An Analysis on Verb Transitivity in Chinese Spoiler Posts by Machine Learning Approaches

余盈蓓

Graduate Institute of Linguistic, National Chengchi University

Research Methods and Academic Writing

黃瓊之 老師

December 28, 2021

Abstract

Spoiler is becoming a problem on social medium nowadays. This paper aimed to investigate verb transitivity in Chinese spoiler posts by machine learning approaches. CKIP Tagger was used to segment and assign POS tag to the posts from movie board of PTT. Support Vector Machine and Binary Logistic Regression were chosen to classify the posts and TF-IDF was used as the baseline feature. The results showed that verb transitivity may be used to make better performance of spoiler detection. This proved that high transitivity should be considered a linguistic feature of spoiler posts and Binary Logistic Regression may perform better than Support Vector Machine on spoiler detecting.

Keywords: spoiler detection, machine learning, verb transitivity

1. Introduction

Spoiler, a word defined as a text or sentence which reveals the main plots and events happening in certain works such as books, movies, and dramas, has become an issue on social medium nowadays. Attributing to internet, people can share their thoughts about various works with others. However, while the internet offers platforms for people to discuss works, the posts may accidentally reveal crucial plot twists in certain works to those who have not watched the works, and ruin their excitement of appreciating those works. To solve this problem, several communities have developed systems that warn readers about the spoilers existing in articles. For instance, a reviewer has to mark the certain sentence he wrote if it can possibly spoil others on IMDB. Despite the truth that IMDB introduced a mechanism to deal with spoiler issues, the mechanism largely depends on a writer's mind whether to report a spoiler sentence or not.

Therefore, to solve the problem of spoilers, researchers have tried to introduce efficient devices that could automatically detect spoiler posts on social medium with machine learning approaches during the past decade. Guo and Ramakrishnan (2010) first proposed an LDA-based model using bag-of-words (BOW) as a supporting linguistic cue that performed well on ranking long comments at text level. Later, some other researchers have tried to introduce machine learning models which focus on sentence level of texts since most of the spoiler appear only in one or two specific sentences of the texts. Graber et al. (2013) selected linear-kernel Support Vector Machine (SVM) and N-gram as their baseline model and feature. They found that such linguistic features could perform better when combined with other meta-data such as genre and length of the plots. Others have focused more specifically on tweets, which usually consists of single or two sentences, targeting to offer personalized spoiler detection technique for social medium users. (Jeon et al., 2016; Sasano et al., 2019) Although many researchers have dedicated themselves to improve spoiler detecting by different approaches, most of them have focused on English spoiler posts. According to Marukatat (2020), the proposed methods useful for English spoiler detecting did not work well on detecting Thai spoiler posts, indicating the need of further investigation in spoiler detection of different languages. Take Chinese as an example, only a piece of past research paper about Chinese spoiler detection has been published until this paper is written.

On the other hand, some of the research done so far has been about calculating similarity of texts with existing keywords or plots of the works. (Guo & Ramakrishnan, 2010; Yang et al., 2019) Others have focused on parsing dependency between words and sentences with machine learning and deep learning models. (Sasano et al., 2019; Wan et al., 2019; Ueno et al., 2019; Chang et al., 2021) Despite the fact that dependency parsing requires linguistic knowledge, little research has been done to analyze the linguistic characteristics of spoilers. Graber et al. (2013) suggested that spoilers may have higher verb transitivity but failed to prove it in their research. Jeon et al. (2016) was the only group of researchers who used linguistic cues as features to train their machine learning model. They found that named entities, high objectivity, and the use of future tense are common features of most spoiler posts.

Given the need of digging in Chinese spoiler detection and looking into linguistic characteristics of spoiler posts, this paper aims to investigate verb transitivity in Chinese spoiler posts by machine learning approaches.

2. Method

2.1. Dataset: PTT Bulletin Board System

PTT Bulletin Board System (PPT) was chosen to be the dataset of this study since it is the biggest bulletin board system in Taiwan and has a long history of 26 years. The long history of PTT provides rich data for researchers. Therefore, it is used very often in the field of Chinese Natural Language Processing (NLP). Moreover, the Movie Board of PTT has a strict rule to prevent people from being spoiled, so users must put spoiler alerts in the titles if their posts contain spoilers. The posts containing spoiler without spoiler alert would be deleted according to the board rule. Thus, the posts are publicly characterized as spoilers or non-spoilers. A total of 926 posts were collected to be used in this study, excluding the announcement posts and advertisements. 624 of them were reviews, and 302 of them were spoiler posts.

2.2. Features

2.2.1. Term frequency – Inverse Document Frequency

Term Frequency (TF) refers to the frequency of a term appearing in a certain document. It is represented by the ratio between the frequency n_x of a term x and the total term count n of the document d in which x appears (Alessa & Faezipour, 2018).

$$(1)\quad \mathrm{TF}\,(\mathrm{x},\mathrm{d}) = \frac{n_x}{n}$$

TF is believed to show the characteristics of a given document. The high frequency terms are surely what the document is about. However, by calculating the frequency of a single document is not enough for NLP since the studies are always about dealing with numerous documents. We are not able to know whether the high frequency words are the characteristics of a document only by weighting TF. For instance, function words are used frequently in every document, so they may not become characteristics of a single document. We must compare the frequency of a word among documents to make sure it is a true characteristic of a document but not a common feature of all documents being analyzed. As a result, Inverse Document-Frequency (IDF) should be calculated to help evaluate the characteristics of analyzed documents. IDF is represented by calculating the ratio between the frequency N_D of the documents D that involve term x and the total number N of documents d in the total documents being analyzed (Alessa & Faezipour, 2018).

$$(2) \quad \mathrm{IDF} = \frac{N_d}{N}$$

Moreover, to keep the advantages of both TF (in-document) and IDF (cross-document), Term Frequency-Inverse Document Frequency (TF-IDF) is calculated by multiplying the value of TF by the value of IDF.

$$(3) \quad \mathrm{TF} - \mathrm{IDF} = \mathrm{TF}(\mathrm{x}, \mathrm{d}) \times \mathrm{IDF}(\mathrm{x})$$

Many of the past studies (Guo & Ramakrishnan, 2010; Graber et al., 2013; Yang et al., 2019) of spoiler detection have chosen TF and TF-IDF to weight the keywords in their corpus. TF-IDF has been viewed as a more consummate weighting method. Therefore, TF-IDF was chosen to be the baseline feature of this study.

2.2.2. Verb Transitivity

Verb transitivity was mentioned by Greber et al. (2013) to be one of spoilers' linguistic feature. Transitivity refers to a characteristic of a clause to show that an action or event is 'carried over' or 'transferred' from an agent to a patient (Hopper & Thompson, 1980). Furthermore, it is found that higher transitivity is more likely to connect the agents, actions, patients, and results in one's schema (Schank & Abelson, 1977). For example, the sentence 'Amy killed John' is considered more possibly to become a spoiler than the sentence 'Judy thinks that Amy killed John' since it carries higher transitivity. Therefore, verb transitivity of spoiler posts was chosen to be evaluated in this study.

Verb transitivity was calculated in this study by dividing the number of transitive verbs N_t with the total number of verbs N appearing in a certain spoiler post.

$$\text{(4)} \quad \text{Verb Transitivity} = \frac{N_t}{N}$$

2.3. POS tagger: CKIP Tagger

CKIP Tagger (Li et al., 2020) is a Chinese segmenting tool developed by Chinese Knowledge and Information Processing (CKIP) Lab. The developers reported that CKIP Tagger reaches a high accuracy of approximately 0.97 while segmenting and assigning part of speech (POS) tags. Moreover, CKIP Tagger was designed to process Chinese used in Taiwan, so it outperforms the jiebaR package, which was designed to process the Chinese used in China, when segmenting posts on PTT. Therefore, CKIP Tagger was chosen to segment the data in this study.

An example of the result of utilizing CKIP Tagger to segment sentences and tag POS is shown in figure 1.

傳達仁(Nb) 今(Nd) 將(D) 執行(VC) 安樂死(Na) ，(COMMACATEGORY) 卻(D) 突然(D) 爆出(VJ) 自己(Nh) 20(Neu) 年(Nf) 前(Ng) 遭(P) 緯來(Nb) 體育台(Na) 封殺(VC) ，(COMMACATEGORY) 他(Nh) 不(D) 懂(VK) 自己(Nh) 哪裡(Ncd) 得罪到(VJ) 電視台(Nc)

Figure 1

The CKIP Tagger assigns a POS tag to every unit it segmented, and the detailed categories of POS tags can be found in their 'Technical Report no. 93-05'. According to Hopper & Thompson (1980), telic and punctual are two of the characteristics of transitive clauses. Thus, the verbs involving agent, goal, and source, VB, VC, VD, and, VE, were chosen to be transitive verbs among the twelve categories by the researcher (CKIP Lab, 1993).

2.4. Model

Linear kernel Support Vector Machine (SVM) model and Binary Logistic Regression model were chosen to be used in this study since they are most used to perform binary classification. The results of these two models was compared to find out which better classifies spoiler posts.

2.5. Evaluation

The evaluation metrics used in this study to measure the performance of the model was accuracy, precision, recall, and, F-score. They are calculated with four measures as a basis, true positive (TP), false positive (FP), true negative (TN), and false Negative (FN). TP refers to the number of the positive cases correctly classified (i.e. correctly classified reviews in the present study); FP refers to the number of incorrectly classified positive cases (i.e. Spoilers classified as reviews in the present study); TN refers to the number of negative cases correctly classified (i.e. correctly classified spoilers in the present study); FN refers to the number incorrectly classified negative cases (i.e. Reviews classified as spoilers in the present study).

2.5.1. Accuracy

Accuracy measures a model's performance by calculating the rate of correctly classified cases.

$$(5)\quad \text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

2.5.2. Precision

Precision measures a model's performance by calculating how precise the positive cases prediction is.

$$(6)\quad \text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$$

2.5.3. Recall

Recall measures a model's performance by calculating how many positive cases it successfully predicts.

$$(7)\quad \mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$$

2.5.4. F-score

F-score takes both precision and recall into account to prevent the deviation an unbalanced corpus may bring about.

$$(8)\quad \mathrm{F-score} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Re\,call}}{\mathrm{Precision} + \mathrm{Re\,call}}$$

2.6. Procedure

The spoiler posts were preprocessed by removing urls, pictures, and special symbols to make the data purer for CKIP Tagger to work on. Considering the completeness of the post and the importance of relationships between words when POS taggers work, no stopwords were removed from the corpus. CKIP tagger was used to segment the posts and assign POS tag to each unit in every post after text cleaning, and the result of it was used to calculate the transitivity and TF-IDF of each text. After doing so, 80 percentage of the data was used to train the models, and 20 percentage of it was used to test the trained models. Finally, the evaluation was done by calculating the accuracy, precision, recall, and F-score of the models.

3. Results and Discussion

3.1. Evaluation Scores

The results of the prediction of the models are shown in table 1 to table 4, whereas the evaluation scores are shown in table 5.

Answer / Prediction	Review	Spoiler
Review	115	18
Spoiler	9	42

Table 1: The prediction of baseline model

Answer / Prediction	Review	Spoiler
Review	119	20
Spoiler	5	40

Table 2: The prediction of SVM model

Answer / Prediction	Review	Spoiler
Review	111	32
Spoiler	13	28

Table 3: The prediction of baseline Logistic Regression model

Answer / Prediction	Review	Spoiler
Review	118	18
Spoiler	6	42

Table 4: The prediction of Logistic Regression model

	Baseline SVM	SVM	Baseline Logistic Regression	Logistic Regression
Accuracy	0.8533	0.8641	0.7554	0.8696
Precision	0.8647	0.8561	0.7762	0.8676
Recall	0.9274	0.9597	0.8952	0.9516
F-score	0.8949	0.9049	0.8315	0.9077

Table 5: The evaluation scores of the models

To follow the natural pattern of online movie reviews, the corpus used in this study was not balanced (approximately 67 percent of the posts were reviews and 33 percent of them were spoilers). Therefore, recall and F-score were chosen to be focused on because they put unbalanced data into consideration. As table 5 shows, both recall and F-score grew after the feature of transitivity was added to train the baseline models. Furthermore, the Logistic Regression model out performed the SVM model after transitivity was added. As a result, a Logistic Regression model was considered a better machine learning method for spoiler detection. More importantly, both the growth of recall and f-score suggested that verb transitivity may be a characteristic of spoiler posts

3.2. Error Analysis

3.2.1. Disadvantage of text level analysis

This study chose to analyze the corpus on text level since the researcher viewed verb transitivity as a characteristic of spoiler posts. That is, in a spoiler paragraph, there must be more transitive verbs used than a non-spoiler paragraph. However, there were still long paragraphs that involve only a few spoiler sentences. Figure 2 displays a spoiler post that was misclassified as a review by the models, and the marked sentences are the sentences which spoil the plots of the movie. In this kind of paragraphs, sentences carrying spoilers seem much less than those that do not. The verb transitivity these sentences carried was lowered by all the other non-transitive ones. Therefore, they became the loopholes that the model in this study could not detect.

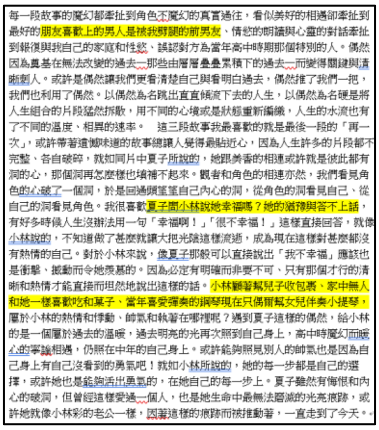
每一段故事的魔幻都牽扯到角色不魔幻的真實過往，看似美好的相遇卻牽扯到最好的朋友喜歡上的男人是被我劈腿的前男友、情慾的朗讀與心靈的對話牽扯到報復與我自己的家庭和性慾、誤認對方為當年高中時期那個特別的人。偶然因為奠基在無法改變的過去—那些由層層疊疊累積下的過去—而變得關鍵與清晰刺人。或許是偶然讓我們更看清楚自己與看明白過去，偶然推了我們一把，我們也利用了偶然。以偶然為名跳出直直傾流下去的人生，以偶然為名硬是將人生組合的片段猛然拆散，用不同的心境或是狀態重新編織，人生的水流也有了不同的溫度、相異的速率。 這三段故事我最喜歡的就是最後一段的「再一次」，或許帶著遺憾味道的故事總讓人覺得最貼近心，因為人生許多的片段都不完整、各自破碎，就如同片中夏子所說的，她跟美香的相連或許就是彼此都有洞的心，那個洞再怎麼樣也填補不起來。觀者和角色的相連亦然，我們看見角色的心破了一個洞，於是回過頭望望自己內心的洞，從角色的洞看見自己、從自己的洞看見角色。我很喜歡夏子問小林說她幸福嗎？她的猶豫與答不上話，有好多時候人生沒辦法用一句「幸福啊！」「很不幸福！」這樣直接回答，就像小林說的，不知道做了甚麼就讓大把光陰這樣流逝，成為現在這樣對甚麼都沒有熱情的自己。對於小林來說，像夏子那般可以直接說出「我不幸福」應該也是衝擊、撼動而令她羨慕的。因為必定有明確而非要不可、只有那個才行的清晰和熱情才能直接而坦然地說出這樣的話。小林顧著幫兒子收包裹、家中無人和她一樣喜歡吃和菓子、當年喜愛彈奏的鋼琴現在只偶爾幫女兒伴奏小提琴，屬於小林的熱情和悸動、帥氣和執著在哪裡呢？遇到夏子這樣的偶然，給小林的是一個屬於過去的溫暖，過去明亮的光再次照到自己身上，高中時魔幻而暖心的爭論相遇，仍照在中年的自己身上。或許能夠照見別人的帥氣也是因為自己身上有自己沒看到的勇氣吧！就如小林所說的，她的每一步都是自己的選擇，或許她也是能夠活出勇氣的，在她自己的每一步上。夏子雖然有悔恨和內心的破洞，但曾經這樣愛過一個人，也是她生命中最無法磨滅的光亮痕跡，或許她就像小林彩的老公一樣，因著這樣的痕跡而被推動著，一直走到了今天。

Figure 2

3.2.2. *Rhetorical Use*

The boxes in figure 3 mark the verbs that were assigned as transitive ones. Nevertheless, none of these verbs talked about real world events that happened. They were in these sentences for rhetorical usage. However, the models used in this study were not capable of recognizing rhetorical sentences, so the disability may be the second possible reason that caused error when the models were classifying texts.

生命(Na) 由(P) 許多(Neqa) 片段(Na) 組合(VC) 而(Cbb) 成(VG) ，(COMMACATEGORY) 片段(Na) 或(Caa) 長(VH) 或(Caa) 短(VH) ，
(COMMACATEGORY) 但(Cbb) 都(D) 是(SHI) 一片片(Neqa) 的(DE) 不(D) 完整(VH) ，(COMMACATEGORY) 由(P) 時間(Na) 和(Caa) 人(Na) 與(Caa)
相遇(VA) 將(P) 其(Nep) 串聯(VC) ，(COMMACATEGORY) 組合成(VG) 有機(A) 的(DE) 生命(Na) 。(PERIODCATEGORY) 長(VH) 的(DE) 片段(Na)
在(P) 重要性(Na) 上(Ng) 不一定(D) 大於(VJ) 短(VH) 的(T) ，(COMMACATEGORY) 短(VH) 的(DE) 片段(Na) 也(D) 可能(D) 重於(VJ) 泰山
(Nc) 。(PERIODCATEGORY) 正(D) 如(P) 故事(Na) 的(DE) 起始(Nv) 紐(Nc) ，(COMMACATEGORY) 往往(D) 也(D) 是(SHI) 被(P) 無心(VH) 的
(DE) 碰觸(VC) 而(Cbb) 啟動(VC) 。(PERIODCATEGORY) 在(P) 偶然(VH) 下(Ng) 被(P) 組合(VC) 而(Cbb) 成(VG) 的(DE) 生命(Na) ，
(COMMACATEGORY) 其中(Nep) 擁有(VJ) 的(DE) 可能性(Na) 是(SHI) 複合式(A) 的(T) ，(COMMACATEGORY) 悠然(VH) 其中(Nep) 千曲百轉(VH)
的(DE) 心情(Na) 與(Caa) 念想(VE) ，(COMMACATEGORY) 是(SHI) 生命(Na) 中(Ng) 最(Dfa) 綿長(VH) 的(DE) 真實(VH) 。(PERIODCATEGORY)
三(Neu) 段(Nf) 故事(Na) 都(D) 各(D) 有(V_2) 其(Nep) 「(PARENTHESISCATEGORY) 魔幻(VH) 」(PARENTHESISCATEGORY) 的(DE) 時刻(Na) ，
(COMMACATEGORY) 「(PARENTHESISCATEGORY) 神奇(VH) 」(PARENTHESISCATEGORY) 的(DE) 一(Neu) 段(Nf) 相遇(VA) 與(Caa) 相處(VA) ，
(COMMACATEGORY) 人生(Na) 在(P) 那(Nep) 一(Neu) 刻(Nf) 又(D) 轉(VAC) 了(Di) 彎(Na) 或是(Caa) 從此(D) 人生(Na) 流淌(VC) 的(DE) 質
地(Na) 不再(D) 相同(VH) 。(PERIODCATEGORY)

Figure 3

3.2.3. *Need of more feature extraction*

Figure 4 shows one of the spoiler posts misclassified as a review by the models. The blocks mark the words that were considered to spoil others about what happened in the movie. The blocks suggested that not only transitive verbs bring about spoilers. Take figure 4 as an example, the sentence blocked is viewed as a spoiler because the name entity, "Jinni" and the word "Jiating geming" reveal the marriage of two main characters. Therefore, one explanation of this error is that the models were not trained to recognize these features.

天狼星(Nb) 死(VH) 後(Ng) 安慰(VC) 哈利(Nb) 的(DE) 是(SHI) 露娜(Nb) 當初(Nd) 唯一(A) 了(Di) 解(VC) 親人(Na) 死去(VH) 的(DE) 悲傷
(VH) 還(D) 能(D) 跟(P) 哈利(Nb) 分擔(VC) 痛苦(VH) 算是(VG) 最(Dfa) 好(VH) 的(DE) 女性(Na) 朋友(Na) 之(DE) 一(Neu) 而且(Cbb) 哈
利(Nb) 還(D) 不(D) 願意(VK) 與(P) 除了(P) 露娜(Nb) 之外(Ng) 的(DE) 任何(Neqa) 人(Na) 談論(VE) 天狼星(Nb) 跟(P) 金妮(Nb) 同(Nes)
屆(Nf) 個性(Na) 直率(VH) 又(Caa) 好(VH) 相處(VA) 的(DE) 可愛(VH) 小(VH) 學妹(Na) 雖然(Cbb) 我(Nh) 知道(VK) 小說(Na) 說(VE) 她
(Nh) 的(DE) 長相(Na) 很(Dfa) 「(PARENTHESISCATEGORY) 怪異(VH) 」(PARENTHESISCATEGORY) 但(Cbb) 靠(P) 小時候(Nd) 電影(Na) 建立(VC)
形象(Na) 的(DE) 我(Nh) 還是(D) 一直(D) 覺得(VK) 她(Nh) 很(Dfa) 正(VH) 混血(A) 王子(Na) 中(Ng) 第一(Neu) 批(Nf) 站出來(VA) 挺
(VC) 哈利(Nb) 的(DE) 隊友(Na) ((PARENTHESISCATEGORY) 還(D) 有(V_2) 妙麗(Nb) 、(PAUSECATEGORY) 榮恩(Nb) 、(PAUSECATEGORY) 金妮
(Nb) 、(PAUSECATEGORY) 奈威(Nb))(PARENTHESISCATEGORY) 如果(Cbb) 沒有(VJ) 榮恩(Nb) 把(P) 妙麗(Nb) 娶走(VC) 的(DE) 劇情(Na) 感覺
(Na) 就(D) 她(Nh) 跟(P) 妙麗(Nb) 其中(Nep) 一(Neu) 個(Nf) 會(D) 是(SHI) 哈利(Nb) 的(DE) 老婆(Na) 要不是(Cbb) 哈利(Nb) 突然(D)
發情(VH) 開始(VL) 喜歡(VK) 金妮(Nb) 不然(Cbb) 我(Nh) 完全(D) 想不到(Dk) 她(Nh) 有(V_2) 任何(Neqa) 機會(Na) 哈利(Nb) 最後(Nd) 還
(D) 把(P) 紀念(VJ) 母親(Na) 命名(VG) 莉莉(Nb) 的(DE) 女兒(Na) 名字(Na) 中(Ng) 再(D) 加入(VC) 了(Di) 好(VH) 閨蜜露娜(Nb) 的(DE)
名字(Na) 為什麼(D) 金妮(Nb) 沒有(D) 鬧(VC) 家庭(Na) 革命(VA) ？(QUESTIONCATEGORY)

Figure 4

4. Conclusion

The current study aimed to verify whether verb transitivity could be said to be a linguistic feature of spoiler posts by machine learning approaches. To do so, Support Vector Machine and Logistic Regression were chosen to be trained for classifying posts on the movie board of PTT Bulletin Board System. TF-IDF was used as the feature of the baseline models, and verb transitivity was added to train the models. The results showed that both SVM model and Logistic Regression model performed better after transitivity was added to the baseline model. Furthermore, the Logistic Regression model outperformed the SVM model. According to the results, high verb transitivity may be considered a feature of spoiler posts on social medium. Also, Logistic Regression model may be a better approach for spoiler detection.

This study has taken a step to investigate Chinese spoiler posts from the angle of linguistics and found that calculating verb transitivity may help elevating the efficacy of spoiler detection. However, the wrong classified cases indicated the disadvantage of training the models at text level. Although text level analysis was chosen because PTT posts are mostly long paragraphs, spoilers appear in short comments, too. Moreover, most of the wrong classified cases were paragraphs that involves a few sentences that may spoil others. Overall, it is worth to mention that there were some problems with the methodology used in this study.

Given the results of the classification, further research is suggested to be worked on sentence level analysis of Chinese posts. In addition, more linguistic features of Chinese spoiler posts should be verified to help machine learning. For example, objectivity and name entities may be study-worthy. Finally, the approach outlined in this study may be merged with the pass studies which focus on dependency parsing to pursue better performance of spoiler detection.

5. References

Boyd-Graber, J., Glasgow, K., & Zajac, J. S. (2013). Spoiler alert: Machine learning approaches to detect social media posts with revelatory information. *Proceedings of the American Society for Information Science and Technology*, *50*(1), 1-9. https://doi.org/10.1002/meet.14505001073

Chang, B., Lee, I., Kim, H., & Kang, J. (2021). " Killing Me" Is Not a Spoiler: Spoiler Detection Model using Graph Neural Networks with Dependency Relation-Aware Attention Mechanism. *arXiv preprint arXiv:2101*.05972.

CKIP Lab. (1993). Zhongwen cilei fenxi, Zhongwenci zhishiku xiaozu jishu baogao #93-05 [An Analysis on Chinese Parts of Speech, CKIP Lab Technological Report #93-05]. Nangang, Academia Cinica.

Guo, S., & Ramakrishnan, N. (2010, August). Finding the storyteller: automatic spoiler tagging using linguistic cues. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)* (pp. 412-420).

Jeon, S., Kim, S., & Yu, H. (2016). Spoiler detection in TV program tweets. *Information Sciences*, *329*, 220-235. https://doi.org/10.1016/j.ins.2015.09.005

Li, P. H., Fu, T. J., & Ma, W. Y. (2020, April). Why attention? Analyze BiLSTM deficiency and its remedies in the case of NER. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 05, pp. 8236-8244). https://doi.org/10.1609/aaai.v34i05.6338

Marukatat, R. (2019, May). A Comparative Study of Using Bag-of-Words and Word-Embedding Attributes in the Spoiler Classification of English and Thai Text. In *International Conference on Applied Computing and Information Technology* (pp. 81-93). Springer, Cham.

Sasano, I., Morisawa, K., & Hirakawa, Y. (2019). Personalized spoiler detection in tweets by using support vector machine. *Journal of Advances in Technology and Engineering Research*, *5*(5), 219-226. https://doi.org/10.20474/jater-5.5.4

Ueno, A., Kamoda, Y., & Takubo, T. (2019). A spoiler detection method for japanese-written reviews of stories. *International Journal of Innovative Computing Information and Control*, *15*(1), 189-198. https://doi.org/10.24507/ijicic.15.01.189

Wan, M., Misra, R., Nakashole, N., & McAuley, J. (2019). Fine-grained spoiler detection from large-scale review corpora. *arXiv preprint arXiv:1905.13416*.

Yang, W., Jia, W., Gao, W., Zhou, X., & Luo, Y. (2019, November). Interactive variance attention based online spoiler detection for time-sync comments. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (pp. 1241-1250). https://doi.org/10.1145/3357384.3357872