

Patients' satisfaction factors mining and sentiment analysis of B2C mail-order pharmacy reviews

Jingfang Liu

Shanghai University

Yingyi Zhou (✉ zhouyingyi@fuji.waseda.jp)

Shanghai University

Xiaoyan Jiang

Tongji University

Wei Zhang

Shanghai University

Research article

Keywords: B2C, mail-order pharmacy, online review, topic mining, sentiment analysis

Posted Date: December 5th, 2019

DOI: <https://doi.org/10.21203/rs.2.18255/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background In recent years, mail-order pharmacies have been accepted by increasingly more patients, and the prospects for mail-order pharmacies are optimistic. This article explores the patients' satisfaction factors addressed in Business to Customer (B2C) mail-order pharmacy reviews and analyzes the sentiments expressed in the reviews. The goal of this work is to help B2C mail-order pharmacy enterprises identify patients' concerns, continuously improve the health services level.

Methods This paper was based on the Latent Dirichlet Allocation (LDA) topic model. From a B2C mail-order pharmacy of integrated e-commerce and a vertical B2C mail-order pharmacy (JD.COM and J1.COM, respectively), 136,630 pieces of over-the-counter (OTC) drug review data posted from January 1, 2015 to December 31, 2018 were selected as samples and used to explore the satisfaction factors of B2C mail-order pharmacy patients regarding the entire drug purchasing process. Then, the sentiments expressed in the drug reviews were analyzed with SnowNLP.

Result Categorization of the 12 factors identified by LDA showed that 5 factors were related to logistics; these 5 factors, which also included the most drug reviews, made up 38.5% of the reviews. The number of factors related to drug prices was second, with 3 factors, and reviews of drug prices made up 25.5% of the reviews. Customer service and drug effects each had two related factors, and a smaller percentage of these reviews (13.95%) were related to drug effects. Patients still maintain positive opinions of JD.COM and J1.COM. However, some opinions on logistics and drug prices are expressed.

Conclusion The most important task for mail-order pharmacies is to improve logistics. It is better to develop self-built logistics. Both types of B2C mail-order pharmacies can improve patient viscosity by implementing marketing strategies. With regard to customer service, focusing on improving employees' service attitudes is necessary.

1. Introduction

1.1. Background

The Internet has completely changed the way in which we live and communicate, and it has also changed the methods and strategies people use to procure necessary items [1]. As Internet access increases, the need to search for health information also increases all over the world [2, 3, 4]. A recent study found that nearly half of Americans first consulted the Internet for information about health or medical problems [5]. The use of mobile devices with portability, mobility, personalization and ubiquity has further amplified this trend [6, 7].

Patients not only retrieve health information from the Internet but also obtain a variety of health services or products [8, 9]. With the continuous expansion of the digital health industry, the pharmaceutical e-commerce has developed rapidly [10]. Mail-order pharmacies may be mail-order branches of offline pharmacies or independently operated Internet sites that represent partnerships between pharmacies [11,

12, 13]. The excellent patient experience and the convenience of transactions during mail-order shopping have contributed to the growing market share of mail-order pharmacies [14, 15]. Mail-order pharmacies have also encountered many problems in their operations [16]. For example, because of the unwillingness of many illegal websites to disclose their actual locations, it is impossible to establish an effective regulatory framework for Internet pharmacy logistics operations [17, 18, 19]. According to The World Health Organization (WHO), 10% of drugs sold globally through mail-order suppliers may be counterfeit [18, 20, 21, 22].

Early reports show that there were very few actual cases in which prescription drugs were purchased online [23]. However, recent reports indicate that the number of people who use Internet pharmacies to purchase drugs and other mail-order health products is increasing [24]. Although the scale of the mail-order drug sales market in China showed a significant increase between 2012 and 2018, it still accounts for 9.1% of the total retail drug market share in 2018. Compared with the drug retail market in the US, of which e-commerce represents 33.3%, China's pharmaceutical e-commerce still has much room for growth.

1.2. Related work

Medical e-commerce is the product of e-commerce development. Patients' reviews on the B2C website are evaluations of the products or services obtained by patients who purchase products or services [25], and patient reviews provide information for other patients to select and purchase products [26]. By reading reviews, a patient can reduce his or her uncertainty about a product or service [27]; at the same time, mail-order pharmacy's reviews can attract more potential patients to the site, increase patient access time on the site, and increase patient stickiness to the site [28, 29, 30, 31]. Current research on mail-order pharmacy reviews includes the use of Analytic Hierarchy Process (AHP) to conduct comprehensive quality assessments of mail-order pharmacies, but the results of AHP are greatly influenced by subjective judgments [32]. The chameleon clustering algorithm was used to cluster hotspot reviews, but the complexity of the algorithm made the calculation too time-consuming to complete [33]. The corresponding analysis method has been used to study the differences among medical e-commerce websites, but the number of samples collected, especially the number of bad reviews, was very small, and this number may have affected the results of the analysis [34]. Examples of cold medicine: "999" reviews on JD.COM and J1.COM are shown in Fig. 1.

2. Methods

2.1. Research framework

The purpose of this study was to mine and analyze reviews of the entire transaction process submitted by patients of two B2C mail-order pharmacies. This article used OTC drug user review data obtained from two mail-order pharmacies, JD.COM and J1.COM; these two mail-order pharmacies are representative of B2C mail-order pharmacies of integrated e-commerce and vertical B2C mail-order pharmacies, respectively. First, the indicators of patient satisfaction posted on the B2C e-commerce websites are summarized according to the literature review. At the same time, LDA, an unsupervised machine learning algorithm based on the topic model, is used to cluster the factors addressed in the patients' reviews. Second, based on a review of the literature, an index of the factors that influence B2C e-commerce website patients' satisfaction is used to classify the review

factors, and four factor categories of B2C mail-order pharmacy drug patient reviews are presented. The factor distributions of reviews posted on the websites of the two mail-order pharmacies are compared and analyzed. Third, through analysis using a sentiment dictionary, this study identifies the emotional tendencies of patients regarding various patients' satisfaction factors and compares the emotional tendencies of the patients in each factor classification. Finally, the conclusions of the study are presented, and the results of the factor discovery, factor classification and sentiment analysis are used to propose rational suggestions for the health services of the two types of B2C mail-order pharmacies. The methodological framework of this paper is shown in Fig. 2.

2.2. Patient Satisfaction Factors by Mail Order

Szymanski defines a review as the consumer's perception of his or her entire mail-order shopping experience [35]. The process of creating a patient review is actually the process in which the patient explicitly expresses his or her degree of satisfaction with the website. Therefore, identifying the factors that influence consumer satisfaction provides a means of classification of the factors addressed in patients' reviews posted on B2C websites, as shown in Table 1 below.

Table 1 Factors influencing consumer satisfaction with B2C websites (from the literature)

Author influence factors	Product factors	Staff factors	Logistics factors	Price factors	Information factors	System factors
Lee[36]		√		√		√
Chun-Chun Lin[37]	√	√	√	√	√	√
Xia Liu[38]	√	√	√		√	√
Yooncheong Cho[39]		√				
Gholamreza Torkzadeh[40]	√	√	√	√		√
Ziqi Liao[41]				√		
Szymanski[35]	√					√
Mckinney[42]					√	√
Kim[43]		√	√		√	√
Wolfenbarger[44]		√	√			
Timo Koivumaki[45]				√		

Through a review of the relevant literature on the factors affecting consumer satisfaction with B2C e-commerce websites, six factors that affect patient satisfaction with B2C e-commerce websites were identified. The influencing factors and their definitions are presented in Table 2 below.

Table 2 Definitions of influencing factors

Influencing factors	Definition
Product factors	Stable product quality; Reliable product brand
Staff factors	Service attitude and service quality of presales, sales, customer service, and logistics staff
Logistics factors	Dispatch speed; Transport speed; Logistics security; Logistics cost
Price factors	Perceived prices; Competitive prices; Promotions
Information factors	Comprehensive product information; Whether or not product matches product description
System factors	Usability of website; Sound payment mechanism; Payment security

2.3. Data collection

In this paper, B2C mail-order pharmacies from which large numbers of patients purchase drugs and that receive a large number of standardized reviews are divided into two categories: B2C mail-order pharmacies of integrated e-commerce and vertical B2C mail-order pharmacies. In this paper, we selected two representative mail-order pharmacies in China, JD.COM and J1.COM, and used their websites to obtain OTC drug reviews posted from January 1, 2015 to December 31, 2018 as a corpus.

In this paper, a total of 136,630 user reviews was obtained using web crawlers; 72,231 of the reviews were obtained from JD.COM, and 64,399 reviews were obtained from J1.COM. The data are cleaned (duplicate, short or meaningless advertisements, symbols, and sentences that are difficult to recognize are deleted) to reduce the interference of the noisy review data on the LDA factor classification results. Finally, 107,198 pieces of data were obtained; 53,306 of these were obtained from JD.COM, and 53,831 were obtained from J1.COM.

2.4. Data-driven analysis

2.4.1 Data preprocessing

In this study, three steps of preprocessing work were performed on the collected Chinese text data. The first step is that a useful Python kit called Jieba was adopted to segment the Chinese sentences into separate words [46]. The second step in preprocessing is the deletion of stopwords whose meaning cannot be recognized from the word segmentation. The third step in preprocessing is the merging of synonyms and phrases such as "express" and "logistics" [47].

When the above three steps of data preprocessing had been completed, 19,127 items remained, and 23% of the items had been deleted.

2.4.2 Factor discovery methods

This study used the LDA (latent Dirichlet allocation) model to classify the factors (topics) of reviews collected from JD.COM and J1.COM. LDA is a Bayesian probability model consisting of a three-layered structure of words, factors, and document collections [48,49]. The LDA model considers that the document collection is a mixture of multiple topics and topic is a polynomial distribution within the fixed word.

The TF-IDF (term frequency- inverse document frequency) model is first used to calculate the weight and the word frequency of each term in the document and to convert each review into a vector. Next, the Gibbs sampling algorithm is used to estimate the posterior of the LDA model parameters[50,51,52].

2.4.3 Sentiment analysis methods

We adopted SnowNLP to carry out sentiment analysis, which is a python kit that specializes in sentiment analyses of Chinese texts. The algorithm of SnowNLP is actually a Naïve Bayes algorithm: a simple probabilistic model often used for binary classification. First, we need to train our data to fit the model. We invited two researchers to select 1000 positive and negative reviews each manually. Then, we used the chosen 2,000 reviews to train the model, and then the trained model was used to perform a sentiment analysis on the rest of reviews.

For better understanding the sentiment analysis results, we converted the sentiment scores range from $[0,1]$ to $[-1,1]$. If the score is above 0, the emotion is regarded as positive; otherwise, it is regarded as negative. The greater the absolute value of the sentiment score, the stronger the emotion.

3. Results

3.1. Factor discovery

3.1.1 Factor discovery results

Blei, the originator of the LDA model, pointed out that the number of topics in the corpus is determined by its perplexity. The perplexity is the predicted average number of equally likely words in certain positions. A lower perplexity means a better predictive performance. Fig.3. shows the predictive power of LDA model in terms of the per-word perplexity as a function of number of factors. Perplexity decreases with the increase of factors, and finally tends to be stable. When T is less than 20, the perplexity reaches the minimum at 12. Perplexity decreases much more slowly when $T > 20$ and it is very difficult to interpret the meaning of factor when the topic number is too large. Therefore, in this study, we set the number of factors T to 12 in order to keep a balance between the perplexity and the interpretability. The first 12 keywords in each of the 12 classified factors are selected for the interpretation of that factor. The drug review factor recognition results are shown in Table 3.

Table 3. B2C mail-order pharmacy review factors discovered by LDA

Factors	Keywords	Interpretation
Factor 1	logistics, packing, professional, attentively, dry glue, drug name, pharmaceutical factory, arrival of goods, paste, protect, standard, send it over	Professional Logistics Packing (PLP)
Factor 2	genuine, brand, no problem, trust, have faith in, next time, purchase, quality of drugs, guarantee, needs, drug, verify	Trustworthy Drug Quality (TDQ)
Factor 3	inside, box, packing, intact, no damage, logistics, awesome, nice, buy medicine, liquid, bubble wrap, protect	Complete Packing in Logistics (CPL)
Factor 4	expensive, price, more than, pharmacy, physical store, elsewhere, dosage, spend, offline, hospital, extra money, profit	Expensive (E)
Factor 5	favorable, price, drug, website, save, registration fee, chronic, bring benefit to, cheap, pharmacy, many times, bottom price	Affordable (A)
Factor 6	dispatch, too slow, speed, too long, wait, unable, receive, drug, bad, for several days, recover, inquire	Slow Dispatch Speed (SDS)
Factor 7	customer service, pass the buck, service, manner, busy, disappointed, solve, problem, adjudicate, irrelevance, unacceptably, cannot understand	Customer service Did Not Solve The Problem (DSP)
Factor 8	slow, logistics, transport, wait, not satisfied with, unable, today, receive, delay, time, for several days, discover	Slow Transport Speed (STS)
Factor 9	fast, speed, logistics, shopping, experience, awesome, pleased, platform, satisfied, receive, today, good	Satisfactory Logistics Speed (SLS)
Factor 10	discount, drug, promotion, satisfactory, high performance-price ratio, awesome, cheap, price, gifts, favorable, bottom price, benefit	Satisfactory Promotion (SP)
Factor 11	take effect, awesome, genuine, confirm, much better, symptom, alleviate, well, pharmacy, same, dose, satisfied	Satisfactory Drug Effects (SDE)
Factor 12	customer service, quickly, answer, awesome, place an order, response, serious, in time, drug, patient, at once, inquire	Quick Response of Customer service (QR)

3.2. Factor classification

3.2.1 Factor classification results

Based on the review of the factors affecting patient satisfaction with B2C e-commerce websites, this study analyzes the 12 factors discussed in the previous section and finds that the 12 factors are mainly discussed from four perspectives –logistics, product, price and staff. The factors identified in the review data do not include factors related to an information and system perspective. The reviews of the pharmaceutical e-commerce websites represented by JD.COM and J1.COM include little discussion of information or system factors. It may be that e-commerce has operated in a mature mechanism and that the e-commerce websites chosen for analysis are readily accessible and easy to use. The use of the websites, the integrity and authenticity of the product information, payment security and information security have reached a certain standard and are relatively mature and stable; because patients are quite accustomed to this, there is little discussion of these factors.

As shown in Fig. 4 above, among the 12 factors, QR, E, and SLS accounted for the greatest proportion of the reviews, and the majority of reviews on JD.COM and J1.COM dealt with one or more of these three factors. QR represents the factor Quick Response of Customer service, E represents the factor Expensive, and SLS represents the factor Satisfactory Logistics Speed.

The proportions of drug reviews that fall into the categories of logistics, drug effects, drug price, and customer service are shown in Fig. 4. The figure shows that 5 of the factors are related to logistics; these factors also yield the most relevant drug reviews and account for 38.5% of the total reviews. The number of factors related to drug prices is second highest, with 3 factors, and reviews related to drug prices make up 25.5% of the total reviews. Customer service and drug effects each have 2 related factors; reviews related to drug effects account for a smaller percentage (13.95%) of the total reviews.

3.2.2 Differences in the factor distributions of reviews posted at JD.COM and J1.COM

A comprehensive analysis of Figs. 4, 5, and 6 shows the following:

- (1) When purchasing medicines, patients pay the most attention to logistics, followed by drug prices and customer service, and they pay the least attention to drug effects.
- (2) The proportion of reviews dealing with the factor of logistics is higher at J1.COM than at JD.COM, mainly because patients engage in extensive discussion of the slow dispatch and transport provided by J1.COM.
- (3) With respect to the evaluation of drug prices, the number of reviews dealing with the factor of drug prices is much greater at JD.COM than at J1.COM, and there are fewer reviews on the Satisfactory Promotion factor at JD.COM than at J1.COM.
- (4) With respect to the evaluation of customer service, the proportion of reviews dealing with the factor Quick Response of Customer service is much larger at JD.COM than at J1.COM, and the proportion of reviews with the factor Customer Service Did Not Solve the Problem is smaller at JD.COM than at J1.COM.

3.3. Sentiment analysis

3.3.1 Sentiment analysis results

The final results in Table 4 shows that patients are really satisfied with the two B2C mail-order pharmacy, as the positive sentiment proportion is approximately 90.71%.

Table 4. The results of the sentiment analysis of the mail-order pharmacy reviews

Sentiment Polarity	Pharmacy	Count	Percentage	Total
Negative Sentiment	JD.COM	5139	9.64%	9.29%
	J1.COM	4818	8.95%	
Positive Sentiment	JD.COM	48167	90.36%	90.71%
	J1.COM	49013	91.05%	

A comprehensive analysis of Table 4, Figs. 7,8 shows the following:

- (1) Patients still maintain positive sentiment for JD.COM and J1.COM. The patients are satisfied with the drug effects and with the customer service provided by JD.COM and J1.COM. However, there are still some opinions on logistics and drug prices.
- (2) The logistics and customer service provided by JD.COM are more satisfying to patients than those provided by J1.COM. The drug prices and drug effects obtained through J1.COM are more satisfying to patients than those obtained through JD.COM.
- (3) Positive sentiment for JD.COM regarding logistics speed and customer service response is far greater than that for J1.COM, but JD.COM's negative sentiment on drug prices is higher than that of J1.COM. The positive sentiment for J1.COM regarding the Satisfactory Promotion factor is greater than that for JD.COM.

3.3.2 Evaluation of sentiment analysis results

To evaluate the accuracy of our model performance, we employed the receiver operating characteristic curve (ROC) to obtain the true positive rate and the false positive rate. The true positive rate means that the rate of positive comments which are correctly identified as positive by the algorithm. While the false positive rate means that the rate of negative comments which are mistakenly identified as positive. Firstly, we randomly selected 500 reviews labeled as positive or negative by two researchers and we used these labeled data as the test set. Then, we used the sentiment scores from SnowNLP as the prediction set. After preparing the test set and the prediction set, the ROC curve could be obtained and Area Under Curve (AUC) could be calculated. AUC represents the accuracy of the classifier. If the value of the AUC is between 0.5 and 1, the accuracy of this classifier is better than that of a random guess. In our case, the AUC is 0.7112, which indicates that the result of the sentiment score is satisfactory. Fig.9. shows the ROC curve of our study.

4. Discussion

In this study, an algorithm based on the use of the LDA topic model to obtain the factors of B2C mail-order pharmacy reviews was proposed. The 12 factors of B2C mail-order pharmacies were mined and classified into four major factors – logistics, drug prices, drug effects, and customer service. The results of data mining show that patients pay the most attention to logistics when purchasing drugs, followed by drug prices and customer service, and that they pay the least attention to drug effects.

In reviews on J1.COM, patients extensively discuss the slow dispatch and transport speed. The logistics of vertical B2C mail-order pharmacies are a problem that needs special attention. Although vertical B2C mail-order pharmacies are professional in terms of medicine and professional packing experience, they must rely on third-party logistics because they do not have their own delivery services. This makes it

difficult to control the delivery time and the logistics speed. For many years, JD.COM has been proud of its self-built logistical system, which uses multiple warehouses and direct distribution, so the speed of its logistics can often satisfy patients.

Concerning the reviews of drug prices, J1.COM is a B2C mail-order pharmacy of integrated e-commerce formed by an offline pharmacy + Internet and offers a greater price advantage than JD.COM. As a major feature of e-commerce, low-cost and varied promotional activities are also of particular concern to patients. Patients often compare the prices of drugs on e-commerce websites with the prices at physical pharmacies, and mail-order pharmacies usually offer a price advantage.

The reviews of customer service reflect the fact that the diversified integrated e-commerce platform has accumulated extensive customer service experience in the sales of home appliances, 3C and other products; its customer service staff is also more adequate and offers better customer service response speed and service quality compared with that of vertical e-commerce B2C mail-order pharmacies. Due to the lifting of the ban on mail-order pharmacy in China for not so long, patients may have questions about the quality of mail-order medicines and the mechanisms of purchase. Because they need timely responses from customer service, patients pay more attention to customer service.

In the reviews of drug effects, patients basically produce positive evaluations for both JD.COM and J1.COM. On one hand, because JD.COM and J1.COM have been well known in China for many years and regardless of whether they are B2C mail-order pharmacies of integrated e-commerce or vertical mail-order pharmacies, they are approved and supervised by the government. They offer genuine guarantees. Since vertical B2C mail-order pharmacies are often professional medical websites, their ability to recommend appropriate drugs based on symptoms is more professional than that of B2C mail-order pharmacies of integrated e-commerce, so patients will be more satisfied.

Patients still maintain positive feelings about JD.COM and J1.COM. The patients are satisfied with the drug effects and with the customer service offered by JD.COM and J1.COM. However, there are still some opinions on logistics and drug prices. The logistics and customer service provided by B2C mail-order pharmacies of integrated e-commerce such as JD.COM are more satisfying to patients than those of vertical mail-order pharmacies such as J1.COM. The prices and drug effects of vertical B2C mail-order pharmacies such as J1.COM make patients more satisfied.

5. Conclusions

This study has many practical theoretical and managerial implications. First, this paper comprehensively uses machine learning methods and theoretical analysis to explore the factor classification and sentiment of B2C mail-order pharmacy patients' reviews.

For unsupervised factor mining, previous studies mainly used predefined theoretical models and structural equations based on questionnaire data or methods using coded text analysis under unscheduled models. These two methods, which are actually artificial or semimanual predefined coding

methods, are time-consuming and laborious, especially when the research includes more than 100,000 pieces of data, and the efficiency of using the manual method is very low. This paper uses an unsupervised machine learning algorithm to automatically identify the factors of B2C mail-order pharmacy patient reviews based on the LDA model. Then, based on a review of the factors affecting patient satisfaction with B2C mail-order pharmacies, the factor recognition results are divided into four major categories.

Second, this article is of great significance with respect to the positioning of patients' needs among the two types of B2C mail-order pharmacies, the continuous improvement of the functions of B2C mail-order drug sales, and the improvement of health services level.

The current work indicates that the most important task for both B2C mail-order pharmacies of integrated e-commerce and vertical B2C mail-order pharmacies is to enhance the logistics level, improve the delivery and transportation speed, and develop self-built logistics as much as possible. At the same time, the mail-order pharmacies can also cooperate with offline pharmacies to realize the O2O mode of medical e-commerce. Due to the denser characteristics of offline pharmacies, the efficiency of distribution can be improved by means of offline pharmacies [56].

B2C e-commerce has an obvious price advantage because it uses flat transaction channels and has fewer circulation links than do physical stores. B2C mail-order pharmacies should continue to maintain their price advantage. Additionally, aging is showing an increasing trend in China. There are many patients with chronic diseases, and the demand for medical products is high. For some chronic diseases that are treated using drugs with high repurchase rates or drugs that need to be kept at home, the two types of B2C mail-order pharmacies can increase the patient viscosity or patient repurchase rate through regular sales.

Customer service should pay attention to cultivating employees' service attitudes. In particular, vertical B2C mail-order pharmacies should improve the timeliness of their customer service responses and their problem-solving abilities. B2C mail-order pharmacies of integrated e-commerce should especially improve the basic expertise on drugs. If necessary, they should hire professional pharmacists to work in customer service who can answer questions in a professional manner and thereby improve patient satisfaction, loyalty and trust. For problems involving these aspects, B2C mail-order pharmacies should analyze the causes of patient concerns and correct their strategies in a timely manner. In the era of big data, a complete customer relationship management (CRM) system should also be established. China has a large population, and the establishment of patient health records still has great room for development and application in the future [57].

Abbreviations

B2C

Business to Customer

LDA

Latent Dirichlet Allocation
OTC
over-the-counter
WHO
World Health Organization
AHP
Analytic Hierarchy Process
TF
term frequency
IDF
inverse document frequency
ROC
operating characteristic curve
AUC
Area Under Curve
CRM
customer relationship management

References

- [1] von Rosen A J, von Rosen F T, Tinnemann P, et al. Sexual health and the internet: cross-sectional study of online preferences among adolescents[J]. *Journal of medical Internet research*, 2017, 19(11): e379.
- [2] Fox S, Duggan M. Pew Research Center's Internet & American Life Project. 2013 Jan 15. Health Online 2013 - 35% of U.S. adults have gone online to figure out a medical condition; of these, half followed up with a visit to a medical professional URL: http://www.pewinternet.org/files/old-media/Files/Reports/PIP_HealthOnline.pdf
- [3] Andreassen H, Bujnowska-Fedak M, Chronaki C, Dumitru R, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. *BMC Public Health* 2007 Apr 10;7:53
- [4] Takahashi Y, Ohura T, Ishizaki T, Okamoto S, Miki K, Naito M, et al. Internet use for health-related information via personal computers and cell phones in Japan: a cross-sectional population-based survey. *J Med Internet Res* 2011 Dec 14;13(4):e110
- [5] Jacobs W, Amuta A O, Jeon K C. Health information seeking in the digital age: An analysis of health information seeking behavior among US adults[J]. *Cogent Social Sciences*, 2017, 3(1): 1302785.
- [6] Gawron LM, Turok DK. Pills on the World Wide Web: reducing barriers through technology. *Am J Obstet Gynecol* 2015 Oct;213(4):500.e1-500.e4.

- [7] Akter S, D'Ambra JP. Service quality of mHealth platforms: development-validation of a hierarchical model using PLS. *Electronic Markets* 2010;20:227.
- [8] Orizio G, Schulz P, Domenighini S, Caimi L, Rosati C, Rubinelli S, et al. Cyberdrugs: a cross-sectional study of online pharmacies characteristics. *Eur J Public Health* 2009 Aug;19(4):375-377.
- [9] Fox S, Duggan M. Health Online. 2013 Jan 15. Information Triage - For one-third of U.S. adults, the internet is a diagnostic tool URL: <http://www.pewinternet.org/2013/01/15/information-triage/>
- [10] Martin K, Papagiannidis S, Li F, et al. Early challenges of implementing an e-commerce system in a medical supply company: A case experience from a knowledge transfer partnership (KTP)[J]. *International Journal of Information Management*, 2008, 28(1):68-75.
- [11] Fung CH, Woo H, Asch S. Controversies and legal issues of prescribing and dispensing medications using the Internet. *Mayo Clin Proc* 2004 Feb;79(2):188-194.
- [12] Orizio G, Merla A, Schulz PJ, Gelatti U. Quality of online pharmacies and websites selling prescription drugs: a systematic review. *J Med Internet Res* 2011 Sep 30;13(3):e74
- [13] US Congress. 2008. Ryan Haight Online Pharmacy Patient Protection Act of 2008 URL: <https://www.govtrack.us/congress/bills/110/hr6353>
- [14] Mackey TK, Nayyar G. Digital danger: a review of the global public health, patient safety and cybersecurity threats posed by illicit online pharmacies. *Br Med Bull* 2016 Jun;118(1):110-126
- [15] Gabay M. Regulation of Internet Pharmacies: A Continuing Challenge. *Hosp Pharm* 2015 Sep;50(8):681-682
- [16] Dudley J. *Entering the Digital Battleground*. In: *Mail Order and Internet Pharmacy in Europe*. Kidderminster UK: James Dudley International Ltd; 2014.
- [17] Cohen JC. Public policy implications of cross-border Internet pharmacies. *Manag Care* 2004 Mar;13(3 Suppl):14-16
- [18] Berkrot B. Fake Avastin shows little protection of drug supply[J]. *Reuters*. March, 2012, 12.
- [19] Blackstone EA, Fuhr JJ, Pociask S. The health and economic effects of counterfeit drugs. *Am Health Drug Benefits* 2014 Jun;7(4):216-224
- [20] Howard D. A silent epidemic: protecting the safety and security of drugs. *Pharmaceutical Outsourcing* 2010 Jul 01;11(4):16-18
- [21] Dorsey PJ, Hellstrom WJ. *Medscape*. 2007. The Illicit Sale of Medications for the Treatment of Erectile Dysfunction URL: http://www.medscape.com/viewarticle/566897_3

- [22]Bate R. The Deadly World of Fake Drugs. Foreign Policy, Washington Post 2008;168:56-65.
- [23]Baker L, Wagner T, Singer S, Bundorf M. Use of the Internet and e-mail for health care information: results from a national survey. JAMA 2003 May 14;289(18):2400-2406.
- [24]Desai K, Chewning B, Mott D. Health care use amongst online buyers of medications and vitamins. Res Social Adm Pharm 2015 Nov;11(6):844-858.
- [25] Chen Y, Xie J. Online Patient Review: Word-of-Mouth as a New Element of Marketing Communication Mix[J]. Management Science, 2008, 54(3):477-491.
- [26] Duan W, Gu B, Whinston A B. Do online reviews matter? – An empirical investigation of panel data[J]. Decision Support Systems, 2008, 45(4):1007-1016.
- [27] Ye Q, Law R, Gu B, et al. The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings[J]. Computers in Human Behavior, 2011, 27(2):634-639.
- [28] Kumar N, Benbasat I. Research Note: The Influence of Recommendations and Patient Reviews on Evaluations of Websites[J]. Information Systems Research, 2006, 17(4):425-439.
- [29] Vermeulen I E, Seegers D. Tried and tested: The impact of online hotel reviews on patient consideration[J]. Tourism Management, 2009, 30(1):123-127.
- [30] Fiske H W M. A long/high view from a stationary geo satellite on project cost control. (A modern birdseye view)[J]. Engineering Costs & Production Economics, 2005, 5(2):81-87.
- [31] Guo Y, Barnes S J, Jia Q. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichletallocation[J]. Tourism Management, 2017, 59:467-483.
- [32] Kahraman C, Onar S Ç, Öztayşi B. B2C marketplace prioritization using hesitant fuzzy linguistic AHP[J]. International Journal of Fuzzy Systems, 2018, 20(7): 2202-2215.
- [33] Barton T, Bruna T, Kordik P. Chameleon 2: An Improved Graph-Based Clustering Algorithm[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2019, 13(1): 10.
- [34] van Horn A, Weitz C A, Olszowy K M, et al. Using multiple correspondence analysis to identify behaviour patterns associated with overweight and obesity in Vanuatu adults[J]. Public health nutrition, 2019: 1-12.
- [35] Szymanski D M, Hise R T. E-satisfaction: an initial examination[J]. Journal of Retailing, 2000, 76(3):309-322.
- [36] Lee M.K.O. A Comprehensive Model of Internet Patient Satisfaction[R].City University of HongKong, 1999.

- [37] Lin C C, Wu H Y, Chang Y F. The critical factors impact online patient satisfaction[J]. *Procedia Computer Science*, 2011, 3:276-281.
- [38] Liu X, He M, Gao F, et al. An empirical study of online shopping patient satisfaction in China: a holistic perspective[J]. *International Journal of Retail & Distribution Management*, 2008, 36(11):919-940.
- [39] Yooncheong Cho, Il Im, Roxanne Hiltz et al. An Analysis of Online Patient Complaints: Implications for Web Complaint Management[C]. *Proceedings of the 35th Hawaii International Conference on System Sciences*, 2002.
- [40] Torkzadeh G, Dhillon G. Measuring Factors that Influence the Success of Internet Commerce[J]. *Information Systems Research*, 2002, 13(2):187-204.
- [41] Liao Z, Cheung M T. Internet-based e-shopping and patient attitudes an empirical study[M]. Elsevier Science Publishers B. V. 2001.
- [42] Mckinney V , Yoon K , Zahedi, Fatemeh "Mariam". The Measurement of Web-Patient Satisfaction: An Expectation and Disconfirmation Approach[J]. *Information Systems Research*, 2002, 13(3):296-315.
- [43] Kim S, Stoel L. Apparel retailers: website quality dimensions and satisfaction[J]. *Journal of Retailing & Customer services*, 2004, 11(2):109-117.
- [44] Wolfinbarger M, Gilly M C. eTailQ: dimensionalizing, measuring and predicting etail quality[J]. *Journal of Retailing*, 2003, 79(3):183-198.
- [45] Koivumaki T, Ristola A, Kesti M. Predicting patient acceptance in mobile services: empirical evidence from an experimental end user environment.[J]. *International Journal of Mobile Communications*, 2006, 4(4):418-435.
- [46] Egbert J, Schnur E. The role of the text in corpus and discourse analysis[J]. *Corpus approaches to discourse. A critical review*, 2018: 158-170.
- [47] Liu W. Automatically refining synonym extraction results: Cleaning and ranking[J]. *Journal of Information Science*, 2018: 0165551518799640.
- [48] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. *J Machine Learning Research Archive*, 2003, 3:993-1022.
- [49] Huang T C, Hsieh C H, Wang H C. Automatic meeting summarization and factor detection system[J]. *Data Technologies and Applications*, 2018, 52(3): 351-365.
- [50] Poria S, Majumder N, Hazarika D, et al. Multimodal Sentiment Analysis: Addressing Key Issues and Setting Up the Baselines[J]. *IEEE Intelligent Systems*, 2018, 33(6): 17-25.

- [51] Slamet C, Atmadja A R, Maylawati D S, et al. Automated text summarization for Indonesian article using Vector Space Model[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2018, 288(1): 012037.
- [52] Fan H, Qin Y. Research on Text Classification Based on Improved TF-IDF Algorithm[C]//2018 International Conference on Network, Communication, Computer Engineering (NCCE 2018). Atlantis Press, 2018.
- [53] Griffiths T L, Steyvers M. Finding scientific factors.[J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101 Suppl 1(1):5228.
- [54] Ku L W, Wu T H, Lee L Y, et al. Construction of an evaluation corpus for opinion extraction[J]. Ntcir, 2005:513–520.
- [55] Rost Virtual Learning Team Rost Sentiment Analysis Tool[EB/ OL][2012-08-31].<http://www.fanpq.com/soft/upload-soft/ROSTEA.rar>
- [56] Su L, Li T, Hu Y, et al. Factor analysis on marketing mix of online pharmacies - Based on the online pharmacies in China[J]. Journal of Medical Marketing, 2013, 13(2):93-101.
- [57] Rosin A J, Sonnenblick M. Autonomy and paternalism in geriatric medicine. The Jewish ethical approach to issues of feeding terminally ill patients, and to cardiopulmonary resuscitation[J]. Journal of Medical Ethics, 1998, 24(1):44.

Declarations

-Ethics approval and consent to participate

Not applicable

-Consent for publication

As a part of the consent process, all participants have agreed that collected information can be used for research and publication.

-Availability of data and material

The datasets used in the current study are available from the corresponding author on reasonable request.

As a part of the consent process, all participants have agreed that collected

information can be used for research and publication.

-Competing interests

The authors declare that they have no competing interests.

-Funding

This research was supported by the Natural Science Foundation of Shanghai under grant number 19ZR1419400.

-Authors' contributions

All authors contributed to the work described in this manuscript. All authors have approved the final version of the manuscript. The detailed division of labor was as follows:

Jingfang Liu provided the original research idea.

Yingyi Zhou and Jingfang Liu performed the data analysis and wrote the manuscript.

Wei Zhang and Xiaoyan Jiang provided advice and expertise throughout the research and creation of the manuscript.

Xiaoyan Jiang prepared the empirical data and wrote part of the manuscript.

-Acknowledgements

The authors wish to thank the National Natural Science Foundation of China (71673169) and the Innovation Project of the Shandong Academy of Medical Sciences for their financial support.

The authors wish to thank the Natural Science Foundation of Shanghai under grant number 19ZR1419400 for their financial support.

Figures

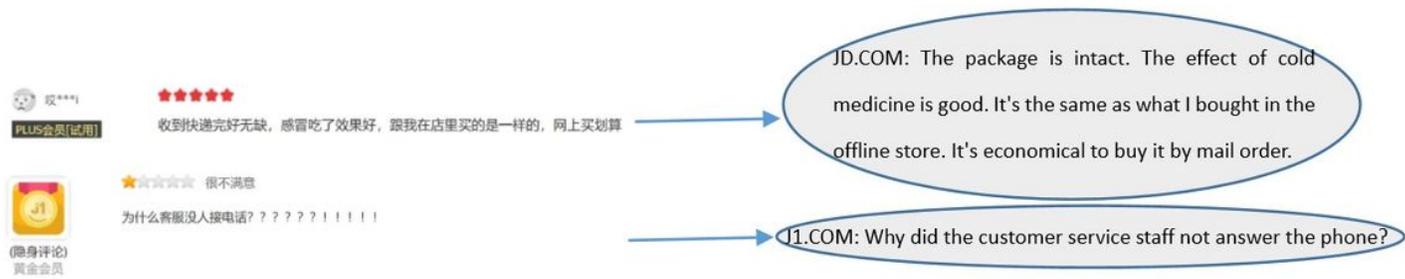


Figure 1

Examples of cold medicine: "999" reviews on two websites

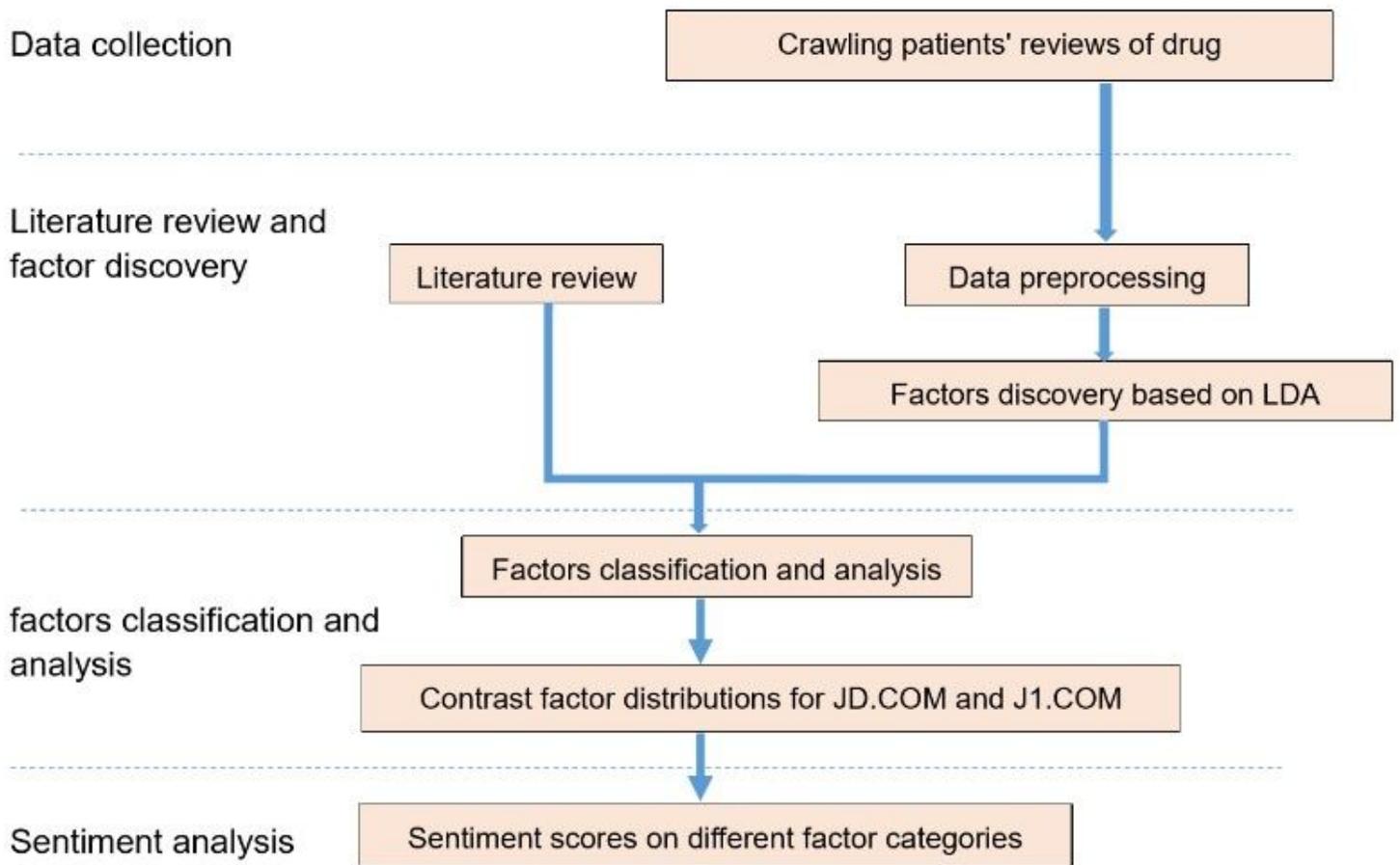


Figure 2

Methodological framework

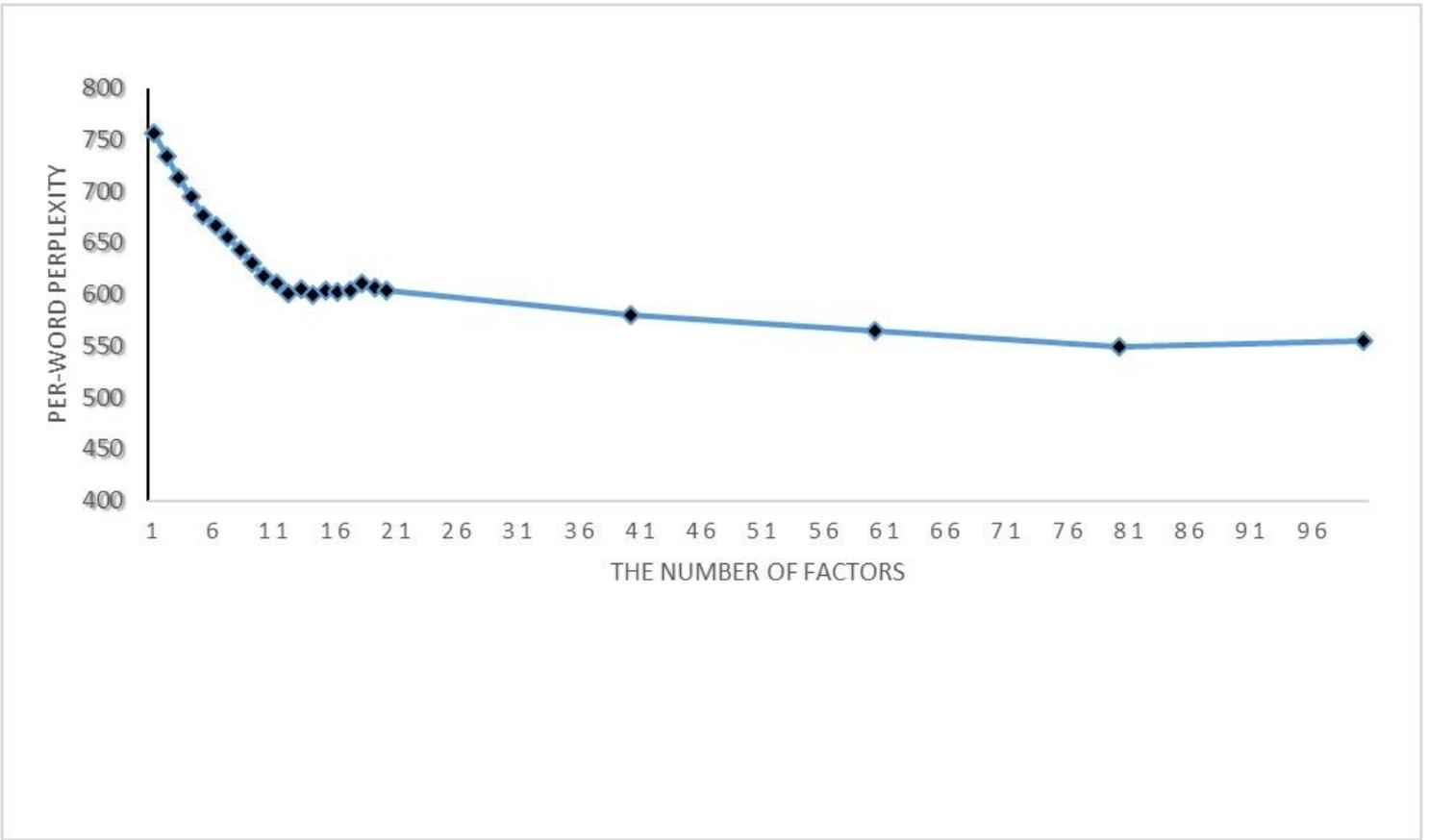


Figure 3

Per-word perplexity as a function of number of factors

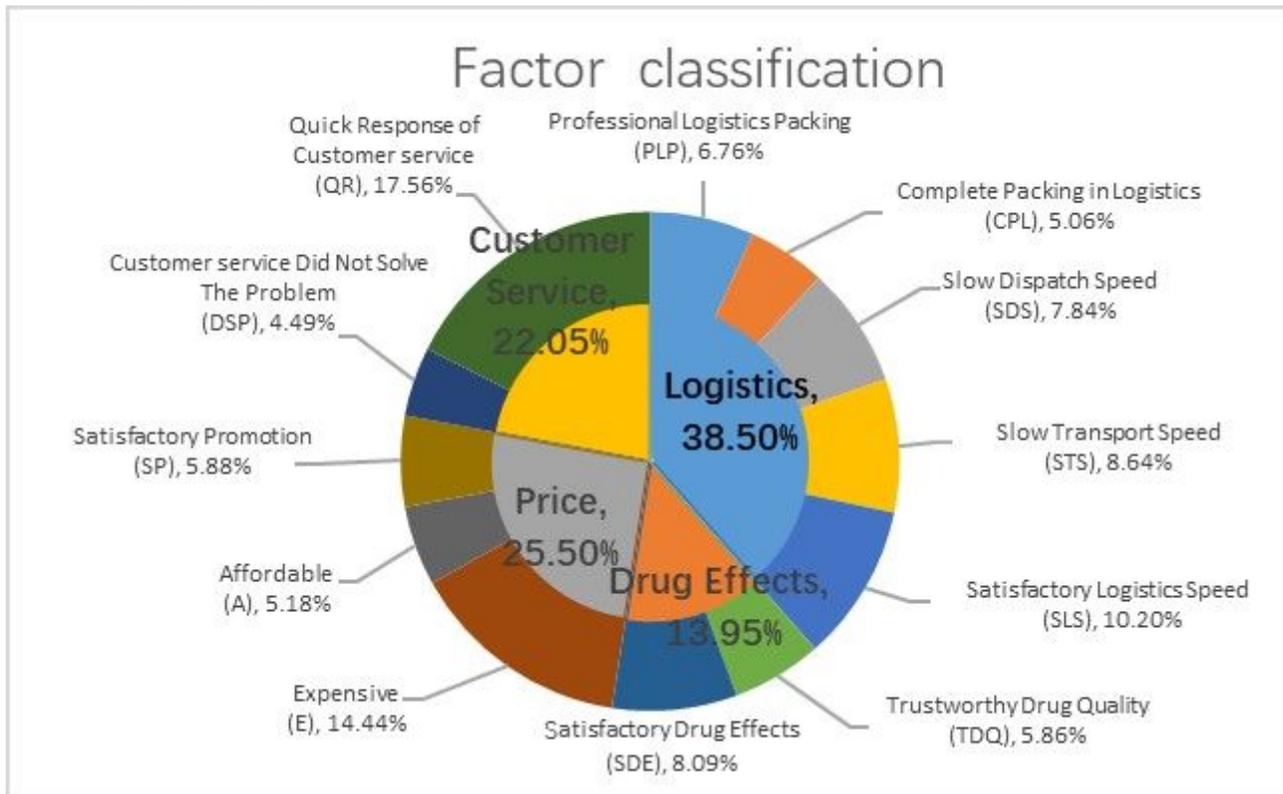


Figure 4

Percentage of reviews in each of the 12 factor categories

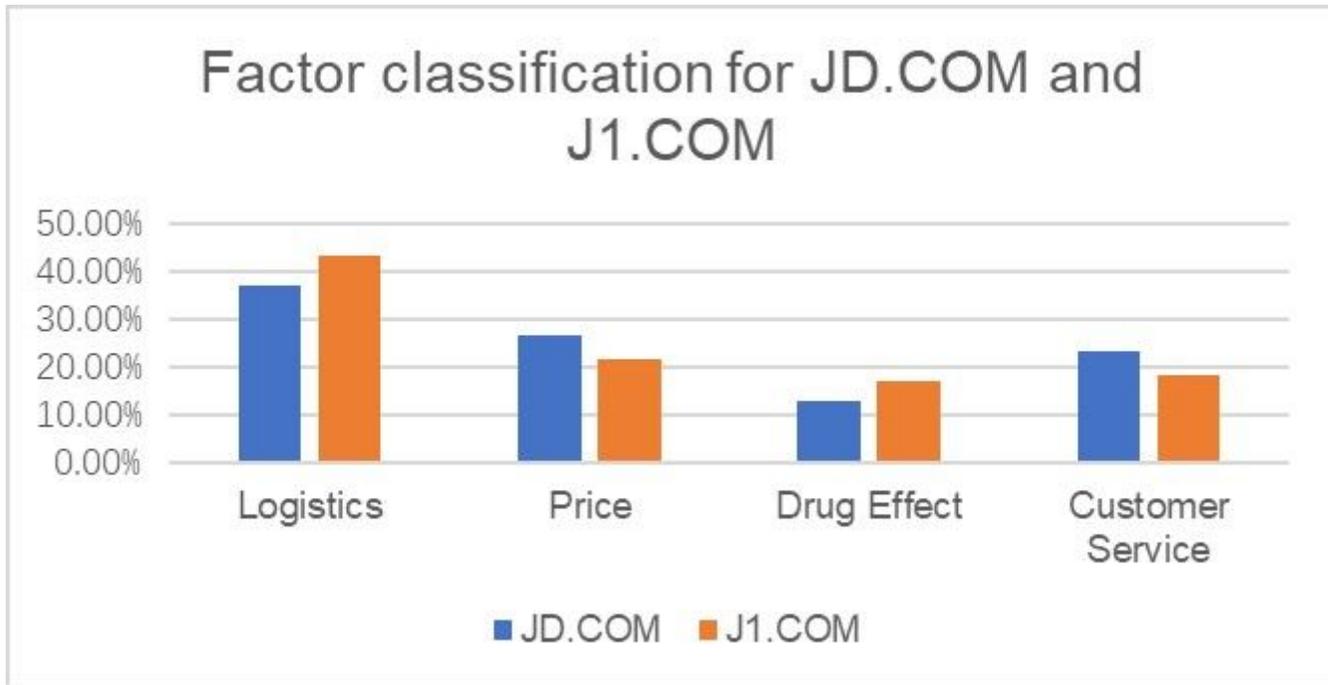


Figure 5

Factor classification for JD.COM and J1.COM

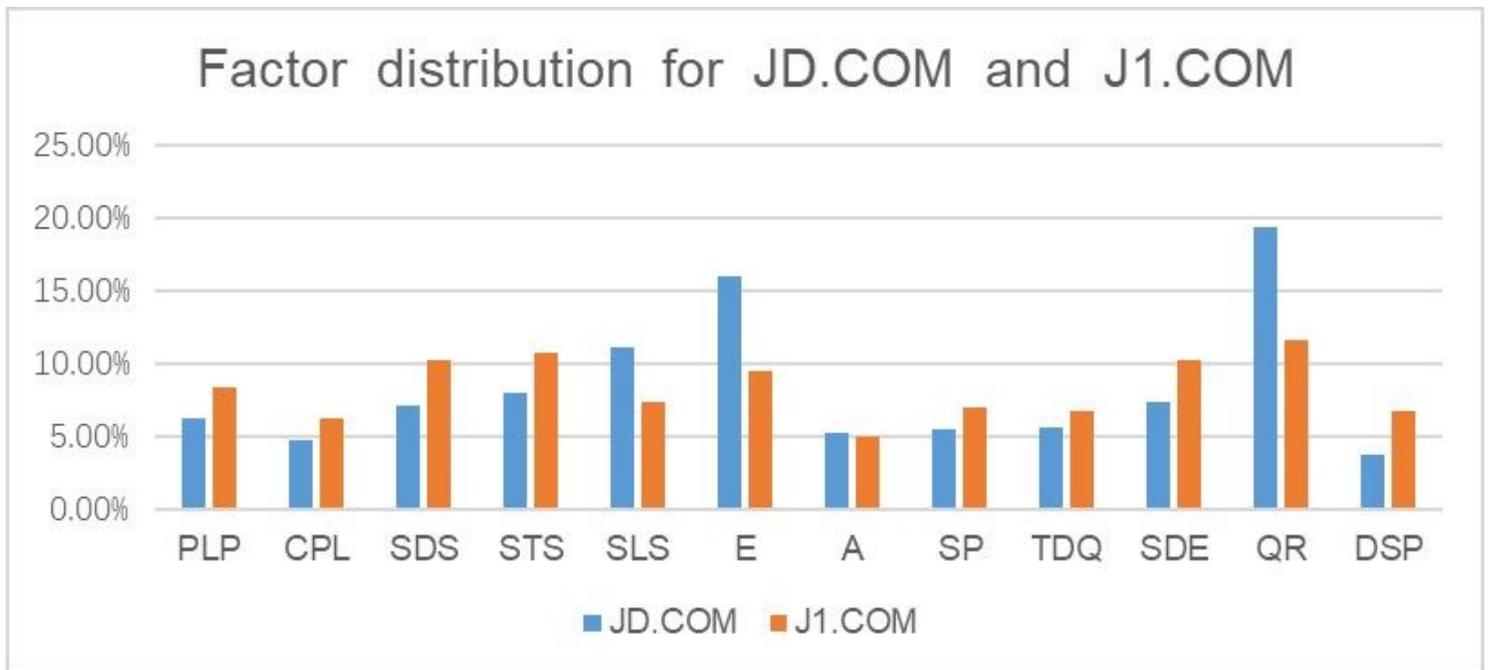


Figure 6

Factor distribution for JD.COM and J1.COM

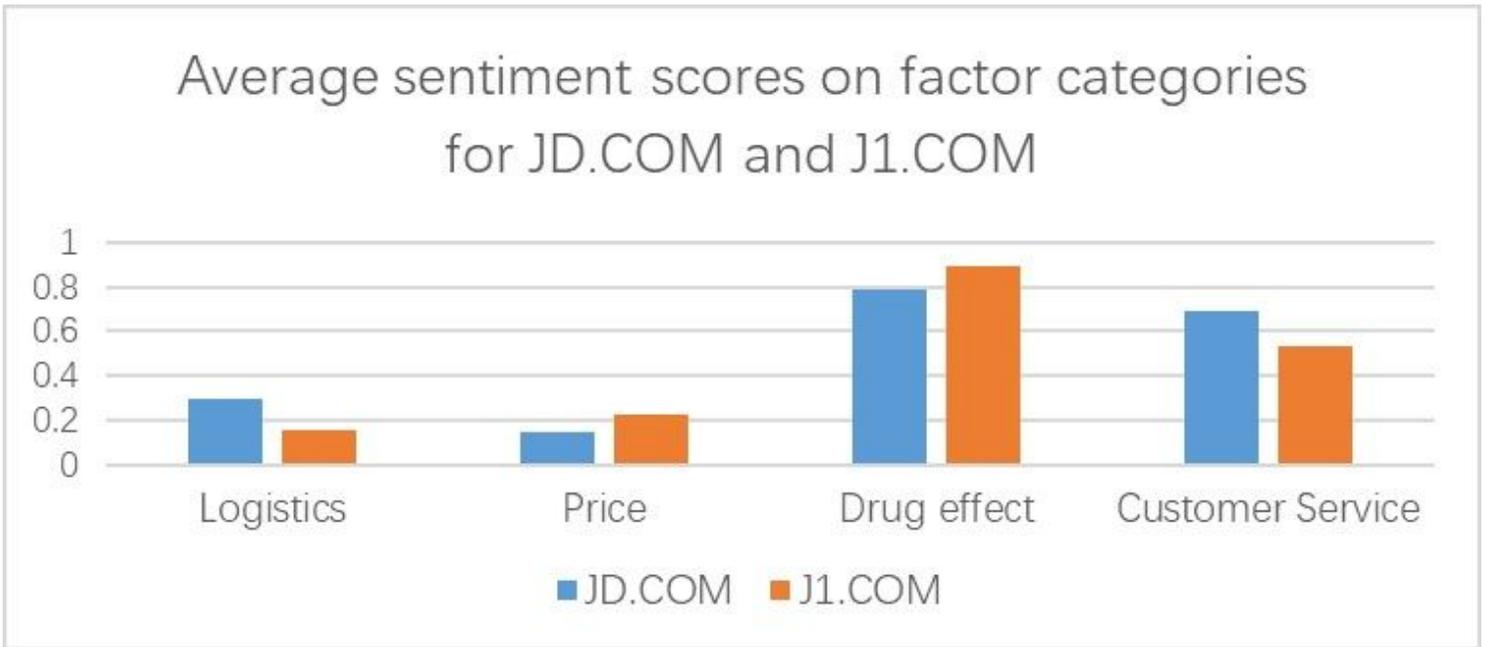


Figure 7

Average sentiment scores on factor categories for JD.COM and J1.COM

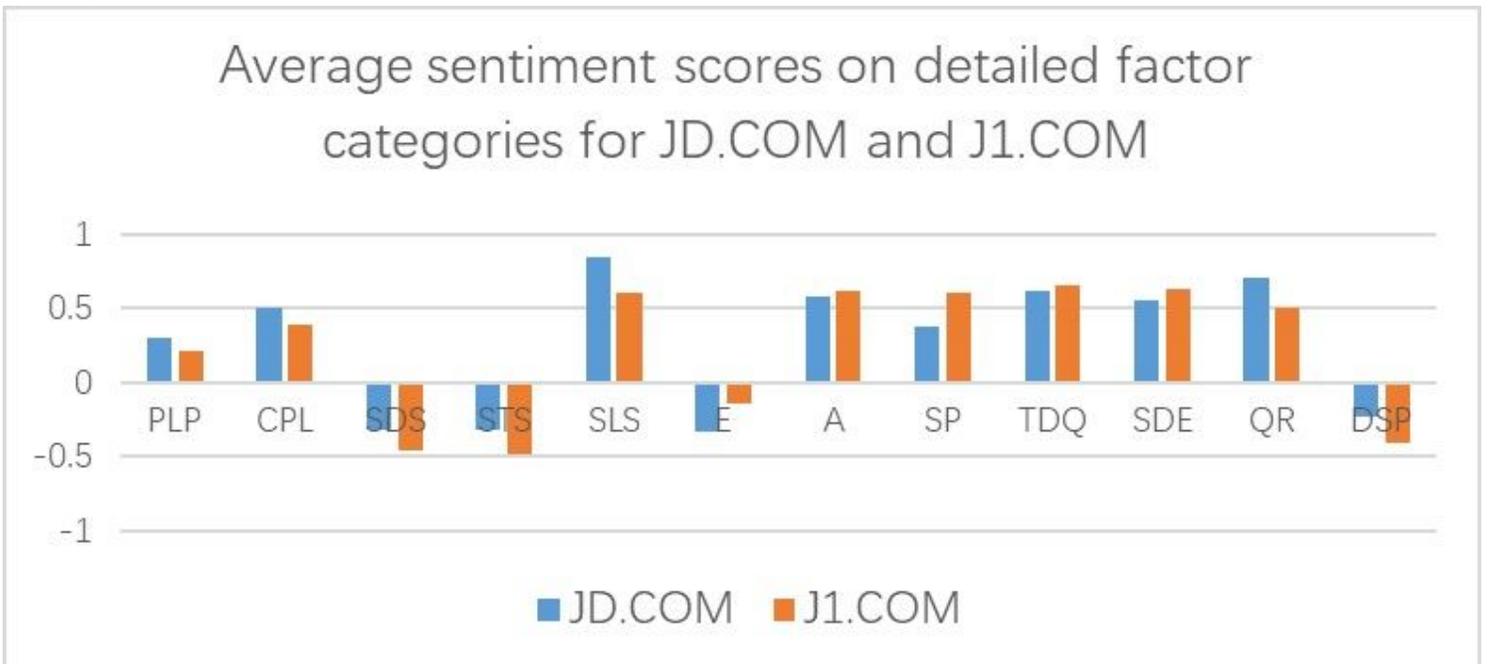


Figure 8

Average sentiment scores on detailed factor categories for JD.COM and J1.COM

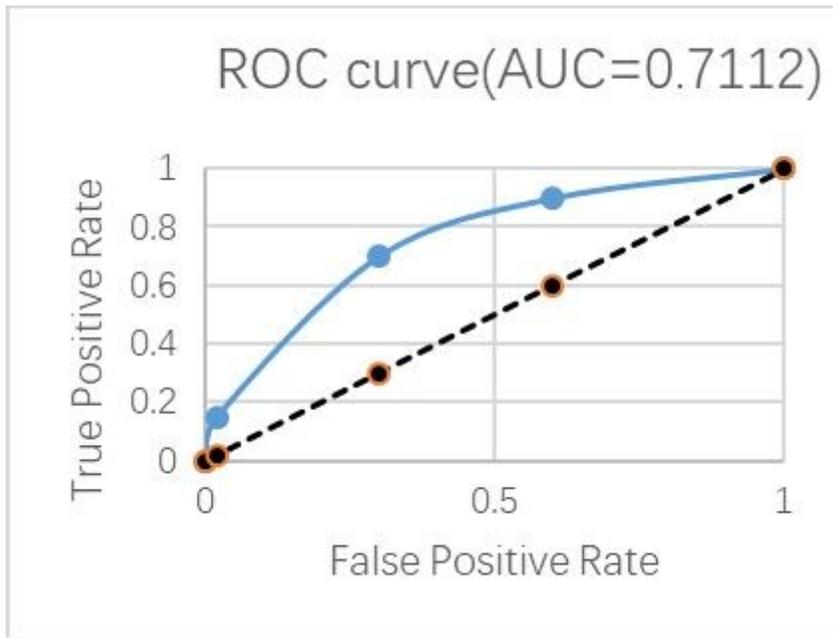


Figure 9

ROC curve for evaluating the sentiment analysis