Automatic Classification of Complaint Reports about City Park

Yuta Sano [*] , Kohei Yamaguchi [†] , Tsunenori Mine [‡]

Abstract

Recently, it has become easy for the growing Goverment 2.0 movement to report complaints. On the other hand, there is a clearly identified and growing delay in responses from the government side due to an overload on government capacity to deal with the increasing number of complaint reports as the movement grows. In this paper, we propose a method of automatically categorizing complaint reports as a first step to reduce the pressure on the government side. We conducted experiments in categorizing the complaint reports. The experimental results showed the following findings: (1) Feature selection is key to improving the accuracy (F-score) of the categorization of complaint reports. The percentage of words that are strongly effective for categorization is about 3.9% of the total of distinct words. (2) Proposed Mutual-Information-based methods outperform a conventional Random-Forest-based method. (3) The city management section seems to classify complaint reports by focusing on demands expressed in the reports. (4) The categorization performance usually high if training data includes various types of categories of data.

Keywords: Categorization, Complaint Report, Government 2.0, Mutual Information, Random Forest

1   Introduction

Recently, useful platforms for reporting complaints have been introduced. FixMyStreet[1] is representative of such platforms. In Japan, some platforms such as the Chiba citizen coordination report (ChibaRepo for short)[2] have also been established. Using such platforms, registered users of the platforms can submit reports on issues involving their local area. The reports can include not only text-based messages, but also photographs and GPS information. Officials in a city management section can reply to the reports. The state of correspondence concerning reported situations can be roughly divided into three states: un-

---

[*]   Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan
[†]   Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan
[‡]   Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan
[1]FixMyStreet:https://www.fixmystreet.com/(2015/11/17 confirmed)
[2]Chiba citizen coordination report (ChibaRepo) (in Japanese):http://chibarepo.force.com/(2015/11/17 confirmed)

responsive, in process and completed. On the platform, everyone can see the local issue in detail, its frequency of occurrence and the state of the correspondence concerning the issue.

On the other hand, it has become recognized that there is frequently a delay before the government side takes action. The answers to a questionnaire concerning the demonstration experiments of ChibaRepo, which were conducted from July 2013 to December 2013, show that 44.4% of registered users felt disappointed with the response from the government side because of delays in taking action, lack of progress in the state of the correspondence, and even incorrect statements concerning progress on the issue, with unsolved issues sometimes claimed as completed [3].

Since these problems result from the overload on the government side, it is indispensable to reduce the burdens on the government side by introducing an automatic mechanism for complaint report processing supporting features such as automatic or semi-automatic judgment of the urgency of complaint reports, recording and classifying the complaint reports, detecting, estimating and responding to demands in the complaint reports, and commissioning professionals to deal with the demands.

Considering these issues, this paper tackles automatic categorization of complaint reports, which is a key issue, since not a small number of complaint reports are wrongly categorized.

The proposed methods begin with feature selection, which detects words strongly effective for categorization of complaint reports, and then categorizes the reports on that basis. Both feature selection and categorization are based on Mutual-Information(MI)-based term weighting methods. We conducted experiments to evaluate our proposed methods using real complaint reports. We compared the proposed methods with a conventional Random-Forest(RF)-based method with TF-IDF (term frequceny - inverse document frequency) term weighting as a base line. The experimental results illustrate the validity of our proposed methods.

In what follows, Section 2 discusses related work; Section 3 describes the data set we used in this research; Section 4 explains the proposed methods; Section 5 describes the evaluation measures; Section 6 discusses the experiments and the results, and we conclude in Section 7.

2 Related Work

Tominaga et al. [1] proposed a method of classifying SNS messages in the field of government. Because of the short length of SNS messages, especially those sent via Twitter, they believed that conventional methods can not provide sufficient precision. Therefore, they used a manifesto, which consists of public promises of political parties, to extract keywords as clues in classifying the SNS messages. They achieved a 7.2% improvement in accuracy compared to a conventional method. Table 1 shows the comparison between Twitter messages and our data (complaint reports). As can be seen from the table, one of the characteristics common to both is the short length of the messages. However there is a big difference between Twitter messages and complaint reports. Twitter messages include richer user information such as user name, the number of tweets, relationship with other users, and a user profile than do complaint reports, which only include the names of the

---

[3]Evaluation reports to Chiba citizen coordination report of demonstration experiment (ChibaRepo [trial]) (in Japanese): https://www.city.chiba.jp/shimin/shimin/kohokocho/documents/chibarepo-hyoukasho.pdf (2015/11/17 confirmed)

submitting users and the officials dealing with the reports. Instead, the complaint reports include such things as e.g. information about area parks and the location of a particular park. In addition, the categories of the reports vary from location to location, so we cannot handle all the data equally. Therefore, classification of complaint reports requires a new method for effectively using the limited information available in complaint reports, which necessarily differs from conventional methods for classifying Twitter messages; we believe that the way one deals with words is the key to classifying complaint reports. Moreover, we should consider the seasonality and locality of the complaint reports to enhance the classification accuracy.

Table 1: Comparison between Twitter Messages and Complaint Reports

Data	average # of characters	regional difference	information besides text data	view points when classifying
Twitter	63[2]	**not present**	**user information**, date, hashtag	**depends on the situation**
Complaint reports	35	**present**	**park information**, date, category	**seasonality, regionality**

Jin et al. [3] compared several classification methods such as random forest (RF) [4], k-nearest-neighbor [5], support vector machine [6], vector quantization method [7], bagging [8], boosting [9] to determine a useful method for detecting an original author of works about literature, essays, and diaries. Jin et al. conducted experiments and showed that an RF-based method had the highest accuracy of all the methods in detecting authors and was the least affected by decreased training data. This paper compares our proposed methods used in estimating categories of the complaint reports about city parks with a conventional RF-based method with a TF-IDF-based word vector.

3   Data Set

In this research, we use complaint reports, which are records of complaint calls from citizens about city parks in Kashiwa city. Complaint reports consists of 6 items: Current Situation (CurS), Demand (Dem), Correspondence Situation (CorS), Date (DT), and Category Number (CN) and Label (CL).

CurS describes the current issues involving a park, and Dem is a demand regarding CurS. Complaint reports often include both CurS and Dem, but sometimes only one. Complaint reports include various kinds of sentence elements: a simple or complex sentence, phrase(s), word(s), or multiple sentences, and identical meanings are often expressed inconsistently in different verbal forms, owing to the richness of written Japanese, with its two syllabaries and large set of logographic characters.

CorS is determined by officials in the city management section and records how they have dealt with the citizen complaint. DT is the date when the complaint was received by the management section. The date entry specifies year, month and day. CN is used to classify complaint reports. There are 14 categories, as shown in Table 2. Multiple CNs can be assigned to a complaint report and the way they are assigned may change from official to official in the city management section. CN_{14}, which means "others", appears when it is not easy to assign a CN to a complaint report. There are some complaint reports that have been wrongly assigned CNs, or not assigned any CN. The Category Label or CL expresses the meaning of a complaint report in a single word or a brief phrase . The correspondences between CNs and CLs are listed in Table 2.

Table 2: Category Number (CN) and Category Label (CL)

CN	CL	CN	CL	CN	CL
1	trees	6	water and sewerage	11	user manners
2	weeding	7	toilet	12	radiation
3	sweeping, fallen leaves	8	equipment	13	usage of park
4	play equipment	9	insect pests and destructive animals	14	others
5	illuminating light	10	abandoned bikes		

4 Proposed Methods

In this paper, we propose five methods. Four of them focus on relationship between CNs and keywords, and only use the value of mutual information (MI) in judging which CNs to assign to each complaint report. The other method combines MI and RF. As a baseline method we used a conventional RF-based method whose term weight is based on TF-IDF, not MI.

The following describes what values each method uses to determine the categories to be assigned to a complaint report. Table 3 shows whether each method uses MI and/or RF.

1. **Max**: This method uses the max value of mutual information.

2. **Sum**: This method uses the summation of mutual information.

3. **Num(Max)**: This method uses the number of keywords and the max value of mutual information.

4. **Num(Sum)**: This method uses the number of keywords and the summation of mutual information.

5. **RF(MI)**: This method uses RF with an MI-based word vector.

6. **RF(TF-IDF)**: This method uses RF with a TF-IDF-based word vector.

Table 3: Methods With (○) or Without (-) MI and RF

	Max, Sum, Num(Max), Num(Sum)	RF(MI)	RF(TF-IDF)
MI	○	○	-
RF	-	○	○

4.1 Preprocessing

4.1.1 Creating Word Vectors

We employed the Japanese parser named Cabocha [4] to analyze complaint reports and to detect words and to identify the part of speech of the words; we selected nouns, verbs, and adjectives, but not pronouns or dependent verbs and adjectives; we transformed the selected words into their base form to absorb the differences between conjugated forms. After that, we created word vectors consisting of pairs of a word and its frequency of occurrence in the report.

---

[4]Cabocha (in Japanese):`https://taku910.github.io/cabocha/` (2015/11/17 confirmed).

4.1.2   Calculation of MI-based Vectors

The five methods other than RF(TF-IDF) use MI for the classification. Therefore, we calculated MI $I(w,c)$ of a pair of word w and CN c by using the following four types of combinations [5]. $I(w,c)$ is calculated only from training data.

- N_{11}: the number of data[6] whose CN is c, and that include word w

- N_{10}: the number of data, whose CN is not c, and that include word w

- N_{01}: the number of data, whose CN is c, and that do not include word w

- N_{00}: the number of data, whose CN is not c, and that do not include word w

We define $I(w,c)$ as $I(w,c) = \sum_{i=0}^{1}\sum_{j=0}^{1} I_{ij}$ and then, define $I_{ij} = \frac{N_{ij}}{N} \cdot log_2 \frac{N \cdot N_{ij}}{(N_{\bar{i}j}+N_{ij})(N_{i\bar{j}}+N_{ij})}$, where $\bar{i}$ is the bit inversion of i. N is the total number of complaint reports in the training data, which is 1,073 in this paper.

For each CN c, words are arranged in descending order of their MI value $I(*,c)$, where $*$ represents any word. The top k words of $I(w,c)$ are defined as k keywords of CN c. k is varied by powers-of-two, such as 2, 4, 8, 16, to a number which is less than the number of distinct words in all the complaint reports.

For RF(MI), we created word vectors whose element weight is the maximum value of $I(w,c)$, e.g. [(tree, 0.089636), (wooly bear, 0.099794) , (pruning, 0.276903)] after extracting keywords.

4.1.3   Calculation of TF-IDF-based Vectors

RF(TF-IDF) uses TF-IDF values as a value of each element in a word vector. We define the TF-IDF weight of a word w in document d as follows: $tf\text{-}idf(w,d) = tf(w,d) \cdot idf(w)$ where $tf(w,d)$ is the term frequency of word w in document d and is defined by the following equation. "Document" here denotes a complaint report.

$tf(w,d) = \frac{n_{w,d}}{\sum_{s \in d} n_{s,d}}$

$n_{w,d}$ is the occurrence frequency of word w in document d and s is a word in document d.

$idf(w)$ is the inverse document frequency of word w and is defined by the following equation.

$idf(w) = \log_2 \frac{N}{df(w)} + 1$

N is the total number of documents, and $df(w)$ is the number of documents in which word w appears. Here N is 1,198, which comprises 1,073 training data and 125 test data in this research.

After calculating TF-IDF values, we applied LDA (Latent Dirichlet Allocation) [10] to TF-IDF-based word vectors and varied the number of dimensions of the vectors by powers-of-two from 2 to n, the highest power which is less than the original number of dimensions of the vectors.

---

[5]Feature Selection with Mutual Information (in Japanese):`http://aidiary.hatenablog.com/entry/20100619/1276950312`(2015/11/17 confirmed)

[6]Here and below "data" means the complaint reports

4.2　Methods of automated classification

4.2.1　MI-based Methods not including RF(MI)

After removing words other than keywords, we automatically classify complaint reports using keywords and their MI values from the following four viewpoints.

Max focuses on the maximum value of MI, and defines the weight of CN c as the maximum value of $I(w', c)$ of word w' in the word vector created from a complaint report, provided that $I(w', c) = 0$ if w' is not in the top k keywords of c. Max ranks CNs in descending order of their weights, selects the CN c with the maximum value of all the CNs and assigns the CN to the complaint report.

Sum focuses on the summation of MI, and defines the weight of CN c as the summation of all the MI $I(w', c)$ of word w' in the word vector created from a complaint report, provided that $I(w', c) = 0$ if w' is not in the top k keywords of c. Sum ranks CNs in descending order of their weights, selects the CN c with the maximum value of all the CNs and assigns the CN to the complaint report.

Num(Max) focuses on the number of keywords and the maximum value of MI. Num(Max) calculates the number of keywords of each CN c in a word vector of a complaint report, defines the number of keywords as the weight of the CN c, and assign CNs to the complaint report in descending order of the value of the CNs. If there are two or more CNs that have the same maximum value, Num(Max) follows Max for the CNs.

Num(Sum) focuses on the number of keywords and the summation value of MI. Num(Sum) calculates the number of keywords of each CN c in a word vector of a complaint report, defines the number of keywords as the weight of the CN c, and assign CNs in descending order of the value of the CNs to the complaint report. If there are two or more CNs that have the same maximum value, Num(Sum) follows Sum for the CNs.

4.2.2　RF-based Methods

The number of items of training data differs according to the CN assigned to the data. If there are big differences in these numbers, the classification of the test data will be biased by the CN with the greatest number of training data [7]. To solve this problem, we equilibrated the size of training data for each CN by selecting the same number of training data for the CN after making two vectors: an MI-based and a TF-IDF-based vector. We used the number of training data for CN_1, 396, in this research because it was the greatest. We then increased the number of training data for the other CNs to 396 by randomly selecting complaint reports from the training data of the CNs regardless of duplication.

Then, we created bootstrap samples by randomly selecting two-thirds of the whole training data set regardless of duplication. We repeated this procedure 1,000 times and created 1,000 sets of bootstrap samples.

After creating bootstrap samples, we built 1,000 decision trees, each of which was built from a bootstrap sample. We randomly chose d' words as explanatory variables so that the correlation between the decision trees became lower [8], where d' is the square root of d, the number of distinct words of relevant parts of speech in the training data. We constructed each decision tree according to [11], as follows:

---

[7]Class separation of imbalanced data (in Japanese):http://www.slideshare.net/sfchaos/ss-11307051 (2015/11/17 confirmed)

[8]Machine Learning for package user(5): Random Forest (in Japanese):`http://tjo.hatenablog.com/entry/2013/12/24/190000` (2015/11/17 confirmed)

We used bootstrap samples as training data, decomposed each node t, and built its decision tree T_m, where the lower limit of the number of a terminal node was 1.

1. Select $d'(=\sqrt{d})$ attributes randomly from d attributes.

2. For d' attributes, determine the attribute that gives the optimum division and its cutoff point.

3. divide node t into two nodes: $left(t)$ and $right(t)$ at the cutoff point.

After constructing an RF model, we applied the model, which consisted of 1000 decision trees, to test data. The RF model ranks CNs in descending order of the number of trees that polled their vote to the CNs and assigns them to complaint reports.

5 Evaluation Measures

5.1 Precision, Recall, F-score

We use Precision, Recall, and F-score as measures to evaluate our methods. The three measures are defined as follows:

$$Precision = \frac{TP}{TP+FP} \qquad Recall = \frac{TP}{TP+TN} \qquad F\text{-}score = \frac{2Recall \cdot Precision}{Recall + Precision}$$

where TP, FP, TN are shown in Table 4.

Since some complaint reports in our data have multiple CNs, we count four values: TP, FP, TN, FN for each CN, not for each complaint report.

TP_c, FP_c, and TN_c are TP, FP, and TN for CN c, respectively.

TP: foreach c $\in S$ if $c \in S \cap S'$, add 1 to TP_c

FP: foreach c $\in S$ if $c \notin S$ and $c \in S'$, add 1 to FP_c

TN: foreach c $\in S$ if $c \in S$ and $c \notin S'$, add 1 to TN_c

where S and S' are a set of actual CNs $\{c_1, \ldots, c_n\}$, and a set of predicted CNs $\{c'_1, \ldots, c'_n\}$, respectively.

Methods for calculating the average include two kinds of average, micro and macro. In this research, the number of data belonging to each CN includes deflection, and we think the evaluation by micro average is not appropriate. Therefore, we take the macro averages Precision P_{macro}, Recall R_{macro}, and F-score F_{macro} as evaluation metrics.

$$P_{macro} = \frac{1}{Q}\sum_{i=1}^{Q} P_i \qquad R_{macro} = \frac{1}{Q}\sum_{i=1}^{Q} R_i \qquad F_{macro} = \frac{1}{Q}\sum_{i=1}^{Q} F_i$$

where Q is the number of CNs, and P_i, R_i, and F_i are Precision, Recall, and F-score of CN i, respectively.

Table 4: TP, FP, TN, and FN

Facts / Prediction	True	False
Predicted	TP	FP
Not Predicted	TN	FN

6   Experiments and Results

6.1   Experiments

We conducted two experiments to answer four research questions (RQs). RQ1, RQ2, and RQ3 are set to investigate the classification performance of the proposed methods. RQ4 are set to consider the influence of seasonality in classifying complaint reports.

6.1.1   Experiment 1

Experiment 1 tackles the following three questions.

- RQ1: Does the number of keywords influence classification performance ?

  To classify complaint reports, we select only useful words as a basis for the classification. We investigate the influence of varying the number of keywords selected on classification performance.

- RQ2: What method achieves the best classification performance ?

  We propose methods using the mutual information between keywords and CNs, and compare the results of the proposed methods with those of a conventional RF-based method.

- RQ3: Which viewpoint achieves the best classification performance ?

  We would like to know which viewpoint enhances the category assignment of a complaint report. To this end, after reassigning CNs to complaint reports from two viewpoints: on the basis of CurS and Dem, we investigate which viewpoint yields better classification performance.

6.1.2   Experiment 2

In Experiment 2, we resolve the following question.

- RQ4: Does the season of the complaint influence classification performance ?

  Analysis of complaint reports shows that the amount of data with some CNs varies from month to month. So, we investigate how the difference of seasons influences classification performance.

The data we used in Experiment 1 consist of one year of data from April 2012 to March 2013 and one month of data from June 2013. The former includes 1,073 complaint reports and was used as training data. The latter consists of 125 complaint reports and was used as test data.

Because complaint reports sometimes include wrongly assigned CNs, we checked the CNs of complaint reports in the test data and, if necessary, reassigned CNs to the complaint reports from the two viewpoints, CurS and Dem. Table 5 shows a comparison of the test data and the original data it corresponds to.

Table 6 shows training data and test data we used in Experiment 2. The training data was collected from April 2011 to March 2013, and the test data is from April 2013 to March 2014. "Months" in the table denotes the month in which the data we used was reported.

Table 5: Comparison of Test Data

data	viewpoint	# of CNs	# of data	average # of CNs
A0	(original data)	133	125	1.06
A1	CurS	240	125	1.92
A2	Dem	213	125	1.70
A3	CurS and Dem	245	125	1.96

Table 6: Training Data in Experiment 2

data	months	# of training data	# of test data
spring	March, April, May	527	288
summer	June, July, August	761	352
autumn	September, October, November	747	407
winter	December, January, February	266	177
all	all months	2,301	1,224

In all experiments, we only used complaint reports assigned a single Category Number other than 14 ("others") as training data, in order to build prediction models effectively. Moreover, we also used complaint reports assigned a CN other than 14 as test data.

6.2   Experimental Results

6.2.1   RQ1: Does the number of keywords influence classification performance ?

Figure 1 shows how the F-score varies according to the value of k, the number of top k keywords which are strongly related to each CN.
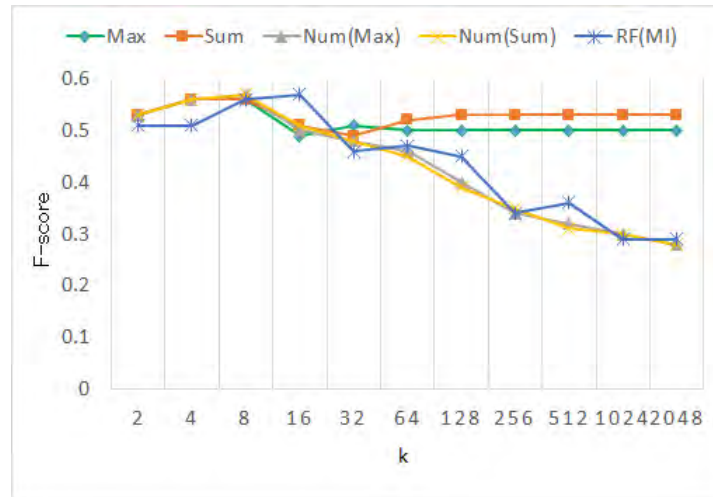

Figure 1: F-score with the top k keywords for varying values of k

Figure 1 shows that all methods had the best results at around $k = 8$. When $k = 8$, the number of distinct keywords is 93, which is 3.9% of the total of 2,398 distinct keywords in the training data.

6.2.2 RQ2: What method achieves the best classification performance ?

Table 7 shows the maximum value of results obtained by each method.

(a) MI-based Methods: Max, Sum, Num(Max), Num(Sum), RF(MI)

All methods other than RF(TF-IDF) are MI-based methods. The number in () brackets is the value of k that yields the maximum value of the results shown, with the smallest value of k selected if there were multiple k values yielding the same maximum value.

(b)TF-IDF-based Method: RF(TF-IDF)

RF(TF-IDF) uses a TF-IDF value of a word in a vector. The number in () brackets is the value of n, which is the number of LDA topics that yielded the maximum value of the results. The smallest value of n was selected if there were multiple n values yielding the same maximum value.

Table 7: Comparison of all methods

	Max	Sum	Num(Max)	Num(Sum)	RF(TF-IDF)	RF(MI)
Precision	**0.71(2)**	**0.71(2)**	**0.71(2)**	**0.71(2)**	0.40(2048)	0.68(16)
Recall	0.67(8)	0.67(8)	0.67(8)	0.67(8)	0.33(1024)	**0.68(16)**
F-score	0.56(4)	0.56(4)	**0.57(8)**	**0.57(8)**	0.31(1024)	**0.57(16)**

The results of all the methods other than RF(TF-IDF) were similar, and only showed slight differences in comparison metrics. From Figure 1 it can be seen that Max and Sum were almost unaffected by the variation in k values and achieved stable results, unlike the other methods. MI-based methods using selected keywords are effective for classification of complaint reports. In other words, each CN strongly depended on keywords.

6.2.3 RQ3: Which viewpoint achieves the best classification performance ?

Figure 2 shows the results when using data set A1 and A2 in Table 5 as test data. Here, RF(TF-IDF) is excluded from the methods because it got the worst results shown in Table 7.
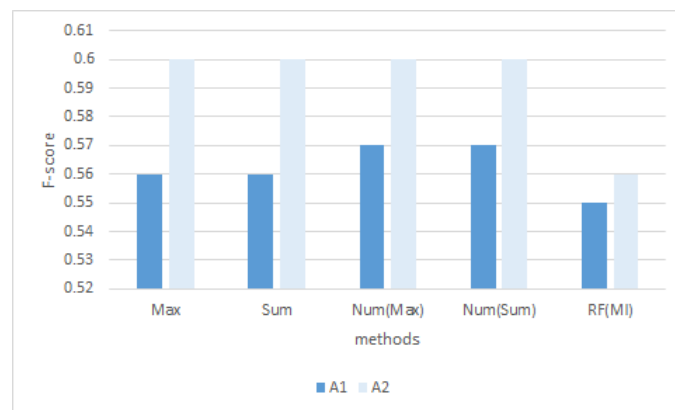

Figure 2: Comparison of F-score for test data set A1 and A2

Figure 2 shows that the results for A2 are higher than those for A1. So, we think that the management section officials classified complaint reports by focusing on the expression of demands because of the importance of these expressions in resolving complaints.

6.2.4 RQ4: Does the season of the complaint influence classification performance ?

Table 8 shows the results for each season of data obtained using Max.

Table 8: Comparison of F-score by season

test \ training	spring	summer	autumn	winter	all
spring	**0.59(64)**	0.51(4)	0.53(4)	0.44(4)	0.56(64)
summer	0.64(32)	0.63(4)	**0.65(16)**	0.52(4)	**0.65(256)**
autumn	0.70(32)	0.69(4)	0.63(16)	0.57(16)	**0.72(8)**
winter	0.67(8)	0.64(4)	0.68(16)	0.67(16)	**0.69(512)**
all	**0.64(64)**	0.57(4)	0.58(8)	0.51(4)	0.63(128)

Table 8 shows that the F-scores are similar in being highest no matter what test data is used, if either spring data or all data is used as training data. From these results, we conclude that the spring data includes various types of categories of data like all data, and that this is why the classification performance is better for spring data than for all other data sets except for all data. In addition, the F-scores are the worst if spring data is used as test data. It also seems that the spring data not only includes various types of categories of data, but also seems to be different from year to year. As mentioned in section 3.2.2 there seems to be correlation between months and CNs. However, carefully selecting training data did not increase the classification performance. We think that the difference in the number of data for each CN may affect the results. Therefore, we should consider the impact of the number of data for each CN as well as the relations between the months and CNs in our future work.

7 Conclusion

This paper discussed classification methods for complaint reports. The results showed us the following three findings. (1) Feature selection is key to improving the F-score in estimating the categories of a complaint report. The percentage of words strongly effective for category estimation is about 3.9% of the total of distinct words. (2) A word-category Mutual-Information-based method outperforms the F-score of the Random-Forest-based methods. (3) City management section officials seem to classify complaint reports by focusing on the expressions of demands in the reports. (4) The classification performance is higher if training data includes various types of categories of data.

We will report methods that estimate and extract sentences related to CurS or Dem from complaint reports in the near future.

Acknowledgments

References

[1] Yuta Tominaga, Hidetsugu Namba, Toshiyuki Takezawa, Automatic Classification of Comments on Government SNS, the 18th ANLP annual symposium, pp.555–558, (2012) (in Japanese)

[2] Papacharissi, Zizi. "Without you, I'm nothing: Performances of the self on Twitter." International Journal of Communication 6 (2012): 18.

[3] Mingzhe Jin and Masakatsu Murakami, Authorship Identification Using Random Forests, Proc. of the Institute of Statistical Mathematics, Vol. 55, No. 2, pp.255–268 (2007) (in Japanese)

[4] Leo Breiman, Random forests, Machine learning 45.1, pp. 5–32 (2001)

[5] T. M. Cover, and P. E. Hart, Nearest Neighbor Pattern Classification, IEEE Transaction on Information Theory, IT-B(1). pp. 21–27 (1967)

[6] V. Vapnik, The Nature of Statistical Learning Theory, Springer, New York.(1995)

[7] Matthews, R. A. J. and Merriam, T. V. N..Neural computation in stylometry I: An application to the works of Shakespeare and Fletcher. Literary and Linguistic Computing, 8(4) pp. 203–210.(1993)

[8] L. Breiman, Bagging predictors, Machine Learning, 24, pp.123–140 (1996)

[9] Yoav Freund, and Robert E. Schapire. Experiments with a new boosting algorithm, ICML, Vol. 96, (1996)

[10] Blei, David M., Andrew Y. Ng, and Michael I. Jordan, Latent Dirichlet Allocation, the Journal of Machine Learning Research 3, pp.993–1022 (2003)

[11] Yuzo Hirai, Introduction of Pattern Recognition, Morikita publisher, p.194 (2012) (in Japanese)