

# A Pilot Study of Mining the Differences in Patterns of Customer Review Text Between US and China AppStore

Lisha Li<sup>(✉)</sup>, Liang Ma, Pei-Luen Patrick Rau, and Qin Gao

Department of Industrial Engineering, Tsinghua University,  
Beijing, People's Republic of China  
li-ls15@mails.tsinghua.edu.cn

**Abstract.** With the fast growing of AppStore market and the developing of techniques in opinion mining, this study was aimed to investigate the sentiment and opinions of customer reviews in both China AppStore and US AppStore, and identify the difference of key term and patterns of apps reviews among different genres and between China AppStore and US AppStore. Results showed that there were small differences in using adjective words used or expressing key opinions. The result of this study could help publisher to extract useful customer feedback from customers reviews when publishing apps in foreign countries.

**Keywords:** Cross-cultural product and service design · Cultural differences · Review mining

## 1 Introduction

In recent years, mobile services and platforms have achieved critical mass in the information and communications technology industry. The key to their success has been mobile app services, including naive softwares and platforms that offer internet-based services with good user experiences [9]. With iOS being one of the major mobile phone operation systems, its app service platform, Apple's App Store (henceforth, AppStore) also prosper in app service market with a growing number of publishers and users. Since AppStore launched with only 500 apps and a dozen developers in July 2008, the market increased to over 2,281,240 apps and 529,078 active app publishers in April 2016 [11]. By April 2016, there were 155 AppStore territories that are available for apps to be sold in the corresponding countries or regions [1]. And up to September 2016, a total of 140 billion of apps were downloaded by users from all over the world [13]. In this rapidly growing market, for publishers that have their apps published in multiple AppStore territories, it is very important to adapt to the local market, and adjust contents of apps accordingly.

AppStore provide a rich source of information about apps, including one app's price, description, technical information, and customer ratings and reviews, which could provide both qualitative and quantitative data about the customer

perception of the apps, and is very important for both customers and apps publishers. On the one hand, customers' ratings and reviews of apps would affect other customers' purchase decisions, this effect is equivalent to the persuasive effect studied in the advertising literature [6]. Meanwhile, online customer review system is one of the most powerful channels to generate online word-of-mouth [5], and earlier studies have found that word-of-mouth may affect others' decisions in different social contexts [10]. According to previous studies, online reviews have significant impact on sales [4,6]. On the other hand, for publishers, customer review is a major source for the feedback. Other feedback source for apps including e-mail feedback and blogs. Feedback could reveal bugs or features of the current version that need to be fixed or improved.

Customer review in AppStore is spontaneous customer feedback, which has rich sources of information. However, these sources are much less structured than traditional surveys for customer satisfaction studies. The information is contained in free-style text, not in a set of answers elicited for a specific set of questions. With the advent of automatic techniques for text mining such as clustering and key term extraction, free-form customer opinions can be processed efficiently and distilled down to essential topics and recurring patterns of content. Researchers have begun to focus on the analysis of opinion typically using supervised machine learning techniques [8]. For example, by analyzing online reviews of computer game, the characteristics of computer games and user experience in game play could be identified [14]. By using linguistic techniques, researchers have extracted and analyzed the most important factor a moviegoer considers when rating a movie online, and found that reviewers mainly discuss their personal evaluation rather than discouraging or encouraging readers to see the movie [12]. By clustering rare textual opinions based on point-wise mutual information and using externally imposed review semantics on a data set from Amazon containing sales data and consumer review data for digital cameras and camcorders over a 15-month period, researchers have analyzed the consumers' relative preferences for different product features and use the textual data to predict future changes in sales [2].

With the fast growing of AppStore market and the developing of techniques in opinion mining, this study was aimed to investigate the sentiment and opinions of customer reviews in both China AppStore and US AppStore, and identify the difference of key term and patterns of content in apps reviews among different genres and between China AppStore and US AppStore. The result of this study could help publisher to extract useful customer feedback from customers reviews when publishing apps in foreign countries, and provide a insight of cultural differences in writing apps reviews between Chinese and American. To be specified, the research questions of this study were:

**RQ1.** Between America AppStore and China AppStore, and among different genres, is there any difference of patterns in customer review for top-selling apps?

**RQ2.** What is the portion of app-review relevant words in the review text for top selling apps, is there any difference among different genres, and between China AppStore and America AppStore?

## 2 Method

### Data Collection

Review text and relative data were collected from the following four genres in both US AppStore and China AppStore: Social Networking, Photo & Video, Games, and Entertainment. The reason of choosing these four genres was that in the top chart for each genre there were enough common apps in both US AppStore and China AppStore, and there were enough reviews for apps in the top chart. For each genre, apps in the top 200 free apps chart and top 200 paid apps chart were collected, therefore the total number of apps that were included in this study were 3200. A web crawler was developed to collect the app information, review text, and other relative data. For each app, the collected app information including region (US or China), app name, genre, release date, overall average rating, overall number of ratings, and app price for paid apps. Fifty most recent reviews were collected for each app, or all of the reviews if total number of reviews was less than 50. Review title was collected together with review text for each review. Apart from the most recent reviews, all reviews of some selected apps were collected as well.

### Natural Language Processing

For reviews wrote in English, processing raw review text including the following steps: removing irregular characters, converting to lower case, word tokenization and part-of-speech tagging, stemming and lemmatizing, and removing stop words, calculating word frequencies, and generating term frequency-inverse document frequency (TF-IDF) matrix. For reviews in Chinese Store, steps of processing raw text were similar to processing English text, with the lack of converting to lower case, and stemming and lemmatizing. We used Natural Language Toolkit (NLTK) [3] for English word tokenization and part-of-speech tagging, and Jieba [7] for Chinese word tokenization and part-of-speech tagging. When generating TF-IDF matrix, each review was treated as a document.

### Reviews Clustering

We selected  $k$ -means clustering algorithm to cluster the reviews. This algorithm is widely used in document clustering and text-mining for it's simplicity. Considering the fact that online reviews have a wide variance in lengths, we chose cosine distance for  $k$ -means algorithm so that the cluster results would be independent to the lengths of reviews. The cosine distance were calculated from TF-IDF matrix, which were calculated during the natural language process

### Noise Point Detection

The density-based spatial clustering of applications with noise (DBSCAN) algorithm views clusters as areas of high density separated by areas of low density. Clusters found by DBSCAN can be any shape, as opposed to  $k$ -means which assumes that clusters are convex shaped. This algorithm can be used to detect

noise point, but the result is heavily related to the input parameters. In this study, we used DBSCAN to find noise reviews which were less relevant other reviews.

### 3 Results

The number of apps and reviews collected were showed in Table 1. The rest of this section showed the results of frequent terms analysis, clustering analysis, and noise review detection. All Chinese have been translated into English for understanding and comparison.

**Table 1.** Number of reviews collected

Genre	Number of reviews collected	
	US	CN
Social networking	13,533	11,865
Photo & video	16,024	14,096
Games	16,771	16,871
Entertainment	14,952	12,651

#### 3.1 Frequencies of Adjectives

Figures 1 and 2 showed the top 20 frequent adjective words of US reviews and Chinese reviews in the whole review collection. Words like “good”, “great”, “fun”, and “easy” were most frequent in both US and Chinese reviews. For US reviews there was no adjectives with negative sentiment in the top frequent adjective words. Similarly, for Chinese reviews, only one adjective with negative sentiment, which was “boring”, occurred in the 20 most frequent adjective words. For Chinese reviews, the term of “not bad” was the most frequent adjective words and had much higher frequency than the rest of adjectives. In US reviews, “great” and “good” were top two frequent words, and compared with Chinese reviews, the gap between the frequency of the most frequent adjective and the frequencies of the rest of the adjectives was smaller. The results suggested that customers in both AppStores were more likely to express positive sentiments. And the high frequent of the term “not bad” in China AppSore may caused by the habit of using “not bad” as a common pet phrase among Chinese people.

For both US reviews and Chinese reviews, top 20 frequent adjective words of the reviews of four genres were similar to that of the whole review collection, Tables 2 and 3 showed the top frequent adjective words of four genres of US reviews, the words that were not common among the four genres were presented in boldface type. For US reviews, the number of common adjective words among four genres was 16, and the number of that for Chinese reviews was 10, which suggested that customers in China AppStore wrote their reviews more specific according to the genre of the app.

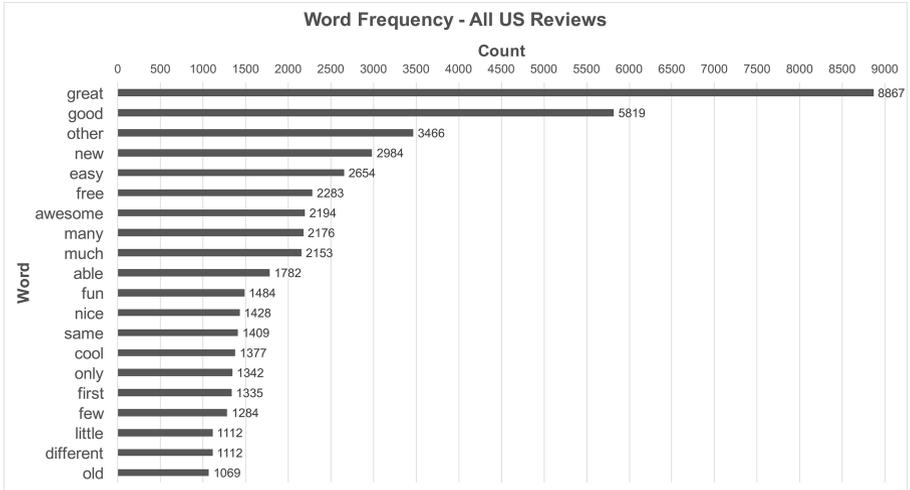


Fig. 1. Adjective frequency, all US reviews

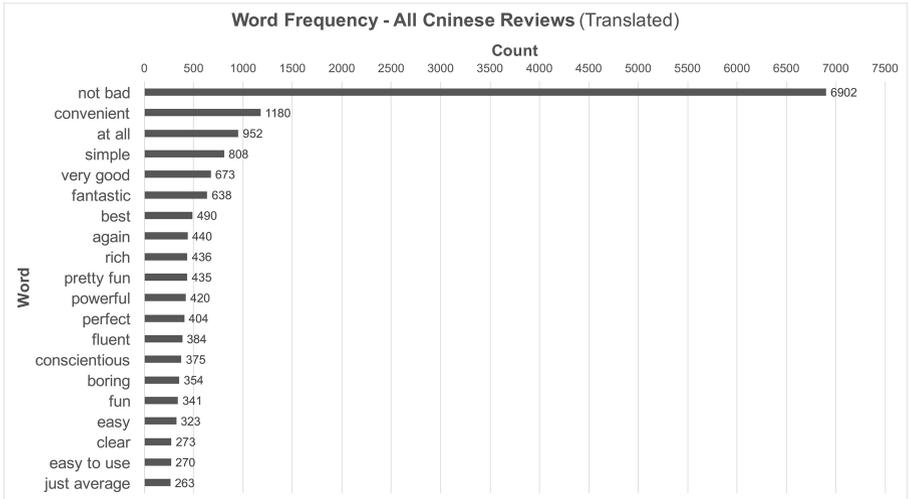


Fig. 2. Adjective frequency, all Chinese reviews

### 3.2 Cluster Results

$K$ -means clustering algorithm was performed on both US reviews collection and Chinese reviews collection. We ran multiple  $k$ -means with  $k$  various from 2 to 16 for both US reviews collection and Chinese reviews collection, and extracted top 50 terms in each cluster for each ran. Then we inspected the results for each ran manually to see if the reviews were clustered by topics or features. The final  $k$

**Table 2.** Top 20 frequent adjective words of four genres of US reviews, words that were not common among the four genres were presented in bold

Genre	Top 20 frequent adjective words
Social networking	Great, good, other, new, free, easy, able, many, much, <b>nice</b> , same, awesome, few, cool, only, <b>update</b> , <b>bad</b> , different, <b>old</b> , first
Photo	Great, good, easy, other, awesome, free, new, many, much, able, <b>nice</b> , cool, only, different, <b>simple</b> , first, same, <b>amazing</b> , few, <b>perfect</b>
Games	Great, good, other, new, <b>fun</b> , awesome, much, many, free, first, same, <b>hard</b> , <b>little</b> , only, able, few, different, easy, cool, <b>bad</b>
Entertainment	Great, good, other, new, free, many, awesome, much, easy, able, cool, <b>old</b> , few, same, only, first, <b>nice</b> , <b>bad</b> , <b>little</b> , different

**Table 3.** Top 20 frequent adjective words of four genres of Chinese reviews, words that were not common among the four genres were presented in bold

Genre	Top 20 frequent adjective words
Social networking	Not bad, <b>convenient</b> , very good, at all, fantastic, best, simple, fun, rich, <b>fluent</b> , <b>easy to use</b> , <b>powerful</b> , <b>clear</b> , <b>boring</b> , again, perfect, <b>concise in visual</b> , <b>pretty fun</b> , <b>beautiful</b> , <b>successful</b>
Photo	Not bad, <b>convenient</b> , simple, <b>powerful</b> , fantastic, at all, best, very good, perfect, <b>easy to use</b> , again, <b>easy</b> , <b>special effects</b> , <b>clear</b> , <b>concise in visual</b> , fun, <b>just average</b> , rich, <b>blurred</b> , <b>important</b>
Games	Not bad, at all, <b>pretty fun</b> , <b>conscientious</b> , simple, <b>boring</b> , again, fantastic, very good, simple, rich, <b>perfect</b> , fun, <b>just average</b> , <b>fluent</b> , best, <b>exquisite</b> , <b>important</b> , <b>delicate</b> , <b>severe</b>
Entertainment	Not bad, <b>convenient</b> , at all, <b>fluent</b> , very good, rich, best, fantastic, simple, <b>clear</b> , again, perfect, <b>powerful</b> , <b>boring</b> , fun, <b>just average</b> , <b>pretty fun</b> , <b>conscientious</b> , <b>easy to use</b> , <b>concise in visual</b>

values were both 12 for US reviews collection and Chinese reviews collection. The extracted top terms and clustered reviews focus were showed in Tables 4 and 5.

Both US reviews and Chinese reviews contained clusters of reviews for complains about compatible and crashes or bugs. Top terms of US reviews contained large amount of “great”, “like/love”, and “fun”, and top terms of Chinese reviews had more specific adjectives or descriptions. Customers in US AppStore complained about the advertisements in apps, while similar complain was not found

**Table 4.** *K*-means result for US reviews collection

Cluster	Number of reviews	Top 10 terms	Key feature
1	188	Upgrade loved app, app recently compatible, app past really, upgraded app recently, unusable, loved app past, upgraded app, made much, spoiled, app past	Compatible
2	839	Way many ad, every, ad pop, play, fun, ad every, good, time, get, great	<b>Advertisements in app</b>
3	855	Game, try, app crash, crashing, work, even, play, keep, time try, crash every	Crash
4	563	Pretty, cool game, app cool, really, game, game cool, seems cool, pretty cool app, cool good, really cool app	Cool game
5	1067	App love, much, fun, use, awesome, really, love love, like, like app, app much	love this app
6	668	Fun play, much, game fun, great fun, really fun, play, super fun, fun fun, nice game, fun use	Fun game
7	917	Great app use, great app work, app use, app work great, great great, work great, app work, great app easy, use, app easy	Great app
8	542	Awesome game, ever played, game much, played, game addicting, game best, addicting, love game addicting, much	Addicting game
9	57	Really, good, like, great original, new feature way, adding new feature, feature way playing, way playing, feature way, playing make	Feature
10	54058	Time, amazing, really, easy, one, would, make, best, please, play	Need fix or update
11	252	Actually work, love work, work perfectly, perfectly, love work great, actually, really, get work, work like, even work	App works good
12	1274	Time, work, refund, buy, work waste, get, even, get money, app waste, want refund	Waste of money, <b>want refund</b>

in Chinese reviews. Customers in US AppStore directly complained about “waste of money” and “want refunding”, while customers in Chinese AppStore expressed dislike for charging for membership. This results suggested that customers in US AppStore were more used to ask customer services for refunds than customers in China AppStore.

### Noise Review Detection

We were also interested in finding “noise” reviews in the review collection. If we consider most reviews were about the apps, then we could use DBSCAN to detect “noise” reviews that had little relevance to the reviewed app. We ran multiple times with different combinations of minimum number of points and distance, and recorded the estimated number of clusters and noise points, and extracted the review text of noise points. The final settings of the two input parameters were that *minimum number of points* = 5, and *distance* = 0.8 for both US reviews collection and Chinese reviews collection. Results showed that in genres of Social Networking and Photo, proportion of irrelevant reviews in US Store would be larger than that in Chinese Store (4.91% and 3.23% for US, 2.82 %

**Table 5.** *K*-means result for Chinese reviews collection (translated)

Cluster	Number of reviews	Top 10 terms (translated)	Key feature
1	36648	Live video, software, like, phone, good, fun, photo, game, effect, easy	Display, UI
2	1670	Version, endless, update, case, phone, photo, bug, album, reason, log in	Compatible
3	1237	First time, download, great, display, functions, good, recommend, find, really, work	Good use experience
4	1293	Update, good, can't open, things, display, uninstall, write reviews, delete, like	Can't use
5	836	Uninstall, free, video player, disgusting, really, annoying, app, wish, right away, bad	<b>Membership, charging</b>
6	692	First try, work good, friends, fun, trustworthy, app, really, display, like, simple	Easy use, interesting
7	1138	Feel, support, indeed, game, childhood, can't stop playing, recommend, originality, real, cute	Playability
8	984	Fun, filter, friends, app, wish, stickers, display, work great, support, recommend	Usability of photo editing
9	3323	Wish, support, friends, great, functions, utility, powerful, really, phone, reviews	Utility
10	789	Trash, effect, supper, friend, fun, support, useful, app, download, live video	Try this app out
11	2370	Wait for, fix it, support, every time, bug, can't log in, trash, system, can't open, video	Crash, bugs
12	4503	Crash, really, good, classic, interesting, time-killer, can't use, player, support, great	Time-killing game

and 2.65% for China), while in genres of Games and Entertainment, proportion of irrelevant reviews in US Store would be larger than that in Chinese Store (2.43% and 2.74% for US, 3.23 % and 3.42% for China), as showed in Table 6. However this difference in number of noise reviews was not that great, as we set same input parameters for both US reviews collection and Chinese reviews collection.

**Table 6.** Number of noise review detected

Genre	Number of noise reviews		Percentage of noise reviews (%)	
	US	CN	US	CN
Social networking	664	334	4.91	2.82
Photo & video	518	374	3.23	2.65
Games	408	545	2.43	3.23
Entertainment	410	433	2.74	3.42

## Discussion

This study investigated the difference of key term and patterns of content in apps review text among different genres and between China AppStore and US AppStore. We presented a preliminary method for mining customer opinions from

free-style review text. This review text mining technique could be used in customer opinion mining and customer satisfaction survey for mobile app publishers and other interested producers with further modification and improvement. The results showed that in general the key term used and opinion expressed in reviews of China AppStore and US AppStore were similar, only minor difference was found. One of the differences was that the reviews wrote by customers in China AppStore were more specifically related to the genres of the reviewed apps. The other differences was that customers in US AppStore were more used to ask customer services for refunds. This differences may caused by the fact that the internet-based services were relatively new to Chinese customers than US customers, and the return policy was more mature in US. As Chinese customers were less used to ask or complain to customer services, they may complain more in the reviews, hence the result of their reviews were more specifically related to the genres. This study was a pilot study, there is still more to explore with review text in this manner, and the comparison between two review collections needed to be more quantified, and works related to culture differences were still needed to investigate further.

**Acknowledgement.** This research was supported by the National Natural Science Foundation of China (NSFC, Grant Number 71471095). This study was also supported by Tsinghua University Initiative Scientific Research Program under Grant Number: 20131089234.

## References

1. Apple: Apple - choose your country or region. <https://www.apple.com/choose-your-country>
2. Archak, N., Ghose, A., Ipeirotis, P.G.: Deriving the pricing power of product features by mining consumer reviews. *Manage. Sci.* **57**(8), 1485–1509 (2011)
3. Bird, S., Klein, E., Loper, E.: *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media Inc., Sebastopol (2009)
4. Chen, Y., Fay, S., Wang, Q.: Marketing implications of online consumer product reviews. *Bus. Week* **7150**, 1–36 (2003)
5. Dellarocas, C.: The digitization of word of mouth: promise and challenges of online feedback mechanisms. *Manage. Sci.* **49**(10), 1407–1424 (2003)
6. Duan, W., Gu, B., Whinston, A.B.: Do online reviews matter?—An empirical investigation of panel data. *Decis. Support Syst.* **45**(4), 1007–1016 (2008)
7. fxsjy: Jie ba - Chinese text segmentation. <https://github.com/fxsjy/jieba>
8. Gamon, M., Aue, A., Corston-Oliver, S., Ringger, E.: Pulse: mining customer opinions from free text. In: Famili, A.F., Kok, J.N., Peña, J.M., Siebes, A., Feelders, A. (eds.) *IDA 2005*. LNCS, vol. 3646, pp. 121–132. Springer, Heidelberg (2005). doi:10.1007/11552253\_12
9. Kim, J., Park, Y., Kim, C., Lee, H.: Mobile application service networks: Apple's app store. *Serv. Bus.* **8**(1), 1–27 (2014)
10. McFadden, D.L., Train, K.E.: Consumers' evaluation of new products: learning from self and others. *J. Polit. Econ.* **104**, 683–703 (1996)
11. PGbiz: app store metrics. <http://www.pocketgamer.biz/metrics/app-store/>

12. Simmons, L.L., Mukhopadhyay, S., Conlon, S., Yang, J.: A computer aided content analysis of online reviews. *J. Comput. Inf. Syst.* **52**(1), 43–55 (2011)
13. Statista: cumulative number of apps downloaded from the apple app store from July 2008 to September 2016 (in billions). <http://www.statista.com/statistics/263794/number-of-downloads-from-the-apple-app-store>
14. Zhu, M., Fang, X.: A lexical approach to study computer games and game play experience via online reviews. *Int. J. Hum.-Comput. Interact.* **31**(6), 413–426 (2015)