Extending associative classifier to detect helpful online reviews with uncertain classes

Zunqiang Zhang, Yue Ma, Guoqing Chen, Qiang Wei

School of Economics and Management, Tsinghua University, Beijing 100084 China

Abstract

While online product reviews are valuable sources of information to facilitate consumers' purchase decisions, it is deemed meaningful and important to distinguish helpful reviews from unhelpful ones for consumers facing huge amounts of reviews nowadays. Thus, in light of review classification, this paper proposes a novel approach to identifying review helpfulness. In doing so, a Bayesian inference is introduced to estimate the probabilities of the reviews belonging to respective classes, which differs from the traditional approach that only assigns class labels in a binary manner. Furthermore, an extended fuzzy associative classifier, namely $GARC_{fp}$, is developed to train review helpfulness classification models based on review class probabilities and fuzzily partitioned review feature values. Finally, data experiments conducted on the reviews from amazon.com reveal the effectiveness of the proposed approach.

Keywords: Review helpfulness, binary classification, class probability, associative classifier

1. Introduction

In recent years online product reviews play a more and more important role in consumers' purchase decision making process [1]. As a kind of digital word-of-mouth, online reviews reflect previous purchasers' experiences and opinions, thus providing potential consumers with valuable information about products [2]. However, the huge amounts of reviews also create a big challenge for consumers to capture useful information. The number of reviews for a single product can easily reach to hundreds or even thousands with various qualities [3]. Considering consumers' limited information processing capacity and time availability, it is deemed desirable and necessary for e-commerce platforms to detect useful reviews and provide them to the consumers as their purchase aid.

In this regard, a recent line of research is conducted to classify reviews based on their features. At the beginning, expert annotators are recruited to assign class labels for training reviews, e.g., *high-quality*, *medium-quality*, *low-quality*, *duplicate*, *spam, etc.* [3, 4], and classification models are built with these exact annotated labels. As a major task in data preprocessing, review labeling by experts guarantees the quality of class labels, but incurs high costs (such as time, labor and money). Thus, this way is often considered impractical/inefficient for real applications [5, 6].

Instead, researchers turn to a social tagging mechanism, namely review helpfulness voting, to seek for re-view labels [7]. To fully use the social characteristics of e-commerce, review helpfulness voting mechanism collects review readers' evaluations to judge the classes of reviews. For example, amazon.com asks review readers the question "*Was this review helpful to you ?*" for each review, and collects review readers' votes, i.e., "*Yes*" or "*No*". Based on the voting results from voluntary non-experts, the class labels of reviews could be judged. More specifically, reviews with enough votes (e.g., at least 10 votes) are usually used to train classification models, and their classes are judged by the proportion of "*Yes*" votes to the total votes [8]. Although suffering from some biases where many reviews still receive very few or even no votes so that their classes could hardly be judged [3], this mechanism enables a much more cost-efficient and larger training dataset, compared with human expert labelling.

Most previous research selects a threshold (e.g., usually set as 0.6) to judge the classes of reviews [5, 8-10]. If the proportion of "*Yes*" votes to total votes is larger than the threshold, the review would be regarded helpful; otherwise, it would be treated as an unhelpful review. Furthermore, research indicates that the error rates would be minimized when the threshold is set to be 0.6 [5, 9].

However, assigning class label for each training review based on the proportion of "*Yes*" votes has some limitations. More concretely, all reviews in the same class are treated equally without considering different levels of confidence. For example, given an excellent review with "*100 of 110 people found the review helpful*" and a fair review with "*70 of 110 people found the review helpful*", the former one should be regarded helpful with higher confidence, although the proportions of "*Yes*" votes for these two reviews are both higher than 60%. In addition, given two fair reviews with "*70 of 110 people found the review helpful*" and "*7 of 11 people found the review helpful*" respectively, the former should also be regarded helpful with higher confidence, since the latter receives fewer votes and its proportion of "*Yes*" votes would be highly influenced by future votes.

To overcome the shortcomings of previous work, a new review labeling approach will be proposed in this paper, which not only judges class labels for training reviews, but also provides the confidence of each review belonging to a particular class (i.e., helpful or unhelpful). It means that a review might belong to more than one class simultaneously with different levels of confidence.

Moreover, an extended fuzzy associative classification method will be introduced to train classification models with uncertain labels. Fuzzy associative classi-

fication has been an active area of research for recent decades in knowledge discovery and business analytics combined with various extensions of fuzzy logic [11-13]. However, most existing extensions are focused on the fuzziness or uncertainty of data attributes [12]. With the consideration of class probabilities, this paper is aimed to present a novel fuzzy associative classification approach called GARC$_{fp}$, which further extends a well-established classifier, namely GARC [14] with its fuzzy extensions [15, 16].

The rest paper is organized as follows. In Section 2, the approach used to label reviews with class uncertainties is introduced. Then, the classifier is developed via training reviews with uncertain classes in Section 3. Subsequently, section 4 conducts data experiments on a review set crawled from amazon.com. Finally, conclusions and future work are stated in Section 5.

2. Review helpfulness class probabilities

Given the voting results for a training review r, i.e., "x of n people found the review helpful", the traditional approach judges the helpfulness class of this review by comparing the proportion of "Yes" votes to total votes, $h=x/n$, with a threshold, i.e., 0.6. If the proportion (h) is larger than the threshold, the class (c) for this review would be "$helpful$" (denoted as 1); otherwise, it would be "$unhelpful$" (denoted as 0). The labeling procedure is calculated as follows:

$$c(r) = \begin{cases} 1(helpful), \text{if } h = x/n \geq 0.6 \\ 0(unhelpful), \text{if } h = x/n < 0.6 \end{cases}. \quad (1)$$

This intuitive approach could provide an exact class label for the review if and only if all people of concern have voted for this review. However, in real contexts, most reviews, even for those reviews used to train classification models, receive a very limited number of votes, which are far away from all people's votes [3]. Thus, it is considered better to estimate the proportion (h) through Bayesian inference based on voting results than directly calculating the proportion [17].

It is reasonable to assume that, for a review with proportion of "Yes" votes $h(0 \leq h \leq 1)$, each review reader would provide "Yes" vote with the probability of h independently. Therefore, the number of "Yes" votes X for this review with the number of total votes n can be regarded as a random variable following a binomial distribution, $X \sim \text{bin}(n,h)$:

$$p_{x|h}(x \mid h) = \binom{n}{x} h^x (1-h)^{n-x}, x = 1, 2, ..., n. \quad (2)$$

Then, the Bayesian posterior distribution of h for this review when "x of n people found the review helpful" is observed can be calculated as

$$p_{h|x}(h \mid x) = \frac{p_{x|h}(x \mid h) p_h(h)}{\int_0^1 p_{x|h}(x \mid h) p_h(h) dh}, \quad (3)$$

where $p_h(h)$ is the Bayesian prior distribution of h for this review, and could be set as an uniform distribution on [0,1] or a beta distribution with parameters (1,1). Thus, the Bayesian posterior distribution of h is a beta distribution with parameters ($x+1$, $n-x+1$), that is,

$$p_{h|x}(h \mid x) = \frac{(n+1)!}{x!(n-x)!} h^x (1-h)^{n-x}, h \in [0,1]. \quad (4)$$

In similar spirit of the notion of review helpfulness class in prior research, the class of a review is dependent on the posterior distribution of h, $p_{h|x}(h|x)$, whether larger or smaller than the threshold (e.g., 0.6) [5, 8-10]. Here, using 1 and 0 to denote the two classes *helpful* and *unhelpful*, respectively, the review helpfulness class probability for a review r is defined on the posterior distribution of h, as

$$\begin{cases} p(c(r)=1) = \int_{0.6}^1 p_{h|x}(h \mid x) dh \\ p(c(r)=0) = \int_0^{0.6} p_{h|x}(h \mid x) dh \end{cases}. \quad (5)$$

In this way, it is possible to express different levels of confidence for reviews in the same class. For example, reviews with "*100 of 110 people found the review helpful*", "*70 of 110 people found the review helpful*" and "*7 of 11 people found the review helpful*" would belong to *helpful* class with probabilities 1.00, 0.77 and 0.56, respectively, as illustrated in Table 1.

Helpfulness voting results	*Helpful* class (1) probability	*Unhelpful* class (0) probability
100 of 110	1.0000	0.0000
70 of 110	0.7741	0.2259
7 of 11	0.5618	0.4382

Table 1: An example of review helpfulness class probabilities.

In addition, a review would no longer belong to only one class. For example, the review with "*70 of 110 people found the review helpful*" would belong to *helpful* class with probability 0.77, and belong to *unhelpful* class with probability 0.23, as illustrated in Table 1.

By discriminating review helpfulness class probabilities, the relationship between review features and review helpfulness classes could be better learned.

3. GRAC$_{fp}$-An extended fuzzy association classifier considering class probabilities

3.1. Gain based Associative Rule Classification

Associative classification in data mining is one of the promising approaches that can make use of the features extracted from reviews to find patterns among them. This type of approaches normally devises classifiers (set of classification rules) that are understandable yet accurate [18].

Our extension will focus on an effective associative classifier, namely GARC (Gain based Associative Rule Classification) [14], which enables to generate a considerably small size of rule set with satisfactory classification accuracy. Briefly speaking, the characteristic of GARC-type approaches [14-16] is threefold: firstly, applying the information gain measure to select the best-split attribute for 1-itemsets, which is to be included in the generation of candidate k-itemsets; secondly, integrating the process of frequent itemsets generation with the process of rule generation; thirdly, defining rule redundancy and rule conflicts, and incorporating corre-

sponding strategies for rule pruning into the mining process. Furthermore, certain fuzzy extensions to GARC have been made in [15, 16] to deal with the numerical attributes using gradual partitionings, so as to solve the "sharp boundary problem" to avoid misclassification.

Based upon GARC and its fuzzy extensions, this paper further extends the scope of research towards building a novel associative classifier (i.e. $GARC_{fp}$) that is capable of classifying the data with class probabilities.

3.2. Fuzzy class association rules with class probabilities

In classification, the resultant rule set (i.e., classifier) consists of fuzzy class association rules (fuzzy CARs).

For a review set R, where F is a set of review features, i.e., $F=\{F_1, F_2, …, F_m\}$. To deal with the numerical feature values, the fuzzy parititioning in [15] is adopted. After the step, each F_k ($1 \leq k \leq m$) can be associated with q_k fuzzy sets defined on domain(F_k), usually labelled as q_k new features in R_f. That is, the new data base R_f is with respect to the feature set $F_f = \{ F_1^1, …F_1^{q_1},\ F_2^1, …F_2^{q_2},$ $F_m^1, …F_m^{q_m} \}$. Let $C=\{1,0\}$ be a set of class labels, where 1 denotes "*Helpful*" and 0 denotes "*Unhelpful*". For each review r in R_f, p_r^c is class probability which indicates the probability of r belonging to c, $c \in C$. Then a fuzzy rule item is of the form $I \Rightarrow c$, where itemset $I \subseteq F_f$, and if $F_k^j \in I$, then $F_k^i \notin I (i \neq j,$ and $1 \leq i \leq q_k, 1 \leq j \leq q_k)$.

In this context, the notions of degrees of support and confidence (i.e., *Dsupp* and *Dconf*) can be extended as follows [14]. For review r in R_f, and any itemset $I = \{f_1, f_2, …f_p\} \subseteq F_f \cup \{c\}$,

$$Dsupp_r(I) = \mu_I(r) = \min(r(f_1), r(f_2), …, r(f_p)) \quad (6)$$

where $r(f_i)$ $i=1,2,…p$, is the corresponding membership degree of r for feature f_i, and the degree of support for itemset I with respect to r is reflected by $\mu_I(r)$ as the membership degree of r for itemset I. Note that t-norm minimum is used here due to its semantics[15]. Other t-norms may also be considered under specific application contexts. Then, with respect to the whole review set R_f, the degree of support for I can be formulated as:

$$Dsupp(I) = \frac{\|I\|}{|R_f|} = \frac{\sum_r \mu_I(r)}{|R_f|} \quad (7)$$

where $|R_f|$ is the cardinality of R_f (i.e., the number of reviews in R_f) and $\|I\|$ is the fuzzy cardinality of I, calculated as a Σcount of $\mu_I(r)$.

Considering the class probability p_r^c, which indicates the certainty degree of the event that the class label c is correct for r, the support degree of itemset Ic with respect to r could be further extended as (8):

$$Dsupp_r(Ic) = Dsupp_r(I) \times p_c^r = \mu_I(r) \times p_r^c \quad (8)$$

where $\|Ic\|$ is the fuzzy cardinality of Ic. Here, $Dsupp_r(Ic)$ is the product of $Dsupp_r(I)$ and the class probability of review r to class c, reflecting the consideration of how certain r belongs to c.

Furthermore, we have the extended notions for the degrees of support and confidence (i.e., (9) and (10)) with respect to the whole review set R_f as follows.

$$Dsupp(I \Rightarrow c) = Dsupp(Ic)$$

$$= \frac{\|Ic\|}{|R_f|} = \frac{\sum_r \mu_I(r) \times p_r^c}{|R_f|} \quad (9)$$

$$Dconf(I \Rightarrow c) = \frac{\|Ic\|}{\|I\|} = \frac{\sum_r \mu_I(r) \times p_r^c}{\sum_r \mu_I(r)} \quad (10)$$

Then, given minimal support α and minimal confidence β ($\alpha, \beta \in [0,1]$), a fuzzy rule item $I \Rightarrow c$ is called frequent if $Dsupp(I \Rightarrow c) \geq \alpha$. Importantly, the fuzzy frequent itemsets also have a well-known anti-monotone property [14]: if a pattern is infrequent, all of its super patterns must be infrequent. Therefore, the pruning strategy of excluded itemsets in [14] is also appropriate in fuzzy context. A fuzzy rule item $I \Rightarrow c$ is called a valid fuzzy CAR if $Dsupp(I \Rightarrow c) \geq \alpha$ and $Dconf(I \Rightarrow c) \geq \beta$.

Suppose the fuzzy CAR set has been learned from the training dataset. For a new review r, the weighted confidence of rule $I \Rightarrow c$ to t is denoted as follows:

$$Wconf_r(I \Rightarrow c) = Dsupp_r(I) \times Dconf(I \Rightarrow c). \quad (11)$$

The helpfulness label predicted is the class label of the rule with maximal weighted confidence.

3.3. The algorithm

The main algorithmic details of $GARC_{fp}$ are shown in Table 2.

Algorithm:

Input: training dataset R
Output: Fuzzy CARs *Frule* = {$rule_1$, $rule_2$, …, $rule_l$}
Begin:
//Fuzzy partitioning using simulated annealing algorithm

1. R_f=Fuzzy(R)
2. *Frule* ={*rule* | *rule* is an 1-item antecedent rule, $Dsupp(rule) \geq \alpha$ and $Dconf(rule) \geq \beta$ }
3. *Fexcluded* ={*erule* | *erule* is an 1-item antecedent rule, and $Dsupp(erule) < \alpha$ }
4. For k from 2 to m do
5. $Candidate_k$ =CandidateGen(*Fexcluded*, k);
6.   *cand*={r | $r \in Candidate_k$, $Dsupp(r) \geq \alpha$ and $Dconf(r) \geq \beta$ }
7.   pruneRules(*cand*)
8.   *Frule*= *Frule* $\cup$ *cand*
9.   *Fexcluded*= *Fexcluded* $\cup$ {e | $e \in Candidate_k$, $Dsupp(e) < \alpha$ }
10. end for

Table 2: Pseudo code of the algorithm.

In Table 2, line 1 performs the fuzzy partitioning operation, so that the original dataset R is then transformed to the fuzzy-partitioned dataset R_f. Lines 2-3

scan the dataset to measure all 1-itemsets rule items. Lines 4-10 perform the consecutive scans of R_f. During each scan, function CandidateGen($Fexcluded$, k) generates k-itemsets from R_f with no super itemsets of ones in $Fexcluded$. Notably, the algorithm will terminate if $cand$ is empty, or otherwise, in at most m passes, where m is the number of features. Since the helpfulness class of reviews are depend on many review features, not dominantly determined by one or few features, more flexible candidate rules are generated with function CandidateGen($Fexcluded$, k).

For the computational aspect of the algorithm, the class probabilities and fuzzy extensions introduced would add little complexity to the efficiency of what GARC takes. As a matter of fact, its worst case is exponential theoretically, while its practical cases are usually near-linear or low-order polynomial, which is comparable with other Apriori-based algorithms (e.g., associative classifier CBA, and association rule (AR) mining) [14].

4. Data Experiments

To evaluate the effectiveness of review helpfulness class probabilities and the performance of the proposed GARC$_{fp}$, real data experiments were conducted on reviews crawled from amaonz.com. To get a representative review sets, main pages of all categories that listed in "Full Store Directory" were firstly collected. Then, the products that had appeared in these category pages were detected. Finally, all reviews and their voting results of these products were crawled during the period of June 2012 to August 2012. In total, 624,940 reviews of 2,647 products were obtained.

Similarly to the experimental setup of prior research, 50,887 reviews with at least 10 votes were selected to conduct experiments [8]. Further, reviews were separated into training sets and testing sets. To clearly evaluate the performance of the methods on classifying reviews, 18,350 reviews with very high class probabilities, $max(p(c(r)=1),p(c(r)=0)) \geq 0.999$, were selected to construct testing sets. The rest 32,537 reviews were used to train classification models.

The review classification process was conducted in three steps. Firstly, certain review features were extracted from review texts, and the reviews were transformed to vector forms, i.e., r: $<f_1, f_2,...f_m>$, where f_j was the value of feature F_j in review r. Secondly, class probabilities were calculated based on voting results for training reviews. After this step, helpfulness class c and corresponding probability p^c were obtained. Thirdly, the training reviews in the form of r: $<f_1, f_2,...f_m : (c, p^c)>$ were used to build the classifier as mentioned in Section 3. The results of each step were described in the following subsections.

4.1. Review features extraction

To train and test review helpfulness classification models, the reviews were represented as vectors of feature values. In this paper, 18 review features that had been widely used in prior research [1, 3-5, 7-9, 19-22] were considered:

Stylistic features, i.e., ratio of uppercase characteristics, ratio of short words (less than 4-letter), number of spelling errors, and readability index.

Content richness features, i.e., review length measured by the number of words, times of product features mentioned, average *tf-idf* values, and average cosine similarity between the review and its previous posted reviews.

Subjectivity features, i.e., number of sentences containing positive, negative and neutral opinions, respectively.

Emotion features, i.e., number of words that express various emotions, such as anger, disgust, fear, joy, sadness and surprise, respectively.

Reviewer expertise features, i.e., number of words which represent time tags that express reviewer's past experience.

In mining review texts, the values of these features were calculated for each review and further used to build classification models. Notably, while feature extraction is the first step in the review classification process, it is not the scope of this paper. Therefore, we determined the related features from prior research and applied respective mainstream techniques for extracting them [1, 3-5, 7-9, 19-22]. Importantly, primary attention of this paper is paid to the other two steps of the process with the notion of class probability and the corresponding algorithm GARC$_{fp}$.

4.2. Review class probability calculation

By calculating on voting results, the review classes were obtained. While the traditional approach simply turns the proportion of "*Yes*" votes to binary classes, the newly proposed approach calculated class probabilities. To illustrate the effectiveness of these two approaches, the distributions of proportions of "*Yes*" votes, binary classes, and *helpful* class probabilities on training reviews were depicted in Figures 1~3.

From these figures, the advantages of our newly proposed approach could be observed. First, the proposed approach performed well on discriminating reviews. As illustrated in Figure 3, most reviews were concentrated on near 0 and 1, revealing that this approach was capable of distinguishing reviews with different helpfulness classes. Comparing with Figure 1 where the proportions of reviews were very near, it was a desirable characteristic.
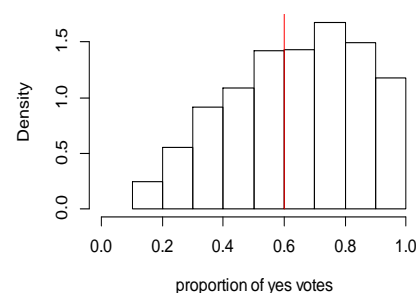

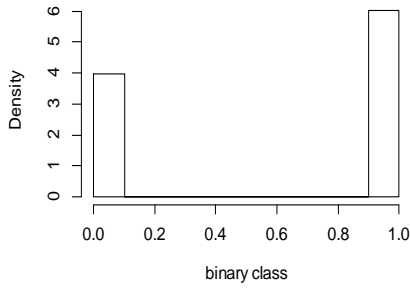Fig. 1: The distribution of proportions of "*Yes*" votes.

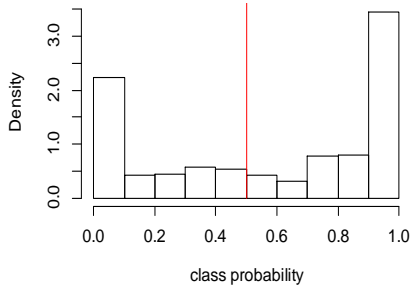Fig. 2: The distribution of binary classes.

Fig. 3: The distribution of *helpful* class probabilities.

Next, the proposed approach was also able to express the uncertainties of review classes. Unlike traditional binary classes as illustrated in Figure 2, the proposed approach also preserved some reviews with relative low class probabilities, near 0.5, which is intuitively appealing. In general, the proposed approach showed a good balance between discriminative power and uncertainty expression.

4.3. Review helpfulness classification results

After obtaining class probabilities for training reviews, a group of experiments were conducted to determine the best parameters, i.e., minimal degree of support α and minimal degree of confidence β. Figure 4 shows the accuracy of the proposed approach with various values of parameter pairs (α, β). It indicated that the approach reached highest accuracy in these experiments when α=0.005 and β=0.35.
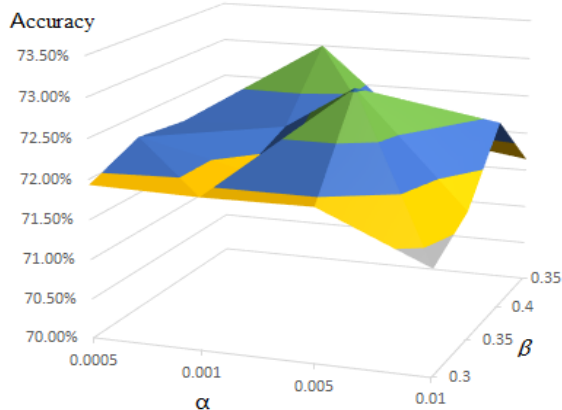

Fig. 4: Sensitivity analysis of parameter pair (α, β).

Then, the performance of the proposed approach was compared with the performance of the approach with binary classes. It is worth noticing that the binary class labels could be regarded as a special case of class probabilities with all probabilities were set to be 1 or 0. Four metrics were used to measure the performances, i.e., accuracy, precision, recall and F-measure. These four metrics were defined on the confusing matrix [23]. The confusing matrix sorts all reviews into four categories, by determining whether the classified classes matched the actual classes as illustrated in Table 3. The number of reviews in each category was counted and displayed in the matrix.

		Classified class	
		helpful (1)	*unhelpful* (0)
Actual	*helpful* (1)	#(TP)	#(FN)
class	*unhelpful* (0)	#(FP)	#(TN)

Table 3: An illustration of confusing matrix.

Accuracy measures the ratio of correctly classified reviews among all testing reviews. Precision indicates the ratio of actual *helpful* reviews among all reviews which are classified as *helpful*. Recall reveals the ratio of actual *helpful* reviews which are classified as *helpful*. These three metrics are calculated as

$$Accuracy = \frac{\#(TP) + \#(TN)}{\#(TP) + \#(TN) + \#(FP) + \#(FN)}, \quad (12)$$

$$Precision = \frac{\#(TP)}{\#(TP) + \#(FP)}, \quad (13)$$

$$Recall = \frac{\#(TP)}{\#(TP) + \#(FN)}. \quad (14)$$

Finally, F-measure is a weighted average of precision and recall.

$$F\_measure = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \quad (15)$$

Figure 5 summarizes the classification performances on testing reviews with both class probabilities and binary classes. Results show that the proposed approach had superior accuracy, precision, F-measure, and a nearly equal recall to the approach with binary classes.
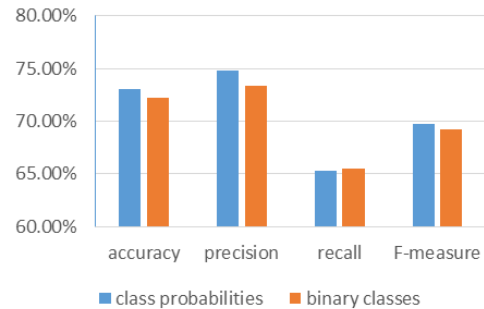

Fig. 5: The performance comparison between class probabilities and binary classes.

5. Conclusions

This paper has proposed an approach to classifying useful reviews in order to support consumers' online product purchase decisions. With the notion of class probabilities, the uncertainty that reviews belong to classes can be measured using a Bayesian inference. This has been further incorporated into the construction of a novel fuzzy associative classifier, namely $GARC_{fp}$, where the values of review features have been fuzzily partitioned. Moreover, real data experiments have been conducted on review sets crawled from amazon.com, with the results showing the advantages of the proposed approach over the traditional one in light of such metrics as accuracy, precision, recall, and F-measure.

Future research will center on applications in other contexts with various types of data and extensions.

References

[1] S. M. Mudambi, and D. Schuff, what makes a helpful online review? A study of customer reviews on amazon.com, *MIS Quarterly,* 34(1):185-200, 2010.

[2] Y. Chen, and J. Xie, Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix, *Management Science,* 54(3): 477-491, 2008

[3] J. Liu, Y. Cao, C.-Y. Lin, Y. Huang, and M. Zhou, Low-Quality Product Review Detection in Opinion Summarization, *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 2007.

[4] C. C. Chen, and Y.-D. Tseng, Quality evaluation of product reviews using an information quality framework, *Decision Support Systems,* 50(4): 755-768, 2011.

[5] X. Zheng, S. Zhu, and Z. Lin, Capturing the essence of word-of-mouth for social commerce: Assessing the quality of online e-commerce reviews by a semi-supervised approach, *Decision Support Systems,* 56: 211-222, 2013.

[6] V. S. Sheng, F. Provost, and P. G. Ipeirotis, Get another label? improving data quality and data mining using multiple, noisy labelers, *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, Las Vegas, Nevada, USA, 614-622, 2008.

[7] A. Y. K. Chua, and S. Banerjee, Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth, *Journal of the Association for Information Science and Technology*, 66(2): 354-362, 2014.

[8] S. Krishnamoorthy, Linguistic features for review helpfulness prediction, *Expert Systems with Applications,* 42(7):3751-3759, 2015.

[9] A. Ghose, and P. G. Ipeirotis, Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics, *IEEE Transactions on Knowledge and Data Engineering*, 23(10):1498-1512, 2011.

[10] Y. Hong, J. Lu, J. Yao, Q. Zhu, and G. Zhou, What reviews are satisfactory: novel features for automatic helpfulness voting, *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, Portland, Oregon, USA, 495-504, 2012.

[11] J. Hühn, and E. Hüllermeier, FURIA: an algorithm for unordered fuzzy rule induction, *Data Mining and Knowledge Discovery,* 19(3):293-319, 2009.

[12] T. Nakashima, G. Schaefer, Y. Yokota, and H. Ishibuchi, A weighted fuzzy classifier and its application to image processing tasks, *Fuzzy sets and systems,* 158(3): 284-294, 2007.

[13] L. A. Zadeh, Fuzzy sets, *Information and control,* 8(3):338-353, 1965.

[14] G. Chen, H. Liu, L. Yu, Q. Wei, and X. Zhang, A new approach to classification based on association rule mining, *Decision Support Systems,* 42(2):674-689, 2006.

[15] G. Chen, Y. Xiong, and Q. Wei, A Fuzzy Extension to Compact and Accurate Associative Classification, *35 Years of Fuzzy Set Theory*, 171-193: Springer, 2011.

[16] Y. Ma, G. Chen, and Q. Wei, A novel fuzzy associative classifier based on information gain and rule-covering. *IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS)*, 2013 Joint. IEEE, 490-495, 2013.

[17] R. Timothy D, Accurate confidence intervals for binomial proportion and Poisson rate estimation, *Computers in Biology and Medicine,* 33(6) : 509-531, 2003.

[18] M. Jabbar, B. Deekshatulu, and P. Chandra, Knowledge discovery using associative classification for heart disease prediction, *Intelligent Informatics*, pp. 29-39: Springer, 2013.

[19] M. P. O'Mahony, and B. Smyth, A classification-based review recommender, *Knowledge-Based Systems,* 23(4):323-329, 2010.

[20] Q. Cao, W. Duan, and Q. Gan, Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach, *Decision Support Systems,* 50(2): 511-521, 2011.

[21] Y. Dezhi, S. D. Bond, and Z. Han, Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews, *MIS Quarterly,* 38(2): 539-560, 2014.

[22] H.-J. Min, and J. C. Park, Identifying helpful reviews based on customer's mentions about experiences, *Expert Systems with Applications*, 39(15): 11830–11838, 2012.

[23] S. V. Stehman, Selecting and interpreting measures of thematic classification accuracy, *Remote Sensing of Environment*, 62(1): 77-89, 1997.