使用了七個文本對特徵預測文本對的蘊涵關係。七項特徵大致分類成詞彙、語意及語法三種。此外，中文文本對在計算特徵前需要斷詞及詞性標記等前處理工作，以便後續演算法使用。前處理以及七個特徵的細節在下列各小節說明。

(一) 前處理

由於中文間詞與詞之間沒有空白分隔，因此中文文本對必須先進行斷詞與詞性標記，將句子以字元表現形式轉換為詞彙表現，並標記每個詞彙的詞性以便後續分析特徵時使用。許多中文斷詞與詞性標記系統已經被提出，也有很好的正確性。在中文文本對的研

究中，未知詞是一個需要處理的問題，因為許多專有名詞大量出現在以知識為主的文本對中。但是由於中文句中詞彙間沒有空白分隔，所以要辨識未知詞是很困難的工作。

另外資料格式不一致的問題，也常發生在文本對中。以下三種是常見的狀況：

1. 用不同的方式表示相同的資料，例如一半、1/2、0.5
2. 縮寫，例如 2003 年、03 年
3. 單位轉換，例如 1kg、1000g

雖然以上這些問題頻繁出現在各種文本對中，但是格式卻多是常見的幾種。在實驗部分本文會進一步說明在前處理階段我們所採用的工具以及前述問題的解決策略。

(二) 詞彙特徵

1. 名詞數量一致性(CNN)

我們觀察到一個現象：當兩個句子中的名詞數量一樣時，這組文本對為雙向及矛盾關係的機會愈高。這是因為名詞是用來表示某些事物，而兩句子名詞數量相同代表兩句子可能在講相同事物的機率較高。舉例來說，以下三個句子 S6、S7 和 S8 中都包含了三個名詞且在敘述同一件事情，因此可能是雙向蘊涵關係，例如文本對(S6,S7)是雙向蘊涵。但是有相同數量名詞的文句對也可能因為句中某些文字導致句子互相矛盾，例如文本對(S7,S8)就是矛盾關係。

S6：H5N1 型病毒株能透過禽類傳染給人體

S7：H5N1 型病毒株是藉由禽類傳染給人體

S8：H5N1 型病毒株並非由禽類傳染給人體

因此本文定義了一項文本對特徵「名詞數量一致性」，簡稱 CNN。若文本對的兩句子名詞數量一致，則該文本對的 CNN 為 1，否則為-1。

2. 詞重疊率差異 (DRO)

我們觀察到當一個文本對中兩個句子使用相同的詞越少，該文本對為獨立關係的機會愈高。因此對於一個文本對(S_i, S_j)，本文定義該文本對的「順向詞重疊率」(RWF)及「反向詞重疊率」(RWB)如下：

$$RWF = \frac{|W_i \cap W_j|}{|W_j|}, \quad RWB = \frac{|W_i \cap W_j|}{|W_i|}$$

其中 W_k 表示句子 S_k 中所有詞的集合， $|W_k|$ 表示集合 W_k 中的詞的數量。對兩個句子而言，若 RWF 與 RWB 兩者同時都低，顯示兩句子的語意可能相差過大，此文本對很有可能是獨立關係，因為兩句話包含了不同的內容。舉例來說，下列文本對(S9,S10)兩句的關係為獨立關係，其 RWF 為 0.05、RWB 為 0.16。

S9：馬來西亞原為日本電子業者眼中最佳的亞洲投資標的，現被中國大陸取代

S10：中國取代美國成為亞洲經濟核心

S11：日本是投資馬來西亞的三大外商之一

S12：日本有投資馬來西亞

若兩句子的 RWF 和 RWB 的差值很大，則表示此文本對有相近的資訊，但兩句提供的資訊量一句較多而另一句較少。這是此文本對可能是正向蘊涵的線索。例如文本對 (S11,S12) 為正向蘊涵，其 RWF 為 0.30 而 RWB 為 0.75。基於上述觀察，本文定義了一個文本對特徵「詞重疊率差異」，簡稱 DRO，定義如下：

$$DRO = \begin{cases} 1, & \text{if } RWF \leq TI \text{ and } RWB \leq TI \\ 0, & \text{(if } RWF \geq TI \text{ or } RWB \geq TI) \text{ and } |RWB - RWF| \geq TD \\ -1, & \text{otherwise} \end{cases}$$

其中 TI 和 TD 是兩個門檻值。根據本文實驗所使用的訓練資料，TI 和 TD 的值分別為 0.6 與 0.2。

(三) 詞法特徵

1. 詞性重疊率差異(DOP)

從 DRO 進一步延伸，我們假設當兩個句子使用的相同詞性越少，文本對就越可能為獨立關係。參照 DRO 的定義，本文定義一個文本對特徵「詞性重疊率差異」，簡稱 DOP。對文本對 (S_i, S_j)，計算 DOP 前先計算該文本對的「順向詞性重疊率」(RPF)及「反向詞性重疊率」(RPB)如下：

$$RPF = \frac{|P_i \cap P_j|}{|P_j|}, \quad RPB = \frac{|P_i \cap P_j|}{|P_i|}$$

其中 P_k 代表 S_k 句子中所有詞性的集合，|P_k| 為集合 P_k 內詞性的數量。若文本對的 RPF 夠高，且 RPF 和 RPB 的差值超過閾值，則這個文本對就有很高的機率是正向關係。因此文本對的 DOP 定義如下：

$$DOP = \begin{cases} 1, & \text{if } RPF \geq TP \text{ and } (RPF - RPB) \geq TK \\ -1, & \text{otherwise} \end{cases}$$

其中 TP 和 TK 是兩個門檻值。根據本文實驗所使用的訓練資料，TP 和 TK 各為 0.7 與 0.2。

(四) 詞意特徵

1. 時間不對稱(OOT)

在一些文本對中，其中一句有提供時間資訊、但另一句沒有，可推論這兩句如果不是獨立關係，就是正向蘊涵關係。以下列文本對 (S13,S14) 為例，兩句有語意上的高度相關，但 S13 中提到了時間「9 世紀」，S14 卻沒有提到任何時間，因此該文本對的關係有可能是正向蘊涵。

S13：韭菜於 9 世紀傳入日本

S14：韭菜曾傳入日本

本文將這個特徵稱為「時間不對稱」，簡稱 OOT。該特徵定義如下：如果文本對中一句有時間資訊而另一句沒有，該特徵值為 1，反之為 -1。

2. 存在否定詞(ENW)

在一些文本對中，兩句話有著高相似度但是兩句話表示的意思卻因為否定詞的出現而造成矛盾。以文本對(S15,S16)為例，句子 S15 比 S16 多出了否定詞「不會」，使得該文本對為矛盾關係。因此本文使用了特徵「存在否定詞」，簡稱 ENW，其值定義為：文本對中若有一句出現否定詞而另一句則無，則該文本對的 ENW 為 1，否則為-1。

S15：阿斯匹靈不會引起不良反應

S16：阿斯匹靈可能引起不良反應

3. 使用同義詞(SYN)

有些文本對的兩個句子使用的詞彙大部分相同，且詞彙出現順序與詞彙詞性也都完全相同，只有少部分相對應位置的詞彙不同。若這些不同的詞彙是同義詞，則該文本對的兩句可能是描述同樣事件或事實的兩個不同說法，因此可能是雙向蘊涵關係。以文本對(S17,S18)為例，兩句只有在「教廷」與「梵諦岡」的位置使用不同詞彙，但詞彙的出現順序與其詞性完全相同。而「教廷」與「梵諦岡」是同義詞，因此該文本對為雙向關係。

我們定義一個文本對特徵「使用同義詞」，簡稱 SYN，其值定義如下。若一個文本對中相對應的位置使用少數不同詞彙，而在同一位置的詞彙互為同義詞，且兩句的詞性順序也相同，該文本對的 SYN 值為 1，反之為-1。

S17：若望保祿二世是教廷領導人

S18：若望保祿二世是梵諦岡領導人

4. 詞序交換(WOE)

和 SYN 相反，有些文本對中使用了完全相同的詞，但只是因順序不同，導致兩句的意思完全相反。和 ENW 不同的是，這樣的文本對中並沒有否定詞。例如下列文本對(S19,S20)，雖然使用的詞完全一樣，但意思完全相反，原因是甘蔗和蔗糖在句法功能的位置不同。這造成該文本對是矛盾關係。

S19：甘蔗是製造蔗糖的原料

S20：蔗糖是製造甘蔗的原料

然而並非所有具有完全相同的詞但詞序不同的文本對都是矛盾關係。以文本對(S21, S22)為例，兩句的用詞完全相同，而「伊普索」與「美聯社」在句子中位置不相同。這個現象卻未導致該文本對為矛盾關係。主要原因是此二者是以連接詞連接，位置交換並未導致語法結構改變，也因此此二句表達完全相同的意思。

S21：美聯社和伊普索斯公司所進行的民調顯示，布希的施政滿意度已首次滑落到 39%

S22：伊普索斯公司和美聯社所進行的民調顯示，布希的施政滿意度已首次滑落到 39%

因此本文定義了特徵「詞序交換」，簡稱 WOE，其值定義如下：若文本對的用詞完全一樣但是詞序不一樣，且順序不同的詞並非以連接詞連接，則該文本對的 WOE 為 1，否則為-1。

(五) 預測模型

本文所定義的上述特徵將成為預測模型的輸入項。預測模型將利用訓練資料中每個文本

對的七項特徵值與已知的文本對蘊涵關係作為訓練預測模型之用。在訓練完成後，對於要預測的文本對，只要計算該文本對的七項特徵值後輸入預測模型，即可得到該文本對關係的預測結果。本文主要目的之一就是比較我們先前採用的 **decision tree** 以及 **SVM** 方法的差異。**SVM** 是一個相當成功的分類方法，已經被廣泛應用在許多領域的研究中。本文將以[15]所提出的 **LibSVM** 作為實作 **SVM** 的系統。另外由於先前採用的 **decision tree** 是由專家建立，本文也將以 **ID3** 最佳化方法自動架構 **decision tree** 並比較效能差異。

四、實驗

本實驗主要目的在觀察本文所提特徵是否具有良好預測性、以及不同預測模型造成的差異。實驗資料是由 **NTCIR-10** 中 **RITE-2** 的任務資料集獲得。訓練資料為任務資料集中的 **development** 子資料集，測試資料為任務資料集中的 **formal run** 子資料集。**RITE-2** 有 **CT-BC** 與 **CT-MC** 兩項任務，在 **CT-BC** 任務中，文本對只需被預測方法歸類為雙向 (**Bidirection**) 或矛盾 (**Contraction**) 兩種關係之一。在 **CT-MC** 任務中，文本對應該被預測方法歸類為雙向 (**Bidirection**)、正向 (**Forward**)、矛盾 (**Contraction**) 和獨立 (**Independent**) 四種關係之一。此實驗將比較不同預測模型在這些任務中的表現。本文所提方法將稱為 **KC99-SVM**。

在實作本文所提方法時，考慮在分析文本對特徵時未知詞的特殊需要，本文以[13]提出的 **WeCAn** 系統為基礎，加以修改後對文本對句子進行斷詞與詞性標記。該系統被修改為先至 **wiki** 蒐集專有名詞，再採用 **SPLR** 方法[14]提高系統辨識未知詞的能力。另外，本文也利用規則式的方法來將數值資料轉換成相同的格式。而對於同義詞的判斷，本文將可能是同義詞的詞送至 **Google** 英譯，若顯示相同的英文詞彙則是為同義詞。另外否定詞則是以列表輔以規則式方式處理。另外在使用 **LibSVM** 時，我們均使用該系統預設值建構本文所使用的 **SVM** 模型，並未進一步進行參數最佳化。

表一是三種不同分類器使用相同的 7 項特徵的結果。這三種分別是專家建立的決策樹 (**Decision tree**)[16]、以 **ID3** 方法自動建立的決策樹以及 **SVM**。由表一可以發現 **SVM** 是表現最好的分類方式，**ID3** 雖然比專家建立 **Decision tree** 的方法表現較佳，但其整體表現仍稍微落後 **SVM**。而 **ID3** 已為最佳化之後的結果，但 **SVM** 僅使用 **LibSVM** 的預設參數值，若進一步進行 **SVM** 參數最佳化，兩者會有更明顯的差距。

表一、評估資料集運行於不同分類器之結果

Tasks		CT-BC		CT-MC				Macro-F1
		Y	N	B	F	C	I	
Decision tree[16]	F1	66.42	48.93	45.48	63.61	16.67	49.24	43.75
	Precision	60.45	57.58	42.94	57.00	15.87	66.08	
	Recall	73.70	42.54	48.34	71.95	17.54	39.24	
ID3	F1	69.80	60.37	55.78	65.72	6.61	56.16	46.07
	Precision	66.99	63.89	57.34	61.38	57.14	51.00	
	Recall	72.86	57.21	54.30	70.73	3.51	62.50	
KC99-SVM	F1	72.78	50.72	61.67	63.48	10.94	49.30	46.35
	Precision	62.96	70.67	53.11	55.03	50.00	58.29	
	Recall	86.42	39.55	73.51	75.00	6.14	42.71	

表二是比較採用本文所提 7 項特徵搭配 SVM 的方法與其他同樣使用 SVM 但不同特徵的方法。本文選擇在 RITE2 中使用 SVM 者效能最佳的 NTOUA[17]系統作為比較對象。從表二可以看出本文系統在整體效能上較[17]為佳。由於該系統使用 20 種特徵而本文僅使用 7 種特徵，因此可知本文所提方法能以較少數量的特徵達到更好的效能。

表二、以 SVM 進行分類但採用不同特徵的方法間效能比較

Tasks		CT-BC		CT-MC				Macro-F1
		Y	N	B	F	C	I	
NTOUA-03[17]	F1	19.39	44.04	61.10	64.21	1.50	52.40	44.80
	Precision	28.81	35.89	50.43	55.00	5.26	70.59	
	Recall	14.61	56.97	77.48	77.13	0.88	41.67	
KC99-SVM	F1	72.78	50.72	61.67	63.48	10.94	49.30	46.35
	Precision	62.96	70.67	53.11	55.03	50.00	58.29	
	Recall	86.42	39.55	73.51	75.00	6.14	42.71	

表三是比較本文所提方法與 RITE2 中效果最好的 IASL[17]方法比較。由表三可知，兩者表現差距非常小。由於本文所提方法採用的架構較單純，採用的指標也較少，因此可能在某些應用上會更適合作為解決方案。

表三、與 RITE2 任務中 Macro-F1 最高的 IASL 之比較

Tasks		CT-BC		CT-MC				Macro-F1
		Y	N	B	F	C	I	
IASL[17]	F1	71.66	62.63	52.35	64.63	29.90	38.41	46.32
	Precision	68.64	66.48	53.06	53.99	36.25	52.73	
	Recall	74.95	59.20	51.66	80.49	25.44	30.21	
KC99-SVM	F1	72.78	50.72	61.67	63.48	10.94	49.30	46.35
	Precision	62.96	70.67	53.11	55.03	50.00	58.29	
	Recall	86.42	39.55	73.51	75.00	6.14	42.71	

五、討論與未來工作

從實驗結果可以看出本文所提 7 項特徵可以用以區別文本對蘊涵關係。而使用 SVM 分類後的整體效能比先前採用的決策樹方法更佳，但是推論矛盾關係的正確率卻比決策樹低。事實上矛盾關係的推論無論是決策樹方法與 SVM 都表現不佳。經過進一步分析，本文所提方法有三項特徵與矛盾關係有關(WOE、ENW、SYN)，然而這三項特徵處理的資料都相當特定、數量有限，因此造成矛盾關係的更深層特徵尚待發掘。另外，本文所提方法雖然也包含語意特徵類別，但都仍屬於語意的間接特徵，並未直接測量語意。這也是推論效能有所侷限的原因。

基於本文所提方法，未來可進一步探討及研究。首先，本文提出的 7 項特徵有些仍待進一步改良，例如使用同義詞、存在否定詞等特徵，所使用的測量方法仍相當簡化。如果能加以改良，效能應可改善。另外，語法特徵在目前提出的七項特徵中僅有一項，但在實驗過程中發現語法特徵有良好的區辨效果。雖然目前已經有中文文法剖析工具提出，但用以分析特徵時錯誤率仍過高。如何使用有效的文法剖析工具發展語法特徵可能是能

大幅提高正確率的途徑之一。

誌謝

本文作者感謝國科會計畫編號 NSC 102-2511-S-151-002 的支持，同時也感謝教育部及國立台灣師範大學「邁向頂尖大學計畫」的支持。

參考文獻

- [1] I. Androutsopoulos and P. Malakasiotis, “A survey of paraphrasing and textual entailment methods,” *Journal of Artificial Intelligence Research*, vol. 38, pp. 135-187, 2010.
- [2] J. Bos and K. Markert, “Recognizing textual entailment with logical inference,” in *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2005, pp. 628-635.
- [3] E. Cabrio, M. Kouylekov and B. Magnini, “Combining specialized entailment engines for rte-4,” in *Proceedings of TAC08, 4th PASCAL Challenges Workshop on Recognizing Textual Entailment*, 2008.
- [4] M. Kouylekov and B. Magnini, “Recognizing textual entailment with tree edit distance algorithms,” in *Proceedings of the First Challenge Workshop Recognizing Textual Entailment*, 2005, pp. 17-20.
- [5] M. Kouylekov and B. Magnini, “Tree edit distance for recognizing textual entailment: Estimating the cost of insertion,” in *Proceedings of the PASCAL RTE-2 Challenge*, 2006, pp. 68-73.
- [6] N. H. Han and L. W. Ku, “The Yuntech system in NTCIR-9 RITE Task,” in *Proceedings of the NTCIR-9 Workshop*, 2011, pp. 345-348.
- [7] H. H. Huang, K. C. Chang, J. M. Haver II and H. H. Chen, “NTU Textual Entailment System for NTCIR 9 RITE Task,” in *Proceedings of the NTCIR-9 Workshop*, 2011, pp. 349-352.
- [8] Stanford Parser: A statistical parser, 2002, Available : <http://nlp.stanford.edu/software/lex-parser.shtml>
- [9] T. H. Chang, C. H. Lee, P. Y. Tsai, and H. P. Tam, “Automated essay scoring using set of literary sememes,” *Information: An International Interdisciplinary Journal*, vol. 12, no. 2, pp. 351-357, 2009.
- [10] C. W. Shih, C. Liu, C. W. Lee and W. L. Hsu, “IASL RITE System at NTCIR-10,” in *Proceedings of the 10th NTCIR Conference*, Tokyo, Japan, 2013.
- [11] Y. Akiba, H. Taira, S. Fujita, K. Kasahara and M. Nagata, “NTTCS textual entailment recognition system for the NTCIR-9 rite,” in *Proceedings of the 9th NII Test Collection for Information Retrieval Workshop*, 2011, pp. 330-334.
- [12] C. J. Lin and Y. C. Tu, “The Description of the NTOU RITE System in NTCIR-10,” in *Proceedings of the 10th NTCIR Conference*, in *Proceedings of NTCIR-10 Workshop Conference*, Tokyo, Japan, 2013, pp. 495-498.

- [13] T. H. Chang, Y. T. Sung and Y. T. Lee, “A Chinese word segmentation and POS tagging system for readability research,” in *Proceedings of Paper presented at 42nd Annual Meeting of the Society for Computers in Psychology*, 2012
- [14] T. H. Chang and C. H. Lee, “Automatic Chinese unknown word extraction using small-corpus-based method,” in *Proceedings of International Conference on Natural Language Processing and Knowledge Engineering*, Beijing, China, 2003, pp. 459-464.
- [15] C. C. Chang and C. J. Lin, “LIBSVM: a library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, 2011.
- [16] T. H. Chang, Y. C. Hsu, Y. C. Hsu, J. I. Chang and C. W. Chang, “KC99: A Prediction System for Chinese Textual Entailment Relation using Decision Tree,” in *Proceedings of the 10th NTCIR Conference*, Tokyo, Japan, 2013, pp. 469-473.
- [17] Y. Watanabe, Y. Miyao, J. Mizuno, T. Shibata, H. Kanayama, C. W. Lee, C. J. Lin, , S. Shi, T. Mitamura, N. Kando, H. Shima and K. Takeda, “Overview of the Recognizing Inference in Text (RITE-2) at the NTCIR-10,” in *Proceedings of NTCIR-10 Workshop Conference*, Tokyo, Japan, 2013.
- [18] L. Ku, E. T. H. Chu and N. Han, “Extracting Features for Machine Learning in NTCIR-10 RITE Task,” in *Proceedings of the 10th NTCIR Conference*, Tokyo, Japan, 2013, pp. 457-461.
- [19] S. H. Wu, S. S. Yang, L. P. Chen, H. S. Chiu and R. D. Yang, “CYUT Chinese Textual Entailment Recognition System for NTCIR-10 RITE-2,” in *Proceedings of NTCIR-10 Workshop Conference*, Tokyo, Japan, 2013, pp. 443-448.
- [20] W. J. Huang and C. L. Liu, “NCCU-MIG at NTCIR-10: Using Lexical, Syntactic, and Semantic Features for the RITE Tasks,” in *Proceedings of NTCIR-10 Workshop Conference*, Tokyo, Japan, 2013, pp. 430-434.