TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 09/10 pp293-305
Volume 20, Number 3, June 2015

# Analysis of Review Helpfulness Based on Consumer Perspective

Yuanlin Chen, Yueting Chai*, Yi Liu, and Yang Xu

**Abstract:** When consumers make purchase decisions, they generally refer to the reviews generated by other consumers who have already purchased similar products in order to get more information. Online transaction platforms provide a highly convenient channel for consumers to generate and retrieve product reviews. In addition, consumers can also vote reviews perceived to be helpful in making their decision. However, due to diverse characteristics, consumers can have different preferences on products and reviews. Their voting behavior can be influenced by reviews and existing review votes. To explore the influence mechanism of the reviewer, the review, and the existing votes on review helpfulness, we propose three hypotheses based on the consumer perspective and perform statistical tests to verify these hypotheses with real review data from Amazon. Our empirical study indicates that review helpfulness has significant correlation and trend with reviewers, review valance, and review votes. In this paper, we also discuss the implications of our findings on consumer preference and review helpfulness.

**Key words:** consumer preference; online decision making; review helpfulness; behavior analysis

## 1 Introduction

With the rapid development of information and network technologies, online transactions are gradually replacing traditional face-to-face transactions and have become the most common trading method for consumers. It is necessary for e-commerce platforms to provide a channel for consumers to express their opinions—which are usually called product reviews—after completing transactions. When consumers decide the product they wish to purchase and the site from which they want to purchase, they tend to retrieve information about alternative products to help make decisions. Generally, this information comes from two main sources. The first source is from sellers who demonstrate their products with descriptions on platforms. The second source is from the consumers who have bought similar products, which is known as Word-Of-Mouth (WOM). However, driven by economic interests, sellers always try to conceal unfavorable information or even exaggerate and fabricate favorable information about products. Consequently, the information asymmetry between sellers and consumers is enhanced until consumers receive real products. Consumers are often disappointed because the product they receive is different from their expectations. Fortunately, after receipt confirmation, consumers have an opportunity to review their transactions from several aspects, such as product quality or promptness of service. Reviews on most platforms comprise numerical ratings and textual comments. Numerical ratings usually range from 1 to 5 stars. Some platforms even provide more than one numerical rating interface for different aspects. For instance, at Tmall, which is the largest B2C e-commerce platform in China, numerical ratings exist for description matching, service attitude, delivery speed,

• Yuanlin Chen, Yueting Chai, Yi Liu, and Yang Xu are with National Engineering Laboratory for E-Commerce Technologies, Tsinghua University, Beijing 100084, China. E-mail: hustom.chen@gmail.com; chaiyt@mail.tsinghua.edu.cn; yiliu@mail.tsinghua.edu.cn; eak@mail.tsinghua.edu.cn.
• Yang Xu is also with DNSLAB, China Internet Network Information Center, Beijing 100190, China.
∗ To whom correspondence should be addressed.
Manuscript received: 2015-02-12; revised: 2015-04-07; accepted: 2015-04-15

and logistics speed. With textual comments, consumers can remark any aspect of purchased products and any detail of their trading experience. Customer reviews are public—the comments posted by customers can be viewed by anyone on the platform including sellers and other customers.

In order to reduce information asymmetry between sellers and consumers before purchase, it is effective to encourage consumers to conduct reviews to the maximum extent possible. Existing work has explored motivations of consumers to review after purchase. The first motivation of consumers is to express feelings. Buying a product online is similar to a trading experience for consumers. This experience influences a consumer's psychological state to cause a desire to share with others, especially when the experience is positive[1]. The second motivation is to provide information that helps other consumers. Trading experience is also a process of obtaining information about products.

If a consumer buys a product and is satisfied with its quality, it will benefit the community if he recommends the product to them. Such an act can reduce alternative costs and effort for others, and build a community that enjoys helping each other. On the contrary, if the customer is not satisfied with product quality, he can have a strong sense of justice to inform others of the poor quality of this product. It can benefit others by preventing them from having a similar experience. No matter which situation consumers encounter, their reviews containing a lot of description and feelings from different aspects of products can provide real information to others. This information can reduce information asymmetry between sellers and consumers to help consumers make better purchase decisions.

Koltler and Keller[2] divided the purchase procedure into several stages, which includes need recognition, information search, evaluation of alternatives, purchase decision, purchase, and post-purchase evaluations. Research has previously demonstrated the significant effect of consumer reviews on sales[3–7]. The third motivation of consumers to review after purchase is to give feedback comments to product providers including producers and sellers. During usage of products, consumers can access a significant amount of first-hand information, which includes both positive and negative feedback on products. This information can be harnessed by providers to improve the performance of products, such as upgrading design and functions. In addition, sellers can also improve their marketing strategy[8] and service quality in time. With the help of such a feedback channel, products can continuously get better.

Besides a numerical rating level, consumers do not have a unified form they can use for review. Reviews could be recommendations with extremely positive attitude or a detailed description of product usage to remind others of its weaknesses. More specifically, owing to different characteristics of consumers—such as character, attitude, experience, expert knowledge, responsibility, economic profits, social needs, and individual worth—different consumers can have different preferences on products. Even on similar products, they can hold divergent opinions. Similarly, when reading reviews, different consumers get varying levels of insight. In order to evaluate the quality of reviews and their influence on other consumers, Amazon provides a choice button beside each review for consumers to vote whether the review is helpful to them. The result of votes is also public to all users and could influence the consumer purchasing intention[9].

In order to find out what consumers think of reviews in terms of perceived helpfulness in their purchase decision making, there is increasing focus in his research field. Existing studies examine several important factors of the perceived helpfulness of online reviews, such as review characteristics[6, 10, 11], product type[12, 13], and comment length[14], which will be discussed in detail in the following section.

However, existing work pays little attention to the latent differentiation trend of helpfulness among different-star ratings. Relatively little is known about whether the helpfulness of different reviews generated by same reviewers has some similarities or not. In addition, there is little research on whether the number of votes for a review has an explicit connection with the helpfulness of the review. However, these questions have immediate and critical influence on constructing a better rating system to provide more precise information to both consumers and sellers.

In order to explore these questions, this paper commits to a better understanding of effect mechanism on helpfulness. First, we construct three hypotheses based on explanatory inferences from three points of view—reviewers themselves, review valance, and the vote number. Subsequently, we verify their reasonability empirically based on real review data collected from Amazon.

This paper is structured as follows. The first section has a brief introduction on the problems studied in this paper. In Section 2, we survey related literatures to review the factors of helpfulness of review. In Section 3, we propose three hypotheses about helpfulness of review and discuss the theoretical foundations of the propositions. Section 4 constructs verification models based on the real review data to test the proposed hypotheses. In Section 5 we discuss our work and explore its applications and the future development.

## 2 Related Works

As one of the most traditional and influential way of communication, WOM allows consumers to exchange opinions and information about products, brands, and services[15]. Researchers have demonstrated that WOM not only influences consumers to make choices and decisions[16], but also has an effect on their expectation and perceptions on products[17]. Existing literature has also shown that there is a tight relationship between WOM and sales. Good product triggers positive WOM and positive WOM promotes sales in turn[7]. Especially along with development of internet, electronic WOM which mainly known as online reviews can be produced more easily by consumers and spread much faster and wider than ever before, and consequently reviews have more powerful influence on consumers and markets. In order to find out the influence mechanism of reviews, research efforts are mainly devoted in three streams.

The first stream is to explore motives of consumers to be engaged in online reviews articulation. Existing research on traditional WOM could provide valuable insights because a new online form may not change its function to be a potential driver of consumer actions[10].

Dichter[18] identified four dimensions of WOM involvement—product, self, others, and message. Compared with Dichter, Hennig-Thurau et al.[15] suggested consumers' desire for social interaction, desire for economic incentives, concern for other consumers, and self-worth enhancement. Sundaram et al.[1] found that expressing positive feelings can be triggered by positive experience. Such feelings can also be classified into the aspect of self in Dichters work.

Anderson[19] developed a utility-based model of WOM to predict whether WOM activity should increase as either satisfaction and/or dissatisfaction increases. Their findings support the proposed asymmetric U-shape for the relationship between consumer satisfaction and WOM activity. Specifically, extremely dissatisfied consumers engage in WOM activity greater than satisfied ones. Chung and Darke[20] suggested consumers are more likely to engage in reviewing products which are relevant to the self-concept rather than utilitarian. There is a bias for consumers to exaggerate benefits of self-relevant products. Tong et al.[21] modeled a set of motivating and inhibiting factors that could influence consumers' intention to contribute product reviews. Their experiments show that perceived satisfaction is associated with helping others and influencing merchants, probability of enhancing self-image, and perceived executional costs. In addition, the presence of an economic rewarding mechanism can promote contribution of reviews in certain conditions.

The second stream is to study the factors that influence perceived helpfulness of reviews to consumers. In related studies, reviewer identity, review valance, product type, and characteristics of review text — including depth, subjective, readability, and spelling errors — are commonly examined.

Review valance generally includes positive, negative, and neutral experiences. In numerical rating of typical five stars, one star indicates an extremely negative view, five stars indicate an extremely positive view, and three stars indicate a moderate view.

In general, products are divided into two basic types—search product and experience product—based on whether consumers can easily obtain accurate measurable objective attributes and information about products prior to purchase[22]. Search products like camera, cellphone, and computers can be known by obtaining detailed parameters from public introduction before using them personally. Experience product like books, music, and food can only be truly understood by real experience.

Review depth, which is usually represented by the number of words in comment text section, can increase information diagnosticity to help consumers obtain information without additional search cost. Longer reviews are believed to contain more information than short reviews. Incremental information can promote confidence of the decision makers[23] and is regarded as more convincing than others. In addition, a review with longer length implies that there is greater involvement of reviewers and greater likelihood of presenting a detailed description of how and where the product was used in a specific context. Review depth has different effect on the helpfulness of review in different types of products. Specifically, review depth has greater different influence on review helpfulness of search goods than experience goods.

Prior researches consider more than one factor simultaneously. Study conducted by Mudambi and Schuff[24] indicated review extremity, review depth, and product type effect on the helpfulness of review from Amazon. Product type can moderate the effect of review extremity. Review depth has a different effect on the helpfulness of a review in different types of products. Pan and Zhang[25] revealed that the review valance and length of the review have positive effects on helpfulness of the review; however, the product type moderated these effects. In addition, they established a curvilinear relationship between reviewer innovativeness and helpfulness. Zhang et al.[13] discovered that promotion consumption goals made consumers perceive positive reviews more persuasively than the negative ones. On the contrary, consumers with prevention consumption goals perceive negative reviews more persuasively. Liu et al.[26] developed models and algorithms to predict the helpfulness of review using three important factors—reviewers expertise, the writing style of review, and the timeliness of review. Some works find that extreme ratings are more influential than moderate ones[5, 27].

Ghose and Ipeirotis[28] explored several aspects of review text and reviewers, including text-level features such as subjectivity levels, readability, and spelling errors and review-level features such as average usefulness of past reviews and self-disclosed identity measures of reviews, and performs an econometric analysis to reveal relations between these aspects and helpfulness. Korfiatis et al.[29] investigated the interplay between review helpfulness, review score, and review text, which are quantized by conformity, understandability, and expressiveness. They also found that review readability influences more on helpfulness than the review length and extremely helpful reviews received higher score than others that were deemed as less helpful. By comparing review data from four national Amazon sites (USA, UK, Germany, and Japan), Danescu-Niculescu-Mizil et al.[30] noted the national differences between reviews collected from different Amazon sites in terms of review variance and review helpfulness.

The third stream is to explore a review's influence on providers with respect to marketing activities, and to consumers around purchase decision-making. The WOM mechanism lets consumers share opinions and experiences on products, companies, and services. Specifically, e-WOM is a lower-cost and more effective channel to enable consumers to express their opinions and be heard. Opinion towards products can influence decision-making of consumers and subsequently influence the product sale. In order to persuade consumers to buy their own products, companies have to pay attention to WOM to understand consumers reactions to their products, such as attitude to certain attributes of products, or different market demand situations in different regions.

According to Simon's classic work[31], a decision process contains three distinct phases — intelligence, design, and choice phase. The first phase is to recognize problem and gather information about problem. The second phase is to structure the problem, develop criteria, and identify alternative solutions. The last phase is to make a final decision to choose the best alternative solution, which meets the criteria. Subsequently, a decision-maker evaluates how well the process was executed using the feedback obtained from the results, which can help stage of posterior intelligence in the future. Based on Simon's decision-making model, Kohli et al.[32] explained that the factors such as consumers cost and time savings lead to consumer satisfaction with online channel, where the wealth of reviews emerges. His research gives instructions to attract buyers and retain them by providing capabilities or tools such as comparing features and price, recommending items to support buyer decision-making process. Kotler and Keller[2] divided purchase decision process into six stages, which include need recognition, information searching, alternatives evaluation, purchase decision, purchase act, and post-purchase evaluation. According to this division, posting comments or reviews on the website is at the last phase of purchase. However, the results extend far beyond this stage. Reviews read by other consumers influence the next purchase decision process. Therefore, websites should assist consumers to explore valuable information more easily so that they can make better purchase decisions.

Liu[8] compared the dynamic patterns of WOM during movie prerelease and opening week with WOM data from Yahoo Movies Web Site and finds that the volume of movie WOM explains box office sales with a significant level both in aggregate and early weeks. His finding highlights the necessity to observe and respond to WOM communications actively, especially during early weeks after the release when most of the revenue is produced.

In addition, reviews could also reveal a product's advantages and disadvantages compared with other products of competitors to improve product quality in time. In addition, purchase intention can also be extracted from comprehensive review valance to predict sale amount in future[10]. According to prediction, manufacturers can arrange production plan and supply chain management flexibly to satisfy consumers demand, lower the costs, and maximize profitability[33].

Existing research described above provides valuable insights on review analysis; however, most of them investigate reviews at a high level. Consumers' individual preference or characteristics on different product categories are seldom discussed and examined. Therefore, this paper focuses on reviews on different categories generated by different reviewers to reveal some meaningful implications, which offer instructions to sellers, producers, and consumers themselves to improve their online activities.

## 3 Review Helpfulness Hypotheses

Reviews are generated by consumers who purchased specific products before, and describe consumers' opinions about products, services, or other relative information. In light of the information diagnosticity theory proposed by Feldman and Lynch[34], helpfulness of a review as perceived by consumers is based on whether the review could reduce their uncertainty when they make purchase decisions. In the paper, helpfulness of review is defined as the ratio of the number of consumers who found it helpful to the total number of consumers who had read and evaluated the review.

Different reviewers have their own characteristic, experience, and preference, which leads to diverse reviews. Similarly, review readers also perceive various levels of helpfulness of reviews. In review systems nowadays, identity of reviewers is public and definite, while identity of review readers and voters is not traced, except for their total votes. This situation makes it impossible to take a stand on the perspective of review readers. Therefore, this paper concentrates on reviewers' characteristic and preference of product type and review valance on review helpfulness. In addition, we also utilize the only information available of review readers, namely the number of votes, to analyze its effect on review helpfulness. We propose three hypotheses in the beginning of the section and make empirical verification based on real review data from Amazon in the next section.

Based on behavioral consistency theory, humans have a tendency to behave with a routine pattern or way. This is the theoretical basis for that behavior to be predicted. Reliable people are trustable because they are accustomed to judge products by objective truth rather than subjective feelings. When they evaluate a product they brought, they mainly rely on objective facts of the product, such as attribute, parameter, or usage situation, instead of subjective feelings. Compared with facts, personal subjective feelings are not as persuasive. Therefore, in order to make their reviews and ratings about products reasonable, consumers utilize real data and evidence to support their opinions. Detailed evidence, such as description of products or how and where to use products in specific situations, can provide others a wealth of information. This information can augment an understanding of the products of consumers and reduce information asymmetry between providers and consumers. In this way, consumer uncertainty can be reduced, which is beneficial for them to make better purchase decisions.

Notwithstanding, different types of products such as books and cellphones might have totally different attributes, parameters, and functionality. At times, professional knowledge is required to get a thorough understanding of certain products. Experts generally provide authoritative advice. However, experts do not necessarily appeal in the same way as most common consumers. The critical concern of consumers is to buy products which can meet their needs within an acceptable price, and not to analyze products like experts. In most cases, products sold online are very simple and foolproof to use. Besides, special and rigid requirement of experts hinders experts to take part in buying popular products which are accepted widely by common consumers online. In other words, reviews from experts might not be able to provide helpful information. Therefore, the decisive factor which influences the helpfulness of reviews from same consumer is not their expertise but their inherent character which helps them understand what others are interested in. Consumer-oriented comments correlated with their own experience appeal more strongly to their peers, and as a result, have greater influence and helpfulness independent of the product category.

Based on the above analysis, we assume that reviewers have a tendency to adopt similar patterns to express their opinions on products regardless of the product category. Therefore, we hypothesize as follows:

**Hypothesis 1: Reviews on different product categories generated by same consumer have approximate helpfulness.**

There are two common valance settings in rating systems. The first setting contains three common levels—low, moderate, and high. The second setting contains five levels, 1–5 stars. Taobao belongs to the first setting, while Amazon and Tmall belong to the second setting. If a consumer gives a high rating valance to a product he purchased, it means that he holds a positive attitude towards the purchased product. Similarly, a low rating implies a negative attitude and a moderate rating implies a reserved attitude. In general, existing statistical methods adopted in e-commerce websites transform rating levels into a rating point initially. For example, low, moderate, and high ratings stand for $-1$, 0, 1 points, respectively, and 1–5 stars stand for 1–5 points, respectively. Subsequently, we calculate average of all rating points together and provide it to consumers as a comprehensive reference in the corresponding page of the product. High mean stands for good rating and vice versa.

Considering the recommend and sorting mechanisms on an e-commerce platform, products with high average rating are recommended first and sorted in front with a greater probability. Consequently, these products have more opportunity to be seen and to be selected as alternatives and be purchased in the end. On the contrary, products with low average ratings have less chance to appear in front of consumers. In addition, under an online transaction scenario, there are numerous products to choose from. That the reviews of a certain product are referred by consumers at least means that this product has been seen and has the chance to be preliminary alternatives. Otherwise it is impossible for consumer referring to relative reviews, not to mention voting to the reviews. Therefore a high valance rating has more chance to be caught into sight.

When consumers are evaluating alternatives, information asymmetry is a major obstacle in making purchase decisions due to uncertainty about products. Researches have demonstrated that high valance has a positive effect on the possibility of being purchased[9, 35, 36]. As discussed above, a review with high rating valance manifests positive attitude on products and can provide strong evidence to support this attitude. Such evidence helps in decision making by eliminating doubt and uncertainty. In other words, such reviews are helpful to consumers.

However, review with low rating valance stands for negative attitude, which is interpreted as a warning. This warning discourages others from buying the corresponding product. Past research[11, 12, 37] finds strong evidence that negative information has more value to a receiver; therefore, people weigh negative information more than positive information during an evaluation, which is called a negative bias. This is because in a social environment, there is greater positive information than negative information. Once negative information does appear, it attracts more attention[38].

Nevertheless, there is a difference between an online environment and a social environment. There is much more information on the internet than in the real world. Due to the virtual character of internet, users tend to say anything they want without undertaking any responsibility. It leads to an overflow of false information. In order to attract more attention on the internet, it is common to fake negative information, even some negative information is deliberate defamation for internal purposes. The situation seems reverse to the real world. Therefore, compared with negative information, positive information is more valuable to the users on the internet.

In addition, considering the special term of referring reviews in purchase procedure, positive reviews help enhance consumers' confidence to buy, while negative reviews increase consumers' uncertainty and can lead to giving up even on alternative products. Surrounded by a lot positive reviews about alternative products, negative reviews appear to lack persuasiveness. In addition, a lot of negative reviews are caused by internal factors of reviewers themselves rather than product quality. Consequently, positive reviews with a high rating valance is more helpful to consumers. Therefore we hypothesize the following:

**Hypothesis 2: Reviews with high rating valance are perceived as more helpful than ones with a low rating valance.**

As discussed above, the basic conditions for being voted helpful in reviews include the following: (1)

Consumers catch sight of the review. (2) Consumers choose corresponding product as an alternative. The first condition is physically necessary and the second condition is responsible for triggering the voting motivation. According to the default mechanism of review organizations with an e-commerce platform such as Amazon, reviews which are voted the most can rank in front pages. When consumers refer to reviews, the most-voted reviews are exposed to consumers first. Given that there are a very large number of reviews on the e-commerce platform, retrieving helpful information requires much time and effort. This method of review organization can decrease consumer search cost and time to a great extent. In this situation, when consumers see dull or useless information, they are inclined to ignore it. However, when they see information, which is particularly distinctive and helpful, they are activated by a strong emotional response and pay more attention. At such a point, voting helpfulness can provide a way to express this emotion.

As we know, first sight has the most powerful effect on impression formation. Therefore, first impressions are very deep-rooted—it takes significant additional effort to change the first impression than to form it. Consumers are accustomed to relying on others' experience to judge the quality of products[22]. Reviews with more votes have more effective impact on a consumer's attitude than alternatives. Since the most voted reviews can rank first, ranking first can result in being seen most, which in turn has an opportunity to be voted most. This procedure is similar to positive feedback. With accumulative time effect, reviews occurring earliest are more likely be voted most. It is difficult for later reviews to rise up to the front, unless it can provide new, valuable, and helpful information. Furthermore, humans have a psychology to follow others, which is called conformist mentality. When humans see others doing one thing, especially when the number of others is pretty large, they like to follow others, including voting. This is because most people believe that majority implies greater belief and less risk. No one wants to stand against the tide. Doing the same as the majority can get more safety and identification. So reviews voted most can attract more attention and in turn get more votes continually.

Besides, under an online scenario, if a review gets attention from many consumers, it means that this review is truly helpful or very controversial. The motivation of consumers to read reviews of products is to know more about products because they are not familiar with the products prior to purchase. In this case, totally opposite opinions about products cannot be formed among consumers. So the possibility of controversial situation is not great. It is the main situation that a review that gets many votes is helpful rather than controversial. For reviews without enough votes, randomness and volatility are strong, and it is hard to find an obvious and consistent relationship between the number of votes and perceived helpfulness. Therefore we hypothesize as follows:

**Hypothesis 3: Reviews with a large number of votes are perceived to be more helpful than ones with fewer votes.**

## 4 Empirical Verification

### 4.1 Dataset and preprocess

This paper utilizes the review dataset, which was collected from one of the biggest e-commerce websites, Amazon, by McAuley and Leskovec[14]. This dataset contains 34 686 770 reviews, 6 643 669 consumers, and 2 441 053 products in total. The number of consumers who reviews more than 50 times in this dataset is 56 772. All reviews in this dataset were generated during the time from Jun 1995 to Mar 2013.

To test the three hypotheses proposed above, we extract the top 100 reviews who reviewed the most times indicated from this dataset based on the listed reasons. (1) In statistics, big errors can be caused within a small number of samples. Moreover, if there are not adequate transaction and review records for consumers, then records distributed into each individual product category are much less. Precision and credibility cannot be guaranteed for statistical analysis of reviews on such data scale. (2) Users who experience a lot of online transactions are more likely to have already formed stable recognition within a community context and have reasonable expectation on product quality. Through long-term accumulation, consumer behavior, including trading and reviewing, can stay relatively stable so that randomness can be reduced. This stability is beneficial to make analysis and prediction of consumer behaviors. After extraction and preprocessing, all reviews of the top 100 reviewers are shown in Fig. 1 and details of part reviews are shown in Table 1.

As Table 1 shows, the number of reviews of the top 100 consumers ranges from 25 297 to 3467. Their

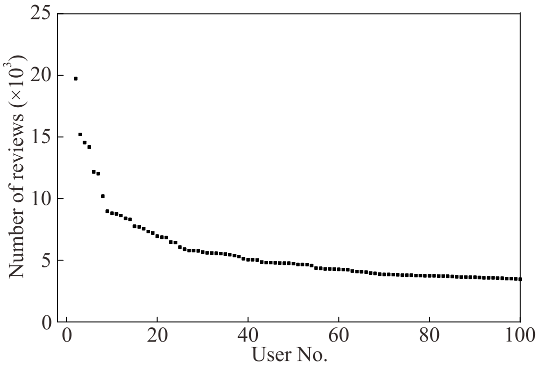


**Fig. 1 Number of reviews of top 100 reviewers.**

reviews demonstrate different distributions in the 1 to 5 rating valance. For example, consumer 1 gives the most rating of 5 with 25 265 times, while consumer 9 gives the most rating of 4 with 2999 times, and consumer 70 gives the most rating of 3 with 1780 times instead. This distinction reflects different rating preferences among consumers.

According to the category list McAuley provides, all reviews of top 100 consumers are classified into 40 types in total. Number of reviews in each category is depicted in Fig. 2. The types containing most reviews are "Movies & TV" (229 011 reviews), "Books" (197 979 reviews), "Music" (110 677 reviews), and "Amazon Instant Video" (24 050 reviews). While "Appliances", "Purchase Circles", "Car Electronics", and "Baby category" have the least reviews, nearly 1. It can also been seen from the target dataset that there is also a preference on product category for consumers. More specifically, for every consumer, the

**Table 1 Reviews statistics of top 100 reviews.**

| No. | Review | | | | | Sum |
|---|---|---|---|---|---|---|
| | Point 1 | Point 2 | Point 3 | Point 4 | Point 5 | |
| 1 | 0 | 0 | 4 | 28 | 25 265 | 25 297 |
| 2 | 571 | 468 | 1065 | 4363 | 13 264 | 19 731 |
| 3 | 0 | 3 | 84 | 4627 | 10 494 | 15 208 |
| 4 | 2073 | 2112 | 1461 | 2603 | 6293 | 14 542 |
| 5 | 3 | 199 | 1517 | 5708 | 6750 | 14 177 |
| 6 | 29 | 88 | 408 | 1769 | 9877 | 12 171 |
| 9 | 1047 | 1582 | 2666 | 2999 | 686 | 8980 |
| 10 | 91 | 92 | 223 | 1499 | 6904 | 8809 |
| 20 | 1 | 49 | 373 | 2516 | 4006 | 6945 |
| 50 | 1 | 2 | 9 | 47 | 4672 | 4731 |
| 70 | 13 | 299 | 1780 | 1417 | 352 | 3861 |
| 80 | 11 | 98 | 294 | 838 | 2500 | 3741 |
| 90 | 21 | 156 | 606 | 1885 | 956 | 3624 |
| 100 | 0 | 0 | 0 | 12 | 3455 | 3467 |

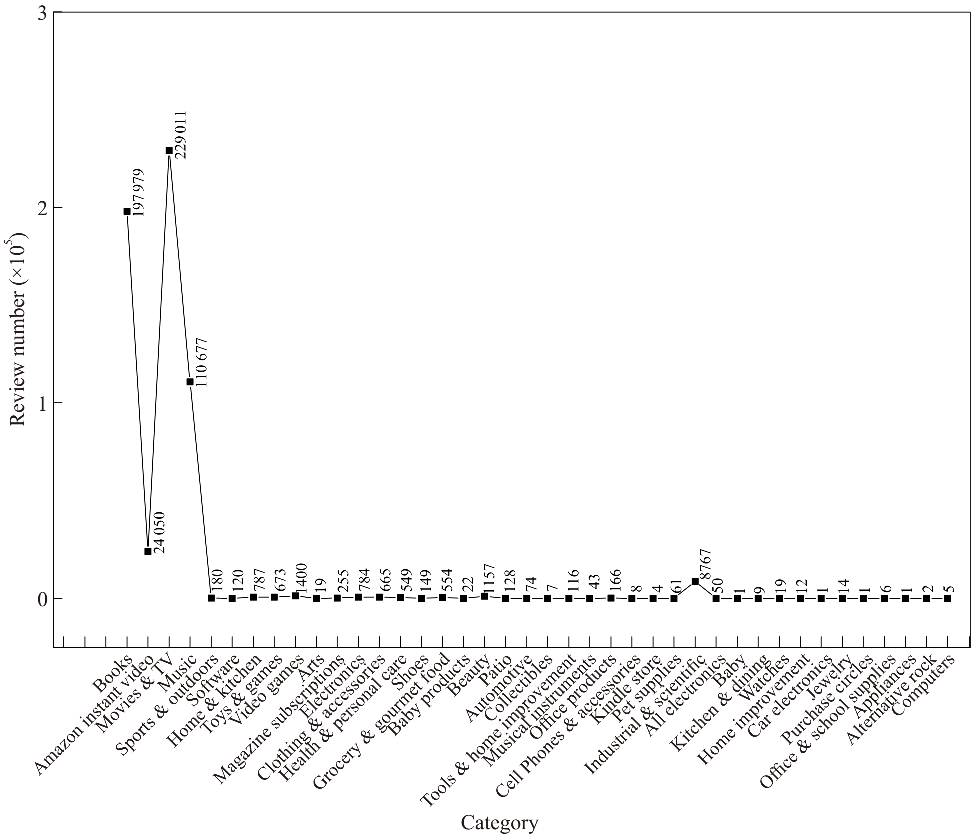


**Fig. 2 Number of reviews in different categories.**

most rated product categories are also different. For example, consumer 1 reviewed most in “Books”, “Movies & TV”, “Music”, and “Amazon Instant Video”, while consumer 83 reviewed most in “Music”, “Books”, “Movies & TV”, and “Electronics”.

## 4.2 Analysis and result

### 4.2.1 Test of Hypothesis 1

Every consumer has a different distribution of reviews in different product categories. Moreover, we find that most consumers’ reviews are centralized in four categories, although the four categories are different for consumers. The reason for this situation might relate to the method of collecting data from Amazon. As discussed before, small samples can cause major errors. In order to reduce errors, the top four categories types, which are most rated and rated at least 20 times for every consumer, are chosen to explore whether there are differences in review helpfulness among different product categories. After extraction, 65 consumers (1, 2, 4, 5, $\cdots$, 99) who satisfy the proposed conditions are left. Then calculate average helpfulness $h_{i,j}$ through dividing the number of all votes allVote in the top $j$-th categories $\text{TopCat}_j$ by the number of helpfulness votes helpVote of consumer $i$ ($i = 1, 2, 4, 5, \cdots, 99$), as Eq. (1) shows:

$$h_{i,j} = \frac{\text{helpVote}_{\text{TopCat}_j}}{\text{allVote}_{\text{TopCat}_j}}, \quad j = 1, 2, 3, 4 \tag{1}$$

Subsequently, we calculate the mean and variance of these 4 helpfulness $h_{i,j}$ of 65 consumers, as seen in Fig. 3.

From Fig. 3, we find that although the means of helpfulness of reviews in the top four product categories generated by different reviewers can vary, the variances of helpfulness of reviews of all consumers remain at a similar low level. This implies that review helpfulness of different categories of the same consumer is approximate. Therefore Hypothesis 1 is verified.

### 4.2.2 Test of Hypothesis 2

Rating valance in our dataset from Amazon contains five options, namely 1–5 score. A 5-score is deemed as extremely high valance, a 3-score as moderate valance, and a 1-score as extremely low valance. Different scores represent different attitudes to a rated product. In order to explore the influence of a review with different attitudes on others, we choose the most rated product category of every consumers as study object to reduce interference caused by product category. First we calculate the average perceived helpfulness $h_{i,v}$ of reviews with valance $v$ in the most purchased product category of $i$-th consumer by Eq. (2) as follows:

$$h_{i,v} = \frac{\text{helpVote}_{\text{valance}=v}}{\text{allVote}_{\text{valance}=v}}, \quad v \in [1, 2, 3, 4, 5] \tag{2}$$

The distribution of 100×5 and their means are shown in Fig. 4. It is intuitive that helpfulness of reviews with higher valance is greater than lower valance.

In order to test whether there is a difference or change trend among the 1–5 point valance, we make a Mann-Kendall trend test[39] on the valance helpfulness matrix. Mann-Kendall trend test was proposed to verify whether there is a trend in a series of data.

Statistical tests are divided into two parts. The first part is to make a Mann-Kendall test on 100 valance helpfulness vectors and find there are 71 valance helpfulness vectors of review which are increasing as valance with a significant level 0.01. The second test part is making an accumulative Mann-Kendall test on 100×5 helpfulness-valance matrix on the whole, which find that the $p$-value is extremely significant as 0.00. This implies that the increasing trend as valance in review helpfulness is significantly supported. Therefore

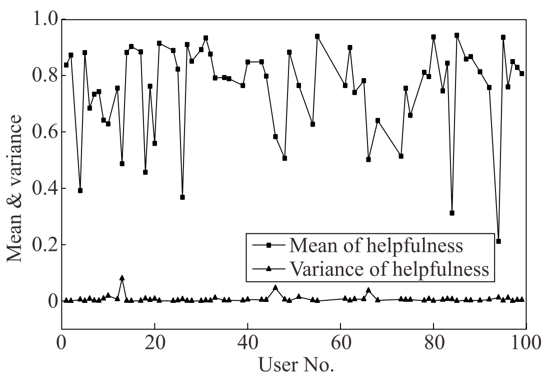


**Fig. 3 Mean and variance of helpfulness.**

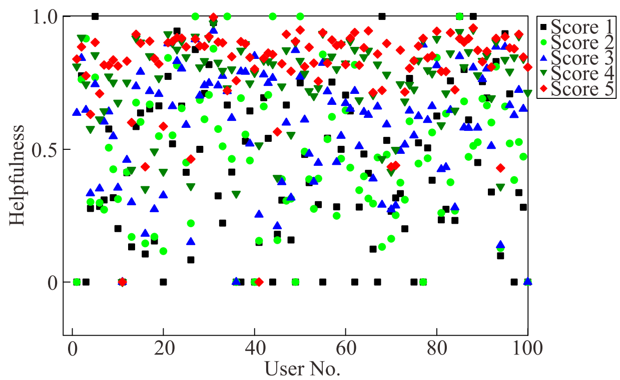


**Fig. 4 Average helpfulness of reviews with 1–5 scores.**

Hypothesis 2 is verified.

### 4.2.3 Test of Hypothesis 3

In order to analyze whether the number of votes of each review has an effect on perceived helpfulness, two review helpfulness vectors are calculated. The first one is the average review helpfulness of all reviews generated by 100 consumers, noted as $\boldsymbol{V}$. The second one is the average review helpfulness of reviews which are voted at least a certain number of times $S$ by others, noted as $\boldsymbol{V}^*$. In our test, $S$ is chosen as 100. There are 67 consumers who satisfy this condition. Subsequently, we compare the two review helpfulness vectors with $1\times67$ dimension and calculate the corresponding ratio of them. The comparison is partly shown in Fig. 5 and Table 2.

From Fig. 5, we find that for all users except the 10th, 72nd, 83rd, and 95th users, average helpfulness of reviews which are voted at least 100 times is larger than that of all voted reviews. The average ratio of the higher percentage of all 67 consumers is 33.5%. In order to test our finding precisely, we make a paired-samples $T$ test on the two helpfulness vectors. As Table 3 shows, the $t$ value of the paired $t$ test is $-8.959$, which means $\boldsymbol{V}$ is much smaller than $\boldsymbol{V}^*$ with a large extent and the $p$ value is 0.00, which means the comparison is significant. In other words, the helpfulness of reviews with at least 100 votes is greater than the one of all reviews with at least one votes. Therefore Hypothesis 3 is verified.

Above all, three proposed hypotheses have been verified by empirical tests respectively. Results show that reviewer, review valance, and review votes have significant correlations and trend with review helpfulness, which coincides with common sense. However, these findings are obtained on statistical tests. And the review helpfulness in this paper is calculated only by votes which have been recorded on the trading platform. There are also a lot of evaluations which have not been recorded. So encouraging more consumers to vote is beneficial to get more accurate helpfulness. In addition, it is also possible for some reviews which come from reviewers who always provide unhelpful reviews before and reviews which are with low valance and few votes to have a high helpfulness actually. So correlation between factors and helpfulness revealed in this paper is only general phenomenons, rather than absolute relationship.

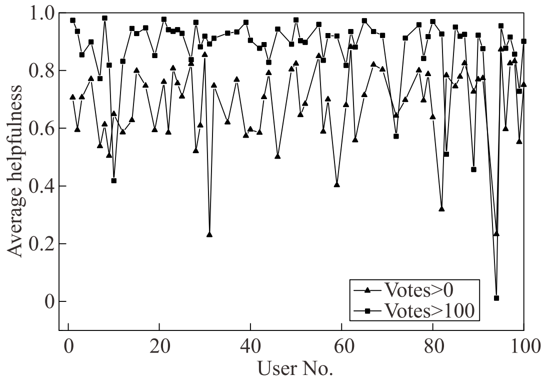


**Fig. 5 Average helpfulness of reviews with different votes.**

**Table 2 Statistics comparison of two helpfulness vectors.**

| | Mean | $N$ | Std. deviation | Std. error |
|---|---|---|---|---|
| $\boldsymbol{V}$ | 0.6801 | 67 | 0.140 24 | 0.017 13 |
| $\boldsymbol{V}^*$ | 0.8690 | 67 | 0.154 75 | 0.018 91 |

## 5 Conclusions and Discussion

In this paper, we explored factors correlated with the helpfulness of reviews from the customers' perspective.

First, we proposed three hypotheses about review helpfulness based on explanatory theory, which includes behavioral theory and psychology theory. In the three hypotheses, we discussed the factors that influence the helpfulness of reviews from different dimensions—such as reviewer, review characteristics, and review votes. In general, reviews are generated by consumers who have purchased specific products. Based on behavioral theory, behaviors of the same actors have some similarity and routine. Reviewing is also a kind of a behavior. Therefore, it is possible that similar helpfulness exists among reviews from the same consumer. Next, review valance stands for consumers' attitude to specific products after a complete purchase procedure. Reviews of high valance represent positive attitude and are considered as indicator of

**Table 3 Result of paired $t$ test of two helpfulness vectors.**

| | Paired differences | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Std. deviation | Std. error mean | 95% confidence interval of the difference | | $t$ | df | Sig.(2-tailed) |
| | | | | Lower | Upper | | | |
| Pair $\boldsymbol{V}$-$\boldsymbol{V}^*$ | −0.188 87 | 0.172 55 | 0.021 08 | −0.230 96 | −0.146 78 | −8.959 | 66 | 0.000 |

recommendation. Information extracted from these reviews can reduce concerns of consumers and help them make decisions faster and with greater ease. Although reviews of low valance can also provide helpful information, given the recommend and sorting mechanisms on websites, these kinds of reviews have limited possibility to be caught in sight. Finally, helpfulness of a review is perceived by consumers who refers to it. More votes a review gets, more attention it attracts, and more votes it gets as a result. Moreover, reviews with more votes are ranked in front and, as a result, are immediately in sight. Since first sight forms a powerful impression, reviews with more votes are more influential.

In order to verify these three hypotheses, this paper made empirical tests respectively on the proposed hypotheses by using real review data from Amazon. The results show that reviews from the same consumers have similar helpfulness, reviews with high valance have higher helpfulness, and reviews with more votes have higher helpfulness. Results from this paper give instructions on several aspects.

First, treating all ratings as equal is not feasible. Existing rating systems weight all ratings equally without considering different characteristics of raters. However, this paper verified distinctions among different raters and similarity of ratings from the same raters. Therefore, treating ratings differently can help evaluate a product more reasonably. E-commerce platforms, such as Amazon, Taobao, and so on, should improve their feedback systems by giving more weight to the reviews with more helpfulness.

Second, current review sorting mechanisms can result in reviews with more votes always being in the front pages. This makes reviews that are generated later to have less chance to be seen or voted on, even if the review was more helpful. This causes absolute first-mover advantage, which can harm the enthusiasm of later reviewers. Therefore it is reasonable to give some chance to reviews from reviewer who has always provided reviews with high helpfulness in the past to sort in front pages although they are produced later.

Third, helpfulness votes are only divided into two categories, helpful and unhelpful. However it is hard to classify which kind of helpful it belongs to—positive helpfulness to choose a purchase or negative helpfulness to give up on a purchase. This information is also useful for consumers to help save time and effort to find desired reviews. Therefore it should be considered that the e-commerce platforms could provide more choices such as “vote to buy” and “vote not to buy” for consumers to distinguish which kind of “helpful” a review belongs to.

Notwithstanding findings above, this paper also has some limitations. First, data used in empirical analysis were collected from Amazon. And review data are centralized into very few categories. Whether reviews from other websites or categories perform similarly is not verified. Second, this paper explores the relationship between helpfulness of a review and several factors. But we do not construct a quantitative calculation model of helpfulness with these factors. Third, review text and browsing behavior of consumers are not considered in our analysis. Existing research has demonstrated that there are relationships between characteristics of review text and helpfulness—such as text length and text errors.

In the future, we expect to consider more factors, greater product categories for classification, more hypotheses, and parameters to construct a quantitative helpfulness model.

**Acknowledgements**

This work was financially supported by DNSLAB, China Internet Network Information Center.

**References**

[1] D. S. Sundaram, K. Mitra, and C. Webster, Word-of-mouth communications: A motivational analysis, *Advances in Consumer Research*, vol. 25, pp. 527–531, 1998.

[2] P. Kotler and K. Kotler, *Marketing Management*. Upper Saddle River, NJ, USA: Prentice Hall, 2005.

[3] V. Dhar and E. A. Chang, Does chatter matter? The impact of user-generated content on music sales, *Journal of Interactive Marketing*, vol. 23, pp. 300–307, 2009.

[4] Q. Ye, R. Law, B. Gu, and W. Chen, The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings, *Computers in Human Behavior*, vol. 27, pp. 634–639, 2011.

[5] C. Forman, A. Ghose, and B. Wiesenfeld, Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets, *Information Systems Research*, vol. 19, pp. 291–313, 2008.

[6] N. Hu, L. Liu, and J. J. Zhang, Do online reviews affect product sales? The role of reviewer characteristics and temporal effects, *Information Technology and Management*, vol. 9, pp. 201–214, 2008.

[7] J. A. Chevalier and D. Mayzlin, The effect of word of mouth on sales: Online book reviews, *Journal of Marketing Research*, vol. 43, pp. 345–354, 2006.

[8] Y. Liu, Word of mouth for movies: Its dynamics and impact on box office revenue, *Journal of Marketing*, vol. 70, pp. 74–89, 2006.

[9] D.-H. Park, J. Lee, and I. Han, The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement, *International Journal of Electronic Commerce*, vol. 11, pp. 125–148, 2007.

[10] P. F. Wu, H. Van der Heijden, and N. Korfiatis, The influences of negativity and review quality on the helpfulness of online reviews, presented in *International Conference on Information Systems*, Shanghai, China, 2011.

[11] J. J. Skowronski and D. E. Carlston, Negativity and extremity biases in impression formation: A review of explanations, *Psychological Bulletin*, vol. 105, p. 131, 1989.

[12] P. M. Herr, F. R. Kardes, and J. Kim, Effects of word-of-mouth and product-attribute information on persuasion: An accessibility-diagnosticity perspective, *Journal of Consumer Research*, vol. 17, no. 4, pp. 454–462, 1991.

[13] J. Q. Zhang, G. Craciun, and D. Shin, When does electronic word-of-mouth matter? A study of consumer product reviews, *Journal of Business Research*, vol. 63, pp. 1336–1341, 2010.

[14] J. McAuley and J. Leskovec, Hidden factors and hidden topics: Understanding rating dimensions with review text, in *Proc. 7th ACM Conference on Recommender Systems*, Hong Kong, China, 2013, pp. 165–172.

[15] T. Hennig-Thurau, K. P. Gwinner, G. Walsh, and D. D. Gremler, Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet? *Journal of Interactive Marketing*, vol. 18, pp. 38–52, 2004.

[16] E. W. Anderson, The formation of market-level expectations and its covariates, *Journal of Consumer Research*, vol. 30, pp. 115–124, 2003.

[17] P. F. Bone, Word-of-mouth effects on short-term and long-term product judgments, *Journal of Business Research*, vol. 32, pp. 213–223, 1995.

[18] E. Dichter, How word-of-mouth advertising works, *Harvard Business Review*, vol. 44, pp. 147–160, 1966.

[19] E. W. Anderson, Customer satisfaction and word of mouth, *Journal of Service Research*, vol. 1, pp. 5–17, 1998.

[20] C. M. Chung and P. R. Darke, The consumer as advocate: Self-relevance, culture, and word-of-mouth, *Marketing Letters*, vol. 17, pp. 269–279, 2006.

[21] Y. Tong, X. W. Wang, C. H. Tan, and H. H. Teo, An empirical study of information contribution to online feedback systems: A motivation perspective, *Information & Management*, vol. 50, pp. 562–570, 2013.

[22] P. Nelson, Information and consumer behavior, *The Journal of Political Economy*, vol. 78, no. 2, pp. 311-329, 1970.

[23] A. Tversky and D. Kahneman, Judgment under uncertainty: Heuristics and biases, *Science*, vol. 185, pp. 1124–1131, 1974.

[24] S. M. Mudambi and D. Schuff, What makes a helpful review? A study of customer reviews on Amazon. com, *MIS Quarterly*, vol. 34, pp. 185–200, 2010.

[25] Y. Pan and J. Q. Zhang, Born unequal: A study of the helpfulness of user-generated product reviews, *Journal of Retailing*, vol. 87, pp. 598–612, 2011.

[26] Y. Liu, X. Huang, A. An, and X. Yu, Modeling and predicting the helpfulness of online reviews, in *Proc. Eighth IEEE International Conference on Data Mining*, Pisa, Italy, 2008, pp. 443–452.

[27] P. A. Pavlou and A. Dimoka, The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation, *Information Systems Research*, vol. 17, pp. 392–414, 2006.

[28] A. Ghose and P. G. Ipeirotis, Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics, *Knowledge and Data Engineering, IEEE Transactions on*, vol. 23, pp. 1498–1512, 2011.

[29] N. Korfiatis, E. Garcła-Bariocanal, and S. Snchez-Alonso, Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content, *Electronic Commerce Research and Applications*, vol. 11, pp. 205–217, 2012.

[30] C. Danescu-Niculescu-Mizil, G. Kossinets, J. Kleinberg, and L. Lee, How opinions are received by online communities: A case study on amazon. com helpfulness votes, in *Proc. 18th International Conference on World Wide Web*, 2009, pp. 141–150.

[31] H. A. Simon, *The New Science of Management Decision*. New York, NY, USA: Harper & Brothers,1960.

[32] R. Kohli, S. Devaraj, and M. A. Mahmood, Understanding determinants of online consumer satisfaction: A decision process perspective, *Journal of Management Information Systems*, vol. 21, pp. 115–136, 2004.

[33] C. Dellarocas, The digitization of word of mouth: Promise and challenges of online feedback mechanisms, *Management Science*, vol. 49, pp. 1407–1424, 2003.

[34] J. M. Feldman and J. G. Lynch, Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior, *Journal of Applied Psychology*, vol. 73, p. 421, 1988.

[35] E. K. Clemons, G. G. Gao, and L. M. Hitt, When online reviews meet hyperdifferentiation: A study of the craft beer industry, *Journal of Management Information Systems*, vol. 23, pp. 149–171, 2006.

[36] C. Dellarocas, N. Awad, and X. Zhang, Exploring the value of online reviews to organizations: Implications for revenue forecasting and planning, in *Proc. International Conference on Information Systems*,Washington DC, USA, 2004.

[37] R. Ahluwalia, How prevalent is the negativity effect in consumer environments? *Journal of Consumer Research*, vol. 29, pp. 270–279, 2002.

[38] D. E. Kanouse and L. R. Hanson Jr, Negativity in evaluations, in *Attribution: Perceiving the Causes of Behavior*. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Inc, 1987, PP. 47–62.

[39] K. H. Hamed, Trend detection in hydrologic data: The Mann-Kendall trend test under the scaling hypothesis, *Journal of Hydrology*, vol. 349, pp. 350–363, 2008.

**Yuanlin Chen** graduated with outstanding graduate award from Huazhong University of Science and Technology in 2010 and earned his bachelor degree in automation. He is currently a PhD candidate in National Engineering Laboratory for E-Commerce Technologies in Department of Automation at Tsinghua University with a focus of consumer behavior and transaction credit evaluation.

**Yueting Chai** earned his PhD degree in automation from Tsinghua University in 2004 and served as deputy chief engineer of National CIMS Engineering Research Center since 1996. He is now the director of National Engineering Laboratory for E-Commerce Technologies with a focus on electronic commerce and modern logistics.

**Yi Liu** earned her MEng degree in automation from Tsinghua University in 1991. She became an associate professor in Department of Automation at Tsinghua University from 1994. She is now focusing on electronic commerce and supply chain management.

**Yang Xu** earned his PhD degree in system engineering from Beijing Jiao-Tong University in 2011 and serves as a postdoctoral researcher from 2012 in the National Engineering Laboratory for E-Commerce Technologies of Tsinghua University with a focus on electronic commerce and modern logistics.