Extracting Emotional Polarity of Words using Spin Model

Hiroya Takamura,† Takashi Inui† and Manabu Okumura†

We propose a method for extracting emotional polarities of words: desirable or undesirable. Regarding emotional polarities as spins of electrons, we use the mean field approximation to compute the approximate probability function of the system instead of the intractable actual probability function. Given only two seed words "good" and "bad", the proposed method extracts 500 emotional polarities with about 75% precision.

1. Introduction

Identification of emotions (including opinions and attitudes) in text is an important task which has a variety of possible applications. For example, we can efficiently collect opinions on a new product from the internet, if opinions in bulletin boards are automatically identified. We will also be able to grasp people's attitudes in questionnaire, without actually reading all the responds.

An important resource in realizing such identification tasks is a list of words with emotional polarity: positive or negative (desirable or undesirable). Frequent appearance of positive words in a document implies that the writer of the document would have a positive attitude on the topic. The goal of this paper is to propose a method for automatically creating such a word list out of glosses (i.e., definition or explanation sentences) in a dictionary. For this purpose, we use *spin model*, which is a model for a set of electrons with spins. Just as each electron has a direction of spin (up or down), each word has an emotional polarity (positive or negative). We therefore regard words as a set of electrons and apply the mean field approximation to compute the average polarity of each word.

We empirically show that the proposed method works well even with a small number of seed words; emotional polarities are given to 500 words with about 75% precision by two seed words "good" and "bad", and with about 85% precision by four seed words "superior", "inferior" and the two words above ☆.

2. Related Work

Kobayashi et al.[4] proposed a method for extracting emotional polarities of words with bootstrapping. The polarity of a word is determined on the basis of its gloss, if any of their 52 hand-crafted rules is applicable to the sentence. Rules are applied iteratively in the bootstrapping framework. Although Kobayashi et al.'s work provided an accurate investigation on this task and inspired our work, it has a drawback: a low recall. In their paper, they reported that the polarities of only 113 words are extracted with precision 84.1% (the low recall would be partly because their set of seed words was too large (1187 words)). This drawback will be removed in our method.

Kamps et al.[3] constructed a network by connecting each pair of synonymous words provided by WordNet[1], and then used the shortest paths to two seed words "good" and "bad" to obtain the semantic orientation of a word. They reported an accuracy around 67% to 77% for adjectives, depending on experimental settings. Limitations of their method are that a synonymy dictionary is required and that how to use a larger set of seed words is unclear. Their evaluation is restricted to adjectives.

Subjective words often have the positive polarity or the negative polarity (not neutral).

† Tokyo Institute of Technology,
Precision and Intelligence Laboratory.
4259 Nagatsuta Midori-ku Yokohama,
JAPAN, 226-8503,
phone: +81-45-924-5295
email: {takamura, oku}@pi.titech.ac.jp,
tinui@lr.pi.titech.ac.jp

☆ In the later experiments, we use Japanese data. However, we write the corresponding English words in text of the paper for readers' convenience. "good", "bad", "superior" and "inferior" are respectively "yoi", "warui", "sugureru" and "otoru" in Japanese.

Wiebe[11] used a learning method to collect subjective adjectives from corpora. Riloff et al.[10] focused on the collection of subjective nouns.

3. Spin Model and Mean Field Approximation

We give a brief introduction to the spin model and the mean field approximation[5], which is a well-studied subject both in statistical mechanics and machine learning communities.

A spin system is an array of N electrons, each of which has a spin with one of two values "+1 (up)" or "−1 (down)". Two electrons next to each other energetically tend to have the same spin. We call this model *the spin model*☆. As a result, the energy function of a spin system can be represented as

$$E(\mathbf{x}, W) = -\frac{1}{2}\sum_{m,n} w_{mn} x_m x_n, \qquad (1)$$

where x_m and x_n ($\in \mathbf{x}$) are spins of electrons m and n, matrix $W = \{w_{mn}\}$ represents weights between two electrons.

In a spin system, the variable vector $\mathbf{x}$ follows the Boltzmann distribution:

$$P(\mathbf{x}|W) = \frac{\exp(-\beta E(\mathbf{x}, W))}{Z(W)}, \qquad (2)$$

where $Z(W) = \sum_{\mathbf{x}} \exp(-\beta E(\mathbf{x}, W))$ is the normalization factor, which is called *partition function* and β is a constant called "inverse-temperature".As this distribution function suggests, a configuration with a higher energy value has a smaller probability.

Although we have a distribution function, computing various probability values is computationally difficult. The bottleneck is the evaluation of $Z(W)$, since there are 2^N configurations of spins in this system.

We therefore approximate $P(\mathbf{x}|W)$ with a simple function $Q(\mathbf{x};\theta)$. θ, a set of parameters for Q, is determined such that $Q(\mathbf{x};\theta)$ becomes as similar to $P(\mathbf{x}|W)$ as possible. As a measure for the distance between P and Q, the variational free energy F is often used, which is defined as the difference between the mean energy with respect to Q and the entropy of Q:

$$F(\theta) = \beta \sum_{\mathbf{x}} Q(\mathbf{x};\theta) E(\mathbf{x}; W) - \left(-\sum_{\mathbf{x}} Q(\mathbf{x};\theta) \log Q(\mathbf{x};\theta) \right). \qquad (3)$$

The parameters θ that minimizes the variational free energy will be chosen. It has been shown that minimizing F is equivalent to minimizing the Kullback-Leibler divergence between P and Q (see (A.1) for proof).

We next assume that the function $Q(\mathbf{x};\theta)$ has the factorial form:

$$Q(\mathbf{x};\theta) = \prod_i Q(x_i;\theta_i). \qquad (4)$$

Simple substitution and transformation leads us to the actual representation of the variational free energy (see (A.2) for details).

$$F(\theta) = -\beta\frac{1}{2}\sum_{m,n} w_{mn}\bar{x}_m\bar{x}_n - \sum_i \left(-\frac{1+\bar{x}_i}{2}\log\frac{1+\bar{x}_i}{2} - \frac{1-\bar{x}_i}{2}\log\frac{1-\bar{x}_i}{2} \right). \qquad (5)$$

From the stationary condition, we obtain the *mean field equation*:

$$\bar{x}_i = \tanh(\beta \sum_j w_{ij}\bar{x}_j). \qquad (6)$$

This equation is solved by the following iterative update rule:

$$\bar{x}_i^{new} = \tanh(\beta \sum_j w_{ij}\bar{x}_j^{old}). \qquad (7)$$

4. Extraction of Emotional Polarity of Words with Spin Model

We use the spin model to extract emotional polarities of words.

Each spin has a direction taking one of two values: up or down. Two neighboring spins tend to have the same direction from a energetic reason. Regarding each word as an electron and its emotional polarity as the spin of the electron, we construct a lexical network by connecting two words if one word appear in the gloss of the other word. Intuition behind this is that if a word has an emotional polarity, then the words in its gloss tend to have the same emotional polarity.

In the following, we explain how to construct a lexical network, compute an approximate probability function and extract emotional po-

☆ This model is also called *Ising model*.

larities.

4.1 Construction of Network

We construct a lexical network by connecting two words if one word appear in the gloss of the other word. We first define $GL_+(t)$ as the set of words in the gloss of word t excluding the words syntactically depended by a negation word in the gloss. We also define $GL_-(t)$ as the antonyms of t and the words syntactically depended by a negation word in the gloss. The adjacency matrix $W = (w_{ij})$ is defined as follows[☆]:

$$w_{ij} = \begin{cases} 1 & (t_i \in GL_+(t_j) \text{ or } t_j \in GL_+(t_i)) \\ -1 & (t_i \in GL_-(t_j) \text{ or } t_j \in GL_-(t_i)) \\ 0 & otherwise \end{cases} \quad (8)$$

For example, for two words t_1 ("exquisite") and t_2 ("beautiful") with the following glosses[8]:

exquisite : extremely beautiful or delicate,
beautiful : delighting the aesthetic senses,

w_{12} is set to 1, because the gloss of "exquisite" contains "beautiful".

4.2 Extraction of Polarity

Average spins (i.e. average emotional polarities) of words are computed as explained in the previous section. Initially, averages of seed words are set according to their polarities and the other averages are set to 0. The words with high average values are classified as positive words. The words with low average values are classified as negative words.

We make two modifications to the original spin model, in order for the model to be better fitted to the task of polarity extraction. One modification is that instead of using update rule (7), we use

$$\bar{x}_i^{new} = \tanh(\beta \frac{1}{\sum_k |w_{ik}|} \sum_j w_{ij} \bar{x}_j^{old}). \quad (9)$$

We require this *normalization factor*, because, with the original update rule, words with longer glosses tend to have extreme averages (very positive or very negative). We should be aware that this modification of normalization is not theoretically justifiable in the sense of minimization of the variational free energy, since the adjacency matrix must be symmetric in the valid spin model.

The other modification is the update rule for seed words. The averages of seed words are reset according to their given polarities at each iteration.

5. Experiments

We evaluate the proposed method using Iwanami Japanese dictionary[7]. For the morphological analysis of glosses, we used ChaSen[6]. We used only content words: nouns, verbs, adjectives, adverbs and auxiliaries. The auxiliaries "nai" and "nu" are regarded as negation words. The words preceding one of these negation words are regarded as "syntactically depended by a negation word". Although dependency analysis would enable a more accurate preprocessing, we use only a simple part-of-speech tagging in order to show that the proposed method works even without a high-performance dependency analyzer.

After deleting isolated words (i.e. words having no connections to other words), we obtain a network consisting of 58185 words. We manually labeled 9790 words with emotional polarities (2491 positive words, 3141 negative words and 4158 neutral words). Sampling of these 9790 words is in some sense biased, because we mainly labeled words with high absolute values of averages. As a result, the number of neutral words is presumably smaller than that of complete random sampling.

The inverse-temperature β is fixed to 0.75 (other values of β ranging from 0.1 to 1.0 caused no significant change in results).

5.1 Results of binary classification

Since we have not included the neutral polarity into the model and only 9790 labeled words are available, we first evaluate the method only for positive labeled words and negative labeled words. The seed words are "good" and "bad". The result is shown in Table 1, which includes the accuracy for each part-of-speech (POS). We can thus automatically determine the polarities of words (especially nouns) with high accuracy, if we know that the words are polarized. However, nouns actually include many physical-object words, which are neutral. We cannot conclude that the classification of nouns

[☆] The condition for w_{ij} being 1 and the condition for w_{ij} being -1 can hold simultaneously. In such cases, w_{ij} is set to 0. We do not explicitly describe those cases for the simplicity.

Table 1 Binary classification accuracy and POS

POS	Accuracy
nouns	0.812
adjectives	0.745
verbs	0.762
others	0.777
all	0.798

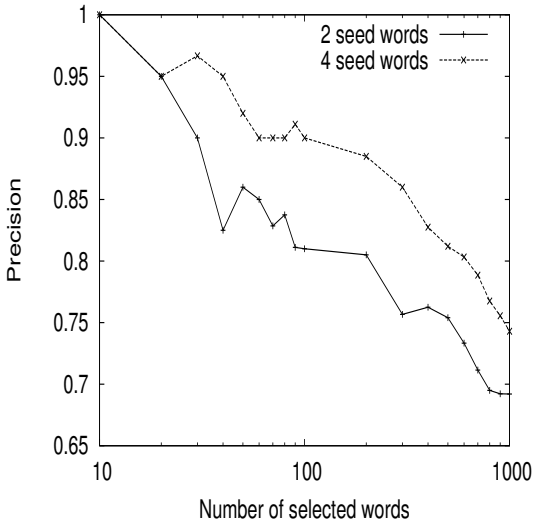

Fig. 1 Precision for the words with high confidence, 2 seed words and 4 seed words.

is easy.

5.2 Precision for the words with high confidence

We next evaluate the proposed method in terms of precision for the words that are classified with high confidence. We regard the absolute value of each average as a confidence measure and evaluate the top 1000 words with the highest absolute values of averages. Unlike the previous subsection, all the 1000 words are included in the evaluation set. If the correct label of a word is neutral and the word is ranked in the top 1000 list, the decision for this word is incorrect.

The result of this experiment is shown in Figure 1. The 2 seed words are "good" and "bad". The 4 seed words are the above 2 words and "superior" and "inferior".

Figure 2 shows the result for each POS. Unlike Table 1, we obtained high precision values for adjectives. We should be aware that in Figure 2, the numbers of adjectives and verbs in the top 1000 list are much smaller than that of nouns.

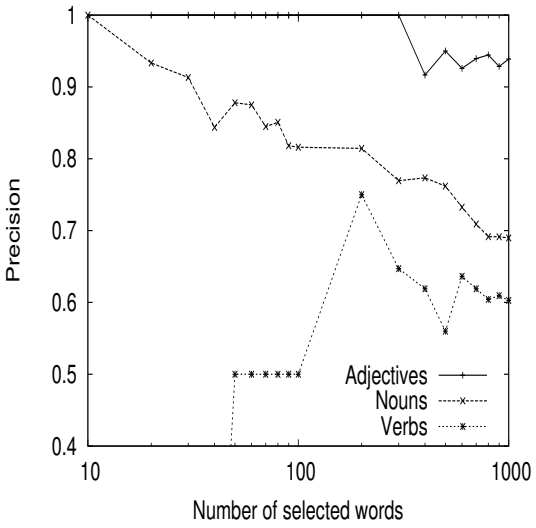

Fig. 2 Precision for each POS, 2 seed words.

6. Future Work

Future work includes the following.

- The weights of edges in the current model are fixed to -1, 0 or 1. Incorporation of more flexible weight-adjusting scheme through training will bring a better performance.
- Importance of each word consisting a gloss to some extent depends on its syntactic role. Therefore, syntactic information in glosses should be useful for classification. Even simple word order information in glosses can influence classification results.
- Although we used only glosses in dictionary, corpus data can also be used in our method. Two words that appear in some special context will have the same emotional polarity[2]. Such words can be connected in the lexical network, as the gloss connects two words in the current method. We can also use synonyms in a thesaurus.
- One deficiency in the current model is that the spin can take only two values, though the actual emotional polarity can take three values: positive, negative or neutral. A promising model that can overcome this deficiency is the *Potts model*[9], in which spins are allowed to take more than two values.
- In order to decrease the amount of manual tagging for seed words, an active learning scheme for this model is desired, in which a small number of *good* seed words are auto-

matically selected.

- We also have to prepare a larger evaluation dataset with high consistency.
- The main part of the proposed method is language-independent. We would like to try other languages as well.

7. Conclusion

We proposed a method for extracting emotional polarities of words. In the proposed method, we regarded emotional polarities as spins of electrons, and used the mean field approximation to compute the approximate probability function of the system instead of the intractable actual probability function. We succeeded in extracting emotional polarities with high precision, even when only a small number of seed words are available.

References

1) Christiane Fellbaum. *WordNet: An Electronic Lexical Database, Language, Speech, and Communication Series*. MIT Press, 1998.
2) Takashi Inui. *Acquiring Causal Knowledge from Text Using Connective Markers*. Ph.D. thesis, Graduate School of Information Science, Nara Institute of Science and Technology, 2004.
3) Jaap Kamps, Maarten Marx, Robert J. Mokken, and Maarten de Rijke. Using wordnet to measure semantic orientation of adjectives. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004), volume IV*, pages 1115–1118, 2004.
4) Nozomi Kobayashi, Takashi Inui, and Kentaro Inui. Dictionary-based acquisition of the lexical knowledge for p/n analysis (in japanese). In *Proceedings of Japanese Society for Artificial Intelligence, SLUD-33*, pages pp.45–50, 2001.
5) David J. C. Mackay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.
6) Yuji Matsumoto, Akira Kitauchi, Tatsuo Yamashita, Yoshitaka Hirano, Hiroshi Matsuda, Kazuma Takaoka, and Masayuki Asahara. *Japanese Morphological Analysis System ChaSen version 2.2.1*, 2000.
7) Minoru Nishio, Etsutaro Iwabuchi, and Shizuo Mizutani. *Iwanami Japanese Dictionary (5th edition)*. Iwanami-shoten, 1994.
8) Judy Pearsall. *The Concise Oxford Dictionary*. Oxford University Press, 1999.
9) Renfrey Burnard Potts. Some generalized order-disorder transformations. In *Proceedings of the Cambridge Philosophical Society*, volume 48, pages 106–109, 1952.
10) Ellen Riloff, Janyce Wiebe, and Theresa Wilson. Learning subjective nouns using extraction pattern bootstrapping. In *Proceedings of the Seventh Conference on Natural Language Learning (CoNLL-03)*, pages 25–32, 2003.
11) Janyce M. Wiebe. Learning subjective adjectives from corpora. In *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI-2000)*, pages 735–740, 2000.

Appendix

A.1 Variational Free Energy and Kullback-Leibler Divergence

$$F(\theta) = -\sum_{\mathbf{x}} Q(\mathbf{x};\theta)\log\exp(-\beta E(\mathbf{x};W)) - (-\sum_{\mathbf{x}} Q(\mathbf{x};\theta)\log Q(\mathbf{x};\theta)) \tag{10}$$

$$= -\sum_{\mathbf{x}} Q(\mathbf{x};\theta)\log\left(\frac{\exp(-\beta E(\mathbf{x};W))}{Z(W)}Z(W)\right) - (-\sum_{\mathbf{x}} Q(\mathbf{x};\theta)\log Q(\mathbf{x};\theta)) \tag{11}$$

$$= -\sum_{\mathbf{x}} Q(\mathbf{x};\theta)\log P(\mathbf{x}|W) - \log Z(W) - (-\sum_{\mathbf{x}} Q(\mathbf{x};\theta)\log Q(\mathbf{x};\theta)) \tag{12}$$

$$= KL(Q||P) - \log Z(W). \tag{13}$$

A.2 Derivation of Variational Free Energy

Since $x_i \in \{+1, -1\}$ holds true and Q is a probability function, we obtain

$$\bar{x}_i = Q(x_i = +1)\cdot 1 + Q(x_i = -1)\cdot -1, \quad 1 = Q(x_i = +1) + Q(x_i = -1). \tag{14}$$

Thus, $Q(x_i;\theta_i)$ can be simply written with its mean $\bar{x}_i$:

$$Q(x_i = +1) = \frac{1+\bar{x}_i}{2}, \quad Q(x_i = -1) = \frac{1-\bar{x}_i}{2}. \tag{15}$$

Since we assume factorial form,

$$\sum_{\mathbf{x}} Q(\mathbf{x};\theta)E(\mathbf{x};W) = \sum_{\mathbf{x}} Q(\mathbf{x};\theta)\left(-\frac{1}{2}\sum_{m,n} w_{mn}x_m x_n\right) \tag{16}$$

$$= -\frac{1}{2}\sum_{m,n} w_{mn}\bar{x}_m\bar{x}_n, \tag{17}$$

$$-\sum_{\mathbf{x}} Q(\mathbf{x};\theta)\log Q(\mathbf{x};\theta) = \sum_i\left(-\sum_{x_i} Q(x_i;\theta_i)\log Q(x_i;\theta_i)\right) \tag{18}$$

$$= \sum_i\left(-\frac{1+\bar{x}_i}{2}\log\frac{1+\bar{x}_i}{2} - \frac{1-\bar{x}_i}{2}\log\frac{1-\bar{x}_i}{2}\right). \tag{19}$$

Therefore

$$F(\theta) = -\beta\frac{1}{2}\sum_{m,n} w_{mn}\bar{x}_m\bar{x}_n - \sum_i\left(-\frac{1+\bar{x}_i}{2}\log\frac{1+\bar{x}_i}{2} - \frac{1-\bar{x}_i}{2}\log\frac{1-\bar{x}_i}{2}\right). \tag{20}$$

A.3 Derivation of Mean Field Equation

We differentiate the variational free energy with respect to $\bar{x}_i$:

$$\frac{\partial F(\theta)}{\partial \bar{x}_i} = -\beta\sum_j w_{ij}\bar{x}_j - \left(-\frac{1}{2}\log\frac{1+\bar{x}_i}{2} + \frac{1}{2}\log\frac{1-\bar{x}_i}{2}\right) \tag{21}$$

By setting the above to 0, we obtain

$$\bar{x}_i = \frac{1-\exp(-2\beta\sum_j w_{ij}\bar{x}_j)}{1+\exp(-2\beta\sum_j w_{ij}\bar{x}_j)} \tag{22}$$

$$= \frac{\exp(\beta\sum_j w_{ij}\bar{x}_j) - \exp(-\beta\sum_j w_{ij}\bar{x}_j)}{\exp(\beta\sum_j w_{ij}\bar{x}_j) + \exp(-\beta\sum_j w_{ij}\bar{x}_j)} \tag{23}$$

$$= \tanh(\beta\sum_j w_{ij}\bar{x}_j). \tag{24}$$