

Disease Event Detection based on Deep Modality Analysis

Yoshiaki Kitagawa[†], Mamoru Komachi[†], Eiji Aramaki[‡],
Naoaki Okazaki[§], and Hiroshi Ishikawa[†]

[†] Tokyo Metropolitan University {kitagawa-yoshiaki at ed., komachi at, ishikawa-hiroshi at}tmu.ac.jp

[‡] Kyoto University eiji.aramaki at gmail.com

[§] Tohoku University okazaki at ecei.tohoku.ac.jp

Abstract

Social media has attracted attention because of its potential for extraction of information of various types. For example, information collected from Twitter enables us to build useful applications such as predicting an epidemic of influenza. However, using text information from social media poses challenges for event detection because of the unreliable nature of user-generated texts, which often include counter-factual statements.

Consequently, this study proposes the use of modality features to improve disease event detection from Twitter messages, or “tweets”. Experimental results demonstrate that the combination of a modality dictionary and a modality analyzer improves the F1-score by 3.5 points.

1 Introduction

The rapidly increasing popularity of Social Networking Services (SNSs) such as Twitter and Facebook has greatly eased the dissemination of information. Such data can serve as a valuable information resource for various applications. For instance, Huberman et al. (2009) investigated actual linked structures of human networks, Boyd et al. (2010) mapped out retweeting as a conversational practice, and Sakaki et al. (2010) detected earthquakes using SNSs.

An important and widespread application of SNS mining is in the public health field such as infection detection. Among various infectious diseases, influenza is one of the most important diseases worldwide.

However, it is difficult to estimate the precise number of influenza-infected patients based on naïve textual features because SNS messages that contain the word “flu” might not necessarily refer

to being infected with influenza. The following tweets are examples of such cases:

- (1) I **might** have the flu.
- (2) **If** I had the flu, I **would** be forced to rest.

“might” in example (1) suggests that there is only a suspicion of having influenza. Similarly, “if” in example (2) shows that the person is not actually infected.

To filter these instances, we propose to integrate two modalities of information into factuality analysis: shallow modality analysis based on a surface string match and deep modality analysis based on predicate-argument structure analysis. The main contribution of this paper is two-fold:

- We annotate a new dataset extracted from Twitter for flu detection and prediction task, and extend the naïve bag-of-words model of Aramaki et al. (2011) and propose several Twitter-specific features for disease event detection tasks.
- We show that modality information contributes to the factuality analysis in influenza-related tweets, which demonstrates the basic feasibility of the proposed approach. All features presented in this paper increase recall.

2 Related work

The task of influenza detection and prediction originates from the work of Serfling (1963) in epidemiology who tried to define a threshold for influenza breakout.

Since then, various approaches have been proposed for influenza detection and prediction (Groendyke et al., 2011; Moreno et al., 2002; Mugglin et al., 2002).

During the last decade, web-mining approaches have been proposed to detect influenza bursts in

Table 1: Examples of annotated data.

label	tweet
positive	やっぱりインフルエンザだったか...こりゃ家族内で蔓延しそうだな... English translation: After all I was infected with flu ... This virus is likely to spread in the family.
negative	まあ俺インフルエンザのワクチンとか打ったことないんですけどね English translation: Well, I'd never got a preventive shot against flu .

their early stages. Two sources of people’s behavior have been mainly employed: (1) web search queries (such as Yahoo! search (Polgreen et al., 2008) and Google search (Ginsberg et al., 2009)), and (2) activity logs of SNSs. This study specifically examines the latter because of the availability and accessibility of data.

Twitter is the SNS that is most frequently used for influenza detection (Achrekar et al., 2012; Aramaki et al., 2011; Ji et al., 2012; Sadilek et al., 2012; Lamb, 2013). Previous research on the subject has revealed a high correlation ratio between the number of influenza patients and actual tweets related to influenza.

It is possible to obtain large amounts of data from Twitter texts, but the main challenge is to filter noise from this data. For example, Aramaki et al. (2011) reported that half of the tweets containing the word “cold (disease)” simply mention some information about a disease, but do not refer to the actual eventuality of having the disease.

To address that problem, a classifier was produced to ascertain the factuality of the disease event. This paper follows that approach, using modality analysis, which provides a strong clue for factuality analysis (Saurí and Pustejovsky, 2012).

Modality has been used and discussed in various places. Li et al. (2014) employ such modality features, although they do not describe the effect of using modality features in web application tasks. Furthermore, several workshops have been organized around the use of specific modalities, such as Negation and Speculation (e.g. NeSP-NLP¹). In this study, we use generic modality features to improve factuality analysis.

¹www.clips.ua.ac.be/NeSpNLP2010/

3 Modality analysis for disease event detection

3.1 Task and data

The disease event detection task is a binary classification task to extract/differentiate whether the writer or the person around the writer is infected with influenza or not. However, because of the inherently noisy nature of tweets, some tweet messages are unrelated to influenza infection even when the messages include the word “flu.” Therefore, we adopt a supervised approach first proposed by Aramaki et al. (2011).

We annotate a tweet with a binary label (influenza positive and negative), as in prior studies (Aramaki et al., 2011)². If a tweet writer (or anybody near the writer) is infected with influenza, then the label is positive. Otherwise, the label is negative. Additionally, we save the time stamp when the tweet was posted online. Table 1 presents some examples. For this study, we use 10,443 Japanese tweet messages including the word “flu.” In this dataset, the number of positive examples is 1,319; the number of negative examples is 9,124.

Because language heavily relies on modality to judge the factuality of sentences, modality analysis is a necessary process for factuality analysis (Matsuyoshi et al., 2010b). In line with this observation, we propose two ways to incorporate modality analysis for factuality analysis.

3.2 Shallow modality feature

In Japanese, multiple words can serve as a function word as a whole (Matsuyoshi et al., 2007). We designate them as “functional expressions.” Even though functional expressions often carry modality information, previous works including Aramaki et al. (2011) do not consider functional expressions that comprise several words. Therefore,

²These data are used for training an influenza web surveillance service “Mr.influ” <http://mednlp.jp/influ/>.

Table 2: Sense ID feature based on Tsutsuji.

tweet	sense ID
インフルエンザですか...びっくりしました。 English translation: You were infected with flu... I was surprised.	で→r32 です→D41 か→Q31 し→n13 (The words such as “were” and “with” are converted to sense IDs.)

Table 3: Extended modality feature based on Zunda.

tweet	extended modality
隣の患者さんがインフルエンザ発覚 English translation: I found out that the patient next to me had the flu.	発覚=成立 found out = happened

we use the hierarchically organized dictionary of Japanese functional expressions, “Tsutsuji³,” as the first approach.

Tsutsuji provides surface forms of 16,801 entries. In addition, it classifies them hierarchically. Each node in the hierarchy has a sense ID. We use the sense ID of Tsutsuji as a shallow semantic feature to capture the modality of the main predicate in tweets. To find functional expressions related to influenza, we use this feature when a functional expression in Tsutsuji is found within 15 characters to the right context of “flu.” Table 2 presents an example of a tweet and the sense ID feature assigned by Tsutsuji.

3.3 Deep modality feature

To incorporate deep modality analysis, we use the output of the Japanese Extended Modality Analyzer, “Zunda,⁴” which analyzes modality such as authenticity judgments (whether the event has happened) and virtual event (whether it is an assumption or a story) with respect to the context of the events (verbs, adjective, and event-nouns). It is trained on the Extended Modality Corpus (Matsuyoshi et al., 2010a) using rich linguistic features such as dependency and predicate–argument structural analysis. It complements the dictionary-based shallow modality feature described in the previous section.

Specifically, Zunda grasps the modality information such as negation and speculation. See the following example:

- (1) インフルエンザにかかって ない。
(English translation: I am **not infected**

³Tsutsuji: Japanese functional expressions dictionary
<http://kotoba.nuee.nagoya-u.ac.jp/tsutsuji/>

⁴Zunda: extended modality analyzer
<https://code.google.com/p/zunda/>

Table 4: Result of binary classification for disease event detection.

feature	Prec.	Rec.	F1-score
BoW	74.0	30.5	43.2
BoW+URL	69.9	31.3	43.2
BoW+Atmark	74.0	30.5	43.2
BoW+N-gram	70.7	34.5	46.4
BoW+Season	72.4	33.3	45.6
BoW+Tsutsuji	76.4	32.1	45.2
BoW+Zunda	69.9	31.3	43.2
baseline	69.7	39.2	50.2
baseline+Tsutsuji	70.2	42.0	52.6
baseline+Zunda	67.9	41.2	51.3
All	68.9	44.0	53.7

with influenza.)

- (2) インフルエンザにかかった かもしれない。(English translation: I **might** be *infected* with influenza.)

For this example, Zunda detects that “infected” is an event and judges the probability of it describing an event. For example (1) and (2), Zunda respectively outputs “not happened” and “high probability happened”.

We consider verbs and event-nouns that follow the word “flu” to be related to influenza infection. In addition, we assign the estimated modality to them as a deep modality feature. Table 3 presents an example of a tweet and the estimated modality feature assigned by Zunda.

4 Experiment of disease event detection

4.1 Evaluation and tools

Considering our purpose of disease event detection, it is important to estimate the number of positive instances for flu correctly. In contrast, it is

Table 5: Contribution and error analysis of shallow modality features.

Example 1 (correct example)	@* 強力なインフルエンザらしくてですね, まだまだ完治しておりませぬ English translation: @* The flu is apparently terrible and I have not recovered yet.
Example 2 (false positive)	@* おかえびもずくさん、インフル流行ってるから手洗いうがいしてね English translation: @* The flu is going around, so you should wash hands and gargle.
Example 3 (false negative)	まさかのインフルエンザ... 全身鳥肌と震え半端ない寒気が... English translation: I can not believe I have the flu! I have goose bumps. I shiver and feel so cold... .

Table 6: Examples of deep modality feature with large weight. English translations are given in parentheses.

feature	weight	feature	weight
罹患=成立 (infection = happened)	0.80	注射=成立 (injection = happened)	-0.62
かかり=成立 (infect = happened)	0.65	対策=成立 (countermeasure = happened)	-0.50
診断=成立 (diagnosis = happened)	0.52	かかり=0 (infection = 0)	-0.48
寝=成立 (sleep = happened)	0.47	なる=成立 (become = happened)	-0.45
発覚=成立 (revelation = happened)	0.47	する=成立 (do = happened)	-0.45
回復=成立 (recovery = happened)	0.44	死亡=成立 (death = happened)	-0.42
ダウン=成立 (down = happened)	0.40	行っ=成立 (perform = happened)	-0.39
うつっ=成立 (give = happened)	0.39	注意=成立 (attention = happened)	-0.38
潜伏=成立 (incubation = happened)	0.37	感染=不成立 (infection = not happened)	-0.37

less important to predict the number of negative instances, although our system has high accuracy (about 91%). Therefore, we computed the precision, recall, and F1-score as the evaluation metrics and conducted five-fold cross-validation.

We used Classias (ver.1.1)⁵ with its default setting to train the model. We applied L2-regularized logistic regression as a training algorithm. We used MeCab (ver.0.996) with IPADic (ver.2.7.0) as a morphological analyzer.

4.2 Feature

The features used for the experiments are presented below. These features are not modality fea-

tures. We selected these features by performing preliminary experiments. Here, we omit the description related to modality features because the details are described in Section 3.

BoW: Bag of Words features of six morphemes around the “flu.”

N-gram (character N-gram): Feature of character N-gram around the word “flu.” The value of N is 1–4.

URL: Binary feature of the presence or absence of URL in messages.

Atmark: Binary feature of the presence or absence of reply in messages.

⁵Classias:<http://www.chokkan.org/software/classias/>

Table 7: Contribution and error analysis of deep modality features.

Example 4 (correct example)	10年ぶりにインフルエンザというものに かかり ました www English translation: It's been 10 years since I last had the flu, but now I have one (LOL).
Example 5 (false positive)	ASPARAGUS 渡邊忍がインフルエンザに 感染 してしまい、本日の 柏 DOME でのライブは中止します。 English translation: Watanabe of ASPARAGUS is infected with flu, and today's concert in Kashiwa Dome has been canceled.
Example 6 (false negative)	インフル とな...おだいじに \ (^o^) / English translation: So you have a flu... . Take care.:)

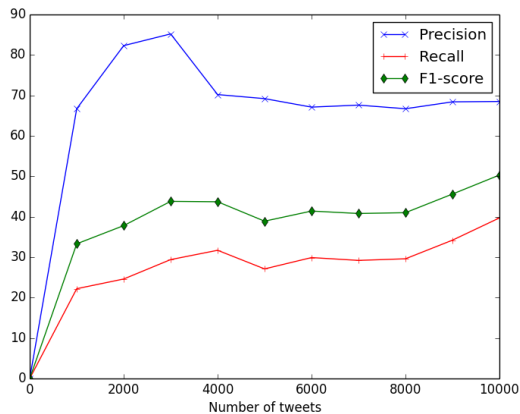


Figure 1: Learning curve for disease event detection.

Season: Binary feature of whether posting time is within December through February or not.

4.3 Baseline

For disease event detection, we follow previous studies Aramaki et al. (2011, 2012) to build the baseline classifier using a supervised approach. The baseline is constructed by combining all features except the modality features.

4.4 Experimental results

The result of disease event detection is shown in Table 4. Overall, they seem to have low recall and F1-Score. However, it turns out to be difficult to achieve high recall because the percentage of positive cases is extremely low (about 12.6%).

As shown, N-gram and Season features improve F1-score. Although the shallow modality feature boosts both precision and recall, the deep modality feature only improves recall in compensation with precision. The highest recall for the F1-score is achieved when using both shallow and deep modality features from Tsutsuji and Zunda (in the case of “All”). This result underscores the utility

of the modality features for classifying a post by its factuality.

In addition, to judge the performance with respect to the amount of data, we plot a learning curve in Figure 1. Although the decision changes only slightly, recall tends to improve by increasing the amount of data.

5 Discussion

As described in this paper, we demonstrate the contribution of modality analysis for disease event detection. In what follows, we conduct error analysis of our proposed method.

5.1 Contribution and error analysis for shallow modality

Table 5 shows the correct and incorrect examples for the shallow modality. Example 1 is a correct example. In this case, we convert “らしく” (“seem”) into sense ID; the classifier outputs an appropriate label. Example 2 is an example of false positive. Example 3 is an example of a false negative. Both examples are incorrect because they are assigned wrong sense IDs. That point illustrates the limitations of a simple string match, which does not take the context into account. It is necessary to perform word sense disambiguation for modality-related words.

5.2 Contribution and error analysis for deep modality

Next, we examine the deep modality features. Table 6 presents results of the deep modality features sorted by weight in descending order.

In many cases, the features can be understood intuitively compared to those of shallow modality features. Among the posts including the word “flu,” posts about disease warnings, posts about

vaccinations, and posts about epidemic news account for a large proportion. This tendency is exhibited clearly when one assigns negative weights. Positive weights include many event-nouns and verbs that are related directly to the disease.

Table 7 presents correct and incorrect examples for deep modality. Example 4 is a correct example. The deep modality feature “infection = happened” makes it possible to judge Example 4 correctly. Deep modality features appear to be critical in many cases, but in some cases they do not function as expected. Example 5 is an example of a false positive. Because of the “infection = happened” feature, the classifier judges it positive. However, not the writer, but a well-known figure (Watanabe of ASPARAGUS) has been infected with influenza. This is a common mistake that the classifier makes. This result indicates the importance of identifying the entity that is involved in a disease event. Furthermore, our classifier is not robust for non-event problems. Example 6 is an example of false positive. This example does not have the argument of an event. It is the characteristics of the colloquial sentence. Such examples can often be found in web documents.

6 Conclusion

This study examined a disease event detection method incorporating both shallow and deep modality features. Results show that the modality features improve the accuracy of the influenza detection. Although we have demonstrated that our method is useful for particular disease event detections, we must still ascertain whether it is applicable for other infectious diseases such as norovirus and dengue.

As future work, we would like to disambiguate functional expressions using sequence labeling techniques (Utsuro et al., 2007); we would also like to identify the predicate–argument structure of disease events (Kanouchi et al., 2015). Apart from that, an information extraction approach that looks for more specific patterns should be verified. Finally, we would like to adopt these findings to improve the prediction of epidemics.

Acknowledgments

We thank anonymous reviewers for their constructive comments, which have helped us to improve the manuscript.

References

- Harshavardhan Achrekar, Avinash Gandhe, Ross Lazarus, Ssu-Hsin Yu, and Benyuan Liu. 2012. Twitter improves seasonal influenza prediction. In *International Conference on Health Informatics*, pages 61–70.
- Eiji Aramaki, Sachiko Masukawa, and Mizuki Morita. 2011. Twitter catches the flu: detecting influenza epidemics using Twitter. In *Proceedings of the 16th Conference on Empirical Methods in Natural Language Processing*, pages 1568–1576.
- Eiji Aramaki, Sachiko Masukawa, and Mizuki Morita. 2012. Microblog-based infectious disease detection using document classification and infectious disease model. *Journal of Natural Language Processing*, 19(5):419–435.
- Danah Boyd, Scott Golder, and Gilad Lotan. 2010. Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. In *Proceedings of the 2010 43rd Hawaii International Conference on System Sciences*, pages 1–10.
- Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, and Larry Brilliant. 2009. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014.
- Chris Groendyke, David Welch, and David R Hunter. 2011. Bayesian inference for contact networks given epidemic data. *Scandinavian Journal of Statistics*, 38(3):600–616.
- Bernardo A Huberman, Daniel M Romero, and Fang Wu. 2009. Social networks that matter: Twitter under the microscope. *First Monday*, 14(1):1–9.
- Xiang Ji, Soon Ae Chun, and James Geller. 2012. Epidemic outbreak and spread detection system based on Twitter data. In *Health Information Science*, pages 152–163.
- Shin Kanouchi, Mamoru Komachi, Naoaki Okazaki, Eiji Aramaki, and Hiroshi Ishikawa. 2015. Who caught a cold? - identifying the subject of a symptom. In *Proceedings of 53rd Annual Meeting of the Association for Computational Linguistics*.
- Hubert Horace Lamb. 2013. *Climate: Present, Past and Future: Volume 1: Fundamentals and Climate Now*. Routledge.
- Jiwei Li, Alan Ritter, Claire Cardie, and Eduard Hovy. 2014. Major life event extraction from Twitter based on congratulations/condolences speech acts. In *Proceedings of the 19th Conferences on Empirical Methods in Natural Language Processing*, pages 1997–2007.
- Suguru Matsuyoshi, Satoshi Sato, and Takehito Utsuro. 2007. A Dictionary of Japanese Functional Expressions with Hierarchical Organization. *Journal of Natural Language Processing*, 14(5):123–146.

- Suguru Matsuyoshi, Megumi Eguchi, Chitose Sao, Koji Murakami, Kentaro Inui, and Yuji Matsumoto. 2010a. Annotating event mentions in text with modality, focus, and source information. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC)*, pages 1456–1463.
- Suguru Matsuyoshi, Megumi Eguchi, Chitose Sao, Koji Murakami, Kentaro Inui, and Yuji Matsumoto. 2010b. Factuality annotation for textual information analysis. *The IEICE Transactions on Information and Systems*, 93(6):705–713.
- Yamir Moreno, Romualdo Pastor-Satorras, and Alessandro Vespignani. 2002. Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 26(4):521–529.
- Andrew S Mugglin, Noel Cressie, and Islay Gemmell. 2002. Hierarchical statistical modelling of influenza epidemic dynamics in space and time. *Statistics in medicine*, 21(18):2703–2721.
- Philip M Polgreen, Yiling Chen, David M Pennock, Forrest D Nelson, and Robert A Weinstein. 2008. Using internet searches for influenza surveillance. *Clinical infectious diseases*, 47(11):1443–1448.
- Adam Sadilek, Henry Kautz, and Jeffrey P Bigham. 2012. Finding your friends and following them to where you are. In *Proceedings of the fifth ACM international conference on web search and data mining*, pages 723–732.
- Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, pages 851–860.
- Roser Saurí and James Pustejovsky. 2012. Are you sure that this happened? Assessing the factuality degree of events in text. *Computational Linguistics*, 38(2):261–299.
- Robert E Serfling. 1963. Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public health reports*, 78(6):494.
- Takehito Utsuro, Takao Shime, Masatoshi Tsuchiya, Suguru Matsuyoshi, and Satoshi Sato. 2007. Chunking and dependency analysis of japanese compound functional expressions by machine learning. In *Proceedings of 10th International Conference on Text, Speech and Dialogue (TSD)*.