An Approach for Reducing the Numeric Rating Bias

Xiu Li
Graduate School at Shenzhen
Tsinghua University
Shenzhen, China
li.xiu@sz.tsinghua.edu.cn

Huimin Wang
Graduate School at Shenzhen
Tsinghua University
Shenzhen, China
wanghm520@gmail.com

Jian Xu
Graduate School at Shenzhen
Tsinghua University
Shenzhen, China
sm_xj@163.com

Abstract—There is bias in customer reviews and the associated ratings. We propose a method to identify and reduce such bias on the part of reviewers. There are three phases in our approach. Firstly, we conduct Rating-aware sentiment analysis for each review text accompanied with the numeric ratings. Then, we extract features of review text and do manual labelling for training learning machine. At last, SVM is used to train the learner to reduce the rating bias. We applied our method on the dataset obtained from amazon.com. Results suggest that biased ratings have effects on customer ratings significantly in recommender systems and that this bias can be substantially reduced by our model.

Keywords—*bias; evaluation; review mining; e-business*

I. INTRODUCTION

After an online shopping, people are always encouraged to give their feedback in the form of numeric ratings accompanied by reviews text, which will be displayed directly and used to do statistical analysis in order to facilitate browsing and selection. Usually, the review text is more detailed and authentic, but it is hard to display as comprehensive statistics. In contrast, a numeric rating is easy to count and visualize. We observe that in many real-life scenarios, providing numeric ratings tends to be more important than representing the review text. Especially when customers need to make a quick decision on several candidate products. However, even given the rating scale, usually from 0 to 5 stars, with 5 being the best, many people are confused about how to do the best choice. As a result, there is bias in the assignment of the reviewers to objects. There are too many possible reason behind the bias in ratings on the part of reviews. In general, the bias of every reviewer are associated with the deviation of its rating mean and the deviation of its rating variation. Intuitively, the average rating values and rating variations could differ from reviews. One of the most important tasks in product evaluation data mining is to mitigate the affection of reviewers' bias on the ratings.

The rating bias mainly reflect in two parts: the bias of rating mean and the bias of rating variation. The bias of rating mean indicates that everyone has different understanding of "neutral comments", and will assign different ratings to the goods accompanied with the same neutral review text. As for the bias of rating variance, for example, in Fig 1, there are two users in www.amazon.com giving different ratings with the same brief reviews – 'love it'. In term of the review text, both of them have shown almost the same satisfaction with the album, but only one of them gave the full stars. When feel satisfied, some people are generous to assign 5 stars, but the others are only assign 4 stars at most.

Fig 1: Typical displays of reviews in www.amazon.com. Biased ratings can cause the same reviews to correspond with different numeric ratings. This paper explores a method for identifying and reducing such bias.

In this paper, we study the notions of "bias" of reviews. We focus on identifying and reducing bias of reviewers. As two different evaluation ways to the same object, the review text and the associated numeric rating should express the same meaning. More intuitively, after carrying out some kinds of transformations, the numeric rating and the associated review text should be able to output the same digital value. Compares with the easy obtained numeric rating, the hard-won review text expresses the reviewer more accurately. Given the biased numeric rating and the relatively unbiased review text, a complete answer to these questions go way beyond computer science. There are some other factors, including click farm and fake reviews, can also account for the biased rating. However, we focus on a data–centric approach, which the rating bias can be quantified and reduced from the review text given by users. In this approach, we assume that the review text represents the ground truth that can be trusted for the study. What's more, if there are reviewers deliberately give a high or low rating, our approach is not able to identify or reduce the biased rating. Essentially, our study also assumes that there are no fraud reviews, but the rating can still be biased.

We aim at identifying and mitigating the effects of bias on the numeric ratings on customer ratings. We propose a novel, extensible and effective rating bias identifying and reducing method based on customer review mining. It will contribute to improve the precision and coverage of recommendation effects. The main phase of our approach are three folds:

Review text processing. To transform the review text to the same space of the numeric ratings, methods of natural language processing are used. Word segment, POS tag, and

syntactic parsing are employed to the text data. After that, a multi-label emotion classification approach, which combines dictionary and rule approach, and machine learning based approach, is further used to assign the review text to one of the five classifications associated the five rating values (from 1 to 5, with 5 being the best).

Regression model training. In this phrase, a regression model is trained to find out the relations between the biased rating and the features extracted. A group of people are invited to assign ratings for the associated review text. The ratings given by people is regarded as unbiased rating y. Gradient Boosting Regression Tree (GBRT) is used to fit the relation between unbiased ratings and features selected from review texts. Several relevant parameters of the model are adjusted to reach an optimal performance.

Bias reducing. The optimal model will be applied on the test data to check the performance compared to the original ratings presented by users. In order to quantify the effect of improvement, several metrics such as mean, variance and RMSE are used

II. RELATED WORK

The numeric rating is an important instrument used to measure people's attitudes towards some target objects, such as products and services. A useful rating has to provide reliable and valid measurements. Hershey et al.'s research showed that biased rating can be caused by several factors, including equal-interval, vague response alternatives, implicit assumptions of the question, rating with unbalanced scales, and context effects [1]. Hawkins' experiment indicated that as the pressure to respond increases, the tendency of providing uniformed responses will increase [2], which, in some way, can give an explanation for the phenomenon of similar review texts with inconsistent ratings. Eric's finding suggests that correcting biased ratings by removing the bias in standard deviation can increase the accuracy of survey research [3]. Customers' judgements of products can be influenced by how they fix their rating means when mapping context-invariant mental representations onto those scale [4]. Lienert et al. believed that the quality criteria of rating should be met by establishing norms [5]. Krishnan S et al. learned and analysed the phenomenon that the tendency for individuals to comply with what is regarded as a norm in society, and proposed a method for mitigating the effect of bias in rating [6]. Bachrach Y et al. proposed a joint probabilistic model based on Bayesian graphical model to predicting grades to texts [7]. Lauw et al. proposed a reinforce model to identify and quantify the bias within evaluation system on the parts of reviewers and products [8]. Li et al. observed that different people may have different sentiment expression preference. And they proposed a learning framework to incorporate reviewer and product information into text based method for rating prediction [9]. McAuley and Leskovec developed statistical methods combining latent dimensions in rating data with topics in review text to justify the ratings with text [10]. However, there are very few researches focusing on reducing bias in ratings on the part of reviewers combining the review

texts, most of the works primarily focus on rating predictions, or rating bias identification.
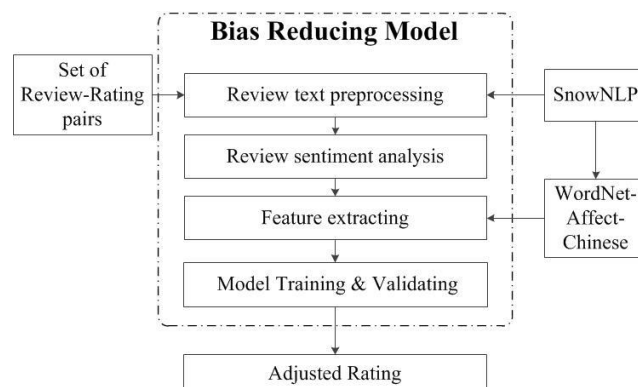

Figure 2: Overview of Bias Reducing Model

Finally, our work also have some elation to sentiment analysis. There are two mainstream categories of text sentiment analysis approaches: dictionary and rule based and machine learning based. The former approach mainly uses emotion lexicon to recognize sentiment words, including attitude words, weight words, and denial words, for emotion index calculation or emotion classification [11, 12, 13, 14, and 15]. Machine learning based approach mainly uses emotion labels as training set to train classifiers for future text emotion classifier [16, 17, 18, 19, and 20]. Multi-Label classification typically uses single-label classifier to solve multi-label classification, or uses adaptive algorithm. Xu et al. proposed an iterative emotion classification approach based on machine learning approach for microblogs [21]. Dave Tried utilized a multi-label classification algorithm to build an automated tag recommender system for isonomy [22]. Trohidis, K et al. compared and evaluated four multi-label classification algorithms including BR, LP, random K-tag set, and MLkNN [23]. Xu J et al. proposed an automatic approach to build Chinese emotion lexicon based on WordNet-Affect [24].

We extend this prior work, and combine dictionary and rule based approach and machine learning approach in our multi-label classification of reviews sentiment. In spite of the similarity between our work and sentiment analysis, there is one critical difference: the goal of such methods is to predict numeric sentimental value at the test stage. In contrast, in the scenario on recommendation systems, our goal is to predict ratings of product.

III. BIAS REDUCING MODLE

Bias reducing model, as shown in Fig.2, begins with review text processing, then feature engineering and manual labeling was employed to train the learner, finally, the adjusted ratings will be obtained. In the following sections, we firstly introduce our Rating-aware sentiment analysis model based on an emotion lexicon and machine learning techniques, and then explain the unbiased rating prediction method.

A. Notions

From a typical online review site, we can have a set of N reviewers assigning numeric ratings accompanied with review texts on a set of M products. We represent the review text written by reviewer *i* using the feature vector. The review text vector is denoted as $X_i \in R^k$, where *k* is the number of extracted feature. For each X_i, there is an associated numeric rating R_i, which is the reviewer's numeric evaluation with bias towards the product. And we denote the adjusted rating as Y_i, which is unbiased. Our goal is to obtain Y_i based on R_i, and X_i and R_i.

B. Feature

Lexicon building. Rating-aware sentiment analysis method is based on lexicon and machine learning. The lexicon was built at first. According to Wong et al.'s approach [25], we built an emotion lexicon called WordNet-Affect-Chinese, or WAC. The generation steps of the lexicon consist of translating all English words by WordNet-Affect into Chinses words, building a bilingual undirected graph to filter all non-emotion words with the help of Chinese synonyms dictionary (Tongyici Cilin), and expanding the obtain Chinese emotion lexicons. Similar to WordNet-Affect, WAC contains 6 emotional categories, namely anger, disgust, fear, joy, sadness, surprise.

Text preprocessing. For refining the input data for further analysis, we filter the duplicative reviews. After that, we focus on Chinese text mining, including words segment, POS tagging and semantic analysis. At this point, we can pass our document through SnowNLP, which is a library written by python, and it has been widely used in handling Chinese text. At this stage, the part of speech of each word and the relationships between parts of the sentence can be yielded. Knowing the part of speech and the relationships between them, we can extract the sentimental feature based on the *WAC*.

Feature selection. After theoretical and experimental analysis, 10 features are used to represent the review text vector. And we denote vector as $X = \{t, r_{et}, r_{an}, r_{ds}, r_{fe}, r_{jo}, r_{sd}, r_{sp}, s_{tm}\}$. Each feature was described in Table 1.

For every review, to extract the features, we count and record the total number of words *t* after conducting words segment. Then we calculate the words matched with those in *WAC* to get N_{st}, the ratio of N_{st} and *t*, or r_{et} is obtained. To obtain the *Anger* index r_{an}, a quaternion need getting first. We denoted the quaternion as $Tr = <a, b, c, d>$, where *a* is the total number of words matched with category *Anger* in *WAC*; *b* and *c* are separately the numbers of the weak and strong gradable adverbs (or gradable phase). For example, *very* is a strong gradable adverb, and *a bit* is a weak gradable phase); and *d* is the total number of negative adverbs related to the sentimental words matched with the category anger of *WAC*. After that, the *Anger* index r_{an} can be calculated as formula (1)

$$r_{an} = a - \frac{1}{2}b + \frac{1}{2}c - d \qquad (1)$$

The same phases can be employed to get *Disgust* index, *Fear* index, *Joy* index, *Sadness* index, *Surprise* index. An example to extract is shown as follow:
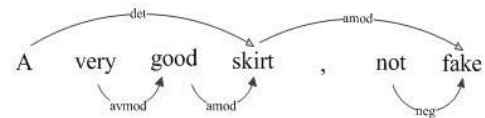

Figure 3: Dependencies representation of a review

Table 1. Feature Notations

Symbol	Description
t	The total number of words in the review text
N_{st}	The total number of words matching with words in *WAC* in the review text
r_{et}	The ratio of N_{st} and t
r_{an}	*Anger* index
r_{ds}	*Disgust* index
r_{fe}	*Fear* index
r_{jo}	*Joy* index
r_{sd}	*Sadness* index
r_{sp}	*Surprise* index
s_{tm}	Sentiment outputs from SnowNLP for each review

Fig.3 shows a typical Syntactic parsing tree of a piece of review text. The word *good* belongs to category *Joy* of *WAC*, and *very* is a strong gradable adverb. Since there are no weak gradable or negative adverbs, the quaternion $Tr = <a, b, c, d>$ can be embodied as <1, 0, 1, 0>. And the *Joy* index is 1.50.

The sentiment outputs for each review, can be obtained by SnowNLP. Based on the above analysis, SnowNLP is a relative effective tool to process text data, especially Chinese text. Its sentiment analysis module has good performance in sentiment analysis of reviews concerning product, for its training corpus is a set of product reviews. We just need to input the piece of review text, and set some parameters, s_{tm} is able to get.

C. Reducing Bias in Ratings

Apart from feature extracting, we also need to do some mental labeling work, which will be discuss later in the experimental part. Now let us consider the following dataset:

$$W \subset X \times R \times Y \qquad (2)$$

Where each element is of the form *(x, r, f(x, r))*, being $f(x, r) \in R$ the predicting unbiased numeric rating given by the review text $x \in X$ accompanied with the original numeric rating $r \in R$. In order to make the function $f(\cdot)$ the review text and the associated rating dependent, the parameter vectors are introduced to capture the effect from the review text and the associated rating on the review text feature vector. So *f* can be of the following form:

$$f(x,r) = (I \cdot r + W)^T x$$
$$= \sum_k (i_k \cdot r + w_k)^T \cdot x_k \qquad (3)$$

Where W and I are k-dimensional parameter vectors. W is to capture the specific effect on every feature of the review text, or each element in X., and I is to capture the effect on the associated rating. Given a training corpus of ratings T, the parameters $\theta = (i_k, w_k)$ are typically chosen by solving the following optimization problems:

$$\hat{\theta} = \arg \min_{\theta} \varsigma(f(x,r), r_0) + \lambda \Omega(\theta) \qquad (4)$$

Where $\Omega(\theta)$ is a regularization term defined in term of the Frobenius norm of the parameter set and plays the role of penalizing overly complex models in order to prevent overfitting. $\varsigma(\cdot)$ is a loss function that measures the deviation between the predicted ratings $f(X,r)$ and the true ratings r_0. Since we focus on the numeric ratings, the loss function can be the least squares loss:

$$\varsigma(x, y) = (x - y)^2 \qquad (5)$$

And the optimization problem can be described in more detail below:

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{|T|} \sum_{r_0 \in T} (f(x,r) - r_0)^2 + \lambda \Omega(\theta) \qquad (6)$$

The parameters I and W can be learned by minimizing the above objective function. We can apply gradient descent algorithm to minimize the objective function.

IV. EXPERIMENT

Given a collection of the review texts accompanied with the numeric ratings, our experiment contains two parts: 1) Presenting and analyzing the bias exiting in the numeric ratings.2) Predicting the "unbiased" rating using Gradient Boosting Regression Tree. And comparing it with several baselines.

The dataset was collected from www.Amazon.com from October 1st, 2013 and January 31st, 2015. We finally obtained 12060 customer reviews accompanied with numeric ratings. Then to train the learning machine, some mental labeling work is needed. The labeling work is distributed to a group of students enrolled in an optional course about e-commerce at a large public university in China. The participants from different backgrounds were invited to give their ratings for the collected reviews without knowing the original numeric ratings, and the average rating of each review is taken as the label. What's more, we need to split the dataset into training sets and test sets, a five-fold cross validation procedure is applied in our experiments, where four folds are used for training, one fold is used for texting.

A. Metrics

In the experiment of bias reducing part, we evaluate the results of different rating adjusting method using Mean Absolute Error (MAS) and Root Mean Squared Error (RMSE):

$$MAE = \frac{\sum_i |S_i - Y_i|}{n} \qquad (7)$$

$$RMSE = \sqrt{\frac{\sum_i (S_i - Y_i)^2}{n}} \qquad (8)$$

Where S_i is the numeric rating for the associated review, Y_i is the adjusted rating, n is the total number of the reviews in the dataset. Both of the metrics can measure how much the adjusted ratings deviated from the true ratings. The smaller the metrics' values are, the better the results will be.
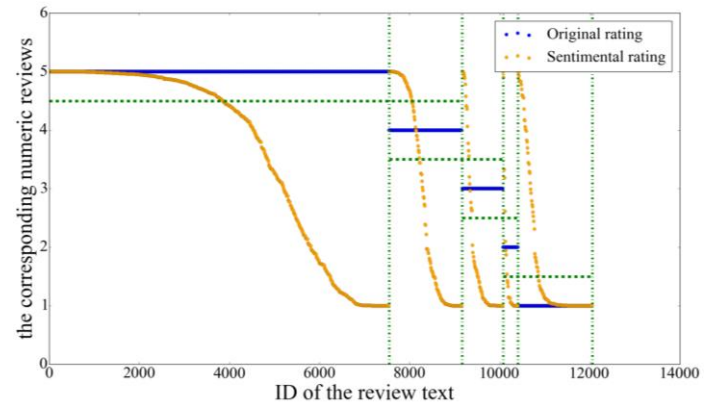

Figure 4: Sentimental Ratings vs Customer Ratings

B. Bias Identification

In the first phrase, all the review texts are mapped into real numbers between zero and five. Fig.4 shows the sentiment ratings sorted in a descending order.

As shown above, the sentiment scores vary a lot compared to the original numeric ratings. To some extent, it can indicates that bias exiting in the customer ratings. But only make a conclusion by presenting the sentiment ratings is arbitrary. We will further take the sentiment rating as an element of the review text feature vector, which is set as the inputs to train the model.

Fig.5 shows the original numeric ratings accompanied with the associated labelled ratings on the whole dataset, where 'o' stands for an original rating, and '+' a labelled rating. Both of the original ratings and labelled ratings are ordered in hierarchical, and the labelled ratings are scattered corresponding to the sentiment ratings sorted in a descending order (in Fig 4) as shown in Fig 5.

From Fig 5, it shows that customers tend to assign a higher rating in one trade. In this case, 62.6% of the reviews assign the numeric ratings of five, which is the highest. And 76.0% present 'positive' ratings (5 stars or 4 stars). However, 53.8% of the labelled ratings are above 4.5 (thus can be rounded to 5-star), and 71.6% are between 3.5 and 5. When it comes to 'moderate' ratings, the variance of the labelled ratings is

relatively large, because there are a thousand "moderate" in a thousand people eyes.
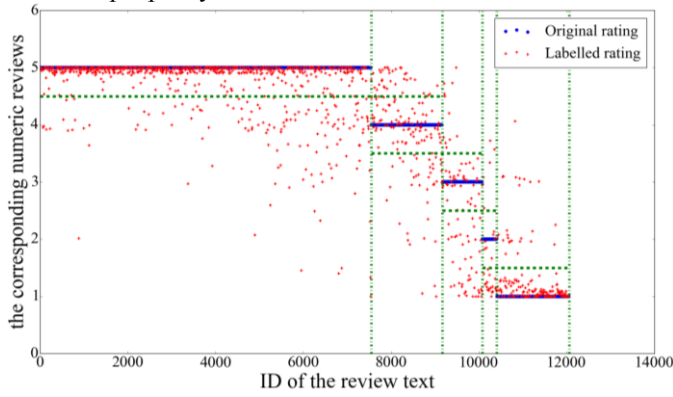

Figure 5: Adjusted Ratings vs Original Ratings

What makes the situation worse is that, a log-in user has access to the rating system on Amazon whether he or she has bought something or not. Some of the users just go to claim their favor, or to give the malicious criticism. Their ratings are commonly deviating a lot from the average ratings.

Fig.6 is a histogram showing the deviation between the original ratings and the labelled ratings. 71.6% of the rating differences are less than 0.5, so the two kinds of ratings are coherent when rounded to an integer. 21.6% of the rating differences are between 0.5 and 1.5, just 1-start bias when rounded to an integer. Other 6.80% have residuals larger than 1.5, which are less tolerable and needed to be improved.
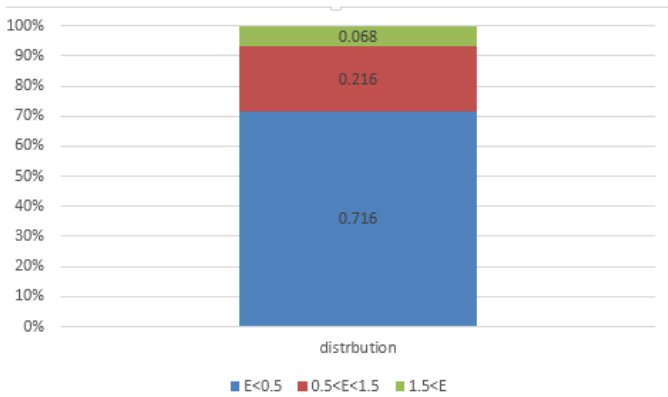

Figure 6: Distribution of Difference between Customer Ratings and Target Ratings on Whole Data Set

C. Bias Reducing

We hold out a random 80% of rating triplets as our training set, and partition another 20% into test set. We have introduced our basic model, linear regression, in the previous section. Comparing with several algorithm, we finally choose our regression model based SVM. The rating prediction results by different regression methods was shown in Table 2.

Compared with the random method, the other four methods, GBRT (Gradient Boosting Regression Tree),

Adaboost regression, Bagging Regression, and SVM regression have got good performances. We also can see that SVM-Regression achieves the best results. Then is GBRT, which is just a little worse than SVM. However, SVM perform better in small training sets, GBRT will be a better choice for large training set.

Table 2. Results on rating adjusting task

	Random	GBRT	Adaboost	Bagging'	SVM
AME	1.7421	0.3486	0.3946	0.3747	0.3411
RMSE	2.0600	0.4982	0.5597	0.5087	0.3857

We use SVM as the regression method. As shown in Table 3 indicates that the adjusted ratings outperform the original ratings. The average original rating is apparently bigger than the average adjusted ratings. The adjusted rating reduces the deviation efficiently. In addition, the adjusted rating has the less variance, which indicates a more robust performance and a better generalization ability. The RMSE of our adjusted ratings is 0. 3857 on test dataset, approximately reducing by 50% compared with the original customer ratings, whose RMSE is 0. 7066.

Table 3. Results on adjusted ratings vs original ratings

	Original ratings' set	Adjusted ratings' set
AME	0.3672	0.3411
RMSE	0.7066	0.3857
Average	0.414	0.397

Fig 7. shows the adjusted ratings on test data set. The regression results are in a descent order on the whole, corresponding to the tendency of labelled ratings. In general, adjusted ratings tend to present the scores lower than the original ratings, except for the 1-star original ratings. It compensates for the potential higher ratings presented by customers as mentioned above.
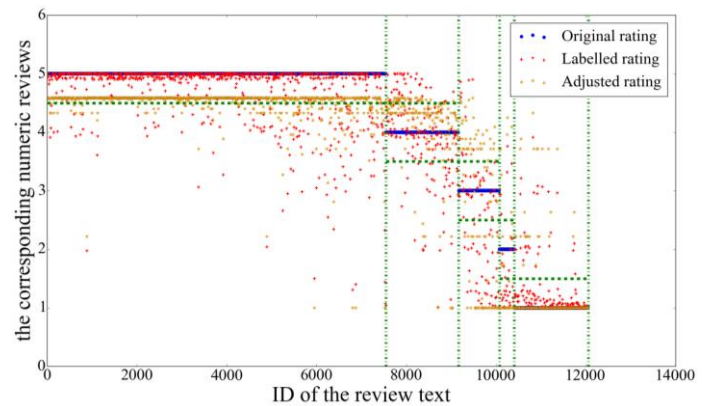

Figure 7: Comparison of Customer Rating, Target Rating, and Adjusted Ratings on Test Data

Fig 8(a) and (b) are histograms that show the distribution of ratings difference from labelled and adjusted ratings under different thresholds. Fig 8(a) takes 0.5 and 1.5 as the thresholds,

while Fig 8(b) extends them to 1 and 2. As shown in the Fig 8(a), the differences between labelled ratings and adjusted ratings are improved a lot, especially for the deviation $E \geq 1.5$, whose percentage reduces from 6.53% to 2.49% significantly. For the part of $E<0.5$, the amount has increased slightly; and for the part of $0.5<=E<1.5$, the amount seems to keep unchanged. In fact, some samples are moved from '0.5<=E<1.5' to 'E<0.5', and meanwhile other samples moved from '1.5<=E' to '0.5<=E<1.5' and 'E<0.5'. It's a coincidence that '0.5<=E<1.5' keeps a balance. This can be proved in another view of Fig 8(b), when we alter the threshold to 1.0 and 2.0.
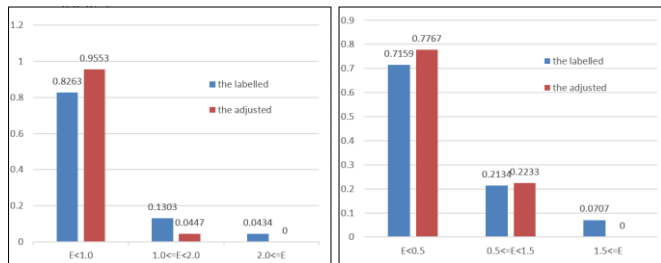

Figure 8: Differences of Labelled and Adjusted Ratings

D. CONLUSION AND FUTURE WORK

Our experiment suggest the affection of biased rating scale on the numeric rating results can be significant in recommender system, and that these bias can be reduced to a large extent with machine learning methods. To apply this method to other recommender system, the key questions for future work are how to make more uniformed and refined rules to label the customer reviews. One idea is clustering the reviews into some representative categories so that only part of the reviews need labeling. We believe that the better the training set are chosen, the better the results will be obtained. We are also interested in more detailed question that rating scales may focus on a single aspect, how to detect and weight it? And we are also exploring methods to detect the false reviews and to study the relationship between customers' level of experience and the rating bias. We would also like to extend our method to recommender system in other domains.

REFERENCES

[1] Friedman H H, Amoo T. Rating the rating scales[J]. Friedman, Hershey H. and Amoo, Taiwo (1999)." Rating the Rating Scales." Journal of Marketing Management, Winter, 1999: 114-123.

[2] Hawkins D I, Coney K A. Uninformed response error in survey research[J]. Journal of Marketing Research, 1981: 370-374.

[3] Greenleaf E A. Improving rating scale measures by detecting and correcting bias components in some response styles[J]. Journal of Marketing Research, 1992.

[4] Lynch Jr J G, Chakravarti D, Mitra A. Contrast effects in customer judgments: changes in mental representations or in the anchoring of rating scales?[J]. Journal of Customer Research, 1991: 284-297.

[5] Lienert G A, Raatz U. Testaufbau und testanalyse[M]. Weinheim: Beltz, 1961.

[6] Krishnan S, Patel J, Franklin M J, et al. Social Influence Bias in Recommender Systems: A Methodology for Learning, Analyzing, and Mitigating Bias in Ratings[J].

[7] Bachrach Y, Graepel T, Minka T, et al. How To Grade a Test Without Knowing the Answers---A Bayesian Graphical Model for Adaptive Crowdsourcing and Aptitude Testing[J]. arXiv preprint arXiv:1206.6386, 2012.

[8] Lauw H W, Lim E P, Wang K. Bias and controversy: Beyond the statistical deviation[C]//Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2006: 625-630.

[9] Li F, Liu N, Jin H, et al. Incorporating reviewer and product information for review rating prediction[C]//IJCAI. 2011, 11: 1820-1825.

[10] McAuley J, Leskovec J. Hidden factors and hidden topics: understanding rating dimensions with review text[C]//Proceedings of the 7th ACM conference on Recommender systems. ACM, 2013: 165-172.

[11] Aman S, Szpakowicz S. Identifying expressions of emotion in text[C]//Text, Speech and Dialogue. Springer Berlin Heidelberg, 2007: 196-205.

[12] Shen Y, Li S, Zheng L, et al. Emotion mining research on micro-blog[C]//Web Society, 2009. SWS'09. 1st IEEE Symposium on. IEEE, 2009: 71-75.

[13] Li X, Wang H, Yan X. Accurate Recommendation Based on Opinion Mining[M]//Genetic and Evolutionary Computing. Springer International Publishing, 2015: 399-408.

[14] Melville P, Gryc W, Lawrence R D. Sentiment analysis of blogs by combining lexical knowledge with text classification[C]//Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2009: 1275-1284.

[15] Pang B, Lee L. Opinion mining and sentiment analysis[J]. Foundations and trends in information retrieval, 2008, 2(1-2): 1-135.

[16] [24] Pang B, Lee L, Vaithyanathan S. Thumbs up?: sentiment classification using machine learning techniques[C]//Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002: 79-86.

[17] Ye Q, Zhang Z, Law R. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches[J]. Expert Systems with Applications, 2009, 36(3): 6527-6535.

[18] Xu J, Ding Y X, Wang X. Sentiment classification for Chinese news using machine learning methods[J]. Journal of Chinese Information Processing, 2007, 21(6): 95-100.

[19] Ohana B, Tierney B. Sentiment classification of reviews using SentiWordNet[C]//9th. IT & T Conference. 2009: 13.

[20] Xia R, Zong C, Li S. Ensemble of feature sets and classification algorithms for sentiment classification[J]. Information Sciences, 2011, 181(6): 1138-1152.

[21] Xu R, Wang Z, Xu J, et al. An Iterative Emotion Classification Approach for Microblogs[M]//Computational Linguistics and Intelligent Text Processing. Springer International Publishing, 2015: 104-113.

[22] Dave K, Lawrence S, Pennock D M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews[C]//Proceedings of the 12th international conference on World Wide Web. ACM, 2003: 519-528.

[23] Trohidis K, Tsoumakas G, Kalliris G, et al. Multi-Label Classification of Music into Emotions[C]//ISMIR. 2008, 8: 325-330.

[24] Xu J, Xu R, Zheng Y, et al. Chinese emotion lexicon developing via multi-lingual lexical resources integration[M]//Computational Linguistics and Intelligent Text Processing. Springer Berlin Heidelberg, 2013: 174-182.

[25] Xu J, Xu R, Zheng Y, et al. Chinese emotion lexicon developing via multi-lingual lexical resources integration[M]//Computational Linguistics and Intelligent Text Processing. Springer Berlin Heidelberg, 2013: 174-182.