Vol.2013-IFAT-110 No.4
Vol.2013-DD-89 No.4
2013/2/28

Feature and Sentiment Based Opinion Mining and Summarizing on Twitter

Wei Chen†1 Koshimura Miyuki†2
Fujita Hiroshi†3 Hasegawa Ryuzo†4

Twitter is one of the most popular social network websites. Through the twitter platform, users share either information or opinions about personalities, politicians, products, companies, events. For a sudden event, there are thousands of tweets related popping out every second. This makes it difficult for a user to read them to manage public opinions and make an informed decision. In this research, we aim to mine and to summarize all tweets of a topic. Our task is performed in three steps: (1) mine topic features that users expressed their opinions on. (2) identify the sentimental words (a specific words or emoticon) in each tweet and decide whether each tweet (which may contains more than one sentimental word) is positive or negative. (3) produce a summary using the discovered information. This paper proposes several techniques to perform these tasks. Our experimental results with three different sets of tweets demonstrate the effectiveness of the techniques.

1. Introduction

With the rapid expansion of social network service, like Twitter[14] and Facebook[15], millions of people post their thoughts and opinions on a great variety of topics. This makes it possible to analyze publicly available data to infer popular opinion in the same manner of an opinion poll. More importantly, comparing with a standard telephone poll, mining public opinions from freely available text content could be a timely and less expensive alternative. Considering the great amount of opinions posted from social media, extracting and summarizing the public opinion provides a challenging task to explore, motivating new research in computational linguistics.

Twitter is one of the most popular social network websites and has been growing at a very fast pace. The number of Twitter users reached an estimated 75 million by the end of 2009, up from approximately 5 million in the previous year[7]. Through the twitter platform, users share either information or opinions about personalities, politicians, products, companies, events etc.

In this paper, we are considering to use several techniques to summarize opinions about a topic in Twitter by performing the following 3 steps. In the first step, we employ both data mining (association mining) and natural language processing techniques (Mecab: Yet Another Part-of-Speech and Morphological Analyzer) to mine topic features. In Step 2, in order to decide the opinion orientation of each tweet, we make use of the combination of sentimental words and emoticons. Firstly identify sentimental words (adjectives, nouns, adverbs and verbs in Japanese) and emoticons in each tweet. Then for each sentimental word or emoticon, determine its semantic orientation (positive or negative) by comparing with a sentimental lexicon. Finally, determine the semantic orientation of every tweet. The last step summarizes the results of previous steps and presents them in a constructed form which is demonstrated in the following paragraph.

The following example illustrates a feature and sentiment based summary on a particular topic(query) , "ドコモ iPhone".

Topic(Query) ドコモ iphone:
  Overall:
    positive 300 <individual tweet>
    negative 190 <individual tweet>
    neutral 1000 <individual tweet>
  Feature 1 : 戦略下、iphone 導入
    positive 153 <individual tweet >
    negative 60 <individual tweet >
    neutral 200 <individual tweet>
  Feature 2: 一人負け、mnp
    positive 134 <individual tweet >
    negative 40 <individual tweet >
    neutral 200 <individual tweet>

Figure 1 An example summary.

In Figure 1, overall there are 300 tweets that express positive opinions about this topic, and 190 that is negative, 1000 that is neutral. The <individual tweet> links to the specific tweets that give personal comments. In all tweets, "戦略下、iphone 導入" and "一人負け、mnp " are the topic features. There are 153 positive tweets about the first topic feature , and 6 that comment negatively. The <individual tweet> links to the specific tweet that comments on the topic. With such a feature and sentiment based summary, anyone interested in this topic can easily see how the people feel about it, especially for customers or manufacturers when making important decisions. If he/she is very interested in a particular feature, he/she can drill down by following the <individual tweet> link to see why they like it and/or what they complain about.

As indicated above, our task is performed in three main steps:

- Mine topic features that have been commented on by users. We first make use of both data mining and natural language processing techniques. This part of the study has been reported in [4]. Then we compact them by adjusting some parameters and adding some restrictions.
- Identify sentimental words and emoticons in each tweet

†1Kyushu University Graduate School Information Science and Electrical Engineering Department of Informatics.
†2 Kyushu University Intelligence Science Department of Informatics Faculty of Information Science and Electrical Engineering
†3 Kyushu University Intelligence Science Department of Informatics Faculty of Information Science and Electrical Engineering
†4 Kyushu University Intelligence Science Department of Informatics Faculty of Information Science and Electrical Engineering

and deciding its sentimental orientation. Note that these sentimental tweets must contain one or more topic features identified above. To decide the orientation of each tweet (whether the orientation expressed in the tweet is positive or negative), we perform three subtasks. First, lexicons of sentimental words and emoticons are created separately. Second, we identify sentimental words through a natural language processing method and recognize emoticons in the tweet. Third, for each sentimental word and emoticon, we determine its semantic orientation by comparison with lexicons built in the previous subtask. Finally, we combine the orientation prediction results to decide the sentimental orientation of each tweet .

- Summarize all the outcomes. This step aggregates the results of previous steps and presents them in a format showed like Figure 1.

This paper is organized as follow. Next we survey recent related work. Then we introduce Twitter briefly and describe our experimental data. Section 4 presents the detailed techniques for performing these tasks. Our experimental results with 3 different sets of tweets show the effectiveness of our proposed methods, which will be presented in Section 5. We conclude in Section 6.

2. Related Work

The sentiment analysis for text mining has attracted an increasing attention [5], especially in the product reviews [6]. Many systems and approaches have been applied to automatically detect sentiment on texts (e.g., news articles, Web reviews and Web blogs) [17][18]. In recent years, as the Twitter becomes more and more important, sentiment detection over twitter data is one of the basic analysis utility functions needed by various applications. On this area, our work is related to Brendan O'Connor 's work in [8]. They connect measures of public opinion measured from polls with sentiment measured from text. They analyze several surveys on consumer confidence and political opinion over the 2008 to 2009 period, and find they correlate to sentiment word frequencies in contemporaneous Twitter messages. In [9], Authors propose an approach to automatically detect sentiments on Twitter messages (tweets) that explores some characteristics of how tweets are written and meta-information of the words that compose these messages. In this paper, they propose a 2-step sentiment analysis classification method for Twitter, which first classifies messages into subjective ones and objective ones, and further classifies the subjective tweets into positive and negative. In [10], They use the chatter from Twitter.com to forecast box-office revenues for movies. They show that a simple model built from rate at which tweets are created about particular topics can outperform market-based predictors. However, these authors do not perform any analysis on emoticons which are widely used in text-based online communication to convey user emotions. In [1], the authors present CAO, a system for affect analysis of emoticons in Japanese online communication. The system achieved nearly ideal scores, outperforming existing emoticon analysis systems.

On the other area our work related, the topic feature detection, some topic modeling methods such as LDA(Latent Dirichlet Allocation) and PLSA(Probabilistic Latent Semantic Analysis) modeled the documents generation and mined the implied targets. However, they did not perform well when applied to very short documents [3]. In [2], The authors propose a new method to extract opinion targets by developing a two-dimensional vector representation for words and a back propagation neural network for classification.

3. Twitter Corpus

In this section, we give some context about Twitter messages and the sources used for our data-driven approach

3.1 Twitter

Twitter is an online social networking service and micro-blogging service that enables its users to send and read text-based messages of up to 140 characters, known as "tweets". It was created in March 2006 by Jack Dorsey and launched that July. The service rapidly gained worldwide popularity, with over 500 million registered users as of 2012, generating over 340 million tweets daily and handling over 1.6 billion search queries per day. Since its launch, Twitter has become one of the ten most visited websites on the Internet, and has been described as "the SMS of the Internet." Unregistered users can read tweets, while registered users can post tweets through the website interface, SMS, or a range of apps for mobile devices.

There are some particular features that can be used to compose a tweet (Figure 2 illustrates some examples): “RT” is an acronym for retweet, which means the tweet was forwarded from a previous post; “@twUser” represents that this message is a reply to the user “twUser”; “#obama” is a tag provided by the user for this message, so-called hashtag, and “http://bit.ly/9K4n9p” is a link to some external source.

Tweet 1: @iskw226 iPhone 売り出しゃ犬は終わるのに RT ドコモ山田社長「渡辺謙さんも犬に勝ちたいと言っている」 http://t.co/RAmKbHRv #news
Tweet 2:【オバマ氏再選】自民・安倍総裁「同盟国として喜び」 - MSN 産経ニュース http://t.co/hXnWeIIm

Figure 2 Examples of tweets

3.2 Data Sources

Twitter is convenient for research because there are a very large number of messages, many of which are publicly available, and obtaining them is technically simple compared to scraping blogs from the web. We use 3 sets of Twitter messages posted in Japanese 2012, collected by querying the Twitter API[19] with 3 different queries as showed in Table 1.

Table 1 Details of the data.

Query ID	Query	Number of Tweets
1	ドコモ iPhone	1447
2	アップル iOS	1447
3	オバマ氏再選	1447

For evaluation, 3 annotators manually read all the tweets and tag them as positive, negative or neutral. Each tweet received 3 annotations, so final label is determined by majority vote. The result is presented in Table 2. If the 3 annotations are different on a tweet, we regard it as neutral. The distribution of the labeled data is showed in Table 2. The first column refers to three annotators have different judgments. The second column refers to tweets where the label is agreed on by two of the three annotators, while the last column requires agreement by all three.

Table 2 Distribution of labels in the annotated data.

Query ID	1/3	2/3	3/3
1	15	553	879
2	5	182	1260
3	0	153	1294

4. Proposed Techniques

In this section, we first talk the pre-process work including trimming and the Mecab(linguistic parser) which is fundamental for our task. Then techniques used in frequent feature identification are introduced in turn. At last sentimental orientation prediction is discussed.

4.1 Pre-Process

Trimming:

- Remove URL links (e.g. http://example.com), Twitter user names (e.g. @alex – with symbol @ indicating a user name), Twitter special words (such as “RT”) which are unlikely to be the topic or sentimental words.
- Remove stop words(some of the most common, short function words, such as "the" ,"a" in English and "これ", "私" in Japanese).
- Find the unique tweets which appear just once in the tweet set in order to process efficiently.

MeCab:

We use the MeCab to split tweet into sentences and to produce the part-of-speech tag for each word (whether the word is a noun, verb, adjective, etc). Considering that topic features are usually nouns or noun phrases, the part-of-speech(POS) tagging is crucial for mining topic features and opinion words. The following shows a sentence with POS tags.

The output format from the left are "表層形,品詞,品詞細分類 1,品詞細分類 2,品詞細分類 3,活用形,活用型,原形,読み,発音".

```
% mecab
すもももももももものうち
すもも	名詞,一般,*,*,*,*,すもも,スモモ,スモモ
も	助詞,係助詞,*,*,*,*,も,モ,モ
もも	名詞,一般,*,*,*,*,もも,モモ,モモ
も	助詞,係助詞,*,*,*,*,も,モ,モ
もも	名詞,一般,*,*,*,*,もも,モモ,モモ
の	助詞,連体化,*,*,*,*,の,ノ,ノ
うち	名詞,非自立,副詞可能,*,*,*,うち,ウチ,ウチ
EOS
```

Figure 3 An sentence with POS tags.

4.2 Topic Features Identification

This step identifies topic features on which many people have expressed their opinions. In this work, we focus on finding features that appear explicitly as nouns or noun phrases in the tweets. Here, we focus on finding frequent features, i.e. those features that are talked about by many users. For this purpose, we first use association mining to find all frequent item sets in our tweet database produced in the pre-process step, here an item set is simply a set of words or a phrase that occurs together in some tweets. Then we remove those unlikely features by using redundancy pruning. At last we add some restrictions to choose topic features from the remaining associated item sets. Figure 4 gives an overview of our frequent features identification process.

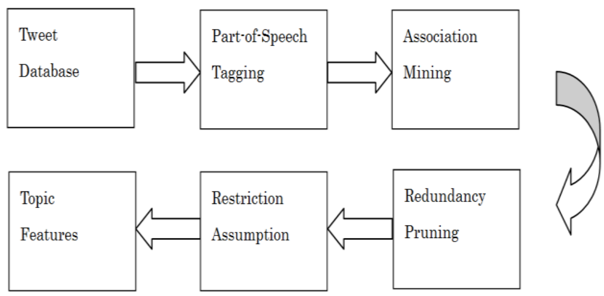

Figure 4 Overview of the topic feature identification.

(1) Association Mining

We use association mining CBA[11] ,which is based on the Apriori algorithm in [12], to find all frequent item sets. We run the association miner CBA, which is based on the Apriority algorithm on the tweet database. Each resulting frequent item set is a possible feature. In our work, we define an item set as frequent if it appears in more than the minimum support which will be set as one percent of the total number in one tweet set. The generated frequent item sets are also called candidate frequent features in this paper.

(2) Redundancy Pruning

In this part, we focus on removing redundant features. To describe the meaning of redundant features, we use the concept of p-support (see Definition 1). We use a minimum p-support value to prune those redundant features. If a candidate feature has a p-support lower than the minimum p-support and the feature is a subset of another feature phrase (which suggests that the feature alone may not be interesting), it is pruned. For instance, "life" by itself is not a useful feature while "battery life" is a meaningful feature phrase.

Definition 1: p-support (pure support)

P-support of feature ftr is the number of tweets that ftr appears in as a noun or noun phrase, and these tweets must contain no feature phrase that is a superset of ftr.

(3)Restriction Assumption

We propose three assumptions in order to get more appropriate topic features. We choose the biggest p-support at which all assumption are satisfied.

- The range of the t-support for should be (1/100,1/10) .
- The more phrase feature the better result. A phrase feature which contains more than one word is more specific than single word feature.
- Every topic feature should be unique(any two topic features should share no more than one word)

4.3 Sentiment Analysis of Opinion Tweets

Previous work on subjectivity has established a positive statistically significant correlation with the presence of adjective, nouns, adverbs and verbs in Japanese. Thus the presence of these words is useful for predicting whether a tweet is subjective. This paper uses these words as sentimental words. On the other hand, emoticons which are strings of symbols widely used in text-based online communication to convey user emotions, are considered as a new breach to analyze sentiment of text. We predict the sentimental orientation of a tweet through sentimental words and emoticons. Let us first define an opinion tweet.

Definition 2: opinion tweet

If a tweet contains one or more topic features and one or more sentimental words, then the tweet is called an opinion tweet.

(1) Sentimental Words Extraction

We now identify sentimental words. These are words that are primarily used to express subjective opinions. The algorithm we are considering to identify the orientation of sentimental words takes three inputs. The first input is a list of sentimental words. Sentimental words are defined by the subjectivity lexicon[13]. The lexicon classify Japanese sentimental words into ten different categories: 喜,安,昂,好,怒,哀,怖,恥,驚,厭. Following previous work, we sort these categories into positive or negative, as Table 3 shows. Table 4 shows some examples of the classified sentimental words.

Table 3 Classification of the ten categories of sentimental words.

Positive	喜,安,昂,好
Negative	怒,哀,怖,恥,驚,厭

The second input is a list of inversion words. There are words like "ない" that invert the sense of the opinion words. When these words occur in the left context of opinion words, they can invert the opinion sense. For example “not good” is a negative opinion. The third input to this algorithm is a list of potential sentimental words. This can be identified using algorithms we talked above. We process each sentence by sentence. For each sentence, we identify each feature and look at the sentimental words in the sentence.

Table 4 Examples from subjectivity lexicon.

Positive	笑う、嬉しい、歓声、ほっと、好き etc.
Negative	怒り、哀しい、恐れる、苦しい、驚く etc.

(2) Emoticon Extraction

An emoticon is a pictorial representation of a facial expression using punctuation marks, numbers and letters, usually written to express a person's feelings or mood. For example, a smiley face :-) usually shows a good mood. In twitter, users tend to express their opinions with some emoticons. Some examples can be seen in Figure 5. We believe that the sentimental orientation of tweets can be detected through emoticons. We query Twitter for two types of emoticons.

Table 5 Examples from emoticon lexicon.

Positive	“(^○^)”, “(＾▽＾)”, “(*'-'*)”, “(o^-^o)”
Negative	“(--)”, “(￣Θ￣;)”, “(￣o￣)”, “(゜○゜)”

Tweet1: "電子マネー対応じゃなきゃ意味がない（−＿−；）RT 現状のドコモの戦略下ではｉＰｈｏｎｅ導入は難しい＝社長"
Tweet2: "(ﾟ◇ﾟ)ｶﾞｰﾝ →現状のドコモの戦略下ではｉＰｈｏｎｅ導入は難しい＝社長|テクノロジーニュース|Reuters"

Figure 5 Two examples of emoticon tweets.

We first built an emoticon database by collecting emoticons from 3 online lexicon websites[16][20][21]. Then we collected eye marks from each emoticon to build an eye set manually, as we show below. Next, we detect emoticons by finding an element of the eye set between each pair of "(" and ")" in the tweet. Finally, compare the detected emoticon with the emoticon database for the orientation detection. In case of no matching for the detected emoticon, we predict its orientation through "eye". In our emoticon database, if more/less positive emoticons contain the "eye" than negative emoticons, the detected emoticon will be considered as positive/negative.

Eye set: "・","┬","゜","︵","^","￣","-","′","~","^","；","｀",">","<","･","︵" ,"T","¯","T",";","＾","°","ʊ","•̀"

(3) Orientation Prediction Method

In our proposition, positive word or emoticon is considered as +1, negative as -1, and neutral as 0. So the orientation of a tweet can be determined by the sum score of the sentimental words and emoticons. If the sum is bigger/smaller than 0, we see this tweet as positive/negative.

The orientation prediction method can be expressed as:

$$S(tweet) = Sum(words) + Sum(emoticons) \quad \textit{Equation 1}$$

5. Experiment

5.1 Summary Generation

The final feature and sentiment based summary can be generated as follow:

- For each discovered feature, related opinion tweets are put into 3 categories: positive, negative, neutral. For each category, a number is counted to show how many tweets are related.
- All features are ranked according to the frequency of their appearances. Feature phrases appear to be more interesting to users than single word feature. Other methods of ranking are also possible. For example, we can also rank features according the number of tweets that express positive and negative opinions.

The following figure shows an example summary for the feature "dvd ドライブ カーナビ、アップル ios6、抹殺".

Feature: dvd ドライブ カーナビ、アップル ios6、抹殺
neutral:18 <link>
1「Google マップ」「DVD ドライブ」「カーナビ」アップル iOS6 に抹殺された 10 の機能|IRORIO（イロリオ）-海外ニュース・国内ニュースで井戸端会議
2 6usersfrom はてブニュース "「Google マップ」「DVD ドライブ」「カーナビ」アップル iOS6 に抹殺された 10 の機能|IRORIO（イロリオ）-海外ニュース・国…"
3 だからどうした感があるのもいなめないけど、まぁいいんでないかな/"「Google マップ」「DVD ドライブ」「カーナビ」アップル iOS6 に抹殺された 10 の機能|IRORIO（イロリオ）-海外ニュース・国内ニュースで井戸端会議"
negative: 2 <link>
1 大丈夫？！チョット不安…/"「Google マップ」「DVD ドライブ」「カーナビ」アップル iOS6 に抹殺された 10 の機能|IRORIO（イロリオ）-海外ニュース・国内ニュースで井戸端会議"
2 出典があるなら正しく訳せ(￣Θ￣;)「Google マップ」「DVD ドライブ」「カーナビ」アップル iOS6 に抹殺された 10 の機能
positive: 1 <link>
1 楽しみすぎるよね（*^_^*）>「Google マップ」「DVD ドライブ」「カーナビ」アップル iOS6 に抹殺された 10 の機能

Figure 6 An example summary for an feature

5.2 Evaluation

We evaluate our proposed method from two perspectives:

- effectiveness of feature extraction
- accuracy of orientation prediction of opinion tweets

Considering that previous work[4] has showed the effectiveness of association mining and p-support pruning, we conduct our experiments mainly on restriction assumptions. One experimental result is showed in Table 6. As we can see, based on our restriction assumption, each generated phrase feature is different from others.

Table 6 Features generated at different frequency.

Random Frequency		Best Frequency	
Feature	P-Support	Feature	P-Support
機能	48	機能	24
対応	36	対応	18
進化	30	進化	15
楽しみ	26	楽しみ	13
発表、facebook 統合、今秋	20	発表、facebook 統合、今秋	20
秋頃	22		
3gs	20		
google マップ、dvd ドライブ、カーナビ、アップル ios6	20		
dvd ドライブ、カーナビ、アップル ios6、抹殺	26	dvd ドライブ、カーナビ、アップル ios6、抹殺	26
		ipad	13

At a random frequency, the last two features are considered to be similar as they share 3 words, " dvd ドライブ", "カーナビ", and "アップル ios6". While at the best frequency based on the assumption, we come out with a better feature phrase without similar features.

Our proposed techniques have a good accuracy in predicting tweets' orientation, which is defined as Emulation 2: the average accuracy for the 3 tweet sets is 80.67%. Table 7 shows that our method is effective.

$$Accuracy = Num(correct)/Num(total) \qquad \textit{Equation 2}$$

Table 7 Results of the orientation prediction.

Query ID	Correct Number	Error Number	Total Number	Accuracy
1	1184	263	1447	81.82%
2	1273	174	1447	87.98%
3	1045	402	1447	72.22%

6. Conclusion

In this article, we proposed a method to provide a feature and sentiment based opinion mining and summarizing of tweets related to a topic. Our experimental result showed the effectiveness of our techniques and indicated that this is a very promising way to solve this kind of task. We believe that as the social network service becomes more and more important, summarizing of opinions posted on social media, like Twitter, is not only necessary to ordinary people, but also crucial to enterprises.

In our opinion, the sentiment analysis could be substantially improved. Besides the need for a more well-suited lexicon of sentimental words and emoticons, the non-formal grammar of spoken languages and the special mode of communication in Twitter should be taken into account. In our future work, we plan to further improve and refine our sentiment analysis algorithm and to expand the lexicons of both words and emoticons.

Vol.2013-IFAT-110 No.4
Vol.2013-DD-89 No.4
2013/2/28

Reference

1) Ptaszynski, M., Maciejewski, J., Dubala, P., Rzepka, R. and Araki, K.: CAO: A Fully Automatic Emoticon Analysis System, in Proceedings of The Twenty-Fourth AAAI Conference on Artificial Intelligence(AAAI-10), pp.1026-1032, Atlanta, Georgia, USA(2010).
2) Lin, S., Haipeng, W., Xinu, D., Mengjie, Z.: Opinion Target Extraction for Short Comments, PRICAI 2012: Trends in Artificial Intelligence, Lecture Notes in Computer Science Volume 7458, 2012, pp 528-539.(2012).
3) Titov, I. and McDonald, R.T.: Modeling Online Reviews with Multi-grain Topic Models. In: WWW, pp. 111‑120 (2008).
4) Hu, M., and Liu, B.: Mining Opinion Features in Customer Reviews. To appear in *AAAI'04* (2004).
5) Morinaga, S., Yamanishi, K., Tateishi, K., Fukushima, T.: Mining Product Reputations on the Web. In: KDD, pp. 341–349 (2002).
6) Popescu, A.M., Etzioni, O.: Extracting Product Features and Opinions from Reviews. In: HLT/EMNLP (2005).
7) Wikipedia, http://en.wikipedia.org/wiki/Main_Page
8) Brendan, O., Ramnath, B., Bryan, R., Noah, A.: From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series, Association for the Advancement of Artificial Intelligence(2010).
9) Luciano, B. and Feng, J.: Robust Sentiment Detection on Twitter from Biased and Noisy Data, Coling 2010: Poster Volume, Pages 36-44 (2010).
10) Sitaram, A. and Bernardo, A.: Predicting the Future with Social Media, Computers and Society(2010).
11) Liu, B., Hsu, W., Ma, Y.: Integrating Classification and Association Rule Mining. KDD'98(1998).
12) Agrawal, R. and Srikant, R.: Fast Algorithm for Mining Association Rules. *VLDB'94*(1994).
13) 中村明: 感情表現辞典, 東京堂出版(1993).
14) Twitter, https://twitter.com/
15) Facebook, http://www.facebook.com/
16) FacemarkParty, http://www.facemark.jp/
17) Wiebe, J. and E. Riloff.: Creating Subjective and Objective Sentence Classifiers from Unannotated Texts. Computational Linguistics and Intelligent Text Processing, pages 486–497(2005).
18) Glance, N., M. Hurst, K. Nigam, M. Siegler, R. Stockton, and T. Tomokiyo. Deriving marketing intelligence from online discussion. In Proceedings of the eleventh ACM SIGKDD, pages 419–428. ACM(2005).
19) Twitter 4J, http://twitter4j.org/ja/index.html
20) 顔文字屋, http://kaomojiya.com/
21) 顔文字図書館, http://www.kaomoji.com/kao/text/index.html

Acknowledgments

We would like to thank all members in Hasegawa-Fujita Lab for their helpful suggestions and assistants, especially Okugawa Takumi, Akashi Yuuko and Kihara Hiroshi for their tremendous cooperation in our experiment. This work is partially supported by JSPS KAKENHI(22650028).