A Review Selection Approach for Accurate Feature Rating Estimation

Chong Long[†] **Jie Zhang**[‡] **Xiaoyan Zhu**[†§]
[†] State Key Laboratory on Intelligent Technology and Systems,
Tsinghua National Laboratory for Information Science and Technology,
Department of Computer Science, Tsinghua University
[‡]School of Computer Engineering, Nanyang Technological University
[§]{Corresponding Author: zxy-dcs@tsinghua.edu.cn}

Abstract

In this paper, we propose a review selection approach towards accurate estimation of feature ratings for services on participatory websites where users write textual reviews for these services. Our approach selects reviews that comprehensively talk about a feature of a service by using information distance of the reviews on the feature. The rating estimation of the feature for these selected reviews using machine learning techniques provides more accurate results than that for other reviews. The average of these estimated feature ratings also better represents an accurate overall rating for the feature of the service, which provides useful feedback for other users to choose their satisfactory services.

1 Introduction

Most of participatory websites such as Amazon (amazon.com) do not collect from users feature[1] ratings for services, simply because it may cost users too much effort to provide detailed feature ratings. Even for a very few websites that do collect feature ratings such as a popular travel website TripAdvisor (tripadvisor.com), a big portion (approximately 43%) of users may still not provide them. However, feature ratings are useful for users to make informed consumption decisions especially in the case where users may be interested more in some particular features of the services. Machine learning techniques have been proposed for sentiment classification (Pang et al., 2002; Mullen and Collier, 2004) based on annotated samples from experts, but they have limited performance especially when estimating ratings of a multi-point scale (Pang and Lee, 2005).

In this paper, we propose a novel review selection approach for accurate feature rating estimation. More specifically, our approach selects reviews written by the users who comprehensively talk about a certain feature of a service - that are comprehensive on this feature, using information distance of reviews on the feature based on Kolmogorov complexity (Li and Vitányi, 1997). This feature is obviously important to the users. People tend to be more knowledgable in the aspects they consider important. These users therefore represent a subset of experts. Statistical analysis reveals that these expert users are more likely to agree on a common rating for the feature of the service. The rating estimation of the feature for these selected reviews based on annotated samples from experts using machine learning techniques is thus able to provide more accurate results than that for other reviews. This statistical evidence also allows us to use the average of the estimated feature ratings to better represent an overall opinion of experts for the feature of the service, which will be particularly useful for assisting other users to correctly make their consumption decisions.

We verify our approach and arguments based on real data collected from the TripAdvisor website. First, our approach is shown to be able to effectively select reviews that comprehensively talk about features of a service. We then adopt the machine learning method proposed in (Pang and Lee, 2005) and the Bayesian Network classifier (Russell and Norvig, 2002) for feature rating estimation. Our experimental results show that the accuracy of estimating feature ratings for these selected reviews is higher than that for other reviews, for both the machine learning

[1]A feature broadly means an attribute or a function of a service.

Coling 2010: Poster Volume, pages 766–774,
Beijing, August 2010

methods. And, the average of these estimated ratings is testified to closely represent the overall feature rating of the service. Our approach is therefore verified to be a successful step towards accurate feature rating estimation.

2 Related Work

Our work aims at estimating feature ratings of a service based on its textual reviews. It is related to sentiment classification. The task of sentiment classification is to determine the semantic orientations of words, sentences or documents. (Pang et al., 2002) is the earliest work of automatic sentiment classification at document level, using several machine learning approaches with common textual features to classify movie reviews. Mullen and Collier (Mullen and Collier, 2004) integrated PMI values, Osgood semantic factors and some syntactic relations into the features of SVM. Pang and Lee (Pang and Lee, 2004) proposed another machine learning method based on subjectivity detection and minimum-cut in graph. However, these approaches focus only on binary classification of reviews.

In 2005, Pang and Lee extended their earlier work in (Pang and Lee, 2004) to determine a reviewer's evaluation with respect to multi-scales (Pang and Lee, 2005). The rating estimation is viewed as multi-class sentiment categorization on documents. They used SVM regression as the multi-class classifier, and also applied a meta-algorithm based on a metric labeling formulation of the problem, which alters a given n-ary classifier's output in an explicit attempt to ensure that similar items receive similar labels. They collected movie reviews from a website named IMDB and tested the performance of their classifier under both four-class and five-class categorization. The five-class sentiment classification is adopted in the evaluation of our method (see Section 5). The performance of their approach is limited. One important reason is that their method considers every review when estimating a feature rating of a movie. However, some reviews do not contain much of the users' opinions about a certain feature simply because the users do not care much or are not knowledgable about the feature. In our work, we study the characteristics of reviews' feature ratings. We investigate which reviews are more useful for us to estimate feature ratings. From some observations stated in the next section, we will see that reviews written by different users reflect their own preferred features of a service.

3 Accurate Feature Rating Estimation

Participatory websites allow users to write textual reviews to discuss features of services that they have consumed. These reviews usually contain words that strongly express the users' opinions about the corresponding features. These words contain important information for estimating a numerical rating for the feature. The estimated ratings can be used for assisting other users when they need to choose which services to consume. Machine learning techniques are often used for training a learner based on annotated samples from experts and estimating a rating for a feature discussed in a review. However, for a review that does not mention a feature or discusses it only in a limited sense, the estimation accuracy is expected to be very low. Besides, the opinion expressed by the user who writes this kind of review is not representative because this user obviously does not care much about the feature. We believe that if we carefully select reviews for estimating feature ratings, the accuracy will be increased and the estimated ratings will be more representative.

We then statistically analyze real data collected from the TripAdvisor website. The results reveal that users who comprehensively discuss a feature of a service in their reviews are more likely to agree on a common rating for this feature of the service. This phenomenon can also be intuitively explained as follows. For the users who comprehensively discuss about a feature, the feature is obviously more important to them. People tend to be more knowledgable in the aspects they consider important. These users therefore represent a subset of experts. Experts likely provide more objective and representative feedback about the feature, and therefore the ratings from them for the feature contain less noise and

are more similar.

Based on the above discussion that experts tend to have similar opinions on a feature of a service, a learner trained by a machine learning technique based on annotated samples from experts should then be able to more accurately estimate the feature ratings from reviews written by other experts. Since the opinions of experts converge, the average of the estimated feature ratings also better represents an overall rating for the feature of the service.

We propose a review selection approach using information distance of reviews on the feature based on Kolmogorov complexity, to select reviews that comprehensively discuss a feature of a service. We rank the reviews based on the comprehensiveness on the feature. The top reviews will be selected for the estimation of feature ratings. Also, the average of these estimated feature ratings will be used for representing the overall rating for the feature. Next, we will first describe in detail how our approach selects comprehensive reviews on a given feature.

4 Our Review Selection Approach

Our review selection approach selects reviews that comprehensively talk about a feature. According to this definition, a review's comprehensiveness depends on the amount of information discussed on a feature. We use Kolmogorov complexity and information distance to measure the amount of information. Kolmogorov complexity was introduced almost half a century ago by R. Solomonoff, A.N. Kolmogorov and G. Chaitin, see (Li and Vitányi, 1997). It is now widely accepted as an information theory for individual objects parallel to that of Shannon's information theory which is defined on an ensemble of objects.

4.1 Theory

Fix a universal Turing machine U. The Kolmogorov complexity (Li and Vitányi, 1997) of a binary string x condition to another binary string y, $K_U(x|y)$, is the length of the shortest (prefix-free) program for U that outputs x with input y. It can be shown that for different universal Turing machine U', for all x, y

$$K_U(x|y) = K_{U'}(x|y) + C,$$

where the constant C depends only on U'. Thus $K_U(x|y)$ can be simply written as $K(x|y)$. They write $K(x|\epsilon)$, where ϵ is the empty string, as $K(x)$. It has also been defined in (Bennett et al., 1998) that the energy to convert between x and y to be the smallest number of bits needed to convert from x to y and vice versa. That is, with respect to a universal Turing machine U, the cost of conversion between x and y is:

$$E(x, y)=\min\{|p|: U(x, p)=y,\ U(y, p)=x\} \quad (1)$$

It is clear that $E(x, y) \leq K(x|y) + K(y|x)$. From this observation, the following theorem has been proved in (Bennett et al., 1998):

Theorem 1 $E(x, y) = \max\{K(x|y), K(y|x)\}$.

Thus, the max distance was defined in (Bennett et al., 1998):

$$D_{\max}(x, y) = \max\{K(x|y), K(y|x)\}. \quad (2)$$

This distance is shown to satisfy the basic distance requirements such as positivity, symmetricity, triangle inequality and is admissible.

Here for an object x, we can measure its information by Kolmogorov complexity $K(x)$; for two objects x and y, their shared information can be measured by information distance $D(x, y)$. In (Long et al., 2008), the authors generalize the theory of information distance to more than two objects. Similar to Equation 1, given strings $x_1, \ldots, x_n$, they define the minimum amount of thermodynamic energy needed to convert from any x_i to any x_j as:

$$E_m(x_1, .., x_n)=\min\{|p|:U(x_i, p, j)=x_j \text{ for all } i,j\}$$

Then it is proved in (Long et al., 2008) that:

Theorem 2 *Modulo to an $O(\log n)$ additive factor,*

$$\min_i K(x_1 \ldots x_n|x_i) \leq E_m(x_1, \ldots, x_n)$$

Given n objects, the left-hand side of Equation 3 may be interpreted as the most comprehensive object that contains the most information about all of the others.

4.2 Review Selection Method

Our review selection method is based on the information distance discussed in the previous section. However, our problem is that neither the Kolmogorov complexity $K(\cdot,\cdot)$ nor $D_{max}(\cdot,\cdot)$ is computable. Therefore, we find a way to "approximate" these two measures. The most useful information in a review article is the English words that are related to the features. If we can extract all of these related words from the review articles, the size of the word set can be regarded as a rough estimation of information content (or Kolmogorov complexity) of the review articles. In Section 5 we will see that this gives very good practical results.

4.2.1 Outline

Our method is outlined in the following. First, for each type of product or service (such as a hotel), a small set of core feature words (such as price and room) is generated through statistics. Then, these feature words are used to generate the expanded words. Third, a parser is used to find the dependent words associated with the occurrence of the core feature words and expanded words in a review. For each review-feature pair, the union of the core feature words, expanded words and dependent words in the review defines the related word set of the review on the feature. Lastly, information distance is used to select the most comprehensive reviews on a feature.

4.2.2 Word Extraction

Feature words are the most direct and frequent words describing a feature, for example, price, room or service of a hotel. Given a feature, the core feature words are the very few most common English words that are used to refer to that feature. For example, both "value" and "price" are used to refer to the same feature of a hotel. In (Hu and Liu, 2004), the authors indicate that when customers comment on product features, the words they use converge. If we remove the feature words with frequency lower than 1% of the total frequency of all feature words, the remaining words, which are just core feature words, can still cover more than 90% occurrences. So firstly we extract those words through statistics; then some of those with the same meaning (such as "value" and "price") are grouped into one feature. They are just "core feature words".

Apart from core feature words, many other less-frequently used words that are connected to the feature also contribute to the information content of the feature. For example, "price" is an important feature of a hotel, but the word "price" is usually dropped from a sentence. Instead, words such as "$", "dollars", "USD", and "CAD" are used. We use information distance $d(.,.)$ based on Google to expand words (Cilibrasi and Vitányi, 2007). Let α be a feature and $\mathcal{A}$ be the set of its core feature words. The distance between a word w and the feature α is then defined to be

$$d(w,\alpha) = \min_{v\in\mathcal{A}} d(w,v)$$

A distance threshold is then used to determine which words should be in the set of expanded words for a given feature.

If a core feature word or an expanded word is found in a sentence, the words which have grammatical dependent relationship with it are called the dependent words (de Marneffe et al., 2006). For example, in sentence "It has a small, but beautiful room", the words "small" and "beautiful" are both dependent words of the core feature word "room". All these words also contribute to the reviews and are important to determine the reviewer's attitude towards a feature.

The Stanford Parser (de Marneffe et al., 2006) is used to parse each review. For review i and feature j, the core feature words and expanded words in the review are first computed. Then the parsing result is examined to find all the dependent words for the core feature words and expanded words, all of which are called "related words".

4.2.3 Computing Information Distance

If there are m reviews $x_1, x_2, \ldots, x_m$, n features $u_1, u_2, \ldots, u_n$, and the related word set S_i is defined to be the union of all the related words that occur in the review x_i. From the left-hand side of Equation 3, the most comprehensive x_i

on feature u_k is such that

$$i = \arg\min_i K(S_1 \dots S_n | S_i, u_k). \quad (3)$$

Let S_i and S_j be two sets of words,

$$K(S_i S_j | u_k) = K(S_i \cup S_j | u_k),$$

$$K(S_i | S_j, u_k) = K(S_i \setminus S_j | u_k),$$

and

$$K(S_i|u_k) = \sum_w K(w|u_k) \approx \sum_w (K(w, u_k) - K(u_k))$$

where $w \in S_i$ and w is in x_i's related word set on feature u_k. For each word w in a set S, the Kolmogorov complexity can be estimated through coding theorem (Li and Vitányi, 1997):

$$K(w, u_k) = -\log P(w, u_k), \;\; K(u_k) = -\log P(u_k)$$

where $P(w, u_k)$ can be estimated by $df(w, u_k)$, which is the document frequency of word w and feature u_k co-exist on the whole corpus. Similarly, $P(u_k)$ can be estimated by feature u_k's document frequency on the corpus. In the next section, Equation 3 will be used to select reviews that comprehensively talk about a feature.

5 Experimental Verification

In this section, we present a set of experimental results to support our work. Our experiments are carried out using real data collected from the travel website TripAdvisor. This website indexes hotels from cities across the world. It collects feedback from travelers. Feedback of each traveler consists of a textual review written by the traveler and numerical ratings (from 1, lowest, to 5, highest) for different features of hotels (e.g., value, service, rooms).

Table 1: Summary of the Data Set

Location	# Hotels	# Feedback	# Feedback with feature rating
Boston	57	3949	2096
Sydney	47	1370	879
Vegas	40	5588	3144

We crawled this website to collect travelers' feedback for hotels in three cities: Boston, Sydney and Las Vegas. Note that during this crawling process, we carefully removed information about travelers and hotels to protect their privacy. For users' feedback, we recorded only the textual reviews and the numerical ratings on four features: Value(V), Rooms(R), Service(S) and Cleanliness(C). These features are rated by a significant number of users. Table 1 summarizes our data set. For each one of the cities, this table contains information about the number of hotels, the total amount of feedback and the amount of feedback with feature ratings. In general, each hotel has sufficient amount of feedback with feature ratings for us to evaluate our work.

Table 2: Comprehensive Reviews on Each Feature (Boston)

Top #	V	R	S	C
1	Y	Y	Y	Y
2	Y	Y	Y	Y
3	N	Y	Y	N
4	Y	Y	Y	N
5	Y	Y	Y	Y
6	Y	Y	N	Y
⋮	⋮	⋮	⋮	⋮

5.1 Evaluation of Review Selection

We first evaluate the performance of our review selection approach using manually annotated data. More specifically, in our data set, for one city, 40 reviews (120 reviews in total) are selected for manual annotation. The annotator looks over each review and decides whether the review is comprehensive on a given feature. Comprehensive reviews on the feature are annotated as "Y", and the reviews that are not comprehensive on this feature are annotated as "N". For the review set of each city, the number of reviews annotated as comprehensive is equal to or less than 20% of the total number of the selected reviews for this city (eight in this experiment). Note that it is possible that one review can be comprehensive on more than one features.

We then use our review selection approach

discussed in Section 4 to rank the reviews for hotels in each city, according to their comprehensiveness on each feature. For example, the most comprehensive review on the feature "Value", which has the minimal information distance to this feature (see Equation 3), is ranked No.1. Table 2 shows the annotated reviews for Boston hotels that are ranked on top six on each feature. It can be obviously seen from the table that most of these top reviews are labeled as comprehensive reviews on respective features. Our comprehensive review selection approach generally performs well.

Table 3: Performance of Comprehensive Review Selection

City	Feature	Precision	Recall	F-Score
Boston	V	0.833	0.714	0.769
	R	1.000	0.875	0.933
	S	0.857	1.000	0.923
	C	0.833	1.000	0.909
Sydney	V	0.667	1.000	0.800
	R	0.600	0.857	0.706
	S	0.667	0.857	0.750
	C	0.750	1.000	0.857
Vegas	V	0.778	1.000	0.875
	R	0.727	1.000	0.842
	S	0.714	0.714	0.714
	C	0.667	0.800	0.727

To clearly present the performance of our comprehensive review selection approach, we use the measures of precision, recall and f-score. The measure f-score is a single value that can represent the result of our evaluation. It is the harmonic mean of precision and recall. Suppose there are n reviews in total. Let p_{jk} $(1 \leq k \leq n)$ be the review ranked the kth comprehensive on feature j. Define

$$z_{jk} = \begin{cases} 1 & \text{if } p_{jk} \text{ is labelled comprehensive on } j; \\ 0 & \text{otherwise.} \end{cases}$$

The precision P, recall R, and f-score F of top k comprehensive reviews on feature j are formalized as follows

$$P_{jk} = \frac{\sum_{l=1}^{k} z_{jl}}{k}, R_{jk} = \frac{\sum_{l=1}^{k} z_{jl}}{\sum_{l=1}^{N} z_{jl}},$$

$$F_{jk} = \frac{2P_{jk}R_{jk}}{P_{jk} + R_{jk}}$$

For each ranked review set on feature j, the maximum F_{jk} and its associated P_{jk} and R_{jk} are listed in Table 3. From this table, it can be seen that for the best f-scores, the precision and recall values are mostly larger than 70%, that is, a great part of reviews that are labeled as comprehensive receive top rankings from our comprehensive review selection approach. Our approach is thus carefully verified to be able to accurately select comprehensive reviews on any given feature.

5.2 Statistical Analysis

A group of users who comprehensively discuss a certain feature are more likely to agree on a common rating for that feature. In this experiment, we use our review selection approach to verify this argument.

Table 4: Deviation of Feature Ratings

City	Feature	20%	50%	All
Boston	V	0.884 (0.0003)	1.030	1.136
	R	0.940 (0.2248)	1.037	1.013
	S	1.026 (0.0443)	1.130	1.144
	C	0.798 (0.0093)	0.892	0.949
Sydney	V	0.862 (0.0266)	1.009	1.054
	R	0.788 (0.0497)	0.932	0.945
	S	0.941 (0.0766)	1.162	1.116
	C	0.651 (0.0037)	0.905	0.907
Vegas	V	0.845 (0.0002)	1.236	1.291
	R	1.105 (0.2111)	1.148	1.175
	S	1.112 (0.0574)	1.286	1.269
	C	0.936 (0.0264)	1.096	1.158

More specifically, for each city, hotels that receive no less than 10 reviews with feature ratings are selected. We use our comprehensive review selection approach to select top 20% and 50% comprehensive reviews on each feature for hotels in each city. We calculate the standard deviation of their feature ratings, as well as that of all feature ratings, for each hotel in a city. We then average these standard deviations over the hotels in the same city. The average values are listed in Table 4. The feature ratings of comprehensive reviews on the feature have smaller average stan-

dard deviations. Standard T-test is used to measure the significance of the results between top 20% comprehensive reviews and all reviews, city by city and feature by feature. Their p-values are shown in the braces, and they are significant at the standard 0.05 significance threshold. It can be seen from the table that although for some items there does not seem to be a significant difference, the results are significant for the entire data set.

Therefore, when these travelers write reviews that are comprehensive on one feature, their ratings for this feature tend to converge. This evidence indicates that the estimation of ratings for the feature from these comprehensive reviews can provide better results, which will be confirmed in Section 5.3. These estimated feature ratings can also be averaged to represent a specific opinion of these travelers on the feature, which will be verified in Section 5.4.

5.3 Feature Rating Estimation

In this section, we carry out experiments to testify that the estimation of feature ratings for comprehensive reviews using our review selection approach provides better performance than that for all reviews. We adopt the approach of Pang and Lee (Pang and Lee, 2005) described in Section 2 for feature rating estimation. In short, they applied a meta-algorithm, based on a metric labeling formulation of the problem to alter a given n-ary SVM's output in an explicit attempt. We also adopt a Baysian Network classifier for feature rating estimation.

Similar to the method of Pang and Lee, we build up a feature rating classification system to estimate reviews' feature ratings. However, the method of Pang and Lee focuses only on single rating classification for a review and assumes that every word of the review can contribute to this single rating. While it comes to feature rating classification, the system has to decide which terms or phrases in the review are talking about this feature. We train a Naive Bayes classifier to retrieve all the sentences related to a feature. Then all the core feature words, expanded words and dependent words are extracted to train a SVM classifier and the Bayesian Network classifier for five-class classification (1 to 5). The eight-fold cross-validation is used to train and test the performance of feature rating estimation on all the reviews and the top 20% comprehensive reviews, respectively.

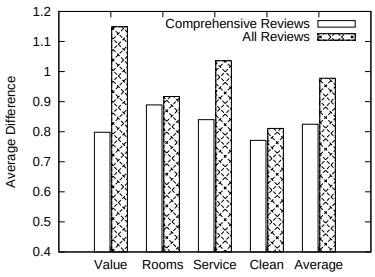

Figure 1: Average Error of Feature Rating Estimation for the Adopted Method of Pang and Lee

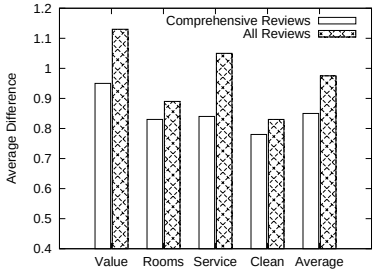

Figure 2: Average Error of Feature Rating Estimation for the Bayesian Network classifier

We formalize a performance measure as follows. Suppose there are n reviews in total. For a test review i ($1 \leq i \leq n$), its real feature rating (given by the review writer) is f_i, and its predicated feature rating (predicted by our classification system) is g_i. Both f_i and g_i are integers between 1 and 5. The performance of the classification on all n reviews can be measured by the average of the absolute difference (d) between each f_i and g_i pair,

$$d = \frac{\sum_{i=1}^{n} |f_i - g_i|}{n}. \quad (4)$$

The lower d is, the better performance the classifier can provide.

Figures 1 and 2 show the results for the performance of feature rating estimation on all reviews versus that on selected comprehensive reviews,

for the adopted approach of Pang and Lee and the Baysian Network classifier respectively. It can be seen that the average difference between real feature ratings and estimated feature ratings on each feature when using selected comprehensive reviews is significantly lower than that when using all reviews, for both the approaches. On average, the performance of feature rating estimation is improved by more than 12.5% using our review selection approach. And, our review selection approach is generally applicable to different classifiers.

5.4 Estimating Overall Feature Rating

Supported by the statistical evidence verified in Section 5.2 that the users who write comprehensive reviews on one feature will more likely agree on a common rating for this feature, we can then use an average of the feature ratings for top 20% comprehensive reviews to reflect a general opinion of knowledgable/expert users. In this section, we show directly the performance of estimating an overall feature rating for a hotel using ratings for the selected comprehensive reviews, and compare it with that for all reviews.

Table 5: Performance of Estimating Overall Feature Rating for Comprehensive Reviews

City	V	R	S	C	AVG
Boston	0.637	0.426	0.570	0.660	0.573
Sydney	0.273	0.729	0.567	0.680	0.562
Vegas	0.485	0.502	0.277	0.613	0.469
Average	0.465	0.552	0.471	0.651	0.535

Table 6: Performance of Estimating Overall Feature Rating for All reviews

City	V	R	S	C	AVG
Boston	0.809	0.791	0.681	0.642	0.731
Sydney	0.433	0.886	0.588	0.593	0.625
Vegas	0.652	0.733	0.502	0.942	0.707
Average	0.631	0.803	0.590	0.726	0.688

Suppose there are m hotels. For each hotel j, we first select the top 20% comprehensive reviews on each feature using our review selection approach. We average the real ratings of one feature provide by travelers for these reviews, denoted as $\bar{f}_j$. We then estimate the feature ratings for these comprehensive reviews using the adopted machine learning method of Pang and Lee. The average of these estimated ratings is denoted as $\bar{g}_j$. Similar to Equation 4, the average difference between all $\bar{f}_j$ and $\bar{g}_j$ pairs on each feature for hotels in each city are calculated and listed in Table 5. From this table, we can see that the average difference between the estimated average feature rating and real average feature rating is only about 0.53. Our review selection approach produces fairly good performance for estimating an overall feature rating for a hotel. We then also calculate the average difference for all reviews. The results are listed in Table 6. We can see that the average difference is larger (about 0.69) in this case. The performance of estimating an overall feature rating is increased by nearly 23.2% through our review selection approach.

6 Conclusion

In this paper, we presented a novel review selection approach to improve the accuracy of feature rating estimation. We select reviews that comprehensively talk about a feature of one service, using information distance of reviews on the feature based on Kolmogorov complexity. As evaluated using real data, the rating estimation for the feature from these reviews provides more accurate results than that for other reviews, independent of which classifiers are used. The average of these estimated feature ratings also better represents an accurate overall rating for the feature of the service.

In future work, we will further improve the accuracy of estimating a general rating for a feature of a service based on the selected comprehensive reviews on this feature using our review selection approach. Comprehensive reviews may contribute differently to the estimation of an overall feature rating. In our next step, a more sophisticated model will be developed to assign different weights to these different reviews.

References

Bennett, C.H., P Gacs, M Li, P.M.B. Vitányi, and W.H. Zurek. 1998. Information distance. *IEEE Transactions on Information Theory*, 44(4):1407–1423, July.

Cilibrasi, Rudi L. and Paul M.B. Vitányi. 2007. The google similarity distance. *IEEE Transactions on Knowledge and Data Engineering*, 19(3):370–383, March.

de Marneffe, Marie Catherine, Bill MacCartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *The fifth international conference on Language Resources and Evaluation (LREC)*, May.

Hu, Minqing and Bing Liu. 2004. Mining and summarizing customer reviews. In *10th ACM International Conference on Knowledge Discovery and Data Mining*, pages 168–177.

Li, M. and P. Vitányi. 1997. *An Introduction to Kolmogorov Complexity and its Applications*. Springer-Verlag.

Long, Chong, Xiaoyan Zhu, Ming Li, and Bin Ma. 2008. Information shared by many objects. In *ACM 17th Conference on Information and Knowledge Management*.

Mullen, Tony and Nigel Collier. 2004. Sentiment analysis using support vector machines with diverse information sources. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 412–418, July.

Pang, Bo and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Annual Meeting of the Association of Computational Linguistics (ACL)*, pages 271–278, July.

Pang, Bo and Lillian Lee. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Annual Meeting of the Association of Computational Linguistics (ACL)*, pages 115–124, June.

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 79–86, July.

Russell, S. and P. Norvig. 2002. *Artificial Intelligence: A Modern Approach*. Second Edition, Prentice Hall, Englewood Cliffs, New Jersey.