Usefulness Individualized Recognition Model of Consumer Online Reviews Based on Transfer Learning

Xianfei Yang\textsuperscript{1,2}\textsuperscript{*}

\textsuperscript{1} College of Economics and Management, Dalian University, Dalian, Liaoning, 116000, China
\textsuperscript{2} School of Electronic and Information Engineering, Taizhou University, Zhejiang, 318000, China
\textsuperscript{*}yangxianfei@dlu.edu.cn

Abstract. Usefulness recognition model of consumer online reviews can filter out useless ones in all reviews, which can reduce cost of searching for commodity information for potential consumers. However, existing recognition models were based on that all consumers judge usefulness of online reviews by same criteria, ignoring differences of theirs. Thus, a new transfer learning support vector machine model was proposed in this paper to realize individualized recognition of online reviews usefulness. Theoretical analysis and experimental results showed that proposed algorithm could achieve better performance with less training data.

1. Introduction

With rapid development of internet, online consumption has become one of the most important ways for Chinese purchasing goods. Meanwhile number of online reviews has also increased exponentially. Online reviews are real feedback after consumers use product, which have great reference value for other potential consumers. Result of survey by CNNIC on consumers online purchasing behaviour showed online reviews were regarded as one of the most important information sources for purchasing decisions by 43.3\% of online consumers\cite{1}. However, not all online reviews contain useful information. More than 20,000 reviews on MP3 at Amazon, only 38\% of them were considered to contain useful information \cite{2}. Therefore, identifying and only displaying online reviews with useful information to consumers can greatly reduce cost of search for commodity information for consumers, and also provide enormous commercial value for e-commerce operators.

Many e-commerce websites have established evaluation system for usefulness of each online review. Amazon will ask other readers to answer, “Is this review useful to you?” Usefulness of them was obtained in this way and reviews that most consumers considered useful were displayed in front. However, such evaluation systems were based on that all consumers had same criteria for judging usefulness of online reviews and which were inconsistent with reality. For example, for online reviews of movies, some consumers pay more attention to reviews which contain description of movie scene, while some focus on movie content. Thus, it has often appeared number of useful votes and useless votes obtained by same review were both high in film website. In order to solve above problem, an individualized review usefulness recognition model was proposed trained by single consumer's historical evaluation data about usefulness of reviews. And this recognition model introduced a new
transfer learning method to solve problem that single consumer evaluation dataset was too small to build a good performance model.

2. Relevant research

Research on usefulness of online review mainly focused on two aspects: influencing factors and recognition model of online review usefulness. Many studies have shown length of review text has an important impact on usefulness of review. Because the longer review text was, the more useful information review contained [3]. However, Gan found if length exceeded a certain value, usefulness of review could decline, because it caused cognitive load of readers [4]. Mudambi pointed out extreme review could impact on sense of usefulness by an empirical study of reviews on Amazon [3]. And extreme negative reviews were more useful than extreme positive ones, because negative reviews tended to point out shortcomings of commodity and thus provided more real and useful information for potential consumers [5]. However Chua draw an opposite conclusion. His research found readers tend to believe that positive reviews were more useful than negative ones [6]. Forman’s research showed real name and address information provided by reviewers leaded to more useful votes for his online reviews [7]. Scholz study discovered the longer reviews appear, the more probability they got useful votes with [8]. To recognize usefulness of online reviews, Jindal defined reviews on brand only, untrue reviews, advertisements and other irrelevant reviews containing no opinions as useless reviews, and filtering repeated reviews and logistic regression models were employed [9]. According to readability, subjectivity and informativeness of product reviews, liu trained a support vector machine to detect low quality product review [10]. Li used support vector machine model to recognize useful reviews based on whether product name, attribute, brand and reviewer name were mentioned in reviews text [11]. Lv employed LDA model to extract topics of reviews text, and used logic regression model to rank usefulness of restaurant online reviews [12].

3. Background

3.1. Support vector machine

On basis of statistical learning theory, Vapnik proposed a classical two class classification method dealing with small sample dataset, namely Support Vector Machine (SVM) [13]. By using structural risk minimization principle, it could obtain better generalization than neural network which regard empirical risk minimization as target. Basic idea of SVM was finding an optimal linear hyperplane in n-dimensional space to maximize minimum distance between two classes’ data and hyperplane. Then finding this optimal hyperplane in SVM was reduced to solve a convex programming problem. SVM mapped sample space to a high dimensional or infinite dimensional feature space by kernel function, so as to solve non-linear classification problem.

Set training dataset of SVM as \( \{(x_i, y_i)\}_{i=1}^{m} \), \( x_i \in \mathbb{R}^n \), \( y_i \in \{-1, 1\} \), optimal linear classification hyper plane could be represented by \( f(x) = w^T x + \beta \). Parameters \( w \), \( \beta \) could be obtained by solving following formulas:

\[
\begin{align*}
\min_{w, \beta} & \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{m} \xi_i \\
\text{s.t.} & \quad y_i(w^T x_i + \beta) \geq 1 - \xi_i, \quad i = 1, \ldots, m \\
& \quad \xi_i \geq 0, \quad i = 1, \ldots, m
\end{align*}
\]  

(1)

Where \( C \) was a penalty factor to achieve tradeoff between cost of data misclassification and complexity of classifier, \( \xi_i \) was slack variable. Lagrange method was utilized to transform formula (1) into dual form:
Lagrange coefficient \( a_i \) was calculated by quadratic programming method. Then \( w, \beta \) could be calculated by:

\[
w = \sum_{i=1}^{m} a_i y_i x_i, \quad \beta = y_j - \sum_{i=1}^{m} a_i y_i (x_i^T g x_j)
\]

For any data \( x \in \mathbb{R}^n \), output of SVM was:

\[
f(x) = w^T g x + \beta = \sum_{i=1}^{m} a_i y_i (x_i^T g x) + \beta
\]

If \( f(x) > 0 \), corresponding \( y = 1 \), otherwise \( y = -1 \). Next, we discussed geometric meaning of value of \( f(x) \) and relationship between posterior probability with it. Supposed \( x = x_p + r_g v/\|w\| \), where \( x_p \) was projection of \( x \) on the optimal classification hyperplane, and \( r_g \) was "distance" from \( x \) to hyperplane, and \( r_g \in \mathbb{R} \). If \( f(x) < 0 \), \( r_g < 0 \). Then \( f(x) = w^T (x_p + r_g v/\|w\|) + b = r_g \|w\|, \quad r_g = f(x)/\|w\| \). Therefore, the larger value of \( f(x) \), the larger "distance" from \( x \) to optimal classification hyperplane. Traditional SVM could only classify data, but could not give posterior probability of data belonging to a certain class. Viewing from geometric meaning of SVM, the larger \( r_g \) was, the greater probability \( x \) belonged to positive class with. So Platt associated output of SVM with posterior probability by using sigmoid function and achieved good performance. Its conversion function was as follow [14]:

\[
P(y=1|f) = 1/(1 + \exp(af + b))
\]

3.2. Transfer learning

In traditional data mining classification model, training data and classification data are independent and identical distribution (i.i.d). Therefore, before classification model with good performance is trained, enough training data satisfying i.i.d should be collected. But they are collected difficulty or costly sometime. Therefore, transfer learning can make full use of known data with similar distribution to solve above problem. Given target domain dataset \( O = \{ (x_1, y_1), L, (x_m, y_m) \} \), which obeys probability distribution \( P_o(x, y) \), and source domain dataset \( S = \{ (x_{m+1}, y_{m+1}), L, (x_{m+n}, y_{m+n}) \} \) with probability distribution \( P_s(x, y) \). Suppose corresponding marginal distribution are \( P_o(x) \) and \( P_s(x) \), and conditional distribution are \( P_o(y|x) \) and \( P_s(y|x) \). \( P_o(x, y) \) and \( P_s(x, y) \) are different but similar. The goal of transfer learning is to make full use of knowledge in \( P_s(x) \) and \( P_s(y|x) \) to train classifiers with \( O \) which is suitable for target domain distribution \( P_o(x, y) \).

4. Individualized recognition model for usefulness of online review based on transfer learning support vector machine

A new transfer learning support vector machine (NTLSVM) was proposed to solve individualized recognition problem about online reviews' usefulness. Let total online reviews of a commodity be \( \{ (x_1, y_1), L, (x_m, y_m), L, (x_m', y_m') \} \), where \( y_i \) was difference value between useful voting and useless voting of ith review. If \( y_i > 0 \), \( y_i' = 1 \). Otherwise \( y_i' = -1 \). According to formula (1) establishing SVM used all \( M \) online review data, corresponding optimal hyperplane parameters \( w_i, b_i \) could be obtained.
Supposed first $m$ data were single consumer’s evaluation dataset for usefulness of online reviews and $\lambda_i \geq \lambda_{i+1}$, $i=1, L, m$. If $x_i$ was evaluated as useful by this consumer, corresponding $y_i = 1$, otherwise $y_i = -1$. Then modeling formula of NTLSVM was as follows.

$$
\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{m} \xi_i + v \|w - w_i\|^2 \\
\text{s.t.} \quad y_i((w^T \mathbf{g}x_i) + \beta) \geq 1 - \xi_i, i=1, L, m \\
\xi_i \geq 0, i=1, L, m \\
w^T \mathbf{g}(x_i - x_{i+1}) \geq 0, i=1, L, m
$$

(6)

Regularization term $v \|w - w_i\|^2$ guaranteed if $m$ is small, difference between parameters $w$ of optimal classification hyperplane and $w_i$ was small. $\lambda_i \geq \lambda_{i+1}$ indicated that difference between useful voting and useless voting obtained by ith review was bigger than that by $i+1$th review. Hence $P(y=1|f(x_i)) \geq P(y=1|f(x_{i+1}))$. According to relationship between posteriori probability and value of $f(x)$ described above, inequality constraints $w^T \mathbf{g}(x_i - x_{i+1}) \geq 0$ could be obtained. Formula (6) was solved by Lagrange method, and corresponding Lagrange function was:

$$
J = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{m} \xi_i + v \|w - w_i\|^2 - \sum_{i=1}^{m} a_i[(w^T \mathbf{g}x_i) + \beta] - 1 + \xi_i] - \sum_{i=1}^{m} \xi_i - \sum_{i=1}^{m} u_i w^T \mathbf{g}(x_i - x_{i+1})
$$

(7)

Derivatives of formula (7) were as follows.

$$
\frac{\partial J}{\partial w} = 0 \rightarrow w = \frac{2v w_j + \sum_{i=1}^{m} a_i y_i x_i + \sum_{i=1}^{m} u_i (x_i - x_{i+1})}{1 + 2v}
$$

$$
\frac{\partial J}{\partial \xi_i} = 0 \rightarrow C - a_i - r_i = 0
$$

$$
\frac{\partial J}{\partial \beta} = 0 \rightarrow \sum_{i=1}^{m} a_i y_i = 0
$$

(8)

Taking formula (8) back to formula (7), dual form of formula (6) could be obtained.

$$
\max \frac{1}{2(1+2v)} \left[ \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j y_i y_j (x_i^T \mathbf{g}x_j) + 2 \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j y_i y_j (x_j^T \mathbf{g}x_i) + \sum_{i=1}^{m} \sum_{j=1}^{m} u_i u_j [(x_i - x_{i+1})^T \mathbf{g}(x_j - x_{j+1})] \right] \\
+ \frac{1}{1 + 2v} \sum_{i=1}^{m} a_i y_i w_i^T \mathbf{g}x_i + \frac{1}{1 + 2v} \sum_{i=1}^{m} u_i w_i^T \mathbf{g}(x_i - x_{i+1}) + \sum_{i=1}^{m} a_i + \frac{v}{(1 + 2v)^2} \|w_i\|^2 \\
\text{s.t.} \sum_{i=1}^{m} a_i y_i = 0 \\
0 \leq a_i \leq C, i=1, L, m \\
u_i \geq 0, i=1, L, m
$$

For each data $x$, output of NTLSVM was shown in equation (4), where $w$ and $\beta$ were respectively:

$$
w = \frac{2v w_j + \sum_{i=1}^{m} a_i y_i x_i + \sum_{i=1}^{m} u_i (x_i - x_{i+1})}{1 + 2v}
$$

$$
\beta = y_j - \frac{2v}{1 + 2v} (w_j^T \mathbf{g}x_j) - \frac{1}{1 + 2v} \sum_{i=1}^{m} a_i y_i (x_i^T \mathbf{g}x_j) - \frac{1}{1 + 2v} \sum_{i=1}^{m} u_j [(x_i - x_{i+1})^T \mathbf{g}x_j]
$$
5. Relevant experiments

Using GooSeeker web crawler software, we obtained 2605 reviews of Alita: Battle Angel at Douban Film Review website. One of them received 3464 votes for usefulness and 119 votes for uselessness, which further verified difference between consumers' criteria for usefulness of online reviews. In order to verify effectiveness of proposed algorithm and reduce difficulty of labeling dataset manually, 200 reviews were randomly selected as experimental dataset. Using this to train SVM, parameters $w_i$, $b_i$ could be obtained. Meanwhile single consumer's label of usefulness of each review in this dataset was obtained by survey by questionnaire. Generally speaking, number of online reviews which were evaluated usefulness by single consumer is small. Therefore $m$ reviews were randomly selected as training dataset of recognition model, and others were used as validation dataset in 200 reviews.

According to previous studies on usefulness of review, textual and semantic features of review could better depict difference between useful reviews and useless ones. Text features included title features, text length and average sentence length et al. It was generally believed that reviews with longer heading and text contained more information, and mean sentence length could measure readability of review. Semantic features were characterized by themes contained in review. In this paper, LDA model was used to identify whether review contains actors’ performances, film pictures, narrative techniques, film content, emotions themes words or not.

In order to validate performance of NTLSVM model in individualized recognition of online Reviews' usefulness, a comparative experiment was conducted with $SVM_m$, $SVM_i$ and TLSVM[15]. $SVM_m$ was built using single consumer evaluation data as training dataset. $SVM_i$ was obtained by using all 200 reviews as training dataset. Kernel functions of all models were Gauss kernel functions. In order to obtain optimal recognition performance of each model, grid method was used to obtain optimal parameters of each model. Accuracy of online reviews usefulness recognition of four models was compared as shown in Figure 1.

![Figure 1. Accuracy comparison of online reviews usefulness recognition of four models with different $m$.](image)

As be seen from above, if $m=6$, recognition accuracy of $SVM_m$ was very low, since training dataset of it contained too little data. TLSVM and NTLSVM which based on transfer learning had higher accuracy than $SVM_m$. Thus used knowledge from similar fields in process of modeling can significantly improve recognition performance of them which were similar to that of $SVM_i$. If $m=10$, with increase of number of training samples, recognition accuracy of $SVM_m$, TLSVM and NTLSVM were improved, and performance of NTLSVM was higher than that of TLSVM. The reason might be NTLSVM not only used regularization term to transfer classification knowledge of similar domains, but also used inequality constraints to transfer posterior probability knowledge of data. If $m=20$, performance of TLSVM and NTLSVM decreased, while $SVM_m$ was the highest. So there was negative transfer learning. It might be train dataset was enough to independently establish a classification with
good performance. However, it is not realistic for single consumer to give useful evaluation to large number of online reviews.

6. Conclusion
Online commodity reviews have become one of the important reference factors for online consumption. However, large number of useless reviews have increased cost of consumer searching for information and become one of the important reasons hindering continued development of online consumption. In this paper, aiming at ignoring differences between consumers i
in existing online reviews usefulness recognition models, an individualized recognition model of online reviews usefulness based on transfer learning was proposed. In new transfer learning process, not only classification knowledge, but also posterior probability knowledge of data in similar fields was transferred, so as to improve recognition efficiency of model under condition of insufficient training dataset.

Acknowledgments
Research funding: Liaoning Natural Science Foundation of china (20170540030).

References
[1] China Internet Information Center (CNNIC). (2009) Research Report on China's Online Shopping Market in 2009. 11: 38 - 42.
[2] Cao Q, Duan W, Gan Q. (2011) Exploring Determinants of Voting for the "Helpfulness" of Online User Reviews: A Text Mining Approach. Decision Support Systems, 50:511-521
[3] Mudambi S M, Schuff D. (2010) What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. Management Information Systems Quarterly, 34: 185-200
[4] Gan Q, Cao Q, Jones D. (2012) Helpfulness of Online User Reviews: More is Less. In: Proceedings of the Americas Conference on Information Systems, Seattle, Washington, pp. 1-8
[5] Liao C L, Cai C J, Y Li. (2013) An Empirical Study on Influence Factors of Online Reviews’ Helpfulness in E-Commerce. Soft Science, 5: 46-50
[6] Chua Y, Banerjee S. (2016) Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. Computers in Human Behavior, 54:547-554
[7] Forman C, Ghose A, Wiesenfeld B. (2008) Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. Information Systems Research, 19: 291-313
[8] Scholz M, Dornier V. (2013) The recipe for the perfect review?An investigation into the determinants of review helpfulness. Business &Information Systems Engineering, 5:141-151
[9] Jindal N, Liu B. (2008) Opinion spam and analysis. In: Proceedings of international conference on Web search and Web data mining. New York. pp.219 - 229
[10]Liu J, Cao Y, Lin C, et al. (2007) Low-quality product review detection in opinion summarization. In: Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning. Prague. pp.334 - 342
[11]Li X, Ding S C. (2013) Research on Review Spam Recognition. Modern Library and Information Technology, 1:63-68
[12]Lv S H, Yang L, Lin H F. (2011) Ranks of Restaurant Reviews Based on LDA Model. Computer Engineering, 10: 62-64, 67
[13]Vapnik V N. (1999) An overview of statistical learning theory. IEEE Trans Neural Netw, 10: 988-999
[14] Lin H T, Lin C J, Weng R C. (2007) A note on Platt’s probabilistic outputs for support vector machines. Machine Learning, 68: 267-276

[15] Xu M, Wang S T, Gu X. (2014) TL-SVM: A transfer learning algorithm. Control and Decision, 29: 141-146