Detection of Football Spoilers on Twitter

Yuji Shiratori[1], Yoshiki Maki[1], Satoshi Nakamura[1], and Takanori Komatsu[1]

[1] Meiji University, 4-21-1 Nakano, Nakano-Ku, Tokyo, Japan
swany181@gmail.com

Abstract. Sports spoilers on SNS services such as Twitter, Facebook and so on spoil viewers' enjoyment when watching recorded matches. To avoid spoilers, people sometimes stay away from SNSs. However, people often use SNSs to habitually check messages posted by their friends and build and maintain their relationships. Therefore, we need an automatic method for detecting spoilers from SNSs. In this paper, we generated a Japanese spoiler dataset on Twitter and investigated the characteristics of the spoilers to create a foothold in construction of automatic spoiler detection system. Consequently, we clarified the relationship between spoilers and the statuses of football matches. In addition, we compared three methods for detecting spoilers and show the usefulness of SVM with Status of Match method.

Keywords: Blocking Spoilers, Machine Learning, Sports, football, SNS, Twitter.

1 Introduction

There are many people who like to watch sports games in real time and feel excited and surprised. However, it is often difficult for them to watch sports games in real time because they are busy with work, studies, etc. In such situations, it is common to schedule a recording in advance and watch it when they have some free time. However, if a viewer gets to know the results of the match before watching it, feelings of excitement and surprise would probably be lost. Since such viewers would like to avoid "spoiler information" such as scores and the winners/losers of the match, they actively choose a self-imposed isolation from their community to block information on the match until they watch it. However, since SNS services such as Twitter, Facebook and so on allow people to habitually check messages posted by their friends and to build and maintain their relationships, the isolation approach should not be taken to keep their good relationship. To avoid deteriorating personal relationships, we need an automatic method for detecting spoilers from SNSs. Here, if potential users want to avoid the status of football games, a simple approach to find football posts may be reasonable. However, the approach blocks all football posts, not excepting posts that viewers don't want to block. Further, we think viewers should be able to enjoy conversation about the target games unless it contains any spoiler information.

In recent years, researches have been conducted to block such spoiler information. For example, Nakamura et al. proposed a method for filtering information on web pages that corresponds to a user's interest on the basis of their e-mail and a TV program guide [1]. In previous researches, the researchers focused on methods for blocking spoilers by interacting with the users and the systems, but it has not been clarified that what characteristics spoilers have and how to detect them with high accuracy.

Therefore, we investigated the characteristics of spoilers by generating a spoiler dataset on posts about football matches on Twitter and examined methods for detecting spoilers with high accuracy.

The contributions of this work are: 1) the generation of the football spoiler dataset about Twitter posts; 2) the experimental proof of the usefulness of SVM with Status of Match method by comparison of the accuracy of spoiler detection with three methods.

The rest of this paper is organized as follows. Section 2 shows a discussion of the related works. Section 3 and 4 are about generating spoiler dataset and analysis of it. Section 5 explains a verification of the effectiveness of three word-based methods. Finally, Sections 6 and 7 are discussion and our conclusions.

2 Related Work

2.1 Influence of Spoilers

Regarding investigation into the influence of spoilers, Leavitt et al. focused on novels and investigated what kinds of differences appeared in terms of user's enjoyment when spoiler information was presented and when it was not [2]. As a result of the experiment, it was claimed that spoiler information does not lower the fun of the content. However, the act of reading a novel and the act of watching sports are essentially different. Moreover, it was only suggested that spoilers help readers understand the content and personal relationships among the characters, resulting in making it easier to read novels by presenting a summary. In addition, Rosenbaum et al. confirmed that those who are not familiar with a novel feel that the story is more interesting with spoilers, and those who are familiar feel that it is more interesting without [3].

Several researchers revealed the bad influence of spoilers. Therefore, there is a need to automatically filter out spoilers from SNSs. To create a foothold to realize it, we generated a spoiler dataset and examined methods for detecting spoilers with high accuracy.

2.2 Blocking Spoilers

As researches into blocking information similar to spoilers, researches on review texts on the Internet have been widely conducted. Ikeda et al. are concerned with the inclusion of spoilers in review texts for story content and eliminate spoilers using machine learning [4]. Pang et al. identified which sentences do not include outline of a story with support vector machine (SVM) for review texts [5]. In the researches on these review texts, all outlines are judged as spoilers, but in sports, outlines (content of the matches) are not directly spoiled (like comments about showing one's happiness and sadness). Therefore, it is slightly different in nature from the spoilers discussed in this paper.

As research on the problem of spoilers on SNSs such as Twitter, Facebook and so on, Boyd-Graber et al. conducted an evaluation of machine learning approaches to find spoilers in social media posts [6]. They targeted movie reviews and used classifiers on multiple sources to determine which posts should be blocked.

Relative to these studies, sports spoilers are often related to game results. Therefore, their contents differ from that of the spoilers dealt with in these studies. We analyze the characteristics of sentences of spoilers about football games and examine methods for detecting spoilers with high accuracy. Jeon et al. proposed a method of detecting spoilers using machine learning, focusing on "named entities", "frequently used verbs", "future tense", etc. in comments on Twitter [7]. By conducting experiments using comments on television programs, they found that it was possible to detect tweets with spoilers with a high precision compared with methods that use keyword matching or latent Dirichlet allocation (LDA), and they confirmed the utility. In addition, they also carried out an experiment about sports spoilers. However, they conducted it for only one match and labeled tweets themselves, so the evaluation was actually not strict. Moreover, it has no applicability to tweets in Japanese because Japanese does not have future tense. We had classifiers construct a spoiler dataset for tweets on nine football matches and we examined methods for detecting spoilers that can be applied to Japanese.

3 Generating Spoiler Dataset

In this section, as a foothold for the construction of automatic spoiler detection system, we analyze the characteristics of spoilers to know what kind of information a spoiler holds. Currently, viewers can encounter spoilers at various forms of media such as news websites, weblogs, and search websites. SNSs like Twitter in particular have increased the chance of encountering spoilers. As for Twitter, there are many people who casually use it because they can learn what their friends are doing just by accessing it and can easily communicate with others; thus, there is a high possibility of seeing spoiler information. Therefore, we collect posts on Twitter related to football matches and analyze the characteristics of spoilers by constructing a spoiler dataset. From now on, a Twitter post is called a "tweet".

To generate a dataset, we first collected tweets on football matches. Here, we focused on matches played by the Japan national football team on which there was a particularly large number of tweets by many fans [8]. The information on the matches is shown in Table 1.

Table 1. Matches for generating dataset

Match	Score	Day
2015 Women's World Cup "Japan vs. England"	JPN 2 – 1 ENG	07/01/15
2015 Women's World Cup "Japan vs. United States"	JPN 2 – 5 USA	07/05/15
2015 EAFF East Asian Cup "Japan vs. South Korea"	JPN 1 – 1 KOR	08/05/15
2015 Women's EAFF East Asian Cup "Japan vs. China"	JPN 2 – 0 CHN	08/08/15

2015 EAFF East Asian Cup "Japan vs. China"	JPN 1 – 1 CHN	08/09/15
World Cup Qualifiers "Japan vs. Cambodia"	JPN 3 – 0 KHM	09/03/15
World Cup Qualifiers "Japan vs. Afghanistan"	JPN 6 – 0 AFG	09/08/15
Friendlies "Japan vs. Iran"	JPN 1 – 1 IRI	10/13/15
World Cup Qualifiers "Japan vs. Singapore"	JPN 3 – 0 SIN	11/12/15

Currently, when tweeting real-time content, symbols called "hashtags" can be used for search/classification in some cases. For example, hashtags such as "#daihyo" and "#JPN" are used for matches of the Japanese national football team. If a hashtag for a target sport is attached to a tweet about a target match, it is sufficient to block all tweets including that hashtag. However, there are many tweets without a hashtag that are actually related to a match. To block them, it is necessary to analyze the contents of the tweets. In addition, it will be possible to cut out the need of changes in the set of hashtags.

However, if we collect all tweets related to a football match, we need to collect all the tweets that are posted at that time and then select tweets related to football matches. This leads to a problem of accuracy in selecting tweets, and also it is not possible to collect private tweets. In addition, streaming APIs provided by Twitter cannot collect all tweets.

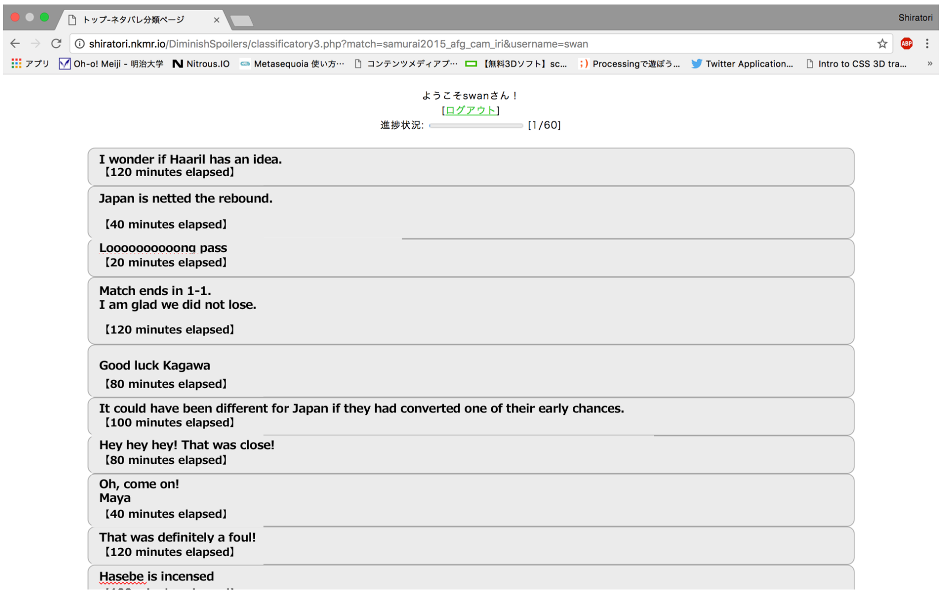
Therefore, tweets with hashtags and tweets without hashtags are considered to have no significant difference in terms of their contents in this paper, although there is a difference in tweets depending on whether they have hashtags or not. We decided to give priority to collect tweets efficiently and collecting tweets with hashtags. Here, some hashtags such as "#daihyo" or "#JPN" which are commonly used for matches of the Japanese national football team were selected before a match, and tweets including the hashtags were collected using the Search API provided by Twitter. Tweets were collected from the start of a match to 2 hours after.

Among the collected data, there were also many tweets that were not appropriate for classification and analysis. Therefore, we removed inappropriate tweets and formatted tweets using the following procedure.

1. Since many tweets from opponent countries are also posted on matches such as the World Cup games, the collected tweets were in multiple languages. Considering that the dataset constructer is a native speaker of Japanese, we removed the tweets in languages other than Japanese. To remove non-Japanese tweets, the language code was acquired when tweets were collected. Japanese tweets were judged depending on whether the language code is "ja" or not.
2. "RT" at the beginning of a tweet is called "retweet". This action can repost other viewers' tweets without modification. This action is taken to send other viewers' tweets to those who are seeing your tweets. Since it overlaps with the original tweet, it was removed by regular expressions.
3. Hashtags were removed from collected tweets. In this case, from "#" to one character short of blank or new line were judged by regular expression. Also, we deleted blank tweets by regular expression (because there are tweets with hashtags only). Here, when there were only spaces or new lines between the beginning of the tweet and the end, that tweet was removed.

4. Because there were many spam tweets unrelated to matches in tweets including URLs, tweets including “http://t.co/” or “https://t.co/” were judged and removed by regular expression.

After doing the procedure above, we developed a web system in order to have tweets in the dataset labeled as spoiler and non-spoiler. Five college students helped us label the tweets. The students were aged 19 to 22 who were interested in watching football matches and regularly use Twitter. Figure 1 shows the web evaluation system (Figure 1 shows tweets we translated). If a labeler feels that a tweet is a spoiler, he/she click it. Also, labelers may find spoilers when tweets such as “Kagawa, get it!” and “Nice!!!” are posted at the same time, when “G”, “O”, “A” and “L” is posted at the same time, when the number of tweets on a match suddenly increases, or others besides independent tweets. However, to avoid the situation that the web system and criteria for classification become complicated and difficult, tweets were labeled based on independent tweets. In addition, since it takes a huge amount of time to classify all tweets, the number of tweets to be presented is 1000 per match. If one match is presented at a time, there is a possibility that the context and content of the match will be clearly transmitted from one tweet. For example, supposing that a match that Kagawa (a football player on the Japan national team) scored points is being labeled, there is a possibility that even tweets that do not clearly include spoilers about a match such as “KAGAWA” will be judged as spoilers if labelers who know the match’s result look at that kind of tweet and assume that the details of a target match can be understood. This is undesirable because labelers actually cannot know the content of the match (they may know preliminary information of the match) even though they look forward to watching the match. Therefore, we decided to present three matches each at random. In other words, 9000 tweets were divided into 3 groups of 3,000 tweets. In addition, labelers can understand the elapsed time roughly from the start of the match without watching. If there is a tweet such as “Defense is meaningless, let's go attack” at the start of the match, it is assumed that many labelers may think that this is a tweet about simple enthusiasm to the match. However, many labelers may regard such a comment as showing their team was losing if the tweet was at the second half of the match. We thought that this is necessary for judging spoilers. Therefore, tweets were presented randomly rather than in chronological order and the approximate elapsed time when the tweet was posted from the start of the match was displayed with each tweet. For example, if “60” was displayed with a tweet, it was a tweet from 51 minutes to 70 minutes after the start of the match. The reason it was set to an approximate time is because the viewer cannot know the exact time when the match was actually recorded.



(a) Entire

(b) Scale-up

Fig. 1. Screenshot of evaluation system

When accessing the web system, tweets were displayed for 50 tweets each page. Also, there were 60 pages per group and there were five labelers per group.

4 Analysis of Spoiler Dataset

In this section, we analyze the contents of the dataset explained in the previous section.

Table 2. Examples of dataset

Tweet	Elapsed time	Label
"Ooooh! Kagawa scores a goal ! ! ! "	20	spoiler
"Already allowed two goals (´Д`)"	0	spoiler
"Now kick off"	0	non-spoiler
"Hmm. A missed pass is no good"	20	non-spoiler

Table 3. Concordance rate of judging spoilers

Number of matching people	Number of tweets	Percentage of tweets	Percentage of tweets in spoiler's
5	351	3.90	14.03
4	680	7.56	27.18
3	620	6.89	24.78
2	217	2.41	8.67
1	634	7.04	25.34
0	6498	72.20	-

Tables 2 and 3 show some examples of the dataset and spoiler matching rates of classified tweets, respectively. Then, Table2 shows tweets translated by us. The tweets that all five judged to be spoilers mostly talked about the final result of the match. This is because the final result of a match is thought to be a spoiler for everyone. Many of the tweets that three or four labelers judged to be spoilers were about how the match went. This indicates that there was a certain number of viewers who considered only the final result of a match as important and did not regard other important moments as spoilers. Also, most of the tweets that one or two people judged to be spoilers indirectly expresses how the match went or were about scenes other than decisive moments. This is probably because the degree of perusing tweets, familiarity with target sports, of sensitivity to spoilers varies from viewer to viewer. There were various tweets that no one judged as spoilers, and they were related to moments that had less to do with the content of a match, simple cheering message, or a moment with low importance.

On the basis of the results, we regarded tweets judged as spoilers by more than half of the participants (= more than 3 people) as spoiler tweets. In other words, there were 1,651 spoiler tweets among the 9,000 tweets.

Many spoiler tweets contained specific pattern descriptions and words of each status of matches. Therefore, ten cases of terms used frequently in spoiler tweets were compared with non-spoiler tweets (TF-IDF [9]), as shown in Table 4. Then, Table4 shows translated terms. Tweets were divided into the winning time zone, losing time zone, and tying time zone for Japan. MeCab was used for the word division. In addition, consecutive nouns of one character were treated as one word. Single-character words that have a basic form other than a noun and were not defined in the dictionary, particles, auxiliary verbs, and meaningless words were eliminated. Also, since repetitive expressions were noise, the dataset was formatted in reference to Brody et al.'s method [10]. Furthermore, since proper nouns vary greatly from match to match, the numbers were mechanically generalized as [num], the player names as [player], the team names as [team], and the coach names as [coach] by pattern matching.

Looking at each time zone from Table 4, terms used frequently in the losing time zone were different from those in the winning time zone and the tying time zone except for the terms "[player]", "[team]", and "Break the deadlock". Also, the winning time zone was different from the tying time zone except for the terms "[player]", "[team]", "Goal", and "Match". There were frequently used terms that directly expressed the statuses of the matches such as "Win" in the winning time zone and "Tie" in the tying time zone. Furthermore, terms on scoring goals, such as "[num]th points", "[num]points", "Point" in the winning time zone, on scoring such as "[num] - [num]" in the tying time zone, and on allowing goals such as "Allowing goals" in the losing time zone, were used frequently, suggesting that the content of spoilers differed depending on the time zone.

Table 4. Terms used frequently in spoiler tweets

Winning time zone		Losing time zone		Tying time zone	
Terms	TF-IDF	Terms	TF-IDF	Terms	TF-IDF
[Player]	0.742	[Player]	0.531	[Player]	0.627
[Team]	0.422	[Team]	0.475	[Team]	0.552
Goal	0.261	Break the deadlock	0.238	Goal	0.226
[Num]th points	0.173	Allowing goals	0.238	Tie	0.201
[Num]points	0.131	Parry	0.224	Match	0.151
[Num]	0.117	Second	0.112	[Num]-[Num]	0.136
Win	0.106	Score	0.112	End	0.125
Match	0.099	Too	0.112	National	0.110
point	0.095	be -ed (passive voice)	0.112	First	0.105

5 Experiment: Spoiler Detection

5.1 Experiment Procedure

In this section, we examined methods for detecting spoilers with high accuracy on the basis of the dataset. According to the previous section, since sports spoilers have prominent characteristics in terms of words, we compared three word-based methods: pattern matching, SVM (frequently terms were used as features), and SVM with Status of

Match (frequently terms were used as features). In addition, we selected a SVM model based on the results of research by Jeon et al [7].

- **Pattern Matching:** Terms used frequently in spoilers such as keywords and tweets containing terms matching the keywords were judged as spoilers. Terms were divided into rules by following the previous section (consecutive tweets with one character for a single word are combined, etc.), and terms with a TF-IDF value of 0.100 or higher were taken as keywords. 0.100 was set as the threshold because the F-measure was the highest at 0.100 as a result of performing an analysis by changing the threshold by 0.050 from 0.000 to 0.300.
- **SVM:** We generated an SVM model using the tweets of matches other than the match to be detected, and we detected the tweets (1000 test data) of the matches using the model. When preparing the model, we adjusted the amount of data by under-sampling because the number of non-spoiler tweets was higher than that of spoilers (Table 5 shows the number of training data for the SVM-method). In addition, vectors for SVM were generated using a BoW (Bag-of-Words) [11] of each tweet. Words were divided by rules from the previous section, and a linear kernel with a learning rate of 0.01 was set as a parameter for learning model generation by grid search. Also, to make the scale of each dimension (word) the same, normalization was performed.
- **SVM with Status of Match**: According to the previous section, since terms used frequently differ by time zone, considering the statuses of matches in generating SVM's model, we generated a winning model for the winning time zone, a losing model for the losing time zone, and a tying model for the tying time zone (Table 6 shows the number of training data for the method of SVM with Status of Match). Then, we detected test tweets with the winning model if Japan was winning at the time of the tweet, with the losing model if Japan was losing at the time of the tweet, and with the tying model if Japan was tied at the time of the tweet (Table 7 shows the number of test data for the method of SVM with Status of Match). We performed word segmentation, under-sampling, SVM parameters (learning rate), and kernels the same way as the SVM method. In addition, since this method detects spoilers for each time zone, there were matches with extremely little or no test data. When the number of spoiler tweets in the test data was 20 or less, it is considered that exceptional tweets would greatly influence the result; therefore, these tweets were excluded from the result (even in the case of 0 because results such as precision cannot be calculated). This method presupposes that, since it is necessary to detect the status of the match at the time of each tweet, it is necessary to have some delay in the display of the tweet during the match, and if it is hard to decide which team (or player) from a domestic league a viewer is cheering, there is a time zone in which it is necessary to use the winning model and the losing model at the same time.

Table 5. The number of training data for the SVM-method

Match	
2015 Women's World Cup "Japan vs. England"	3248
2015 Women's World Cup "Japan vs. United States"	3240
2015 EAFF East Asian Cup "Japan vs. South Korea"	3704
2015 Women's EAFF East Asian Cup "Japan vs. China"	3544
2015 EAFF East Asian Cup "Japan vs. China"	3710
World Cup Qualifiers "Japan vs. Cambodia"	3424
World Cup Qualifiers "Japan vs. Afghanistan"	3350
Friendlies "Japan vs. Iran"	3624
World Cup Qualifiers "Japan vs. Singapore"	3596

Table 6. The number of training data for the method of SVM with Status of Match

Match	Winning	Losing	Tying
2015 Women's World Cup "Japan vs. England"	1704	674	870
2015 Women's World Cup "Japan vs. United States"	2106	954	180
2015 EAFF East Asian Cup "Japan vs. South Korea"	2106	764	834
2015 Women's EAFF East Asian Cup "Japan vs. China"	1794	880	870
2015 EAFF East Asian Cup "Japan vs. China"	2106	770	834
World Cup Qualifiers "Japan vs. Cambodia"	1624	930	870
World Cup Qualifiers "Japan vs. Afghanistan"	1528	952	870
Friendlies "Japan vs. Iran"	2106	756	762
World Cup Qualifiers "Japan vs. Singapore"	1774	952	870

Table 7. The number of test data for the method of SVM with Status of Match

Match	Winning	Losing	Tying
2015 Women's World Cup "Japan vs. England"	328	672	0
2015 Women's World Cup "Japan vs. United States"	0	12	988
2015 EAFF East Asian Cup "Japan vs. South Korea"	0	897	103
2015 Women's EAFF East Asian Cup "Japan vs. China"	345	655	0
2015 EAFF East Asian Cup "Japan vs. China"	0	838	162
World Cup Qualifiers "Japan vs. Cambodia"	842	158	0
World Cup Qualifiers "Japan vs. Afghanistan"	932	68	0
Friendlies "Japan vs. Iran"	0	797	203
World Cup Qualifiers "Japan vs. Singapore"	855	145	0

The three methods above were compared in terms of precision, recall, and F-measure. For all three, the experiment was conducted for nine matches (the number of matches in the dataset), and the average of the nine matches was calculated as a result. Here, for each method, precision means “the ratio of tweets that were detected correctly to detected tweets”, recall means “the ratio of tweets that were detected correctly to spoiler tweets” and F-measure is expressed by equation (1).

$$F - measure = \frac{2 \cdot precision \cdot recall}{precision + recall} \quad (1)$$

5.2 Experimental Results

Table 8 shows the average of the precision, recall, and F-measure of each method for each match.

SVM with Status of Match had the highest F-measure. The highest of precision was SVM with Status of Match, but that for recall was SVM.

Table 8. Accuracy in detecting spoilers for each method

Method	Precision	Recall	F-measure
Pattern matching	0.270	0.668	0.372
SVM	0.617	0.601	0.598
SVM with Status of Match	0.698	0.565	0.611

6 Discussion

The F-measure of SVM with Status of Match was the highest; thus, this method was superior to the others. In particular, the precision was better than the others. The reason the precision of SVM and SVM with Status of Match was superior to pattern matching is that pattern matching detected a spoiler only from the player name. For example, pattern matching detected a spoiler about a tweet such as “Kagawa’s missed pass is scary because of the heavy turf” because of the word “Kagawa” in the tweet, but SVM detected not only the players’ names but also words such as “goal” that appeared alone with the names. The reason the precision of SVM with Status of Match was superior to SVM seems to be that mistakenly learned tweets by SVM were no longer learned for every time zone by SVM with Status of Match. In fact, tweets such as “It’s been a while since I felt refreshed last” and “I saw a sweeping victory for the first time in a very long time” in the winning time zone were able to be detected correctly, so it is considered that tweets such as “Attacking midfielder Kagawa maybe after a long time” and “I saw a national team match for the first time in a very long time!!” in the tying time zone at the start of the match were fitted as non-spoilers in SVM but were not learned by SVM with Status of Match.

In comparison, SVM was superior to SVM with Status of Match for recall. This is because the training data for SVM with Status of Match were divided into three, so it is assumed that the amount of training data was simply less than for SVM. Therefore, it is possible that the recall was improved by increasing the number of matches of the training data and also for the F-measure in SVM with Status of Match. Regarding recall, pattern matching was the most excellent. It is considered that when the threshold of TF-IDF was set to 0.100, there are many spoiler words among the matches. However, the precision was low as a result.

As a result, the F-measure was not that high for any of the methods. This may be because we labeled tweets as spoiler or non-spoiler directly. Therefore, unimportant tweets such as "Nagatomo got a cramp!" were judged to be spoilers because we don't set up the standard for labeling. We need to focus on crucial spoilers at first and set up the standard for labeling. This may also be because tweets such as "I want to see Honda score a goal" and "We will win" were judged as spoilers. Tweets of hope and enthusiasm need to be judged as non-spoilers, but it is difficult to judge from the grammar because the Japanese language does not have a future tense; therefore, it is necessary to use a method different from morphological analysis. Examining methods for detecting tweets for the future is a future problem. In addition, the fact that the number of training data was small for the two methods using SVM is also considered to be the reason the F-measure is not that high. In particular, as shown in Table 5, the number of training data for SVM with Status of Match in the losing model was small. In fact, the accuracy of detecting spoilers with this method for each model is shown in Table 9. The F-measure for the winning model was 0.664, and that for the losing model was 0.447. It is suggested that the number of training data may have been an influence. Moreover, we plan to examine separation of SVM with Status of Match model because there is a possibility that the accuracy of SVM with Status of Match may be improved by separating the timing of the goal from the model.

Figure 2 shows Precision-Recall curve for SVM with Status of Match in the winning model. It appears that precision was about 0.3 if keeping high recall. We need to think other methods to design a high-recall model first and then create models realizing higher precision because it may shock viewers even if a spoiler detection system cannot block just one spoiler tweet.

Table 9. Accuracy of detecting spoilers with SVM with Status of Match for each model

Model	Precision	Recall	F-measure
Winning model	0.716	0.646	0.664
Tying model	0.656	0.528	0.585
Losing model	0.773	0.315	0.447

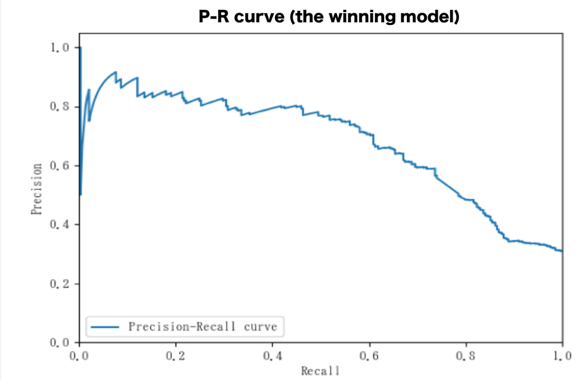


Fig. 2. Precision-Recall curve

7 Conclusion

We investigated the characteristics of spoilers by generating a spoiler dataset for football matches. As a result of analyzing the dataset, it was revealed that the content of spoilers varies depending on the status of a match. Furthermore, we compared the accuracy of spoiler detection by pattern matching, SVM, and SVM with Status of Match. Consequently, we showed that SVM with Status of Match was superior to the other methods in terms of F-measure. The method can be applied to other languages because feature values were frequencies of used terms and status of matches were language-neutral.

In the future, we will improve the accuracy of spoiler detection by increasing the amount of training data and devising better data preprocessing for construction of automatic spoiler detection system such as a Twitter client in order to realize smoother collaborative communication. Furthermore, we plan to conduct experiments for other sports genres.

Acknowledgments. This work was supported in part by JST ACCEL Grant Number JPMJAC1602, Japan.

References

1. Nakamura, S., Tanaka, K.: Temporal Filtering System for Reducing the Risk of Spoiling a User's Enjoyment. In: Proceedings of the 12th international conference on Intelligent user interfaces, pp. 345-348. ACM, Honolulu (2007).
2. Leavitt, J, D. and Christenfeld, N, J, S.: Story Spoilers Don't Spoil Stories. In: Psychological Science, vol. 22, pp. 1152–1154 (2011).
3. Rosenbaum, Judith, E. and Johnson, Benjamin, K.: Who's afraid of spoilers? Need for cognition, need for affect, and narrative selection and enjoyment. In: Psychology of Popular Media Culture, vol. 5, pp. 273–289. (2016).

4. Ikeda, K. Hijikata, Y. and Nishida, S.: Proposal of Deleting Plots from the Reviews to the Items with Stories. In: Proc. of SNSMW'10, vol. 6193, pp. 346-352. CDROM, (2010).
5. Pang, B. and Lee, L.: A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In: Proc. of ACL'04, pp. 271-278. (2004).
6. Boyd-Graber, J. Glasgow, K. and Zajac, J. S.: Spoiler Alert: Machine Learning Approaches to Detect Social Media Posts with Revelatory Information. In: Proceedings of the 76th Annual Meeting of the American Society for Information Science and Technology, vol. 50, pp. 1-9. ASIST, (2013).
7. Jeon, S. Kim, S. and Yu, H.: Spoiler Detection in TV Program Tweets. In: Information Sciences, vol. 329, pp. 220-235. (2016).
8. Top 10 Most Watched Sports In The World, http://top-10-list.org/2010/10/04/10-most-watched-world-sports/, last accessed 2017/01/27.
9. Baeza-Yates, R. A. and Ribeiro-Neto, B. A.: Modern information retrieval: the concepts and technology behind Search. 2nd Edition. Addison-Wesley Professional, USA (2011).
10. Brody, S. and Diakopoulos, N.: Cooooooooooooooollllllllllllll!!!!!!!!!!!!!!: using word lengthening to detect sentiment in microblogs. In: Proc.Conference on Empirical Methods in Natural Language Processing, pp. 562-570. Association for Computational Linguistics, Stroudsburg (2011).
11. Manning, C.D. and Schtze, H.: Foundations of Statistical Natural Language Processing. The MIT Press, USA (1999).