

Extracting Troubles from Daily Reports based on Syntactic Pieces

Kakimoto Yoshifumi and Kazuhide Yamamoto

Nagaoka University of Technology,
1603-1, Kamitomioka, Nagaoka, Niigata 940-2188 Japan
{kakimoto, ykaz}@nlp.nagaokaut.ac.jp

Abstract. It is expensive for companies to browse daily reports. Our aim is to create a system that extracts information about problems from reports. This system operates in two steps. First, it records expressions involving troubles in a dictionary from training data. Second, it expands the dictionary to include information not included in the training data. We experimentally tested this extraction system; in the tests, a two-values classifier attained an F-value of 0.772, and experimental extraction of troubles attained a precision of 0.400 and a recall of 0.827.

Keywords: syntactic piece, extracting troubles, knowledge dictionary

1 Introduction

In recent years, many companies request daily reports of text data from company members. The text data is Email and web form. However, daily reports are browsed by human, it is expensive for companies. Our aim is to cut back the workload of daily reports browsing.

This paper discusses a type of information extraction. Ichimura et al. (2001) developed a system that reduces the cost of access to reports and assists decision-making for users. This system extracts ‘good’ or ‘bad’ practices using a human-created knowledge dictionary. However, the knowledge dictionary depends on the specific enterprise because it is constructed using reports from that enterprise. Saito and Watabe (2001) developed a system that extracts and visualizes information indicating troubles using extraction rules defined by humans. They chose the extraction categories ‘trouble’, ‘causality’ and ‘countermeasure’, which their system evaluated with a precision of 0.878, 0.701 and 0.703, respectively. This system depends on the specific domain, because they considered only printer problems. Both systems produce rules or a dictionary only from training data. This approach has two problems. First, because the dictionary and rules are made by humans, the cost is considerable. Second, this system cannot extract troubles that is not included in the training data. Our approach corrects these problems. Our system automatically creates a dictionary, and can extract troubles by expanding the dictionary to add data not included in the training data.

2 Definition of the Trouble

In this paper, we define the trouble as ‘content regarding some problem in daily reports’. Troubles must take into account the context of the problem. A single word does not provide adequate troubles because it is too short. For example, when the word ‘壊れる(break)’ occurs in daily reports, we can identify a trouble word. But ‘椅子が壊れる(The chair breaks)’ and ‘サーバーが壊れる(The server breaks)’ have different meanings, calling for different responses and involving different degrees of risk. Aoki and Yamamoto (2007) investigate syntactic pieces that mine units of syntactic structure. Syntactic pieces are pairs consisting of a modifier and a modificand, as shown in Figure 1.

入力文(input sentence):

古いパソコンのバッテリーがいきなり爆発した。
(The battery of the old computer exploded suddenly.)

出力構文片(output syntactic pieces):

古い(old) ⇒ パソコン(computer)

バッテリーが(battery) ⇒ 爆発する(explode)

いきなり(suddenly) ⇒ 爆発する(explode)

Figure 1: Example of syntactic pieces

Syntactic pieces have the following characteristics:

- They are easy to extract
- They yield statistical information readily
- They are amenable to matching analysis
- They can deal with a chunk of meaning

We can apply these characteristics to troubles. Therefore, we found it more comfortable to deal with syntactic pieces than with words. Furthermore, we dealt with continuous modification of the syntactic pieces because troubles must take into account the context.

3 Method

3.1 System Overview

Figure 2 shows an overview of the system.

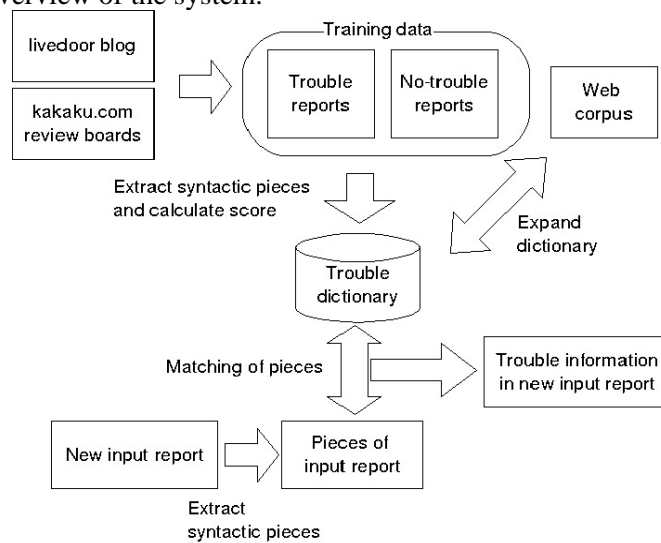


Figure 2: System overview

This paper defines weblogs and bulletin boards as daily reports. Our system uses weblogs and bulletin boards as sources of training data. First, we extract syntactic pieces from the training data, compute scores and construct a dictionary of troubles (the trouble dictionary). Second, we expand this dictionary because we must process troubles that do not appear in the training data. Third, we extract troubles from new reports using the trouble dictionary.

3.2 Training Data

The weblogs we used are part of the ‘livedoor blog’⁽²⁾, and have original tags and titles supplied by each author. We accumulated reports that mentioned problems (trouble reports) in the tag and title text. We defined trouble reports as those with the term ‘トラブル(trouble)’ in tags or

titles. We defined no-trouble reports as those that did not include ‘トラブル(trouble)’ in tags, titles or sentences.

Bulletin boards are ‘kakaku.com review boards’⁽³⁾. These boards have context tags supplied by users. We defined trouble reports as those with the term ‘悪い(bad)’ in the tags. No-trouble reports did not include ‘悪い(bad)’ or ‘質問(question)’ in the tags.

3.3 Trouble Dictionary

We constructed a trouble dictionary using these reports. We collected syntactic pieces from trouble and no-trouble reports and assigned a score to each one as troubles. The score observes deviation between trouble and no-trouble reports. Common pieces appear with similar frequency in trouble and no-trouble reports. Troubles has a higher score because it occurs more often in trouble reports. Therefore, we employed the method of Fujimura et al. (2004). We calculate the trouble score as follows:

1. Prepare the corpus, classified as trouble reports or no-trouble reports
2. Extract the syntactic piece
3. Count the frequency of each piece in each reports
4. Compute the trouble score with the following equation

$$S(w_i) = \frac{P'(w_i) - N(w_i)}{P'(w_i) + N(w_i)} \quad (1)$$

$$P'(w_i) = P(w_i) \times \frac{N_{doc}}{P_{doc}} \quad (2)$$

where w_i is a syntactic piece, $P(w_i)$ is the frequency of trouble reports containing w_i , $N(w_i)$ is the frequency of no-trouble reports containing w_i , P_{doc} is the total number of trouble reports and N_{doc} is the total number of no-trouble reports. Expression (2) defines a population parameter because there really is difference of frequency between P_{doc} and N_{doc} . If the score computed by expression (1) is a positive number, the syntactic piece occurs frequently in trouble reports. However, expression (1) does not consider the frequency of syntactic pieces in the training data. For example, in the two cases

Case1. frequency: 100, score: 0.9 and

Case2. frequency: 10000, score: 0.9,

One may regard the second case is more reliable than the first. We applied the confidence interval estimation method to fix this problem. Expression (3) shows this method (Agresti and Coull, 1998).

$$score(w_i) = S'(w_i) - 2 * 1.96 \sqrt{\frac{S'(w_i)(1 - S'(w_i))}{P'(w_i) + N(w_i) + 4}} \quad (3)$$

$$S'(w_i) = \frac{P'(w_i) + 2}{P'(w_i) + N(w_i) + 4} - \frac{N(w_i) + 2}{P'(w_i) + N(w_i) + 4} \quad (4)$$

The second term of expression (3) is the confidence interval. This paper considers only negative values for the confidence interval because we treat syntactic pieces with positive values for $score(w_i)$. The second term of expression (3) is the doubled confidence interval, because $score(w_i)$ is the computed difference of two probabilities. This paper considers only negative values for the confidence interval, because we treat syntactic pieces with positive values for $score(w_i)$. As mentioned previously, we computed the trouble score of syntactic pieces, and added syntactic pieces having positive scores to the trouble dictionary.

3.4 Expansion of Trouble Dictionary with Syntactic Pieces

Extraction of troubles from new input reports was conducted by comparison with the trouble dictionary created as described in section 3.3. However, we cannot extract all troubles in the new reports using the trouble dictionary because it contains troubles from the training data only. Therefore, we expanded the trouble dictionary, tackling troubles not included in the training data, as described in the following sections.

3.4.1 Expansion of Target

We considered syntactic pieces that only included verbal nouns, and we expanded only verbal nouns in syntactic pieces. Verbal nouns means nouns that used in the same manner as verbs. We consider that if a verbal noun changes, the meaning of the syntactic piece changes too. For example, ‘メッセージが出ない (don’t output message)’ is a trouble, but ‘クレームが出ない (don’t output complaint)’ is not a trouble. Therefore, this paper uses only verbal nouns for expanded objects. We do not expand verbs because verbs have many conjugations, and handling them would be too complex. Below, ‘expansion’ refers only to verbal nouns.

3.4.2 Expansion Method

As we expand the trouble dictionary, the plausibility of troubles in syntactic pieces in the dictionary must not change. We must find verbal nouns that easily apply in the context of the expanded object. Figure 3 outlines the expansion method.

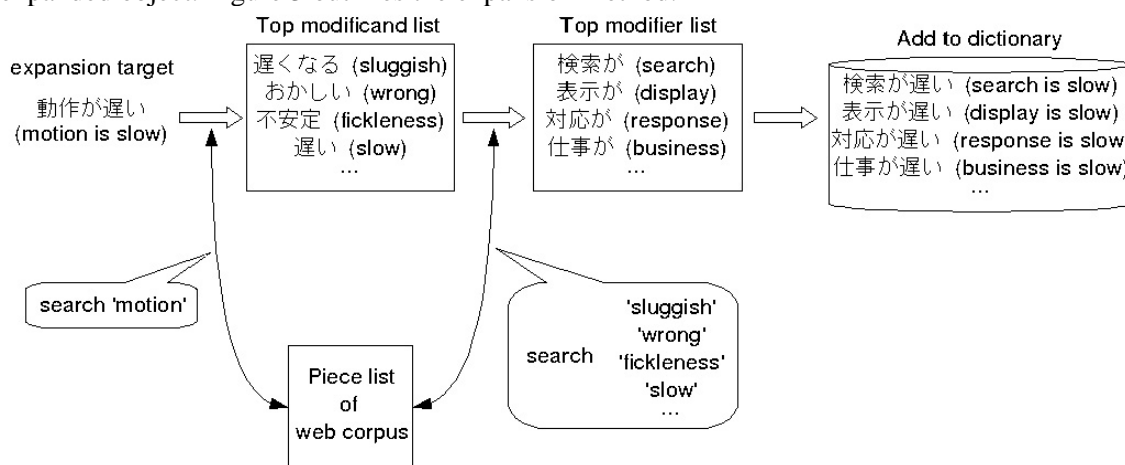


Figure 3: Expansion overview

First, we arrange a large web corpus that does not include the training data. Second, we extract syntactic pieces from this corpus, making a piece list of syntactic pieces. In the piece list, we note the frequency of syntactic pieces in the web corpus. Third, we use one of two methods to expand the dictionary. One method is to expand the modifiers (modifier expansion); the other is to expand the modificands (modificand expansion). Following are details of modifier expansion. For modificand expansion, substitute the word ‘modifier’ for ‘modificand’ and vice versa.

1. Search the piece list for a modifier (motion), and compute the frequency of each modificand (sluggish, wrong, fickleness, and slow). The 10 most frequent modificands are added to the ‘top modificand list’.
2. Search the piece list for each modificand (sluggish, wrong, fickleness, and slow) in the top modificand list, and compute the frequency of each modifier (search, display, response, and business). The 10 most frequent modifiers are added to the ‘top modifier list’.
3. Modifiers on the top modifier list are considered highly likely to occur along with a modifier of the expansion target. We connect the modifiers on that list with a modificand of the expansion target, and add these to the dictionary as new syntactic pieces (search is slow, display is slow, response is slow, and business is slow) having the same trouble score as the expansion target piece.

In this section, we construct the trouble dictionary. We extract troubles from new reports using this dictionary. Following are details of extracting troubles.

1. Extract syntactic pieces from the input report.

2. Set thresholds in the dictionary, and used syntactic pieces that scored higher than the thresholds as troubles.
3. If syntactic pieces in the input report occur in pieces of dictionary, we correct input pieces as troubles.

4 Evaluation

We extracted troubles from new input reports with the trouble dictionary described above. We used two evaluation methods.

1. Evaluation of a two-values classifier

We designed a two-values classifier that differentiated trouble reports from no-trouble reports.

We decided that input reports that extracted troubles were trouble reports, and other cases were no-trouble reports. We classified input reports as having troubles or no-troubles, and evaluated this result.

2. Evaluation of troubles extracted from input reports

We evaluated troubles for plausibility.

4.1 Evaluation Data

We prepared reports that did not include training data, and chose three human evaluators. Evaluators sorted reports into trouble reports and no-trouble reports. The classification basis is whether a report includes expressions indicating troubles. Classification reports include 400 reports having ‘トラブル(trouble)’ in the title, and 400 not having the word ‘トラブル(trouble)’ in the title or main text. These 800 reports are judged in trouble reports or no-trouble reports by evaluators. The evaluators classified 133 as trouble reports and 253 as no-trouble reports. Therefore we used 133 trouble reports and 253 no-trouble reports as evaluation data.

4.2 Evaluation of Two-Values Classifier

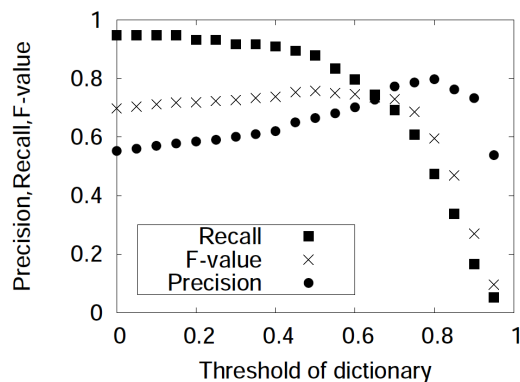


Figure 4 :

Figure 4: Two-values classifier results (unexpanded dictionary)

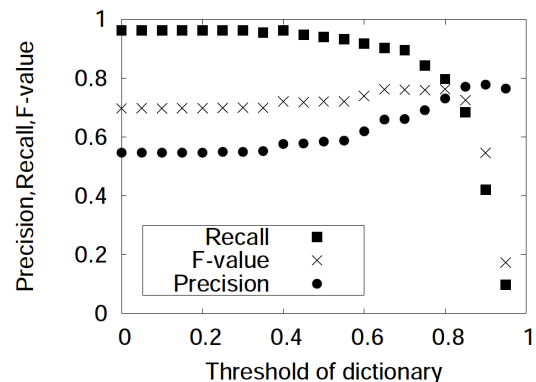


Figure 5:

Figure 5: Two-values classifier results (expanded dictionary)

Figure 4 shows the result of analysis using the dictionary created in section 3.3. Evaluation of the two-values classifier used precision, recall and F-value. The highest F-value is 0.757, which occurs when the threshold is 0.5. In this case, the precision and recall are 0.665 and 0.880, respectively.

Figure 5 shows the result using the expanded dictionary created in section 3.4. The highest F-value is 0.772 when the threshold is 0.780. In this case, the precision and recall are 0.724 and 0.827, respectively.

4.3 Evaluation of Troubles

This section evaluates troubles extracted from the evaluation reports. We evaluate the result obtained using the expanded dictionary. The threshold of the expanded dictionary is 0.780. This

score is the highest point in Figure 5. Four hundred seven pieces of troubles were extracted from the evaluation reports. We evaluated the results by hand, based on the following:

1. The piece is about a problem.
2. The piece is not about a problem, but we can associate it with a problem.
3. The piece is not about a problem, and we cannot associate it with a problem.

The number of pieces corresponding to (1), (2) and (3) is 116, 47 and 244, respectively, indicating that the system can identify troubles but with a precision of only about 0.300. If we consider both (1) and (2) as being the right answer, the precision is about 0.400. The precision is shown expression (5). We cannot evaluate the recall of the extracted troubles, but section 4.2 indicates a recall of 0.827. We believe it represents the recall of extracted troubles. This value of recall is difficult to attain with heuristic rules and a dictionary.

$$Precision = \frac{\text{number of troubles}}{\text{whole number of output pieces}} \quad (5)$$

5 Discussion

5.1 Extracted Troubles

We show part of the extracted troubles from the evaluation data in Table 1.

Table 1: Example of extracted troubles

basis	trouble information
(1)	画面-が ⇒ 表示-され-ない (don't appear on the screen) 遅延-が ⇒ 発生-する (is delayed) 音-が ⇒ 途切れる (sound is interrupted)
(2)	サポート-に ⇒ 電話-する (call for support) 販売-店-に ⇒ 返品-する (return goods to selling office) 原因-を ⇒ 特定-する (identify reasons)
(3)	コンセント-を ⇒ 抜く (pull out a plug) 電源-を ⇒ 入れる (turn on power) 一度-も ⇒ 繋がる (connect only once)

Basis (1) included lots of troubles that had a word meaning ‘ない(not)’. Syntactic pieces including ‘ない(not)’ are easily identified as troubles. The syntactic piece ‘音が⇒途切れる (interrupt the sound)’ clearly indicates the nature of the troubles. In Basis (2), we regarded terms as indicating troubles when they occur with Basis (1) terms. In fact, ‘サポートに⇒電話する (call for support)’ and ‘画面が⇒表示されない(do not appear on the screen)’ occur in the same reports. Therefore, if we connect Basis (2) and Basis (1), we can look on Basis (2) as belonging to Basis (1), with Basis (2) terms clarifying the meaning. Basis (3) includes terms that cannot be considered troubles. As a result, the rate of Basis (3) terms is higher than that of Basis (1) and (2). But there are valid pieces for which the task of two-values classification would work if the pieces cannot be judged as troubles manually. However, we must consider a score that can separate Basis (1) and (2) from Basis (3).

5.2 Expanded Dictionary

We show part of the extracted troubles obtained using the expanded dictionary. In Basis (1) and (2), we obtain the pairs ‘サービス(service)→イメージ(imagery)’, ‘検索する(search)→表示する(display)’, ‘連絡する(inform)→相談する(consult)’ and ‘エラーが(error)→マークが(mark)’. These word usages resemble each other.

Also, troubles in Basis (1) and (2) retains the attribution of troubles in both the unexpanded and expanded dictionaries. In Basis (3), our system can expand the pieces ‘報告する(inform)→取引する(have a deal)’ and ‘連絡を(contact)→返事を(reply)’; these words resemble each other. However, the expanded pieces are not troubles because the pieces of the expansion target are not troubles. Most pieces in Basis (3) belong to this case. This result indicates that the expansion is good. In other words, our expansion method by syntactic pieces is found to be useful for this task. In this paper, the expanded pieces are few because we expanded only verbal

nouns. We must consider other parts of speech: verbs, nouns and adjectives. If other parts of speech are expanded, we can create bigger dictionary.

Table 2: Example of troubles extracted using expanded dictionary

basis	trouble information
(1)	悪い ⇒ サービス (bad service) → 悪い ⇒ イメージ (bad image) 検索-が ⇒ できない (can't search) → 表示-が ⇒ できない (can't display)
(2)	サポートに ⇒ 連絡-する (inform for support) → サポート-に ⇒ 相談-する (consult for support) エラー-が ⇒ 出る (get the error) → マーク-が ⇒ 出る (get the mark)
(3)	私-は ⇒ 報告-する (I inform) → 私-は ⇒ 取引-する (I have a deal) 連絡-を ⇒ くれる (receive the contact) → 返事-を ⇒ くれる (receive the reply)

6 Conclusion

In summary, we developed a system that extracts troubles from reports. Our dictionary is constructed using training data involving syntactic pieces, and is expanded to accommodate unknown troubles. We extract troubles from input reports with the dictionary. We evaluated our system using two methods. The two-values classifier had an F-value of 0.772, and the extracted troubles had a precision of 0.400 and recall of 0.827.

7 Future Work

We are currently working on two future projects. The first problem involves the treatment of Basis (2) in section 5.1. These syntactic pieces can be considered troubles but do not directly indicate a problem. We think that if we can connect Basis (2) and Basis (1) pieces, Basis (2) pieces can be called troubles. The second problem is about the expanded dictionary. The expansion method cannot add many pieces because we consider only verbal nouns. We must consider expanding other parts of speech.

Tools and language resources

- (1) CaboCha, Ver.0.53, Matsumoto Lab., Nara Institute of Science and Technology.
<http://chasen.org/~taku/software/cabocha/>
- (2) livedoor Blog, <http://blog.livedoor.com/>
- (3) kakaku.com review boards, <http://bbs.kakaku.com/bbs/>

References

- Alan Agresti and Brent A. Coull. 1998. Approximate is better than “exact” for interval estimation of binomial proportion. *The American Statistician*, 52:119-126.
- Suguru Aoki and Kazuhide Yamamoto. 2007. Opinion Extraction based on Syntactic Pieces. *The 21st Pacific Asia Conference on Language, Information and Computation*, pages 76-86.
- Shigeru Fujimura, Masashi Toyota, and Masaru Kitsuregawa. 2004. A Consideration of Extracting Reputations and Evaluative Expressions from the Web. *Technical Report of The Institute of Electronics, Information and Communication Engineers*, 104:141-146.
- Yumi Ichimura, Yasuko Nakamura, Toshio Akahane, Miyoko Miyosi, Toshikazu Sekiguchi, and Yousuke Fujiwara. 2001. Text Mining System for Analysis of a Salesperson’s Daily Reports. *Pacific Association for Computational Linguistics*, pages 127-135.
- Takahiro Saito and Isamu Watabe. 2001. The mining from trouble control data(in Japanese). *Information Processing Society of Japan SIGNL Note*, 93:145-152.