

Integrating Collaborative Filtering and Sentiment Analysis: A Rating Inference Approach

Cane Wing-ki Leung and Stephen Chi-fai Chan and Fu-lai Chung¹

Abstract. We describe a rating inference approach to incorporating textual user reviews into collaborative filtering (CF) algorithms. The main idea of our approach is to elicit user preferences expressed in textual reviews, a problem known as *sentiment analysis*, and map such preferences onto some rating scales that can be understood by existing CF algorithms. One important task in our rating inference framework is the determination of sentimental orientations (SO) and strengths of opinion words. It is because inferring a rating from a review is mainly done by extracting opinion words in the review, and then aggregating the SO of such words to determine the dominant or average sentiment implied by the user. We performed some preliminary analysis on movie reviews to investigate how SO and strengths of opinion words can be determined, and proposed a relative-frequency-based method for performing such task. The proposed method addresses a major limitation of existing methods by allowing similar words to have different SO. We also developed and evaluated a prototype of the proposed framework. Preliminary results validated the effectiveness of various tasks in the proposed framework, and suggest that the framework does not rely on a large training corpus to function. Further development of our rating inference framework is ongoing. A comprehensive evaluation of the framework will be carried out and reported in a follow-up article.

1 INTRODUCTION

Collaborative Filtering (CF) is a promising technique in recommender systems. It provides personalized recommendations to users based on a database of user preferences, from which users having similar tastes are identified. It then recommends to a target user items liked by other, similar users [5, 10].

CF-based recommender systems can be classified into two major types depending on how they collect user preferences: user-log based and ratings based. User-log based CF obtains user preferences from implicit votes captured through users' interactions with the system (e.g. purchase histories as in Amazon.com [12]). Ratings based CF makes use of explicit ratings users have given items (e.g. 5-star rating scale as in MovieLens [6]). Such ratings are usually in or can easily be transformed into numerical values (e.g. A to E).

Some review hubs, such as the Internet Movie Database (IMDb), allow users to provide comments in free text format, referred to as user reviews in this paper. User reviews can also be considered a type of "user ratings", although they are usually natural language texts rather than numerical values. While research on mining user preferences from reviews, a problem known as *sentiment analysis*

or *sentiment classification* (e.g. [18, 23, 7, 17]), is becoming increasingly popular in the text mining literature, its integration with CF has only received little research attention. The PHOAKS system described in [21] classifies web sites recommended by users in newsgroup messages, but it does not involve mining user preferences from texts.

This paper describes our proposed framework for integrating sentiment analysis and CF. We take a rating inference approach which infers numerical ratings from textual reviews, so that user preferences represented in the reviews can easily be fed into existing CF algorithms. The contributions of this approach are two-fold. Firstly, it addresses the well-known data sparseness problem in CF by allowing CF algorithms to use textual reviews as an additional source of user preferences. Secondly, it enables extending CF to domains where numerical ratings on products are difficult to collect, or where preferences on domain items are too complex to be expressed as scalar ratings. An example of such domains is travel and tourism, in which most existing recommender systems are built upon content- or knowledge-based techniques [20]. "Reviews" written by travelers on tourism products or destinations, for instance, are available as travel journals on TravelJournals.net [22]. Eliciting travelers' preferences from their travel journals may contribute to the development of more advanced and personalized recommender systems.

We use an example to introduce the idea of rating inference and the key tasks it involves. The paragraph below is extracted from a movie review on IMDb:

This movie is quite good, but not as good as I would have thought.....It is quite boring.....I just felt that they had nothing to tell.....However, the movie is not all bad.....the acting is brilliant, especially Massimo Troisi.

In the above paragraph, the user (author) stated that the movie being reviewed is "quite good" and "not all bad", and the acting of Massimo Troisi is "brilliant". The user, however, also thought that the movie is "is quite boring" and "had nothing to tell". Given all these positive and negative opinions, rating inference is about determining the *overall sentiment* implied by the user, and map such sentiment onto some fine-grained rating scale (the user-specified rating of the above review was 5/10). This involves a few tasks:

1. User reviews are unstructured, natural language texts. Interesting information, such as opinion words, have to be extracted from the reviews for further analysis. This is usually done with the aid of various natural language processing (NLP) techniques.
2. The sentimental orientations (SO) of the expressed opinions must be identified [4]. This task is usually based upon a database of opinion words with their predicted SO. This is a key step to the correct elicitation of users preferences.

¹ Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, HKSAR. Email: {cswkleung, csschan, cskchung}@comp.polyu.edu.hk.

3. The strengths of the SO of the opinion words must also be determined [4]. For instance, both “excellent” and “good” represent positive sentiments, but we know that the sentiment implied by “excellent” is much stronger.
4. An overall rating can then be inferred from a review. This can be done by aggregating or averaging the SO of the opinions words it contains to determine the dominant or average sentiment.

The rest of this paper is organized as follows. Section 2 describes related work on sentiment analysis. Section 3 outlines the proposed framework for inferring ratings from user reviews. Section 4 discusses preliminary results related to the modeling of SO and strengths of opinion words, and finally Section 5 concludes the paper by outlining our ongoing and future work.

2 RELATED WORK

Sentiment analysis aims at classifying user reviews according to their sentiments [18, 23, 3, 16, 8, 4]. This section describes a few sentiment analysis algorithms that are related to our work.

Turney [23] described the PMI-IR algorithm which computes SO as the mutual information between phrases in user reviews and two seed adjectives, “excellent” and “poor”. A phrase is considered positive (resp. negative) if it is strongly associated with the word “excellent” (resp. “poor”), and the overall SO of a review is determined by the average SO obtained by the phrases it contains. The PMI-IR algorithm was tested on a few domains, and it was found that movie reviews are more difficult to classify than the others. This motivates some later work, including ours, to use the movie domain for testing sentiment classifiers. The PMI-IR algorithm is capable of determining the overall SO of user reviews, which is important for rating inference, but is computationally very expensive.

Pang et al. [18] investigated whether it is appropriate to perform sentiment analysis by using standard topic-based classification techniques. They concluded that sentiment analysis is a more difficult task than topic-based classification, and that some discourse analysis of reviews might improve performance. Our proposed framework attempts to address such issue by tagging features that may be important for determining the overall sentiments of reviews.

Hu and Liu [7, 8] presented an interesting approach to predicting the SO of opinion words. Their approach first defines a small set of seed adjectives having known sentiments, and automatically grows the set using the synonym and acronym sets in WordNet [14], *assuming that similar meanings imply similar sentiments*. The major shortcoming of their approach with regard to rating inference is that the strengths of opinion words cannot be determined. Our approach addresses this by determining SO of opinions using a relative-frequency-based method as described in Section 3.3.

The algorithms introduced above classify sentiments as either positive or negative. The rating inference problem has been discussed recently in Pang and Lee [17], which proposed to classify reviews into finer-grained rating scales using a multi-class text classifier. A classifier assigns similar labels to similar items based on some similarity functions. While term overlapping is a commonly used similarity function, it does not seem effective in distinguishing documents having different sentiments [17]. They therefore proposed an alternative item similarity function, known as positive-sentence percentage (PSP), defined as the number of positive sentences divided by the number of subjective sentences in a review. We found that the statistical distributions of opinion words in user reviews may also be useful for rating inference as discussed in Section 4.2.

We distinguish two possible approaches to rating inference based on the related work. The first approach addresses rating inference as a classification problem as proposed in Pang and Lee [17]. The second approach is a simple “score assignment” approach similar to Turney’s work [23], although such work only classifies reviews as *Recommended* or *Not Recommended*. Our framework takes the score assignment approach to investigate whether such approach is effective for rating inference.

3 THE PROPOSED FRAMEWORK

The proposed framework consists of two components. The first component is responsible for analyzing user reviews and inferring ratings from them, while the second one is a collaborative filter that generates item recommendations based on the ratings inferred. Fig. 1 depicts an overview of the proposed framework.

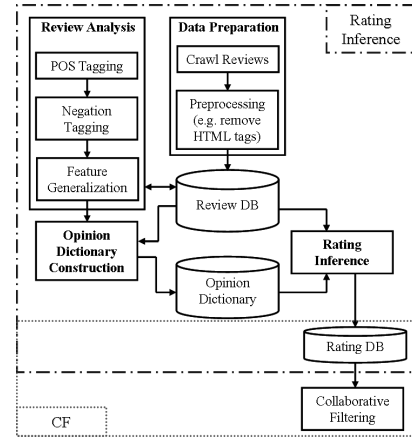


Figure 1. Overview of the proposed framework

The following subsections provide further details about the rating inference component. It includes four major steps (highlighted in boldface in Fig. 1), namely data preparation, review analysis, opinion dictionary construction and rating inference. Our discussions only focus on the rating inference component because, as noted, ratings inferred from user reviews can be fed into existing CF algorithms.

3.1 Data Preparation

Data preparation involves collecting and preprocessing user reviews for the subsequent analysis. Different preprocessing steps may be required depending on the data sources. For example, if user reviews are downloaded as HTML pages, the HTML tags and non-textual contents they contain are removed in this step.

A user review is likely to be a semistructured document, containing some structured headers and an unstructured text body. A movie review on IMDb, for example, contains structured headers including a user (author) identity and a one-line summary, and unstructured blocks of text, which are the user’s comments on the movie being reviewed, written in natural language.

Sentiment analysis algorithms usually do not use information other than the comments and the original ratings given by the users (e.g. for performance evaluation), if any. Our framework, however, extracts also the identities of users and the subject matters being reviewed because they are useful for performing CF, and such

information are retained to facilitate our future work. Since we focus on rating inference in this paper, the term “reviews” hereafter refers to the comments given by the users on the relevant subject matters.

3.2 Review Analysis

The review analysis step includes several tasks that help identifying interesting information in reviews, which are unstructured, natural language texts. Some essential tasks include:

- **Part-of-speech (POS) tagging.** POS tagging is an important task in our framework. As discussed in related studies, product features in a user review are usually nouns or noun phrases, while user opinions are usually adjectives or verbs (e.g. [3, 7]). POS tagging therefore helps extracting such information from reviews. There exist a variety of POS taggers in the NLP literature, and we adopted a tool known as MontyLingua [13]. It was developed based on the most well-known POS tagger by Brill [1] but with improved tagging accuracy (around 97%). Note that this task is language dependent, and our work only deals with English reviews at the moment.
- **Negation tagging.** Some words have negation effects on other words, and negation tagging aims at identifying such words and reflecting their effects when determining the SO of reviews [2, 18, 3]. For example, “good” and “not good” obviously represent opposite sentiments. Negation tagging identifies the existence of the word “not”, and adds the tag “NOT_” to other words in the same sentence based on some linguistic rules. We used fuzzy string matching (using regular expressions) when identifying negation words to handle word variants. For example, “cannot”, “can’t” and “cant” are considered the same term.
- **Feature generalization.** Feature generalization, or metadata substitution [3], is about generalizing features that may be overly specific. This task can be performed when attributes of domain items are available. For the movie reviews domain, for example, a sentence “*Toy Story is pleasant and fun.*”, in which “Toy Story” is the name of the movie being reviewed, is generalized to “*MOVIE is pleasant and fun.*”. This facilitates rating inference which involves assigning weights to product features as discussed in Section 3.4.

3.3 Opinion Dictionary Construction

Opinion dictionary construction is an important step in the proposed framework as the overall SO of a review is computed based on those of the opinion words it contains. An opinion dictionary contains opinion words, their estimated SO and the strengths of their SO.

Determining the SO and strengths of opinion words is done by answering the question: “Given a certain opinion word, *how likely* is it to be a positive sentiment, and *how likely* is it to be a negative one?”. Some related studies, such as [7, 9], adopted *word-similarity-based* methods to do this. Such methods assume that similar meanings imply similar sentiments, thus the SO of a word is the same as that of its synonyms and is opposite to its antonyms.

We found that this assumption is not necessarily true based on our analysis on movie reviews as discussed in Section 4.2. We use an example to illustrate this. The terms “terrible” and “frightening” are synonyms in WordNet [14], and both terms seem to be negative sentiments. However, we found that “terrible” appeared in negative reviews 75% of the time, whereas “frightening” appeared in negative reviews only 29% of the time (consider the case where *a horror*

movie is frightening!). This reveals that similar meanings may not imply similar sentiments.

We propose to construct an opinion dictionary using a *relative-frequency-based* method. More specifically, this method estimates the strength of a word with respect to a certain sentiment class as the relative frequency of its occurrence in that class:

$$OS(a_n, c) = \frac{F(a_n, c)}{\sum_{c_i \in C} F(a_n, c_i)} \quad (1)$$

In Eq. (1), $OS(a_n, c)$ denotes the strength of an opinion word a_n with respect to a particular sentiment class c . For instance, $OS(\text{“good”}, \text{Positive})$ denotes the strength of the word “good” with respect to the sentiment class *Positive*. c and c_i are elements in C , which is the set of sentiment classes that are used for computing relative frequencies of opinion words (e.g. $C = \{\text{Positive}, \text{Negative}\}$). $F(a_n, c)$ and $F(a_n, c_i)$ denote the frequency count of a_n in reviews in c and c_i respectively.

3.4 Rating Inference

A review usually contains a mixture of positive and negative opinions towards different features of a product, and rating inference aims at determining the *overall* sentiment implied by the user. We attempted to perform such task by aggregating the strengths of the opinion words in a review with respect to different sentiment classes, and then assigning an overall rating to the review to reflect the dominant sentiment class.

Our score assignment approach to rating inference enables assigning weights to different opinion words according to their estimated importance. Such importance may be determined by, for example, the positions in the reviews where the opinion words appeared. In [18, 15], for instance, opinions appeared in the first quarter, the middle half and the last quarter of a given review are added the *position tags* “F_”, “M_” and “L_” respectively. It was, however, found that position information do not improve performance.

The product features on which opinions are expressed may also be useful for determining weights of opinions, and this is facilitated by the feature generalization task described in Section 3.2. Generally speaking, opinions towards a product as a whole may be more useful for determining the SO of a review. This also allows easy integration with user-specified interest profiles if necessary (e.g. to address the new user problem in CF [19]). For example, if a certain user of a movie recommender system is particularly interested in a certain actor, then the acting of that actor in a movie may have stronger influence on the his overall sentiment towards the movie.

4 PRELIMINARY EXPERIMENTS

This section describes the dataset we used, and then reports our results on two groups of experiments. As noted, we performed an analysis on movie reviews to assist the opinion strengths determination task. Our observations are discussed in Section 4.2. We also developed and evaluated a prototype of the proposed framework. Some preliminary results are reported in Section 4.3.

4.1 Dataset

We collected movie reviews from IMDb for the movies in the MovieLens 100k dataset, courtesy of GroupLens Research [10]. The MovieLens dataset contains user ratings on 1692 movies. We

removed movies that are duplicated or unidentifiable (movies without names), and crawled the IMDb to download user reviews for the remaining movies. We filtered out contributions from users who have provided fewer than 10 reviews and reviews without user-specified ratings, which will later be used for evaluating our proposed framework. The resulting dataset contains approximately 30k reviews on 1477 movies, provided by 1065 users.

Each review contains a number of headers and a text body. The headers include *movie ID*, *user ID*, *review date*, *summary*, which is a one-line summary in natural language text written by the user, and a *rating*, which is a user-specified number ranging from 1 (awful) to 10 (excellent). The text body is the user’s comments on the movie.

4.2 Analysis on the Use of Opinion Words

Determining opinion strengths would be an easy task if a certain opinion word *always* appears in reviews with a certain rating, for example, if the word “brilliant” always appears in reviews rated as 10/10. This is, however, not likely to be true. A review may contain both positive and negative opinions. This means a movie receiving a high rating may also have some bad features, and vice versa.

We performed some preliminary experiments to analyze the use of opinion words in user reviews. By doing so, we hope to discover interesting usage patterns of opinion words that can help determining opinion strengths. We first performed the tasks described in Sections 3.1 and 3.2 on the dataset. We then randomly sampled three training sets, namely *T10*, *T5* and *T1*, each containing 500 reviews whose user-specified ratings were 10/10, 5/10 and 1/10 respectively. These ratings were chosen as they seem to be appropriate representative cases for *Positive*, *Neutral* and *Negative* sentiments. We used a program to extract opinion words, which are words tagged as adjectives [7], and compute their frequency counts in each of the training sets. Some frequent opinion words were further analyzed.

The number of distinct opinion words appeared in the training sets is 4545, among which 839 (around 18.5%) appeared in two of the three training sets, and 738 (around 16.2%) appeared in all three. We further examined opinion words that appeared in more than one training set. Table 1 lists, due to space constraint, the 10 most frequent opinion words (top 10) of this kind in each training set in descending order of their frequency counts. In the table, the number in brackets following an opinion word is its relative frequency in the particular training set. Boldface is used to highlight words having the highest relative frequency among the three training sets.

Table 1. Top 10 opinion words with relative frequencies.

Training set	Opinion words with relative frequencies
T10	best (0.68) , great (0.66) , good (0.33), many (0.47) , first (0.38) , classic (0.71) , better (0.30), favorite (0.75) , perfect (0.75) , greatest (0.85)
T5	good (0.39) , more (0.54) , much (0.51) , bad (0.35), better (0.41) , other (0.32), few (0.73) , great (0.21), first (0.34), best (0.19)
T1	bad (0.65) , good (0.28), worst (0.89) , much (0.49), more (0.46), other (0.28), first (0.28), better (0.29), many (0.24), great (0.14)

Our observations are summarized as follows. Firstly, the relative frequencies of positive opinion words are usually, but not always, the highest in T10 and the lowest in T1, and vice versa for negative

opinion words. Table 2 lists as examples the relative frequencies of the most frequent opinion word (top 1) in each training set. Boldface is used to highlight the highest relative frequency of each opinion word. Such observation suggests that relative frequencies of opinion words may help determining their SO and strengths. For example, the word “best” appeared in T10 68% of the time. It may therefore be considered a positive opinion word with the strength 0.68.

Table 2. Top 1 opinion words with relative frequencies.

Opinion word	Understood SO (strength)	Relative frequencies in:		
		T10	T5	T1
best	positive (strong)	0.68	0.19	0.13
good	positive (mild)	0.33	0.39	0.28
bad	negative (strong)	0	0.35	0.65

Secondly, almost 35% of all opinion words, including those having clear and strong understood SO (e.g. “best”), appeared in more than one training set. We model this fact by adopting the fuzzy set concept [24], which means that an attribute can be a member of some fuzzy sets to certain membership degrees in the range [0,1], determined by some membership functions. In the context of our work, the “membership degree” of a word with respect to a sentiment class is determined by the relative frequency of the word in the corresponding training set. For instance, the word “best” has SO *Positive*, *Neutral* and *Negative* with strengths 0.68, 0.19 and 0.13 respectively. The use of fuzzy sets to model *user ratings* in CF has recently been proposed in [11], but our work deals with a different problem as we adopt the fuzzy set concept to model SO and opinion strengths.

Thirdly, the SO of opinion words determined by the relative-frequency-based method may not agree with their generally understood SO. An example is the word “frightening” which seems to be a negative sentiment. Its relative frequency in T1, however, is only 0.29. Based on this observation, we further conclude that synonyms do not necessarily have the same SO. For instance, “terrible” is a synonym of “frightening” in WordNet, but its relative frequency in T1 is 0.75. Recall that word-similarity-based methods make use of a set of seed adjectives and the similarities between word meanings to determine SO of opinion words [7, 9]. Our analysis, however, indicates that *similar meanings may not imply similar sentiments*. This suggests that our relative-frequency-based method overcomes a major limitation of the word-similarity-based methods, because it allows similar words to have different SO.

To sum up, we propose a relative-frequency-based method for determining the SO and strengths of opinion words. This method overcomes a major limitation of word-similarity-based methods, and allows opinion words to have multiple SO.

4.3 Evaluation of the Proposed Framework

We developed and evaluated a prototype of the proposed framework. The evaluation aims at investigating the effects of the various key tasks in the framework, and exploring future research directions towards more accurate rating inference. We measured the accuracies produced by the proposed framework in rating reviews with respect to a 2-point and a 3-point scale. The 2-point scale was chosen to facilitate our future work on comparing our work with related work, and the 3-point scale was used as it was suggested in [17] that

human raters would do well in classifying reviews into 3 classes. The original ratings in the dataset were transformed from a 10-point scale into 2-point and 3-point scales to facilitate our experiments.

Only major observations from the experiments are summarized in this paper due to space constraint. Firstly, our hypothesis that opinions towards a product as a whole are more important for determining the overall SO of reviews is proven correct. Double-weighting features representing movies produces higher accuracies. Secondly, addressing the contextual effects of negation words using negation tagging improves accuracies slightly. This suggests that the task is useful, but we shall consider more advanced opinion extraction techniques in our future work. Thirdly, excluding opinion words having weak or ambiguous SO improves accuracies. More specifically, a rating of a review is determined by the strengths of the SO of the opinion words it contains. When computing the strengths of opinion words with respect to a certain SO, considering only opinion words having strengths above a certain threshold resulted in higher accuracies than using all opinion words. Lastly, our framework does not rely on a large training corpus to function, and its performance improves as more reviews become available. We performed a group of experiments to rate reviews based on opinion dictionaries built using different sizes of training set. When the size of training set reduced greatly from 1000 to 200, accuracies achieved using the 2-point and the 3-point scales dropped only by 2% and 1.5% respectively. Such finding is encouraging because it suggests that our framework can be applied to domains where only a small number of labeled reviews (reviews with ratings) are available.

5 CONCLUSIONS AND FUTURE WORK

We propose a rating inference approach to integrating sentiment analysis and CF. Such approach transforms user preferences expressed as unstructured, natural language texts into numerical scales that can be understood by existing CF algorithms.

This paper provides an overview of our proposed rating inference framework, including the steps it involves and the key tasks and design issues in each step. It also discusses our preliminary analysis on the use of opinion words. The purpose of such analysis was to investigate how opinion strengths can be determined and modeled in our proposed framework. We conclude that opinion words can have multiple SO and strengths, and propose a relative-frequency-based method to determine SO and strengths of opinion words. This paper also outlines preliminary results of an evaluation of the proposed framework. Further development of the framework is still ongoing. More detailed descriptions about the framework and comprehensive results will be reported in a follow-up article.

As noted, our rating inference approach transforms textual reviews into ratings to enable easy integration of sentiment analysis and CF. We nonetheless recognize the possibility to perform text-based CF directly from a collection of user reviews. A possible solution is to model text-based CF as an information retrieval (IR) problem, having reviews written by a target user as the “query” and those written by other similar users as the “relevant documents”, from which recommendations for the target user can be generated. This remains as an interesting research direction for future work.

ACKNOWLEDGEMENTS

The first author would like to thank Sau-ching Tse and Ada Chan for interesting discussions on this work. This work is partially supported by the Hong Kong Polytechnic University Research Grant A-PE35.

REFERENCES

- [1] E. Brill, ‘A simple rule-based part-of-speech tagger’, in *Proceedings of the 3rd Conference on Applied Natural Language Processing*, pp. 152–155, (1992).
- [2] S. Das and M. Chen, ‘Yahoo! for Amazon: Extracting market sentiment from stock message boards.’, in *Proceedings of the Asia Pacific Finance Association Annual Conference*, (2001).
- [3] K. Dave, S. Lawrence, and D. M. Pennock, ‘Mining the peanut gallery: Opinion extraction and semantic classification of product reviews’, in *Proceedings of the 12th International World Wide Web Conference*, pp. 519–528, (2003).
- [4] A. Esuli and F. Sebastiani, ‘Determining the semantic orientation of terms through gloss classification.’, in *Proceedings of the ACM International Conference on Information and Knowledge Management*, pp. 617–624, (2005).
- [5] D. Goldberg, D. Nichols, B. Oki, and D. Terry, ‘Using collaborative filtering to weave an information tapestry’, *Communications of the ACM*, **35**(12), 61–70, (1992).
- [6] GroupLens. MovieLens dataset. <http://www.grouplens.org/>.
- [7] M. Hu and B. Liu, ‘Mining and summarizing customer reviews’, in *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 168–177, (2004).
- [8] M. Hu and B. Liu, ‘Mining opinion features in customer reviews’, in *Proceedings of the 19th National Conference on Artificial Intelligence*, pp. 755–760, (2004).
- [9] J. Kamps, M. Marx, R.J. Mokken, and M. de Rijke, ‘Using Wordnet to measure semantic orientations of adjectives’, in *Proceedings of the 4th International Conference on Language Resources and Evaluation*, pp. 1115–1118, (2004).
- [10] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl, ‘Grouplens: Applying collaborative filtering to usenet news’, *Communications of the ACM*, **40**(3), 77–87, (1997).
- [11] C. W. K. Leung, S. C. F. Chan, and F. L. Chung, ‘A collaborative filtering framework based on fuzzy association rules and multiple-level similarity’, *Knowledge and Information Systems*, (forthcoming).
- [12] G. Linden, B. Smith, and J. York, ‘Amazon.com recommendations: Item-to-item collaborative filtering’, *IEEE Internet Computing*, **7**(1), 76–80, (2003).
- [13] H. Liu. MontyLingua: An end-to-end natural language processor with common sense, 2004. <http://web.media.mit.edu/~hugo/montylingua>.
- [14] G. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller, ‘Introduction to Wordnet: An online lexical database’, *International Journal of Lexicography (Special Issue)*, **3**(4), 235–312, (1990).
- [15] R. Mukras, *A comparison of machine learning techniques applied to sentiment classification*, Master’s thesis, University of Sussex, Brighton, UK., 2004.
- [16] T. Nasukawa and J. Yi, ‘Sentiment analysis: Capturing favorability using natural language processing’, in *Proceedings of the 2nd International Conference on Knowledge Capture*, pp. 70–77, (2003).
- [17] B. Pang and L. Lee, ‘Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales’, in *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pp. 115–124, (2005).
- [18] B. Pang, L. Lee, and S. Vaithyanathan, ‘Thumbs up? Sentiment classification using machine learning techniques’, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 79–86, (2002).
- [19] A.M. Rashid, I. Albert, D. Cosley, S.K. Lam, S. McNee, J.A. Konstan, and J. Riedl, ‘Getting to know you: Learning new user preferences in recommender systems’, in *Proceedings of the 2002 International Conference on Intelligent User Interfaces*, pp. 127–134, (2002).
- [20] F. Ricci, ‘Travel recommender systems’, *IEEE Intelligent Systems*, **17**(6), 55–57, (2002).
- [21] L. Terveen, W. Hill, B. Amento, D. McDonald, and J. Creter, ‘Phoaks: A system for sharing recommendations’, *Communications of the ACM*, **40**(3), 59–62, (1997).
- [22] Traveljournals.net. <http://www.traveljournals.net>.
- [23] P.D. Turney, ‘Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews’, in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 417–424, (2002).
- [24] L. A. Zadeh, ‘Knowledge representation in fuzzy logic’, *IEEE Transactions on Knowledge and Data Engineering*, **1**(1), 89–100, (1989).