FU-01 Team's Classification of Fact-checkable Opinions in NTCIR-14 QA Lab-PoliInfo Task

Souichi Furukawa[1], Yuto Naritomi[1], Hokuto Ototake[1], Toshifumi Tanabe[1], and Kenji Yoshimura[1]

Fukuoka University, Fukuoka, Japan
{tl151293,tl151294}@cis.fukuoka-u.ac.jp,
{ototake,tanabe,yosimura}@fukuoka-u.ac.jp

Abstract. This paper reports on the achievements of Classification subtask of the NTCIR-14 QA Lab-PoliInfo task of FU-01 team. We proposed two different methods, rule-based and MaxEnt based methods, for classifying pros and cons of a political topic and whether an utterance sentence includes fact-checkable reasons or not. The results of formal run of the subtask shows our MaxEnt based method achieved higher accuracy than the rule-based method.

Team Name. FU-01

Subtasks. Classification Task (Japanese)

Keywords: Classification · Rule-based method · Maximum entropy classifier

1 Introduction

We, FU-01 team, participated in Classification subtask of the NTCIR-14 QA Lab-PoliInfo task [1]. Classiffication task aims at finding pros and cons of a political topic and presenting their fact-checkable reasons. In the subtask, a political topic such as "The Tsukiji Market should move to Toyosu" and an utterance sentence in assembly minutes are given. Participants including us classify four kinds of labels on given sentences, *Fact-checkability*, *Relevance*, *Stance* and *Class*. *Fact-checkability* means whether or not a sentence contains fact-checkable reasons. *Relevance* means whether or not a given sentence refer to a given topic. *Stance* means whether or not a speaker of the sentence agrees on the topic. A label of *Class* depends on labels of the other three class, as shown in Table 1. *Fact-checkability* and *Relevance* labels are regarded as 2-class classification, which their values can be 0 (absence) or 1 (existence). *Stance* labels are regarded as 3-class classification, which their values can be 0 (other), 1 (agree) or 2 (disagree).

For classifying these labels, we proposed two different methods and submitted their respective results. One of the proposed methods is based on handmade rules, which is described in Section 2. Another uses a maximum entropy classifier described in Section 3. In Section 4, we describe results of the two methods. Finally, Section 5 is the conclusion.

Table 1. Relationship between *Class* and the other three labels

Class	*Fact-checkability*	*Relevance*	*Stance*
1	1	1	1
2	1	1	2
0	All other combinations		

2 Rule-based Classification Methods

In this task, labeled sentences of the same topic as formal run test data are distributed to us in advance (hereinafter, this is called training data). In this section, we describe our rule-based methods for estimating three kinds of labels, *Fact-checkability*, *Relevance* and *Stance* respectively. For extracting words and recognizing their part-of-speech, we use MeCab morphological analyzer [2] with mecab-ipadic-NEologd dictionary [3] in the methods.

2.1 Fact-checkability

We consider an utterance sentence as fact-checkable one if the utterance sentence includes at least one of the fact-checkable keywords as described below. We looked for characteristic words from fact-checkable sentences in training data. As a result, we selected the following five words as a characteristic word of a fact-checkable sentence, "*ori*(おり: and)", "*rei*(例: example)", "*kara*(から: because)", "*de-ari*(であり: and)" and "*riyu*(理由: reason)".

2.2 Relevance

We consider an utterance sentence as being relevant with the topic of the utterance if the sentence includes at least one of the nouns in the topic.

2.3 Stance

We implement a dictionary-based method for judging whether a utterance sentence means approval for or opposite to its topic. In this method, we use Japanese Sentiment Dictionary (Volume of Verbs and Adjectives [4] and Volume of Nouns [5]). This dictionary has pairs of a word and its sentiment polarity (positive or negative). The number of records of the dictionary is about $13,500$ ($5,000$ verbs and adjectives, and $8,500$ nouns).

We assume a sentiment polarity score that is $+1$ or -1 if a recorded sentiment polarity in the dictionary is positive or negative respectively. When the method get an utterance sentence, it calculates a sum of sentiment polarity scores from words of the sentence included in the dictionary. The method considers the sentence as approving stance if the sum is greater than or equal to $+1$. It considers the sentence as opposite stance if the sum is less than or equal to -1.7. Otherwise it outputs *other* stance.

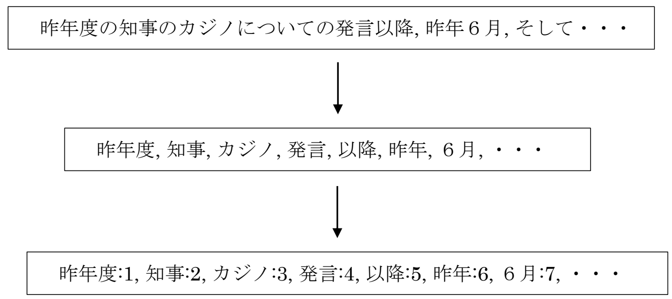

Fig. 1. An example of noun numbering.

3 MaxEnt Classification Methods

In addition to our rule-based methods described in section 2, we also propose a method based on a maximum entropy classifier for estimating the three kinds of labels. We created each model of *Fact-checkability*, *Relevance* and *Stance* for each topic.

Prior to models creation, we numbered nouns as shown in Fig. 1. The sentence at the top of Fig. 1 represents an utterance sentence to be input. We extract nouns from the result of morphological analysis of the sentence as shown at the middle of Fig. 1. As in the rule-based methods, we use MeCab morphological analyzer with mecab-ipadic-NEologd dictionary in this preprocessing. Finally, we number different collected nouns as shown at the bottom of Fig. 1. We constructed a noun number dictionary from the training data.

Using the noun number dictionary, we obtain sentence vectors from input utterance sentences. The n-th dimension of the vector represents the n-th noun in the sentence, and the value of each dimension represents the noun number of the dictionary. Please note that this vectorization method is different from bag-of-words in which the n-th dimension of the vector represents the noun number of the dictionary. We selected this vectorization method because of consideration of word order of sentences. We construct models from pairs of such a sentence vector and the label using "scikit-learn" machine learning library[1].

4 Results of Formal Run

In this section, we describe the results of Formal run. In Formal run, the number of utterance sentences in the training data is $10,291$. The test data's one is $3,412$. The number of topics is 14 as shown in Table 2.

[1] https://scikit-learn.org/stable/

Table 2. Topics

#	abbreviated name	topic sentence
1	カジノ (Casino)	カジノを含む統合型リゾートを推進するべきである
2	集団的自衛 (Self defense)	集団的自衛を認めるべきである
3	八ッ場ダム (Yanba)	八ッ場ダムの建設を進めるべきである
4	高齢者 (Elderly people)	高齢者への医療助成を増やすべきである
5	私学助成 (Private school grants)	私学助成を推進するべきである
6	中京都構想 (Medium Kyoto)	中京都構想を推進するべきである
7	オスプレイ (Osprey)	オスプレイを配備する
8	特定秘密保護法 (Secret protection)	特定秘密保護法案を進めるべきである
9	道州制 (Do-Shu-system)	道州制を導入するべきである
10	子ども医療 (Children medical expenses)	子ども医療費を無料化にするべきである
11	教員増加 (Regular faculty members)	正規の教員を増やすべきである
12	生活保護 (Welfare)	生活保護の基準額を引き下げるべきである
13	東京オリンピック (Tokyo Olympics)	東京にオリンピックを招致するべきである
14	空き家 (Vacant houses)	行政の判断で空き家を処理できるようにするべきである

4.1 Results of Rule-based Method

Fig. 2 shows the results of accuracy of rule-based *Fact-checkability* classification for each topic. The best result achieves only 57.59% accuracy in "Casino" topic. From these results, it can be seen that the fact-checkable keywords described in **2.1** are not much included in the test data.

Fig. 3 shows the results of accuracy of rule-based *Relevance* classification for each topic. In Fig. 3, the accuracy of "Self defense" topic is very low compared to others. The reason of this is the difference between the results of the two morphological analysis, "集団的自衛" and "集団的自衛権". The word "集団的自衛" is included in the topic sentence and is divided into "集団的" and "自衛" as a result of the analysis. On the other hand, the word "集団的自衛権" is included in utterance sentences a lot, which is judged as proper noun and not divided as a result of the analysis. As a result, the noun matching rate of topic sentence and utterance sentence is lowered, and the accuracy has decreased.

Fig. 4 shows the results of accuracy of rule-based *Stance* classification for each topic. The best result achieves only 47.50% accuracy in "Medium Kyoto" topic. We examined the recall rate, it was found that the recall rate of only approving stance was high: our rule-based method tends to output approving stance rather than opposite and other stances.

Fig. 5 shows the results of accuracy of rule-based *Class* classification for each topic.

4.2 Results of MaxEnt Classification Method

Fig. 6–9 show the results of accuracy of maximum entropy based *Fact-checkability*, *Relevance*, *Stance* and *Class* classification respectively for each topic. Maximum entropy classification is strongly affected by labels distribution of the training

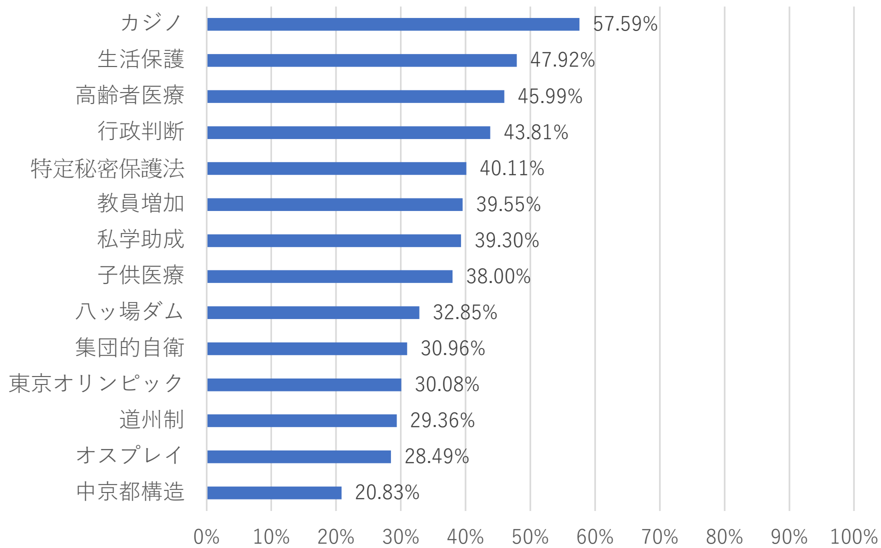

Fig. 2. Accuracy of rule-based *Fact-checkability* classification

data. Table 3 shows labels distribution of the training data. At the column headers of Table 3, *Rl*, *Fc* and *St* represent *Relevance*, *Fact-checkability* and *Stance* labels respectively. *Rl* 0 or 1 means that a relevance is absence or existence respectively. *Fc* 0 or 1 means that a fact-checkability is absence or existence respectively. *St* 0, 1 or 2 means that a stance is other, agree or disagree respectively. In Table 3, every topic has bias of data size of labels. As one of the reasons for lowering accuracy, it is conceivable that the bias of data size between training data and test data is different.

5 Conclusions

We proposed two different methods, rule-based and MaxEnt based methods, for classifying pros and cons of a political topic and whether an utterance sentence includes fact-checkable reasons or not. In our rule-based method, *Fact-checkability* are classified based on the keywords. *Relevance* existence is judged by nouns in a topic sentence. *Stance* are classified by the sentiment polarity score of an utterance sentence. However, it was doubtful whether the sentiment polarity of words could be used for *Stance* classification or not. Our MaxEnt based method achieved higher accuracy than the rule-based method. This MaxEnt based method consists of classifiers of the three labels for each topic and uses the same features for all the classifiers. There is a possibility of improvement in accuracy by using more suitable features for each label.

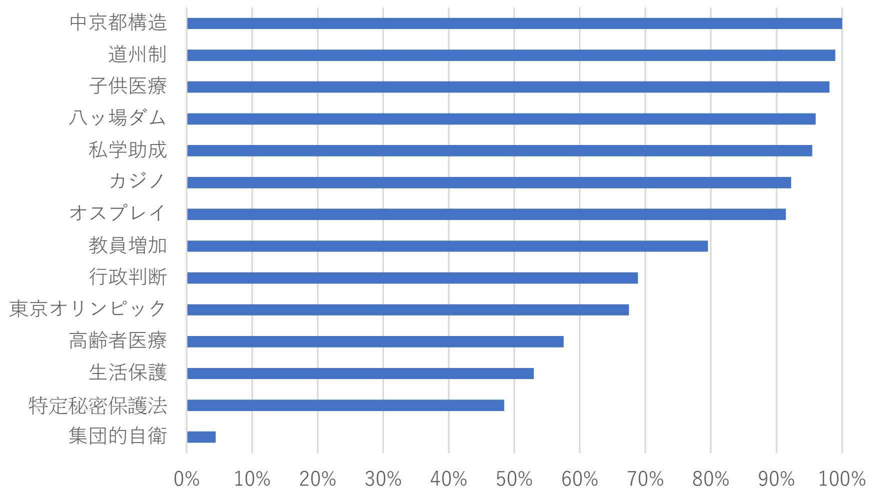

Fig. 3. Accuracy of rule-based *Relevance* classification

References

1. Kimura, Y., Shibuki, H., Ototake, H., Uchida, Y., Takamaru, K., Sakamoto, K., Ishioroshi, M., Mitamura, T., Kando, N., Mori, T., Yuasa, H., Sekine, S., Inui, K.: Overview of the NTCIR-14 QA Lab-PoliInfo Task. In: Proceedings of the 14th NTCIR Conference, 2019.
2. Kudo, T., Yamamoto, K. and Matsumoto, Y.: Applying Conditional Random Fields to Japanese Morphological Analysis. In: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, Vol. 4, pp.230–237, 2004.
3. Sato, T., Hashimoto, T. and Okumura, M.: Implementation of a word segmentation dictionary called mecab-ipadic-NEologd and study on how to use it effectively for information retrieval (in Japanese). In: Proceedings of the Twenty-three Annual Meeting of the Association for Natural Language Processing, NLP2017-B6-1, 2017.
4. Kobayashi, N., Inui, K., Matsumoto, Y. and Tateishi, K.: Collecting Evaluative Expressions for Opinion Extraction. In: Journal of Natural Language Processing 12(3), pp.203-222, 2005.
5. Higashiyama, M., Inui, K. and Matsumoto, Y.: Learning Sentiment of Nouns from Selectional Preferences of Verbs and Adjectives. In: Proceedings of the 14th Annual Meeting of the Association for Natural Language Processing, pp.584-587, 2008.

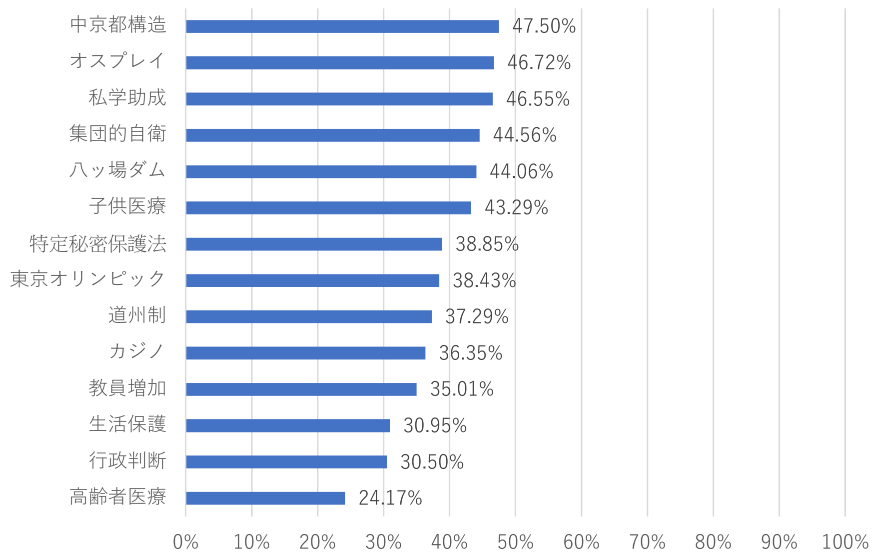

Fig. 4. Accuracy of rule-based *Stance* classification

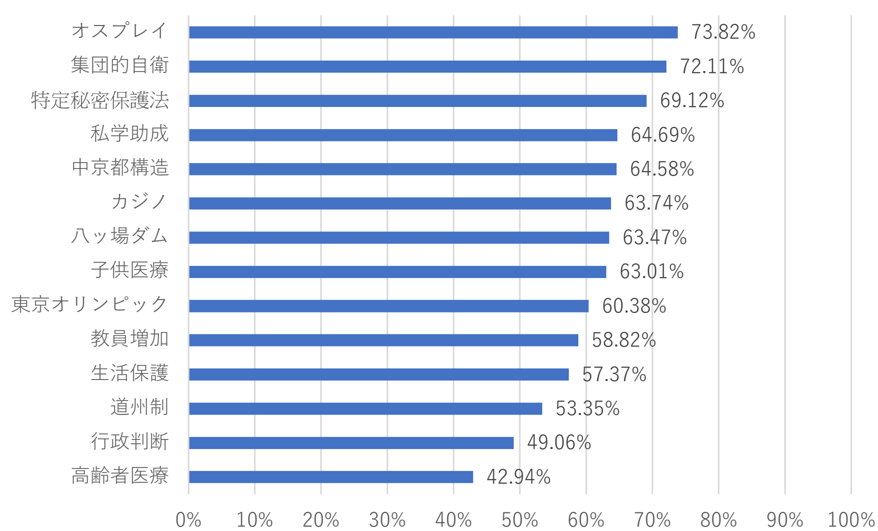

Fig. 5. Accuracy of rule-based *Class* classification

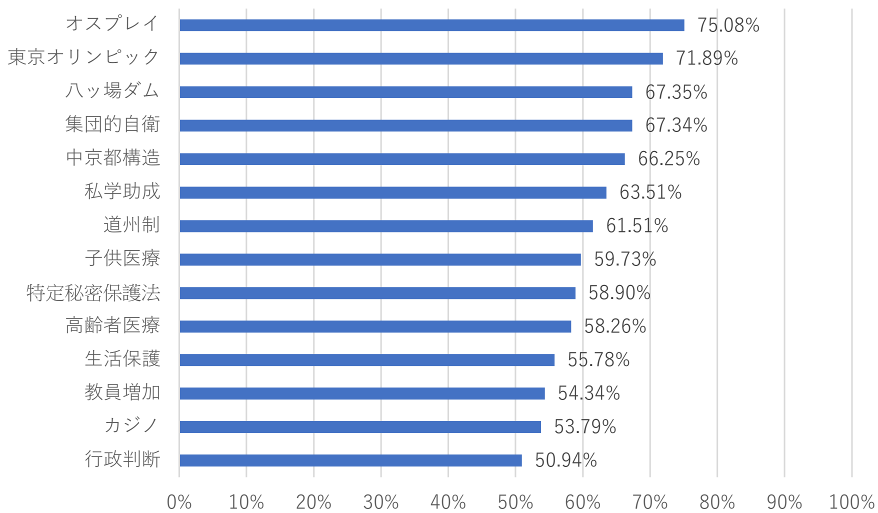

Fig. 6. Accuracy of MaxEnt *Fact-checkable* classification

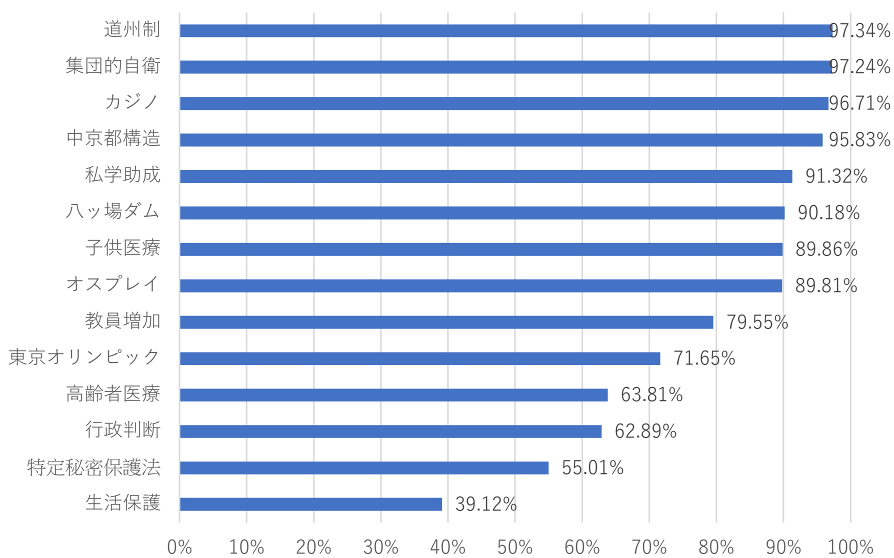

Fig. 7. Accuracy of MaxEnt *Relevance* classification

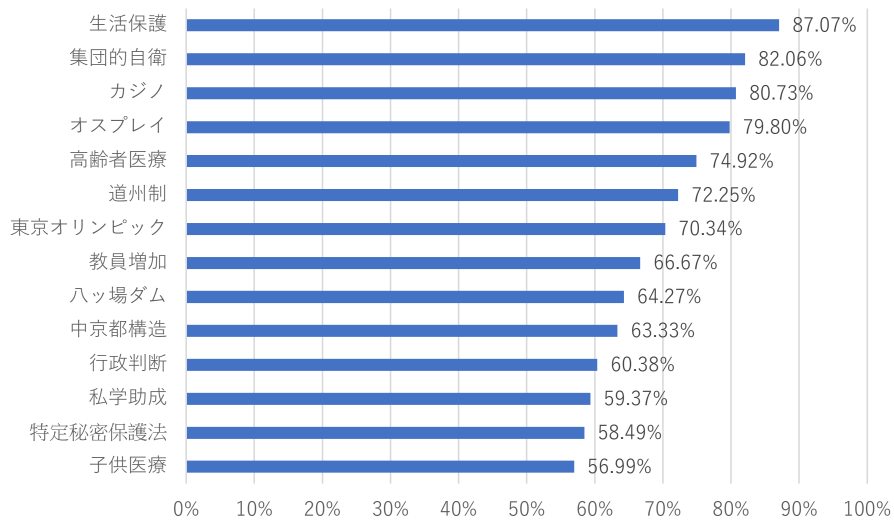

Fig. 8. Accuracy of MaxEnt *Stance* classification

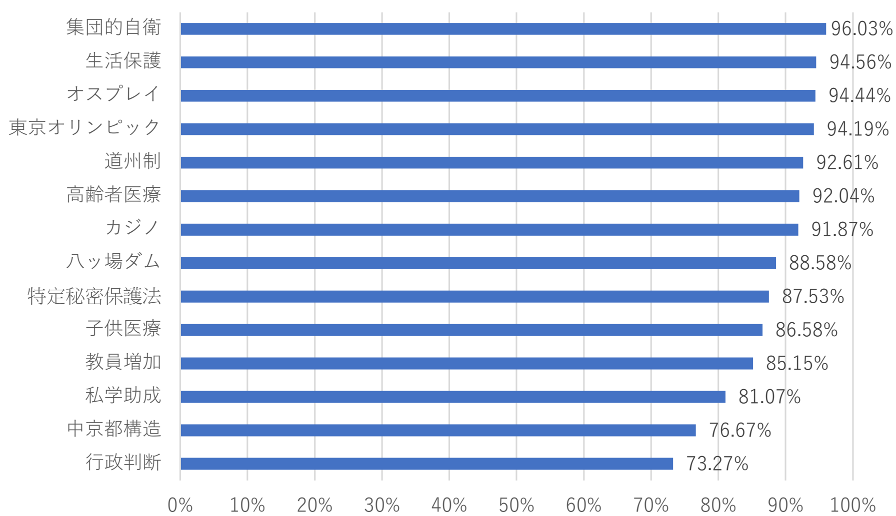

Fig. 9. Accuracy of MaxEnt *Class* classification

Table 3. Labels distribution in the training data

Topic name	*Rl* 0	*Rl* 1	*Fc* 0	*Fc* 1	*St* 0	*St* 1	*St* 2
カジノ (Casino)	110	1,055	745	420	1,068	52	45
集団的自衛 (Self defense)	65	280	171	174	243	97	5
八ッ場ダム (Yanba)	87	799	625	261	721	117	48
高齢者 (Elderly people)	104	560	533	131	617	47	0
私学助成 (Private school grants)	96	406	314	188	399	100	3
中京都構想 (Medium Kyoto)	6	220	136	90	195	27	4
オスプレイ (Osprey)	241	933	1,090	84	1,072	2	100
特定秘密保護法 (Secret protection)	178	284	283	179	249	34	179
道州制 (Do-Shu-system)	6	1,146	578	574	1,043	81	28
子ども医療 (Children medical expenses)	4	237	211	30	157	79	5
教員増加 (Regular faculty members)	89	1,426	1,364	151	1,417	28	70
生活保護 (Welfare)	626	201	696	131	821	0	6
東京オリンピック (Tokyo Olympics)	14	820	741	93	587	228	19
空き家 (Vacant houses)	53	245	89	209	187	85	26