

Extracting Semantic Orientations of Phrases from Dictionary

Hiroya Takamura

Precision and Intelligence Laboratory
Tokyo Institute of Technology
takamura@pi.titech.ac.jp

Takashi Inui

Integrated Research Institute
Tokyo Institute of Technology
inui@iri.titech.ac.jp

Manabu Okumura

Precision and Intelligence Laboratory
Tokyo Institute of Technology
oku@pi.titech.ac.jp

Abstract

We propose a method for extracting semantic orientations of phrases (pairs of an adjective and a noun): positive, negative, or neutral. Given an adjective, the semantic orientation classification of phrases can be reduced to the classification of words. We construct a lexical network by connecting similar/related words. In the network, each node has one of the three orientation values and the neighboring nodes tend to have the same value. We adopt the Potts model for the probability model of the lexical network. For each adjective, we estimate the states of the nodes, which indicate the semantic orientations of the adjective-noun pairs. Unlike existing methods for phrase classification, the proposed method can classify phrases consisting of unseen words. We also propose to use unlabeled data for a seed set of probability computation. Empirical evaluation shows the effectiveness of the proposed method.

1 Introduction

Technology for affect analysis of texts has recently gained attention in both academic and industrial areas. It can be applied to, for example, a survey of new products or a questionnaire analysis. Automatic sentiment analysis enables a fast and comprehensive investigation.

The most fundamental step for sentiment analysis is to acquire the semantic orientations of words: positive or negative (desirable or undesirable). For example, the word “beautiful” is positive, while the word “dirty” is negative. Many researchers have developed several methods for this purpose and obtained good results. One of the next problems to be solved is to acquire semantic orientations of phrases, or multi-term expressions, such as “high+risk” and “light+laptop-computer”. Indeed the semantic orientations of phrases depend on context just as the semantic orientations of words do, but we would like to obtain the orientations of phrases as basic units for sentiment analysis. We believe that we can use the obtained basic orientations of phrases for affect analysis of higher linguistic units such as sentences and documents.

A computational model for the semantic orientations of phrases has been proposed by Takamura et al. (2006). However, their method cannot deal with the words that did not appear in the training data. The purpose of this paper is to propose a method for extracting semantic orientations of phrases, which is applicable also to expressions consisting of unseen words. In our method, we regard this task as the noun classification problem for each adjective; the nouns that become respectively positive (negative, or neutral) when combined with a given adjective are distinguished from the other nouns. We create a lexical network with words being nodes, by connecting two words if one of the two appears in the gloss of the other. In the network, each node has one of the three orientation values and the neighboring nodes expectedly tend to have the same value. For

example, the gloss of “cost” is “a sacrifice, loss, or penalty” and these words (cost, sacrifice, loss, and penalty) have the same orientation. To capture this tendency of the network, we adopt the Potts model for the probability distribution of the lexical network. For each adjective, we estimate the states of the nodes, which indicate the semantic orientations of the adjective-noun pairs. Information from seed words is diffused to unseen nouns on the network.

We also propose a method for enlarging the seed set by using the output of an existing method for the seed words of the probability computation.

Empirical evaluation shows that our method works well both for seen and unseen nouns, and that the enlarged seed set significantly improves the classification performance of the proposed model.

2 Related Work

The semantic orientation classification of words has been pursued by several researchers. Some of them used corpora (Hatzivassiloglou and McKeown, 1997; Turney and Littman, 2003), while others used dictionaries (Kobayashi et al., 2001; Kamps et al., 2004; Takamura et al., 2005; Esuli and Sebastiani, 2005).

Turney (2002) applied an internet-based technique to the semantic orientation classification of phrases, which had originally been developed for word sentiment classification. In their method, the number of hits returned by a search-engine, with a query consisting of a phrase and a seed word (e.g., “*phrase* NEAR *good*”) is used to determine the orientation. Baron and Hirst (2004) extracted collocations with Xtract (Smadja, 1993) and classified the collocations using the orientations of the words in the neighboring sentences. Their method is similar to Turney’s in the sense that cooccurrence with seed words is used. In addition to individual seed words, Kanayama and Nasukawa (2006) used more complicated syntactic patterns that were manually created. The four methods above are based on context information. In contrast, our method exploits the internal structure of the semantic orientations of phrases.

Wilson et al. (2005) worked on phrase-level semantic orientations. They introduced a polarity shifter. They manually created the list of polarity shifters. Inui (2004) also proposed a similar idea.

Takamura et al. (2006) proposed to use based on latent variable models for sentiment classification of noun-adjective pairs. Their model consists of variables respectively representing nouns, adjectives, semantic orientations, and latent clusters, as well as the edges between the nodes. The words that are similar in terms of semantic orientations, such as “risk” and “mortality” (i.e., the positive orientation emerges when they are “low”), make a cluster in their model, which can be an automated version of Inui’s or Wilson et al.’s idea above. However, their method cannot do anything for the words that did not appear in the labeled training data. In this paper, we call their method the latent variable method (LVM).

3 Potts Model

If a variable can have more than two values and there is no ordering relation between the values, the network comprised of such variables is called *Potts* model (Wu, 1982). In this section, we explain the simplified mathematical model of Potts model, which is used for our task in Section 4. The Potts system has been used as a mathematical model in several applications such as image restoration (Tanaka and Morita, 1996) and rumor transmission (Liu et al., 2001).

3.1 Introduction to the Potts Model

Suppose a network consisting of nodes and weighted edges is given. States of nodes are represented by c . The weight between i and j is represented by w_{ij} .

Let $H(\mathbf{c})$ denote an energy function, which indicates a state of the whole network:

$$H(\mathbf{c}) = -\beta \sum_{ij} w_{ij} \delta(c_i, c_j) + \alpha \sum_{i \in L} -\delta(c_i, a_i), \quad (1)$$

where β is a constant called *the inverse-temperature*, L is the set of the indices for the observed variables, a_i is the state of each observed variable indexed by i , and α is a positive constant representing a weight on labeled data. Function δ returns 1 if two arguments are equal to each other, 0 otherwise. The state is penalized if c_i ($i \in L$) is different from a_i . Using $H(\mathbf{c})$, the probability distribution of the network is represented as $P(\mathbf{c}) = \exp\{-H(\mathbf{c})\}/Z$, where Z is a normalization factor.

However, it is computationally difficult to exactly estimate the state of this network. We resort to a

mean-field approximation method that is described by Nishimori (2001). In the method, $P(\mathbf{c})$ is replaced by factorized function $\rho(\mathbf{c}) = \prod_i \rho_i(c_i)$. Then we can obtain the function with the smallest value of the variational free energy:

$$\begin{aligned}
F(\mathbf{c}) &= \sum_{\mathbf{c}} P(\mathbf{c})H(\mathbf{c}) - \sum_{\mathbf{c}} -P(\mathbf{c}) \log P(\mathbf{c}) \\
&= -\alpha \sum_i \sum_{c_i} \rho_i(c_i) \delta(c_i, a_i) \\
&\quad -\beta \sum_{ij} \sum_{c_i, c_j} \rho_i(c_i) \rho_j(c_j) w_{ij} \delta(c_i, c_j) \\
&\quad - \sum_i \sum_{c_i} -\rho_i(c_i) \log \rho_i(c_i). \tag{2}
\end{aligned}$$

By minimizing $F(\mathbf{c})$ under the condition that $\forall i, \sum_{c_i} \rho_i(c_i) = 1$, we obtain the following fixed point equation for $i \in L$:

$$\rho_i(c) = \frac{\exp(\alpha \delta(c, a_i) + \beta \sum_j w_{ij} \rho_j(c))}{\sum_n \exp(\alpha \delta(n, a_i) + \beta \sum_j w_{ij} \rho_j(n))}. \tag{3}$$

The fixed point equation for $i \notin L$ can be obtained by removing $\alpha \delta(c, a_i)$ from above.

This fixed point equation is solved by an iterative computation. In the actual implementation, we represent ρ_i with a linear combination of the discrete Tchebycheff polynomials (Tanaka and Morita, 1996). Details on the Potts model and its computation can be found in the literature (Nishimori, 2001).

After the computation, we obtain the function $\prod_i \rho_i(c_i)$. When the number of classes is 2, the Potts model in this formulation is equivalent to the mean-field Ising model (Nishimori, 2001).

3.2 Relation to Other Models

This Potts model with the mean-field approximation has relation to several other models.

As is often discussed (Mackay, 2003), the minimization of the variational free energy (Equation (2)) is equivalent to the obtaining the factorized model that is most similar to the maximum likelihood model in terms of the Kullback-Leibler divergence.

The second term of Equation (2) is the entropy of the factorized function. Hence the optimization problem to be solved here is a kind of the maximum entropy model with a penalty term, which corresponds to the first term of Equation (2).

We can find a similarity also to the PageRank algorithm (Brin and Page, 1998), which has been applied also to natural language processing tasks (Mihalcea, 2004; Mihalcea, 2005). In the PageRank algorithm, the pagerank score r_i is updated as

$$r_i = (1 - d) + d \sum_j w_{ij} r_j, \tag{4}$$

where d is a constant ($0 \leq d \leq 1$). This update equation consists of the first term corresponding to random jump from an arbitrary node and the second term corresponding to the random walk from the neighboring node.

Let us derive the first order Taylor expansion of Equation (3). We use the equation for $i \notin L$ and denote the denominator by Z_β , for simplicity. Since $\exp x \approx 1 + x$, we obtain

$$\begin{aligned}
\rho_i(c) &= \frac{\exp(\beta \sum_j w_{ij} \rho_j(c))}{Z_\beta} \\
&\approx \frac{1 + \beta \sum_j w_{ij} \rho_j(c)}{Z_\beta} \\
&= \frac{1}{Z_\beta} + \frac{\beta}{Z_\beta} \sum_j w_{ij} \rho_j(c). \tag{5}
\end{aligned}$$

Equation (5) clearly has a quite similar form as Equation (4). Thus, the PageRank algorithm can be regarded as an approximation of our model. Let us clarify the difference between the two algorithms. The PageRank is designed for two-class classification, while the Potts model can be used for an arbitrary number of classes. In this sense, the PageRank is an approximated Ising model. The PageRank is applicable to asymmetric graphs, while the theory used in this paper is based on symmetric graphs.

4 Potts Model for Phrasal Semantic Orientations

In this section, we explain our classification method, which is applicable also to the pairs consisting of an adjective and an unseen noun.

4.1 Construction of Lexical Networks

We construct a lexical network, which Takamura et al. (2005) call the gloss network, by linking two words if one word appears in the gloss of the other word. Each link belongs to one of two groups:

the same-orientation links SL and the different-orientation links DL .

If a negation word (e.g., *nai*, for Japanese) follows a word in the gloss of the other word, the link is a different-orientation link. Otherwise the link is a same-orientation link¹.

We next set weights $W = (w_{ij})$ to links :

$$w_{ij} = \begin{cases} \frac{1}{\sqrt{d(i)d(j)}} & (l_{ij} \in SL) \\ -\frac{1}{\sqrt{d(i)d(j)}} & (l_{ij} \in DL) \\ 0 & otherwise \end{cases}, \quad (6)$$

where l_{ij} denotes the link between word i and word j , and $d(i)$ denotes the degree of word i , which means the number of words linked with word i . Two words without connections are regarded as being connected by a link of weight 0.

4.2 Classification of Phrases

Takamura et al. (2005) used the Ising model to extract semantic orientations of words (not phrases). We extend their idea and use the Potts model to extract semantic orientations of phrasal expressions.

Given an adjective, the decision remaining to be made in classification of phrasal expressions concerns nouns. We therefore estimate the state of the nodes in the lexical network for each adjective. The nouns pairing with the given adjective in the training data are regarded as seed words, which we call seen words, while the words that did not appear in the training data are referred to as unseen words.

We use the mean-field method to estimate the state of the system. If the probability $\rho_i(c)$ of a variable being positive (negative, neutral) is the highest of the three classes, then the word corresponding to the variable is classified as a positive (negative, neutral) word.

We explain the reason why we use the Potts model instead of the Ising model. While only two classes (i.e., positive and negative) can be modeled by the Ising model, three classes (i.e., positive, negative and neutral) can be modelled by the Potts model. For the semantic orientations of words, all the words are sorted in the order of the average orientation value, equivalently the probability of the word being positive. Therefore, even if the neutral class is

¹For English data, a negation should *precede* a word, in order for the corresponding link to be a different-orientation link.

not explicitly incorporated, we can manually determine two thresholds that define respectively the positive/neutral and negative/neutral boundaries. For the semantic orientations of phrasal expressions, however, it is impractical to manually determine the thresholds for each of the numerous adjectives. Therefore, we have to incorporate the neutral class using the Potts model.

For some adjectives, the semantic orientation is constant regardless of the nouns. We need not use the Potts model for those unambiguous adjectives. We thus propose the following two-step classification procedure for a given noun-adjective pair $\langle n, a \rangle$.

1. if the semantic orientation of all the instances with a in L is c , then classify $\langle n, a \rangle$ into c .
2. otherwise, use the Potts model.

We can also construct a probability model for each noun to deal with unseen adjectives. However, we focus on the unseen nouns in this paper, because our dataset has many more nouns than adjectives.

4.3 Hyper-parameter Prediction

The performance of the proposed method largely depends on the value of hyper-parameter β . In order to make the method more practical, we propose a criterion for determining its value.

Takamura et al. (2005) proposed two kinds of criteria. One of the two criteria is an approximated leave-one-out error rate and can be used only when a large labeled dataset is available. The other is a notion from statistical physics, that is, *magnetization*:

$$m = \sum_i \bar{x}_i / N. \quad (7)$$

At a high temperature, variables are randomly oriented (*paramagnetic phase*, $m \approx 0$). At a low temperature, most of the variables have the same direction (*ferromagnetic phase*, $m \neq 0$). It is known that at some intermediate temperature, ferromagnetic phase suddenly changes to paramagnetic phase. This phenomenon is called *phase transition*. Slightly before the phase transition, variables are locally polarized; strongly connected nodes have the same polarity, but not in a global way. Intuitively, the state of the lexical network is locally polarized.

Therefore, they calculate values of m with several different values of β and select the value just before the phase transition.

Since we cannot expect a large labeled dataset to be available for each adjective, we use not the approximated leave-one-out error rate, but the magnetization-like criterion. However, the magnetization above is defined for the Ising model. We therefore consider that the phase transition has occurred, if a certain class c begins to be favored all over the system. In practice, when the maximum of the spatial averages of the approximated probabilities $\max_c \sum_i \rho_i(c)/N$ exceeds a threshold during increasing β , we consider that the phase transition has occurred. We select the value of β slightly before the phase transition.

4.4 Enlarging Seed Word Set

We usually have only a few seed words for a given adjective. Enlarging the set of seed words will increase the classification performance. Therefore, we automatically classify unlabeled pairs by means of an existing method and use the classified instances as seeds.

As an existing classifier, we use LVM. Their model can classify instances that consist of a seen noun and a seen adjective, but are unseen as a pair. Although we could classify and use all the nouns that appeared in the training data (with an adjective which is different from the given one), we do not adopt such an alternative, because it will incorporate even non-collocating pairs such as “green+idea” into seeds, resulting in possible degradation of classification performance. Therefore, we sample unseen pairs consisting of a seen noun and a seen adjective from a corpus, classify the pairs with the latent variable model, and add them to the seed set. The enlarged seed set consists of pairs used in newspaper articles and does not include non-collocating pairs.

5 Experiments

5.1 Dataset

We extracted pairs of a noun (subject) and an adjective (predicate), from Mainichi newspaper articles (1995) written in Japanese, and annotated the pairs with semantic orientation tags : positive, neutral or negative. We thus obtained the labeled dataset

consisting of 12066 pair instances (7416 different pairs). The dataset contains 4459 negative instances, 4252 neutral instances, and 3355 positive instances. The number of distinct nouns is 4770 and the number of distinct adjectives is 384. To check the inter-annotator agreement between two annotators, we calculated κ statistics, which was 0.640². This value is allowable, but not quite high. However, positive-negative disagreement is observed for only 0.7% of the data. In other words, this statistics means that the task of extracting neutral examples, which has hardly been explored, is intrinsically difficult.

We should note that the judgment in annotation depends on which perspective the annotator takes; “high+salary” is positive from employee’s perspective, but negative from employer’s perspective. The annotators are supposed to take a perspective subjectively. Our attempt is to imitate annotator’s decision. To construct a classifier that matches the decision of the average person, we also have to address how to create an average corpus. We do not pursue this issue because it is out of the scope of the paper.

As unlabeled data, we extracted approximately 65,000 pairs for each iteration of the 10-fold cross-validation, from the same news source.

The average number of seed nouns for each ambiguous adjective was respectively 104 in the labeled seed set and 264 in the labeled+unlabeled seed set. Please note that these figures are counted for only ambiguous adjectives. Usually ambiguous adjectives are more frequent than unambiguous adjectives.

5.2 Experimental Settings

We employ 10-fold cross-validation to obtain the averaged classification accuracy. We split the data such that there is no overlapping pair (i.e., any pair in the training data does not appear in the test data).

Hyperparameter α was set to 1000, which is very large since we regard the labels in the seed set is reliable. For the seed words added by the classifier, lower α can be better. Determining a good value for α is regarded as future work.

Hyperparameter β is automatically selected from

²Although Kanayama and Nasukawa (2006) that κ for their dataset similar to ours was 0.83, this value cannot be directly compared with our value because their dataset includes both individual words and pairs of words.

{0.1, 0.2, ..., 2.5} for each adjective and each fold of the cross-validation using the prediction method described in Section 4.3.

5.3 Results

The results of the classification experiments are summarized in Table 1.

The proposed method succeeded in classifying, with approximately 65% in accuracy, those phrases consisting of an ambiguous adjective and an unseen noun, which could not be classified with existing computational models such as LVM.

Incorporation of unlabeled data improves accuracy by 15.5 points for pairs consisting of a seen noun and an ambiguous adjective, and by 3.5 points for pairs consisting of an unseen noun and an ambiguous adjective, approximately. The reason why the former obtained high increase is that pairs with an ambiguous adjective³ are usually frequent and likely to be found in the added unlabeled dataset.

If we regard this classification task as binary classification problems where we are to classify instances into one class or not, we obtain three accuracies: 90.76% for positive, 81.75% for neutral, and 86.85% for negative. This results suggests the identification of neutral instances is relatively difficult.

Next we compare the proposed method with LVM. The latent variable method is applicable only to instance pairs consisting of an adjective and a seen noun. Therefore, we computed the accuracy for 6586 instances using the latent variable method and obtained 80.76%. The corresponding accuracy by our method was 80.93%. This comparison shows that our method is better than or at least comparable to the latent variable method. However, we have to note that this accuracy of the proposed method was computed using the unlabeled data classified by the latent variable method.

5.4 Discussion

There are still 3320 (=12066-8746) word pairs which could not be classified, because there are no entries for those words in the dictionary. However, the main cause of this problem is word segmenta-

³Seen nouns are observed in both the training and the test datasets because they are frequent. Ambiguous adjectives are often-used adjectives such as “large”, “small”, “high”, and “low”.

tion, since many compound nouns and exceedingly-subdivided morphemes are not in dictionaries. An appropriate mapping from the words found in corpus to entries of a dictionary will solve this problem. We found a number of proper nouns, many of which are not in the dictionary. By estimating a class of a proper noun and finding the words that matches the class in the dictionary, we can predict the semantic orientations of the proper noun based on the orientations of the found words.

In order to see the overall tendency of errors, we calculated the confusion matrices both for pairs of an ambiguous adjective and a seen noun, and for pairs of an ambiguous adjective and an unseen noun (Table 2). The proposed method works quite well for positive/negative classification, though it finds still some difficulty in correctly classifying neutral instances even after enhanced with the unlabeled data.

In order to qualitatively evaluate the method, we list several word pairs below. These word pairs are classified by the Potts model with the labeled+unlabeled seed set. All nouns are unseen; they did not appear in the original training dataset. Please note again that the actual data is Japanese.

positive instances	
noun	adjective
cost	low
basic price	low
loss	little
intelligence	high
educational background	high
contagion	not-happening
version	new
cafe	many
salary	high
commission	low

negative instances	
noun	adjective
damage	heavy
chance	little
terrorist	many
trouble	many
variation	little
capacity	small
salary	low
disaster	many
disappointment	big
knowledge	little

For example, although both “salary” and “commission” are kinds of money, our method captures

Table 1: Classification accuracies (%) for various seed sets and test datasets. ‘Labeled’ seed set corresponds to the set of manually labeled pairs. ‘Labeled+unlabeled’ seed set corresponds to the union of ‘labeled’ seed set and the set of pairs labeled by LVM. ‘Seen nouns’ for test are the nouns that appeared in the training data, while ‘unseen nouns’ are the nouns that did not appear in the training dataset’. Please note that seen pairs are excluded from the test data. ‘Unambiguous’ adjectives corresponds to the pairs with an adjective which has a unique orientation in the original training dataset, while ‘ambiguous’ adjectives corresponds to the pairs with an adjective which has more than one orientation in the original training dataset.

seed \ test	seen nouns		unseen nouns		total
labeled	68.24 (4494/6586)		73.70 (1592/2160)		69.59 (6086/8746)
	unambiguous	ambiguous	unambiguous	ambiguous	
	98.15 (1166/1188)	61.65 (3328/5398)	94.85 (736/776)	61.85 (856/1384)	
labeled+unlabeled	80.93 (5330/6586)		75.88 (1639/2160)		79.68 (6969/8746)
	unambiguous	ambiguous	unambiguous	ambiguous	
	98.15 (1166/1188)	77.14 (4164/5398)	94.85 (736/776)	65.25 (903/1384)	

Table 2: Confusion matrices of classification result with labeled+unlabeled seed set

		Potts model							
		seen nouns				unseen nouns			
		positive	neutral	negative	sum	positive	neutral	negative	sum
Gold standard	positive	964	254	60	1278	126	84	30	240
	neutral	198	1656	286	2140	60	427	104	591
	negative	39	397	1544	1980	46	157	350	553
	sum	1201	2307	1890	5398	232	668	484	1384

the difference between them; “high salary” is positive, while “low (cheap) commission” is also positive.

6 Conclusion

We proposed a method for extracting semantic orientations of phrases (pairs of an adjective and a noun). For each adjective, we constructed a Potts system, which is actually a lexical network extracted from glosses in a dictionary. We empirically showed that the proposed method works well in terms of classification accuracy.

Future work includes the following:

- We assumed that each word has a semantic orientation. However, word senses and subjectivity have strong interaction (Wiebe and Mihalcea, 2006).
- The value of α must be properly set, because lower α can be better for the seed words added by the classifier,
- To address word-segmentation problem discussed in Section 5.3, we can utilize the fact that the heads of compound nouns often inherit the property determining the semantic orientation when combined with an adjective.
- The semantic orientations of pairs consisting of a proper noun will be estimated from the named entity classes of the proper nouns such as person name and organization.

References

- Faye Baron and Graeme Hirst. 2004. Collocations as cues to semantic orientation. In *AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*.
- Sergey Brin and Lawrence Page. 1998. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117.
- Andrea Esuli and Fabrizio Sebastiani. 2005. Determining the semantic orientation of terms through gloss analysis. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management (CIKM'05)*, pages 617–624.
- Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the Association for Computational Linguistics*, pages 174–181.
- Takashi Inui. 2004. *Acquiring Causal Knowledge from Text Using Connective Markers*. Ph.D. thesis, Graduate School of Information Science, Nara Institute of Science and Technology.
- Jaap Kamps, Maarten Marx, Robert J. Mokken, and Maarten de Rijke. 2004. Using wordnet to measure semantic orientation of adjectives. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04), volume IV*, pages 1115–1118.
- Hiroshi Kanayama and Tetsuya Nasukawa. 2006. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'06)*, pages 355–363.
- Nozomi Kobayashi, Takashi Inui, and Kentaro Inui. 2001. Dictionary-based acquisition of the lexical knowledge for p/n analysis (in Japanese). In *Proceedings of Japanese Society for Artificial Intelligence, SLUD-33*, pages 45–50.
- Zhongzhu Liu, Jun Luo, and Chenggang Shao. 2001. Potts model for exaggeration of a simple rumor transmitted by recreant rumormongers. *Physical Review E*, 64:046134,1–046134,9.
- David J. C. Mackay. 2003. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press.
- Mainichi. 1995. Mainichi Shimibun CD-ROM version.
- Rada Mihalcea. 2004. Graph-based ranking algorithms for sentence extraction, applied to text summarization. In *The Companion Volume to the Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, (ACL'04)*, pages 170–173.
- Rada Mihalcea. 2005. Unsupervised large-vocabulary word sense disambiguation with graph-based algorithms for sequence data labeling. In *Proceedings of the Joint Conference on Human Language Technology / Empirical Methods in Natural Language Processing (HLT/EMNLP)*, pages 411–418.
- Hidetoshi Nishimori. 2001. *Statistical Physics of Spin Glasses and Information Processing*. Oxford University Press.
- Frank Z. Smadja. 1993. Retrieving collocations from text: Xtract. *Computational Linguistics*, 19(1):143–177.
- Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2005. Extracting semantic orientations of words using spin model. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 133–140.
- Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2006. Latent variable models for semantic orientations of phrases. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL'06)*.
- Kazuyuki Tanaka and Tohru Morita. 1996. Application of cluster variation method to image restoration problem. In *Theory and Applications of the Cluster Variation and Path Probability Methods*, pages 353–373. Plenum Press, New York.
- Peter D. Turney and Michael L. Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems*, 21(4):315–346.
- Peter D. Turney. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)*, pages 417–424.
- Janyce M. Wiebe and Rada Mihalcea. 2006. Word sense and subjectivity. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics (COLING-ACL'06)*, pages 1065–1072.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of joint conference on Human Language Technology / Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP'05)*, pages 347–354.
- Fa-Yueh Wu. 1982. The potts model. *Reviews of Modern Physics*, 54(1):235–268.