A Correction Method of Online Product Scores from the Perspective of Discourse Markers Theory

Weijun Wang[1, a], Yanqiu Song[2,b *], Boyang Chen[3,c]

1(School of Information Management, Central China Normal University, Wuhan 430079, China; Key Laboratory of Adolescent Cyber psychology and Behavior of Ministry of Education, Central China Normal University, Wuhan, 430079,China)

2(School of Information Management, Central China Normal University, Wuhan 430079, China; Key Laboratory of Adolescent Cyber psychology and Behavior of Ministry of Education, Central China Normal University, Wuhan, 430079,China)

3(School of Information Management, Central China Normal University, Wuhan 430079, China; Key Laboratory of Adolescent Cyber psychology and Behavior of Ministry of Education, Central China Normal University, Wuhan, 430079,China)

[a]email:wangwj@mail.ccnu.edu.cn, [b]email:yanqiusong@yeah.com,[c]email:by_chen1991@163.com

Keywords: online product reviews; score deviation; Discourse Markers

Abstract: Online product reviews is an important information source which affects consumer purchasing decision-making deeply, but the inconsistency between reviews and scores brings troubles in the process of decision-making for consumers. So the importance of solving reviews' authenticity is focused on. This paper attempts to introduce the Discourse Markers Theory to make a correction on review scores by analyzing the content of reviews, achieving a breakthrough in the semantic and pragmatic level. Firstly, we build a library of discourse markers in the field of online product reviews and design a correction system including credibility, position, emotional attitude. Then, we get the weight of each index through the questionnaire and propose a algorithm to correct deviation of reviews scores. In the last, the verification results from Jingdong mall show the method has better feasibility and validity.

Introduction

Due to the rapid development of e-commerce and Web2.0, online product reviews as a new form of online reputation has been one of the important information source of decision-making for both consumers and sellers In general, the score of reviews provides consumers an overall sense about the products or services , meanwhile the contents of reviews provide a detailed perception. However, in actual situations, the reviews scores and contents which are inconsistent have great influenced the authenticity and validity of online product reviews. As consumers have fell into a vague and contradictory mental state, the consumer purchase decision-making are disturbed.

Studies in both China and abroad have concentrated on usefulness of bias deviation and comments score deviation. Usefulness of bias deviation includes imbalance vote bias, winner circle bias and early bird bias[1], and comments score deviation mainly includes estimating sequential bias[2],self-selection bias[3,4] and seller manipulating bias[5]. In fact, this type of deviation of score inconsistent with content of product reviews has recognized by the scholars, but it hasn't get enough attention, however. In addition, the existing ways are mainly modifying some suggestions and computer algorithms to improve the quality and credibility of comments in order to correct the deviation, so that it can achieve the purpose of rectifying the deviation indirectly.

This paper will focus on the deviation type of score inconsistent with content of review. We try to introduce Discourse Markers Theory of linguistic science into information science, and achieve the goal of fine-grained deviation modification from the perspective of semantics and pragmatics, and then provide consumers with a more objective and real review score.

Discourse Markers Theory

As an important kind of linguistic expressions in speech systems, discourse markers (DM) which are often independent of syntactic can be used for semantic hints and pragmatic constrain, as to show the various types of dialogue relations[6].

Product reviews which belongs to the real application of natural language in online shopping are often made at random, so it has become an important source of discourse markers corpus. For different types of discourse markers have different pragmatic functions, they convey different information as well, which is shown in Table1.

Table 1 main types of discourse markers

Types	pragmatic functions
Discourse Markers of Admitted Type	Highlight the authenticity of information; Emphasize the presentation is frank, fair and impartial;
Discourse Markers of Elaboration Type	Highlight the speaker hold the relatively subjective position to comment;
Discourse Markers of Motivation Type	Enhanced the function of reasoning and make presentation rational;
Discourse Markers of Assertion Type	Highlight the speaker hold the absolutely subjective position to comment;
Discourse Markers of Evaluation Type	Express the speaker's different emotional attitudes such as joy, glad, regret, irony, etc

Score Correction System Based on Discourse Markers Theory

First of all, we build discourse markers library in the field of online product reviews, and write programs to automatically identify the discourse markers in reviews; secondly, we obtain different types of discourse markers' weight in credibility, position, emotional attitudes from questionnaire; finally, according to our proposed algorithm, we calculate the final scores of online product reviews which contain discourse markers**.**

Construction of discourse markers library

Try to build the discourse markers library, this paper has mainly chose two ways. On the one hand, Some discourse markers have been extracted from relevant papers on DM from both China and abroad. On the other hand, we have randomly collected 5000 ordinary goods comments from Jingdong Mall by web crawler program, and then we have adopted the method of frequency statistics to filter low frequency of discourse markers and keep those highs (frequency ≥50).At the same time, considering the difference between online product reviews and traditional literature, we eliminate some discourse markers which are written language not spoken language.

Discourse markers library consist of credibility-discourse markers, position-discourse markers and emotional attitudes-discourse markers. Specifically, position-discourse markers involves the subjectivity or objectivity of the speaker's comment position, and emotional attitudes-discourse markers involves the positivity or negativity of the speaker's attitude.

Weight of Credibility, Position and Emotion Attitude

In order to obtain weight of credibility, position and emotion attitude in different types of discourse markers, we have conducted a questionnaire survey, 28 total discourse markers in which five types such as the discourse markers of admitted type, elaboration type, motivation type, assertion type and evaluation type have bee covered. A total of 203 persons have participated in this survey, and 162 valid questionnaires have been collected at last. In the test, subjects who have 1 years experience of online shopping at least are aged between 20 to 35 years old, among them, male accounted for 41.6% and female accounted for 58.4%.

Aimed at 28 discourse markers, the questionnaire adopts double sentence form. Sentence II which contains certain discourse marker has been compared with sentence I. The test subjects would read these sentences and then check these discourse markers whether it enhance the expression of credibility , subjective/objective stance, and emotional tendency or not. Let's take "to tell the truth" for example.

Sentence I Quality of clothes is xx. There are shortcomings, color is xx.

Sentence II Quality of clothes is xx. *To tell the truth,* there are shortcomings, color is xx.

After questionnaire collecting, we have calculated the mean value of statistical data and obtain the weight of credibility, position (subjective, objective) and emotion attitude(positive, negative).The results are shown in Table 2.

Table 2 Weight of credibility, position and emotion attitude

Discourse Markers	Some examples	credibility	position		emotion attitude	
			subjective	objective	positive	negative
Credibility-DM	tell the truth,to be honest, to be fair	0.684	0.135	0.181	/	/
Position-DM	personally speaking, on a personal level, as far as I'm concerned	0.509	0.406	0.085	/	/
	I dare say, I'm sure, definitely,completely	0.531	0.420	0.049	/	/
	it is said that, in general	0.437	0.074	0.489	/	/
Positive emotion-DM	satisfying, amazing, to one's surprise	0.175	0.235	0.021	0.569	/
	well enough, not bad	0.192	0.208	0.048	0.48	0.072
	fortunately, luckily, thankfully	0.214	0.07	0.095	0.443	0.178
Negative emotion-DM	ironically, frustrating	0.149	0.11	0.066	/	0.675
	intolerable, awful, terrible	0.146	0.157	0.029	/	0.668
	to one's upset, to one's disappointment	0.19	0.191	0	/	0.619
	it's a pity, regretfully ,unfortunately	0.239	0.148	0.084	0.026	0.503

Correction Algorithm Based on DM

Discourse markers is defined as the five-dimensional vector $word_m=(a_m, b_m, c_m, d_m, e_m)$,and the alphabetical a-e represents the credibility, subjective position, objective position, positive emotion, and negative emotion respectively. Assume that a single review contains multiple discourse markers, each discourse marker is expressed with five-dimensional vector, so the vectors from $word_1$ to $word_n$, as shown below.

$$\mathrm{word}_1 = (a_1, b_1, c_1, d_1, e_1) \quad \mathrm{word}_2 = (a_2, b_2, c_2, d_2, e_2) \quad \ldots \quad word_n = (a_n, b_n, c_n, d_n, e_n)$$

We have determined and compared different DM vectors' values in the same property, and respectively acquired the maximum value of credibility, position, positive emotion, negative emotion, which is represented by q_1, q_2, q_3, q_4. In particular, q_2 is the bigger one between maximum value of subjective position and maximum value of objective position.

$$q_1 = (a_1, a_2, ..., a_n) \qquad q_2 = (b_1, b_2, ..., b_n, c_1, c_2, ..., c_n)$$
$$q_3 = (d_1, d_2, ..., d_n) \qquad q_4 = (e_1, e_2, ..., e_n)$$

Credibility and position can promote the expression of emotional attitude.For example,if the speaker takes the firm stand and tells the authentic information about the goods or service, then what he has said is considered convincing ,and the emotion expressed can be strengthened.

At the same time, based on the statistical results of the previous survey, we have found that there is a certain proportion relationship in the credibility, positions and emotional attitude, namely credibility: position: positive emotion: negative: emotion = 0.291:0.279:0.175:0.256. On the basis of the foregoing, in this paper, we have designed formula "W_{DM}" which can calculate the total weights of multiple discourse markers in single review, as follows.

$$a=\mathrm{sgn}(q_3+q_4)$$

$$\beta=|1-(1-q_1)\times(1-0.291)|+|1-(1-q_2)\times(1-0.279)|$$

$$\gamma=|[1-(1-q_3)\times(1-0.175)]-[1-(1-q_4)\times(1-0.256)]|$$

$$W_{DM}=a\cdot(\beta+\gamma)$$

Taking the polarity of emotion and original rating state of review into account, the calculation method of final score of review is designed, as follows.

$$\text{Praise: } if\,(q_3+q_4)>0, Score=3+W_{DM};\ if\,(q_3+q_4)<0, Score=5+W_{DM};$$

$$\text{Neutral: } if\,(q_3+q_4)>0, Score=2+W_{DM};\ if\,(q_3+q_4)<0, Score=4+W_{DM};$$

$$Bad\text{: } if\,(q_3+q_4)>0, Score=1+W_{DM};\ if\,(q_3+q_4)<0, Score=3+W_{DM};$$

Examples Illustration

To verify the effectiveness of the proposed algorithm, we have randomly selected 525 online product reviews, about five categories such as clothing, skincare, digital products, books, and furniture. Then, we get original website score, DM revised score and manual score of each review. Specially,we have gotten manual scores by recruiting 40 volunteers from School of Information Management and School of Psychology in Central China Normal University. In addition, reliability of manual scores of commodities have been are examined, and output of testing proves that manual scores are consistent, stable and reliable (Cronbach's Alpha>0.8),which is shown in Table 3.

Table 3 Reliability test of manual scores

Number	Category	Cronbach's Alpha
1	Clothing	0.953
2	Skincare	0.940
3	Digital Products	0.870
4	Books	0.836
5	Furniture	0.951

A comparison of original website score(t_1), DM revised score(t_2) and manual score (t_3) are conducted through Independent samples T-test to verify the effectiveness of the correction algorithm.The results are summarized in Table 4.

Table 4 Results of Independent Samples Test

Category	comparative item	Independent Samples Test						
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
Clothing	t1 – t2	2.12	139	0.036	0.472	0.223	0.032	0.912
	t2—t3	-1.18	118	0.239	-0.23	0.194	-0.615	0.155
Skincare	t1 - t2	2.08	104	0.04	0.419	0.201	0.02	0.818
	t2 – t3	0.78	104	0.439	0.135	0.173	-0.209	0.478
Digital Products	t1 - t2	2.03	80	0.045	0.59951	0.29501	0.01242	1.18661
	t2—t3	-0.57	80	0.568	-0.14146	0.24658	-0.63218	0.34925
Books	t1 - t2	2.06	115	0.041	0.25268	0.12246	0.01011	0.49524
	t2—t3	0.80	140	0.423	0.07239	0.09016	-0.10586	0.25065
Furniture	t1 - t2	2.284	126.089	0.024	0.28971	0.12685	0.03869	0.54074
	t2—t3	1.553	138	0.123	0.14257	0.09181	-0.03896	0.3241

Note:t1-original website score, t2-DM revised score, t3-manual score; t1-t2:T test on original website score and DM revised score; t2-t3:T test on DM revised score and manual score.

The significance level P values of manual scores and original website scores are less than 0.05, there is a significant difference; The P values of manual scores and revised scores are greater than 0.05, the difference is not significant. Manual scores are closer to revised scores, which verify the validity and applicability of Discourse Markers' application to revise the score deviation.

Conclusion

Product reviews and scores have a direct impact on consumers purchase behavior. However, there are inconsistency between reviews and scores. Based on the Discourse Markers Theory, this paper puts forward online product reviews-scores correction method, achieving a breakthrough in the semantic and pragmatic level. The results of examples illustration shows the method can not only get accurate scores match with the content of reviews, but also help consumers make better decisions and improve the network evaluation mechanisms.

Acknowledgements

This study is supported by Program of National Natural Science Fund "Service Selection Optimization of SaaS based on User Preference-aware in Cloud Computing" (No: 71271099) and Natural Science Foundation of Hubei Province for Innovative Research Team in University "Knowledge Integration and Service based on Cloud Computing" (No: 2011CDA116).

References

[1]Liu J, Cao Y, Lin C-Y, et al. Low-Quality Product Review Detection in Opinion Summarization[C]// Proceedings of the 2007 Joint Conference on Empirical Methods in Natural

Language Processing and Computational Natural Language Learning, Prague, Association for Computational Linguistics. 2007:334-342.

[2]Sikora R T, Chauhan K. 2012. Estimating sequential bias in online reviews: a Kalman filtering approach[J]. Knowledge-Based Systems, 2012,27:314-321.

[3]Kapoor G, Piramuthu S. Sequential Bias in Online Product Reviews[J].Journal of Organizational Computing and Electronic Commerce, 2009,19(2):85-95.

[4]Li X X, Hitt L M. 2008. Self-Selection and Information Role of Online Product Reviews[J].Information Systems Research, 2008,19(4):456-474.

[5]Yan Zhao Literature Review of Online Fake Review[J].Journal of Shanghai management science.2014,04:85-88.(in Chinese)

[6] Liyan Liu. Study on Chinese Discourse Markers,Beijing,2011.(in Chinese)