6R-8

Affect Annotation of Narrative Text Using Crowdsourcing

Lei Duan, Satoshi Oyama, Haruhiko Sato, Masahito Kurihara †
Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan †

***Abstract*—This paper focuses on the collection of training data for affect analysis research using crowdsourcing, and then applies some statistical methods in order to make better use of crowdsourcing.**

***Index Terms*—affect analysis, annotation, narrative text, crowdsourcing**

1. Introduction

TEXT based affect analysis is one of the hotspot in artificial intelligence. Meanwhile, crowdsourcing has already played an important role in this research area. This paper discusses the utility of crowdsourcing for building an affect annotated corpus of narrative text.

2. Affect analysis

Up to now, many researchers are concentrating on affect analysis research. (Alm, et al, 2005) discussed the importance of different features in a sentence for affect analysis with SNoW learning architecture. (Ptaszynski el at, 2012) did an experiment about affect analysis of certain roles in narrative text.

As the core of affect analysis research, the analysis module can be divided into two types: one is unsupervised learning module, which only needs some machine learning algorithms; the other is supervised learning module. Besides learning algorithms, it also needs a large amount of training data. (Alm, 2010) is a study about characteristics of high agreement affect annotation of training data for affect analysis.

3. Crowdsourcing

For supervised learning module, the quality of training data has direct bearing on the success or failure of affect analysis research. Usually, training data are often made up by manual annotation. There are two kinds of annotator: one is expert annotators, who are expensive, good, but not so fast; the other is non-expert annotators, who are cheap, fast, but not so good. According to (Snow et al, 2008), on average, it requires only 4 non-expert annotators to achieve the performance as a single expert annotator. However, the cost for one non-expert annotator is much less than one-quarter of the cost for one expert annotator. So to collect data from non-expert annotators will be more effective.

Crowdsourcing is a helpful tool to collect data from non-expert annotators. For a complete general overview of algorithms used to improve the quality of crowdsourcing data, see (Kashima et al, 2012).

4. Task design

This section describes the general design of our experiment after introducing narrative text and emotion set used in the experiment.

4.1. Narrative text

To evaluate the utility of crowdsourcing, it is important to choose appropriate narrative text. Thanks to the simplicity and clarity, children's fairy stories are popular in affect analysis researches.

So in this task, a Japanese children's fairy story called “Little Masa and a red apple (政ちゃんと赤いりんご)” is chosen as the narrative text. Resumptively, it is a story about two brothers fight for a better apple of two. Anyone who is interested in the story can download it from Aozora Library freely[1].

4.2. Affect set

Depending on different purpose, the standard of suitable affect set changed. Sometimes, affect set with only “emotive, neutral” or “positive, negative, neutral” is used in some simple affect analysis researches. In the other hand, “the Big Six” (Cornelius, 2000), including happiness, fear, anger, surprise, disgust and sadness, is widely used in affect analysis researches. In our experiment, we choose an affect set with 10 affect types on the basis on “Emotive Expression Dictionary” (Nakamura, 1993), see Table 1. These redundant affect types allow us to perform an in-depth study on the relationships between affects.

4.3. Experiment

In this experiment, only the actor’s lines are annotated. All the 78 lines in this story are presented to 10 non-expert annotators gathered from a crowdsourcing marketplace, Lancers[2]. Every annotator has to read each line, then presume and check the actor’s emotions reflected by the line spontaneously. The response for each line can be multiple-choice. If annotator has a feeling of that none of the emotions is reflected by the line, he (or she) should check “neutral”. Finally, ten independent annotations are collected separately.

[1] Aozora Library, http://www.aozora.gr.jp/
[2] Lancers, http://www.lancers.jp/

Table 1 Frequency of affects

Affect	Freq	Affect	Freq
anger (A)	192	disgust (D)	85
relief (R)	164	surprise (Su)	85
happiness (H)	131	excitement (E)	83
sadness (Sa)	125	fear (Fe)	57
neutral (N)	*98*	shame (Sh)	32
fondness (Fo)	90	**Total**	**1122**

5. Statistical results

This section lists some statistical results of the affect annotation for the purpose of evaluating the utility of crowdsourcing.

5.1. Frequency & Distribution of Affects

After annotating, we got 1122 affect labels in total (including 98 neutrals). Maybe because the emotion expression of this story is not so complicated, on average, one sentence is annotated only 1.32 affects (except neutral) by one annotator.

Table 1 shows the specific frequency of each affect descendingly. For all affects, anger is the most frequent of all. Probably because there are many quarrel words owing to the story's topic.

Figure 1 shows the ratio of sentences given each affect. However, sentences given relief, not anger, occupy the biggest fraction of all. Perhaps anger is easier to be distinguished by annotators. So the large numbers of "angers" center on a relatively small number of sentences. According to the plot of this story, the hero, little Masa, realized some essence about interpersonal relationship at last. Therefore, it is no wonder that relief is the most common affect among sentences.

5.2. (Dis)agreement of Annotation

Sentences with mixed neutral and affect occupied 62%. The rest 38% are sentences with mixed affects. Perhaps 10 annotators are too many, none of sentences with only neutral or high-agree affects is found.

Let p_{ij} be the probability that sentence i is annotated affect j. Furthermore, "agreement" can be scored via entropy:

$$\mathrm{H}(i) = -\sum_{j\in J} p_{ij} \log_2 p_{ij}$$

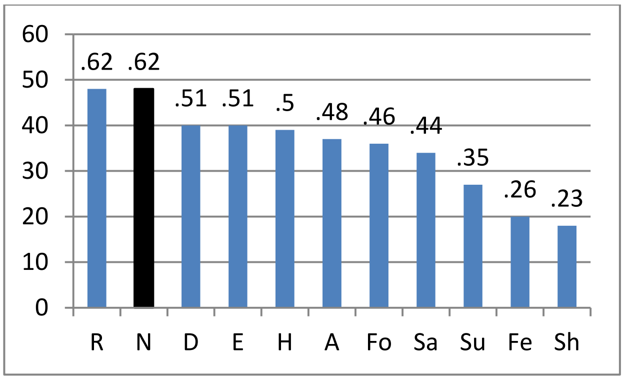

Figure 1 Ratio of sentences given each affect

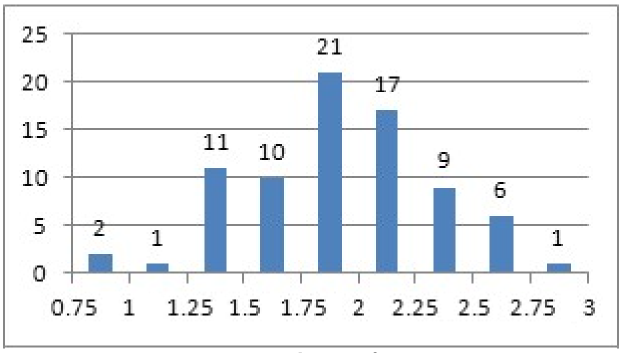

Figure 2 Distribution of entropy

Entropy is the average uncertainty of a single random variable. The more entropy is, the less agreeable the sentence is. Distribution of entropy is in Figure 2, which is suitably modeled by Gaussian distribution. The average entropy of all sentences is merely 1.93. Due to the distinct emotion expression of the story, annotators' determination is relatively consistent.

6. Conclusion

Inferior training data have the direct bearing on the failure of affect analysis research. This paper brought attention to the collection of training data for affect analysis research using crowdsourcing, and discussed some statistical results of affect annotation on a Japanese story called "Little Masa and a red apple". In future research, we plan to explore more useful properties of crowdsourcing, like inspecting characteristics of ambiguous and agreeable sentences, and so on. We hope that it'll be helpful to avoid some shortages of crowdsourcing and improve the quality of training data for affect analysis research finally.

References

[1] C. O. Alm, D. Roth, R. Sproat, "Emotions from text: machine learning for text-based emotion prediction", *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2005, pp. 579-586.

[2] M. Ptaszynski, H. Dokoshi, S. Oyama, R. Rzepka, M. Kurihara, K. Araki, Y. Momouchi, "*Affect Analysis in Context of Characters in Narratives*", Expert Systems with Applications, *Elsevier*, 2012, pp. 0957-4174.

[3] C. O. Alm, "Characteristics of high agreement affect annotation in text", *Proceedings of the Fourth Linguistic Annotation Workshop*, Association for Computational Linguistics, 2010, pp. 118-122.

[4] R. Snow, B. O'Connor, D. Jurafsky, A. Y. Ng, "Cheap and fast---but is it good?: evaluating non-expert annotations for natural language tasks", *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2008, pp. 254-263.

[5] H. Kashima, H. Kajino, "Crowdsourcing and Machine Learning", *Journal of the Japanese Society for Artificial Intelligence*, The Japanese Society for Artificial Intelligence, 2012, vol. 27, no. 4 pp. 381-388.

[6] Cornelius, R. Randolph, "Theoretical approaches to emotion", *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, 2000.

[7] A., Nakamura. "Kanjo hyogen jiten [Dictionary of Emotive Expressions] (in Japanese)", *Tokyodo Publishing*, Tokyo.