ICEB 2013 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-1-2013

App Review Analytics Of Free Games Listed On Google Play

Chao Chang Chiu

Re Jiau Sung

Yu Ren Chen

Chih Hao Hsiao

APP REVIEW ANALYTICS OF FREE GAMES LISTED ON GOOGLE PLAY

Chao Chang Chiu, Yuan Ze University, Taiwan, 101chiu@gmail.com
Re Jiau Sung, Yuan Ze University, Taiwan, rejiausung@gmail.com
Yu Ren Chen, Yuan Ze University, Taiwan, spiralyu34@gmail.com
Chih Hao Hsiao, Yuan Ze University, Taiwan, imhaur@gmail.com

ABSTRACT

Smartphones have become popular in recent years; in turn, the number of application developers and publishers has grown rapidly. To understand users' app preferences, many platforms such as Google Play provide different mechanism that allows users to rank apps. However, more detailed insights on user's feelings, experiences, critiques, suggestions, or preferences are missing due to a lack of additional written comments. This research attempts to investigate the review analytics of Android games listed on Google Play using a proposed text analytic approach to extract all user reviews from game apps in Chinese. A total of 207,048 reviews of 4,268 free games from February to March 2013 are extracted and analyzed according to various metrics including game type and game attribute. The findings indicate there is high dependency between users' gender and game type, males and females have differing opinions on game attributes. In particular, users of different game types prefer different game attributes. The results reveal product usage insights, as well as best practices for developers.

Keywords: Android smartphone, text analytic approach, game apps, Google Play.

INTRODUCTION

In addition to serving as communication devices, smartphones also act as a venue for entertainment or commerce. The smartphone application (app) market has become one of the most important consumer electronic resources. This market emerged rapidly and is highly competitive, due to the lower costs associated with developing apps as compared to developing traditional software. Google Play provides different types of information including the app descriptions, screenshots, user reviews, and star ratings information to help users select apps. Prior researches about user generated contents (UGC) in the web have shown that they are useful as a marketing tool and effective in increasing competitive advantage. To understand users' app preferences, many platforms such as Google Play provide different mechanism that users can use to review downloaded apps. The most commonly mechanism is a ranking scale from 1 to 5. However, more detailed insights on user's feelings, experiences, critiques, suggestions, or preferences are missing due to a lack of additional written comments.

According to the report from app store analytics firm Distimo, "the uptake in the types of quick downloads are most common with games, which are the most downloaded and revenue generating types of apps" [1]. Game reviews, especially those written by fans and nonprofessionals, are a relevant source of information that can help us understand how players describe games, gameplay, and so on [2]. It undoubtedly can influence potential users' decision to purchase or download the game.

This research investigates Chinese app review analytics associated with Android smartphone games listed on Google Play using an opinion mining approach. Google Play provides different types of information including app descriptions, screenshots, user reviews, and star ratings to help users select apps. Opinion mining helps to identify the subjectivity, sentiment, appraisal or feeling of user expressed comments in unstructured texts on specific topics, or the overall context of a review, using certain analytic approaches [3]. In this paper, a heuristic n-phrase rule technique is proposed to automatically extract all user reviews crawled from game apps in Google Play. These texts are analyzed according to various metrics including game types (arcade & action, casual, brain & puzzle, cards & casino, sports, and racing) and game attributes (gameplay, aesthetics, musicality, stability, developer). The findings indicate there is high dependency between users' gender and game type, males and females have differing opinions on game attributes. In particular, users of different game types prefer different game attributes. The findings offer insights into product usage as well as best practices for developers.

BACKGROUND

App Reviews

Reviews add value by providing feedback to both the developer and the user community. App user reviews are important because they communicate information that may influence product-purchasing decisions via polarized sentiment and user expressed opinion. User reviews can warn people about undesirable or privacy-invasive apps [4]. For developers, reviews represent user generated and crowd-sourced content regarding user preferences and app quality; they also facilitate developers in terms of design priorities and marketing strategies. UGC offers the opportunity to learn from successful apps that are similar to the one being designed and developed [14].

Typically, app users provide a numerical rating (often as stars) and a brief text comment. Hu et al. [9] demonstrated that product reviews have a J-shaped distribution with mostly 5-star ratings, some 1-star rating, and hardly any ratings in between. People tend to write reviews only when they are either extremely satisfied or dissatisfied.

Prior researchers have attempted to apply data mining techniques to deal with context dependent opinion words from app reviews. Ha & Wranger [7] examined a majority of reviews focused on the quality of apps available through Google Play and found that people often described an app using an adjective, wrote about its features/functionality, and clarified whether the app worked or not. The star ratings were generally positive and had a J-curve distribution. Huang & Ting [11] suggested that recipients' ability to search for information, the relationship between the message receiver and communicator, and opinion leader's viewpoints are the most important factors influencing the word-of-mouth marketing strategies for apps. Raison et al. [16] extracted user opinions posted at Gamespot by applying co-clustering to an adjective-context co-occurrence matrix. From the derived co-clusters, they discovered that game users tend to care about the overall look and feel of the game more than the concrete elements used in the game.
In the current paper, we employ a sentiment analysis approach to extract users' sentiments associated with polarized ratings of game apps on Google Play.

Opinion Mining
Opinion mining is a useful mechanism to determine the subjectivity, sentiment, appraisal or feeling of an author expressed in texts on specific topics or the overall contextual polarity of a document [3]. The essential issues in opinion mining include identifying how sentiments are expressed and whether the expressions indicate positive or negative opinions toward the subject. Opinion mining effectively reduces the amount of manual effort required to identify, store, and analyze business intelligence. Opinion mining can be classified into two major tasks: information extraction and sentiment classification. Information extraction focuses on the extraction of opinions consisting of information about particular aspects of interest in a structured form from a set of unstructured text data. Sentiment classification aims to explore effective techniques for classifying an opinionated text as expressing a positive or negative opinion that is treated as a text classification problem [13].
Sentiment classification can be performed at three different levels of text granularity: document-, sentence- and feature/attribute-level. Document-level sentiment classification attempts to classify an entire document as either positive or negative according to the overall sentiment expressed in the text. Sentence-level sentiment classification attempts to classify the positive or negative polarity of each sentence. Feature-level sentiment classification is intended to identify opinions expressed about individual features or attributes. While document-level and sentence-level sentiment classification can determine the overall sentiment in a document or sentence, they are unable to indicate positive/negative meanings for specific object attributes; feature-level classification is more appropriate for this task.
This study employs feature-level opinion mining to identify the specific product aspects users reviewed positively or negatively. Prior studies have shown that feature-level opinion mining has achieved valuable performance [12, 15]. Feature-level opinion mining consists of two primary subtasks: extracting information on various product attributes and associating each attribute with a corresponding opinion. For example, Hu and Liu [8] extracted nouns and noun phrases as features, such that nearby adjectives were extracted and considered as an opinion words. Turney [17] used two-word phrases that contained adjectives or adverbs in particular part-of-speech patterns from reviews. The opinion is determined by the semantic orientations of a group of words and/or phrases corresponding to the feature in the text.

THE PROPOSED METHODOLOGY

Users commonly seek quality information from online user reviews prior to downloading or purchasing an app, while many developers use online user reviews as an important resource for product development and marketing management. As illustrated in Figure 1, an opinion mining method is proposed to extract attribute-opinion pairs that reveal insights associated with user reviews for free downloadable Android games from Google Play. Content analysis accompanied by correspondence analysis (CA) is also utilized to summarize and visualize users' reviews according to various metrics including game type and game attributes.

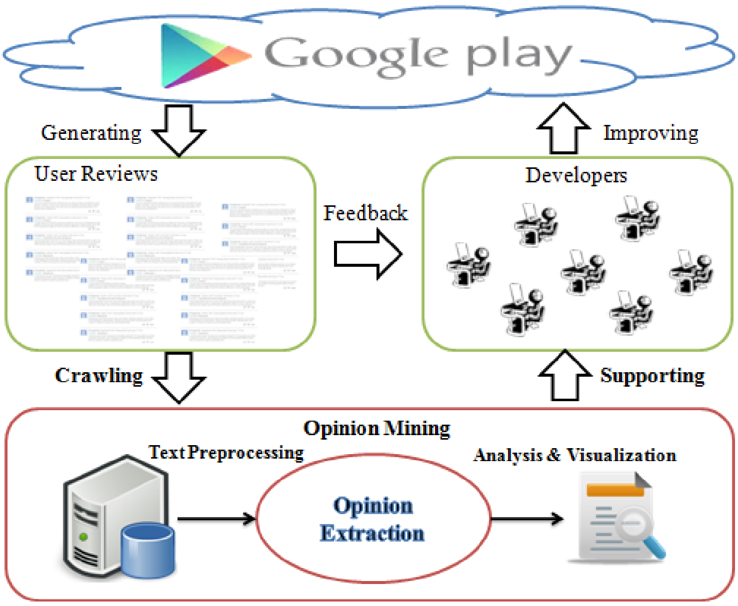


Figure 32. App Review Analysis Architecture

As shown in Figure 2, the app review mining process consists of four main steps: (1) data acquisition, (2) text preprocessing, (3) opinion extraction, and (4) analysis & visualization.

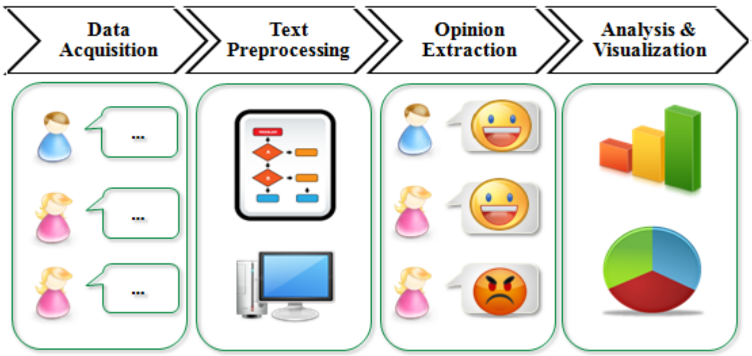


Figure 2. The App Review Mining Process

Data acquisition and text preprocessing

A web crawler program is developed to crawl UGC about free games across six game types listed on Google Play in Taiwan from February 1 to March 31, 2013; the game types include: arcade & action, casual, brain & puzzle, cards & casino, sports, and racing. This study focuses on free apps—Google Play only received permission to sell apps in Taiwan on February 27, 2013. The number of downloaded paid apps is obviously less than the number of downloaded free apps. The app and user information, including the name of the app and its developer, category, description, user account, gender, rating, comments, device, and post time are retrieved and stored in the initial repository.

Due to differences in contextual understanding between Chinese and English, existing text preprocessing techniques for English cannot be directly applied to Chinese sentiment classification. However, Academic Sinica developed a Chinese knowledge information processing (CKIP, http://ckip.iis.sinica.edu.tw/CKIP//index.htm) system used for word segmentation and part-of-speech tagging that is more applicable for processing Chinese reviews.

Chinese terms or phrases often contains more than two part-of-speech tags; as such, full parsing is inappropriate during mining because it is difficult to achieve high accuracy [20]. This study proposes a Chinese noun phrase chunking module to identify commonly used noun phrases. Most noun phrases contain at least one noun. The other word or words may be nouns, verbs or adjectives. This module automates the process of collecting and comparing keyword suggestions from Google to build a noun phrase list. In addition, stop words that contain little or no content information, such as pronouns, prepositions, conjunctions, interjections, digitals, and articles, need to be filtered out.

Opinion extraction

Attribute-opinion pairs extraction

In general, reviews for mobile apps are relatively short. Further, Vasa et al. [18] observed that users tend to leave significantly shorter reviews for games than for other categories. All sentences containing either attribute- or opinion-based information that expresses users' positive or negative opinions about a game are collected and manually labelled within a training data set. All nouns, noun phrases, verbs, adjectives, and adverbs of degree are then filtered as candidate attributes.

For complex and multi-faceted objects such as game apps, single words are often insufficient to describe different app attributes. In addition, some reviews focus on praising or complaining about the company that developed the app [7]. Further, game reviews vary in terms of the attributes covered. This study focuses on the following five game attributes:

- Gameplay: the specific way that players interact with a game such as play, weapons, levels, tasks, ending, challenges.
- Aesthetics: the app's overall look or interface such as style, picture quality, color scheme, resolution, appearance.
- Musicality: the quality of the sound effects, sounds, voices, tones, songs, music style.
- Stability: the stability of the game such as connection, server, reaction time.
- Developer: praise or complaints about the app developer.

The polarity of opinion words is measured based on the sentiment words set by HowNet and the National Taiwan University Sentiment Dictionary (NTUSD). Words not listed on HowNet or NTUSD are manually labeled. A minor complication regarding opinion word labeling is that attributes within a sentence can be stated explicitly or implicitly. For example, 'stability' is an implicit attribute corresponding to negative opinions including crashes, hang ups, break downs, blank screens, lag, and bugs that may not appear in the review, but are still implied. All the opinion words are contained in the opinion lexicon. Further, negation operators such as no, not, and never that reverse sentence meanings are adopted.

Each matched degree word has a predefined strength value that is used to compute the strength sentiment of phrase. The following is a list of Chinese degree words defined on HowNet: (1) 欠 (insufficiently); (2) 稍 (-ish); (3) 較 (more); (4) 很 (very), (5) 最 (most); (6) 超 (over) [10]. The list contains 228 degree words, and each matched degree word has a predefined strength value used to compute the strength sentiment of the word.

Heuristic n-phrase rule

We assume that an attribute-opinion pair contains an attribute and an opinion word that tend to be found close to each other. In addition, Chinese reviews commonly place the subject before the opinion word rather than after it. A heuristic n-phrase rule is proposed to identify the opinion polarity of an attribute in review sentences. An n-phrase is a contiguous slice of *n* words or phrases of a longer sentence. An n-phrase of size 2 is referred to as a bi-phrase; size 3 is a tri-phrase; size 4 is a four-phrase, and so on. The steps are as follows:

Step 1: Identify the attribute in a processed sentence—that is, whether it is on the predefined game attribute word list.

Step 2: Check the first phrase after the attribute to see whether it is in the opinion lexicon. If so, these two words "attribute" + "opinion polarity" are put together as an attribute-opinion pair. If the first phrase after the attribute is not in the opinion lexicon, check the first phrase prior to the attribute. If this matches, these two words are put together as an attribute-opinion pair.

Step 3: If the first phrase prior to the attribute is not in the opinion lexicon, check the second phrase after the attribute. If no, check the second phrase prior to the attribute. Continue this procedure until a qualified attribute-opinion pair is located, or until none can be found.

Step 4: If a negation operator precedes the opinion word by one or two words, the opinion polarity is reversed.

Step 5: If a degree word precedes the opinion word by one or two words, the strength sentiment of the phrase is computed based on the predefined level of strength of the degree word.

Sentiment with Opinion Scoring

The sentiment is computed based on the matched opinion words, degree words and negations. Let *G* denote a set of free Android games and *R* denote a set of user reviews of a game. To extract and evaluate the opinion for user reviews, the scoring function is defined as follows:

$$\mathrm{OS}_i = \sum_{j=1}^{n}\left(\omega \times \sum_{k=1}^{m}\mathrm{SO}\left(w_k\right)\right) \tag{1}$$

where

OS_i: opinion score of *i*-th game in *G*

r_i: *j*-th user review in *R*

n: the number of user reviews

w_k: *k*-th opinion word

m: the number of identified opinion words in the review

$\mathrm{SO}(w_k)$: the polarity (+1 or -1) of an opinion word

ω: the degree weights assigned to six category of degree words: (1) insufficiently: 0.83; (2) -ish: 1.67; (3) more: 2.50; (4) very: 3.33, (5) most: 4.17; (6) over: 5.00.

Visualization

User reviews of a game can be subdivided into specific attributes that can be graphically illustrated on coordinate axes known as a perceptual map. CA is a technique for representing categorical data on a low-dimensional map. It is one of the most commonly used visualization techniques to produce perceptual maps and has been successfully applied in different fields. CA can be used to identify a brand's competitive strengths, as well as ideas to improve a brand's competitive position [19]. In this study, CA is employed on a two-dimensional perceptual map to visually display the relationships between the game and the users' opinions with respect to different attributes, where the distances on the map represent correspondence.

Basically, CA takes the frequency of co-occurring features and converts them to distances, which are then plotted, revealing how things are related in terms of how close to or far from each other they are in a two- or three-dimensional visualization. CA allows for a clear and graphic presentation of interdependence among a set of categorical variables. For visualization purposes and ease of interpretability, a large percentage of total variance is accounted for by CA's first two principal axes. The distance on the map represents the closeness: a closer proximity means greater perceived similarity [5, 6].

THE EXPERIMENTS

Data Description

This study mined and analyzed 207,048 reviews of 4,268 game apps from 2,181 developer free games across the six game types (arcade & action, casual, brain & puzzle, cards & casino, sports, and racing) listed on Google Play from February 1 to March 31, 2013. Further, since Google Plus is integrated with Google Play, any reviews written by Google Plus members appear on their individual profile, which includes their name and gender, as well as other information. Through the integration to Google Plus, the users' individual profiles are contained.

Table 1 shows the descriptive statistics of the data, which includes the numbers of reviews made by men (59.5%), women (32.3%), and those with no specified gender (8.2%). In addition, more than 75% of the reviews focused on either arcade & action or casual games. More specifically, males reviewed more arcade & action, sports, and racing games, while women reviewed more brain & puzzle and casual games, suggesting significantly different game type preferences between genders. Furthermore, the reviews were overwhelmingly positive: the number of positive reviews exceeded the negative ones by a factor of 2.8, particularly for the arcade & action and sports games.

Table 1. Descriptive Statistics of the Data

Game Type	No. of Apps	No. of Developers	No. of Reviews	Gender				Opinion		
				Men	Women	Unknown	M/W	Positive	Negative	P/N
Arcade & Action	1,159	519	78,501	55,616	16,853	6,032	3.3	62,486	16,015	3.9
Casual	1,235	557	77,578	37,986	32,444	7,148	1.2	55,984	21,594	2.6
Brain & Puzzle	852	472	28,432	13,472	12,715	2,245	1.1	18,024	10,408	1.7
Cards & Casino	428	224	9,891	5,979	3,339	573	1.8	7,229	2,662	2.7
Sports	268	177	8,190	6,711	938	541	7.2	6,554	1,636	4.0
Racing	326	232	4,456	3,472	610	374	5.7	2,261	2,195	1.0
Total	4,268	2,181	207,048	123,236	66,899	16,913	1.8	152,538	54,510	2.8

* M/W: men/women ratio; P/N: positive/negative ratio.

J-Shaped Distribution of Users' Reviews

As shown in Figure 3, the distribution of users' reviews exhibited a J-shaped distribution with mostly 5-star ratings across all samples, followed by 1-star ratings. 80.0%, 81.7%, 81.5%, 83.7%, and 77.6% of the ratings, developers, reviews, reviews from male, and reviews from female are greater or equal to four stars that confirming review ratings of free Android game apps are overwhelmingly positive. However, the J-shaped distribution for female reviews was flatter than that for males. This suggests that males tend to write reviews when they are either extremely satisfied or extremely unsatisfied accompanied with the extremely star rating. In contrast, females are more prudent in terms of star ratings than are males. In addition, the polarized developer ratings (77.0% 5-star; 14.1% 1-star) demonstrate that users tended to express their sentiments using extreme ratings.

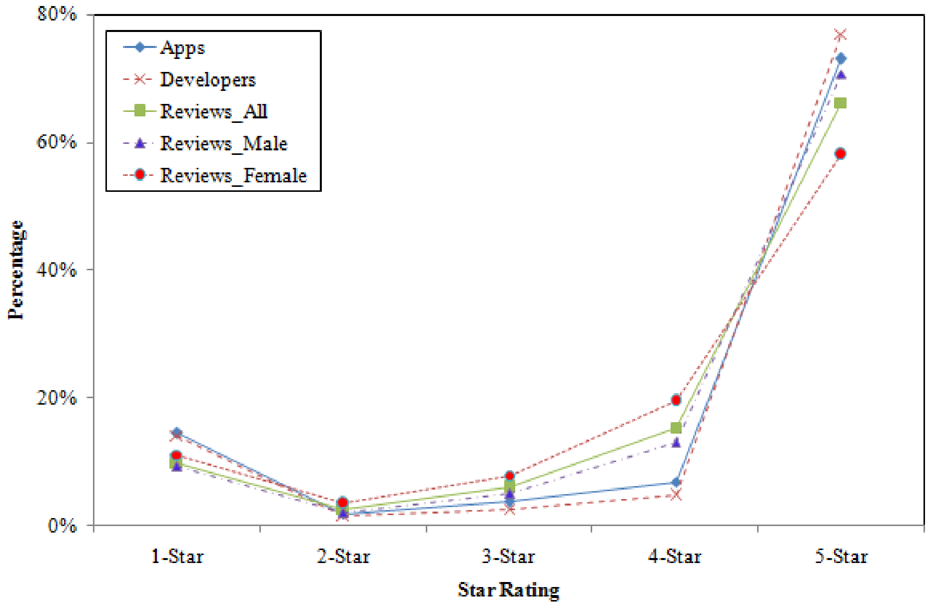
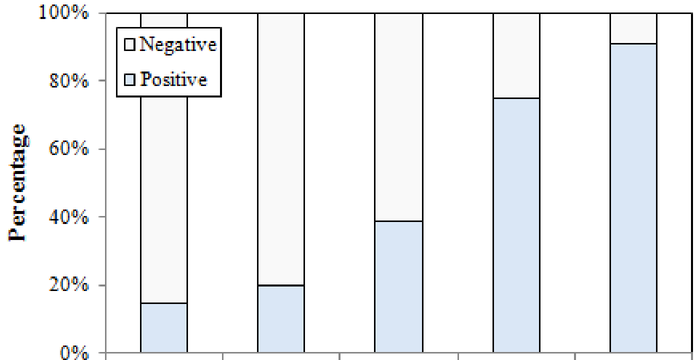


Figure 3. Distribution of Star Rating Reviews by Game, Developer, and Gender

Figure 4 presents the distribution of sentiments calculated using the opinion scoring technique across the ratings. A strong correlation of star ratings with positive reviews can be observed. Based on the distribution of sentiment polarity, 91.3% of the 5-star rating reviews had positive polarity, while 85.6% of the 1-star rating reviews exhibited negative polarity.



	1_Star	2_Star	3_Star	4_Star	5_Star
Negative	17,296	4,268	7,916	9,022	14,963
Positive	2,913	1,050	4,981	27,157	156,492

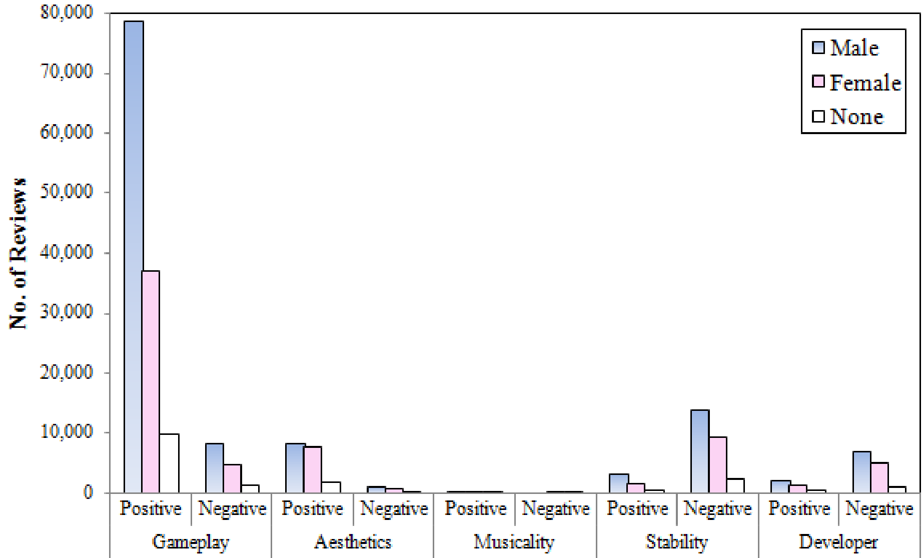
Star Rating

Figure 4. Distribution of Sentiments among the Ratings

Sentiment Mining for Game Attributes

Figure 5 shows the distribution of reviews among the game attributes based on gender and opinion polarity. Most reviews (67.5%) focus on gameplay discussion, while 17.6% focus on stability. Only 0.4% discuss musicality. Male positive reviews for game attributes nearly double those from females; however, male negative reviews exceed female negative reviews by 50%.

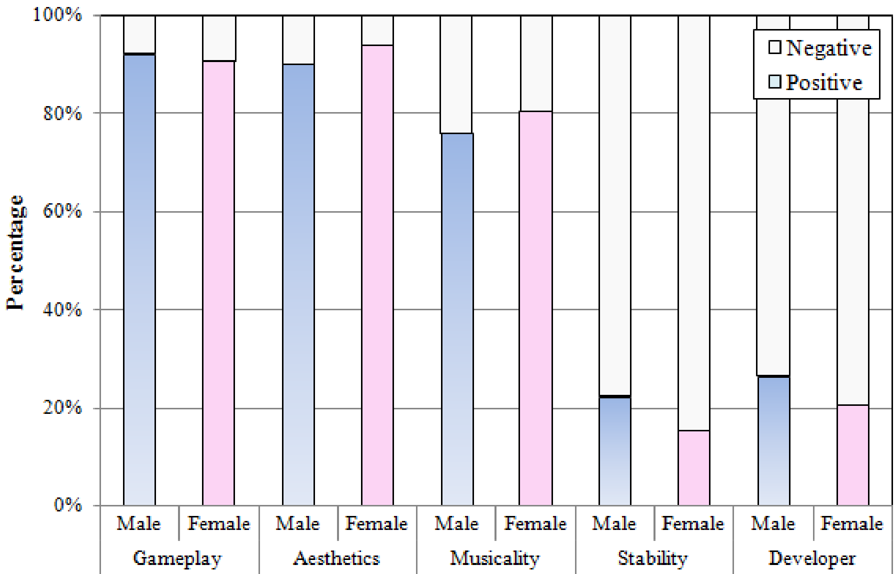
Figure 6 presents the distribution of sentiments among the game attributes. Based on the distribution of sentiment polarity, most of the reviews associated with gameplay, aesthetics and musicality have positive polarity; however, those associated with stability and the developer have negative polarity. More specifically, 84.7% of female stability reviews by women are negative, while the figure for men is 77.7%; 79.5% of female reviews regarding the developer are negative, as are 73.5% of male reviews.



	Gameplay		Aesthetics		Musicality		Stability		Developer	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
Male	78,814	8,286	8,242	1,109	375	141	3,275	13,814	2,272	6,908
Female	36,921	4,783	7,637	693	243	82	1,393	9,092	1,147	4,908
None	9,711	1,182	1,644	199	80	34	446	2,209	338	1,070

Attributes

Figure 5. Distribution of Reviews among the Game Attributes



	Gameplay		Aesthetics		Musicality		Stability		Developer	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Negative	9,073	5,243	1,281	778	165	90	14,771	9,724	7,224	5,117
Positive	106,97	51,354	11,643	11,819	519	366	4,241	1,755	2,602	1,316

Attributes

Figure 6. Distribution of Sentiments among the Game Attributes

In this study, a commercial software "XLSTAT" is used to conduct CA to better understand the relationships among the different game types, star ratings, and users' sentiment polarities on game attributes within a two-dimensional perceptual map. Figure 7 shows the perceptual map for the six game types on the relative proximities of the five attributes with sentiment polarities (P: positive, N: negative) in corresponding space. The horizontal axis (Dimension 1) accounted for 63.80% of the total variance, and the vertical axis (Dimension 2) for 30.66%. As such, the associations between each attribute and user sentiment polarity are will explained by Dimensions 1 and 2 (94.46%). Further, as shown in Figure 7, three groups emerge: (a) arcade & action and sports games are mostly rated positively in terms of gameplay and musicality, (b) cards & casino games are rated positively in terms of stability, and (c) brain & puzzle, casual, and racing games are mostly rated negatively in terms of gameplay, aesthetics, musicality, and stability, but developer ratings show polarized sentiment.

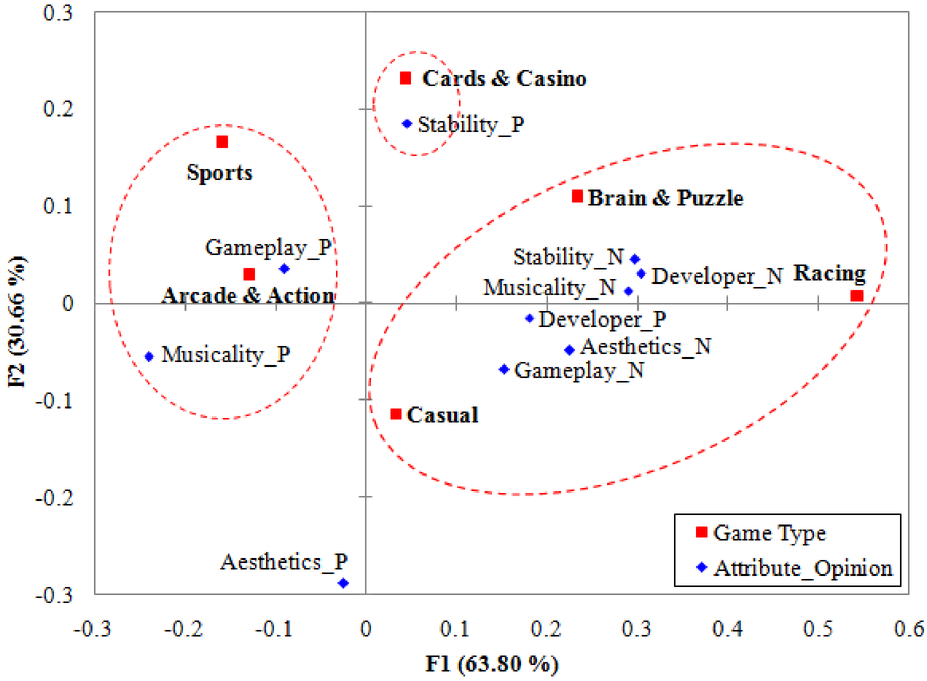


Figure 7. Perceptual Map for Game Type and Attribute

Figure 8 depicts the perceptual map for the star ratings and game attributes. The first two principal components collectively explain 99.85% of the variance, with 98.80% accounted for by the first dimension and 1.05% accounted for by the second dimension. The 5-star ratings are mostly related to positive sentiment in terms of gameplay and musicality; the 3-star ratings are mostly related to negative sentiment in terms of gameplay, musicality, and aesthetics; and the 1- and 2-star ratings are mostly related to negative sentiment in terms of stability and the developer. These findings suggest that the gamplay and musicality of game apps are generally viewed positively by users; however, users are unlikely to tolerate game instability.

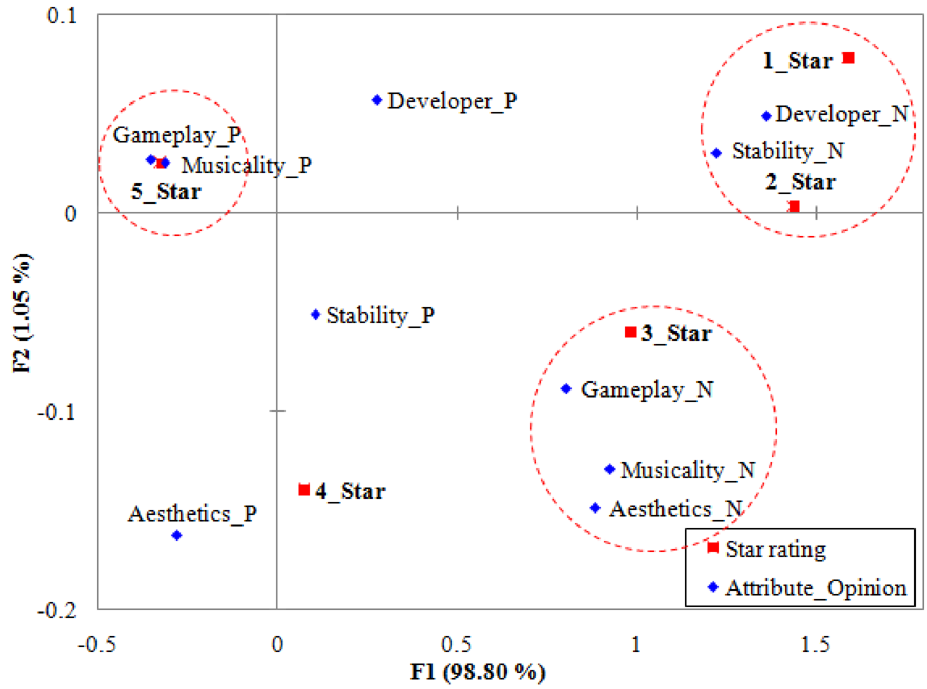


Figure 8. Perceptual Map for Star Rating and Attribute

Review Analytics for Popular Games

Table 1 shows the top 5 games among men, women and all respondents based on sentiments calculated using the opinion scoring technique. These most popular games generally fall into the arcade & action and casual game types; however, the *9 Innings: 2013 pro baseball* sport game is the highest rated game overall, which correlates with the fact that baseball is the national sport in Taiwan.

Table. 2 Top 5 Games among All Respondents, Males, and Females

	Ranking	1	2	3	4	5
All	Name	9 Innings: 2013 Pro Baseball	Chick Kitchen	Cat War2	Undead Slayer	Mandora
	Game Type	Sports	Casual	Arcade & Action	Arcade & Action	Arcade & Action
	Developer	Com2uS	iDT Digital	WestRiver	NHN	Rayark
	Opinion Score	5,566	5,263	4,983	4,737	4,390
Males	Name	9 Innings: 2013 Pro Baseball	Undead Slayer	Cat War2	Little Commander - WWII	Battle Cats
	Game Type	Sports	Arcade & Action	Arcade & Action	Casual	Casual
	Developer	Com2uS	NHN	WestRiver	Cat Studio HK	PONOS
	Opinion Score	4,628	4,150	3,748	3,337	2,369
Females	Name	Chick Kitchen	Mandora	LINE Play	Hotel Story	Ovenbreak
	Game Type	Casual	Arcade & Action	Casual	Casual	Arcade & Action
	Developer	iDT Digital	Rayark	NAVER	Happy Labs	Com2uS
	Opinion Score	3,554	2,395	1,758	1,470	1,434

Figures 9-11 show the perceptual maps for the top five popular games among all respondents, men, and women, respectively, in terms of the relative proximities of the five attributes with sentiment polarities in the correspondence space. In Figure 9, the horizontal axis (Dimension 1) accounts for 86.67% of the total variance, and the vertical axis (Dimension 2) for only 6.82%; accordingly, user sentiment polarity seems to be explained by Dimensions 1 and 2 (95.49%). Further, *Cat War 2*, *9 Innings: 2013 Pro Baseball*, and *Undead Slayer* are rated similarly by users: polarized sentiment in terms of gameplay and the developer, negative sentiment in terms of musicality and aesthetics, and positive sentiment in terms of stability. *Mandora* and *Chick Kitchen* show positive sentiment in terms of musicality and aesthetics, but negative sentiment in terms of stability.

In Figure 10, the first two principal components collectively explain 81.23% of the variance, with 51.39% accounted for by the first dimension, and 29.84% by the second dimension. *9 Innings: 2013 Pro Baseball*, *Undead Slayer*, and *Little Commander – WWII Battle* are mostly rated positively in terms of gameplay and stability, negatively in terms of the developer, and show polarized results in terms of aesthetics and musicality.

Figure 11 shows that the first two principal components explain 93.14% of the variance, with 48.62% accounted for by the first dimension and 44.52% by the second dimension. The five games are located in four separate quadrants: *Ovenbreak* and *Mandora* show polarized sentiment in terms of gameplay and musicality; *LINE Play* is rated negatively in terms of gameplay but shows polarized sentiment in terms of stability and the developer; *Chick Kitchen* is rated positively in terms of aesthetics; but *Hotel Story* is rated negatively on aesthetics.

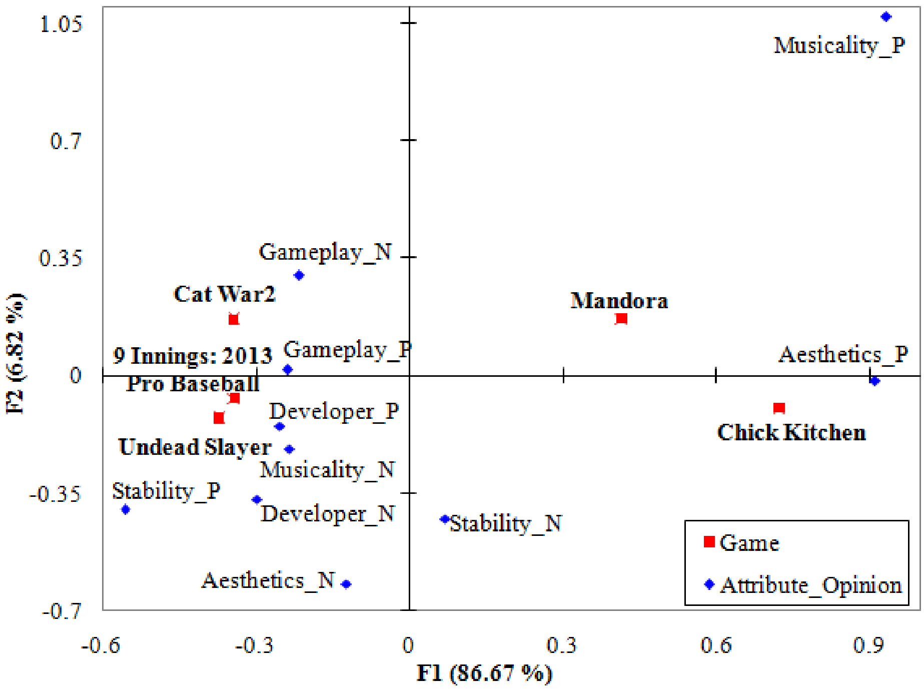


Figure 9. Perceptual Map for Top 5 Games among All Respondents

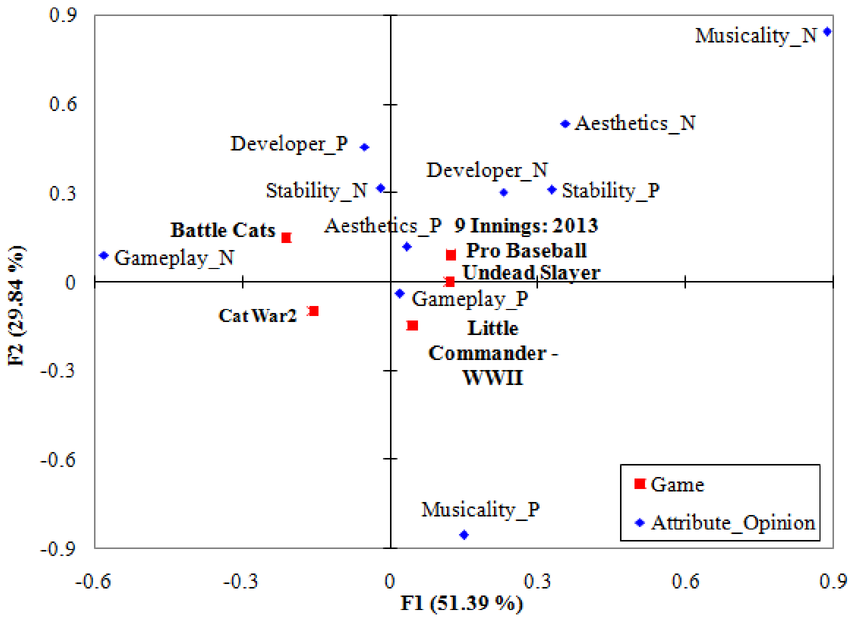


Figure 10. Perceptual Map for Top 5 Games among Males

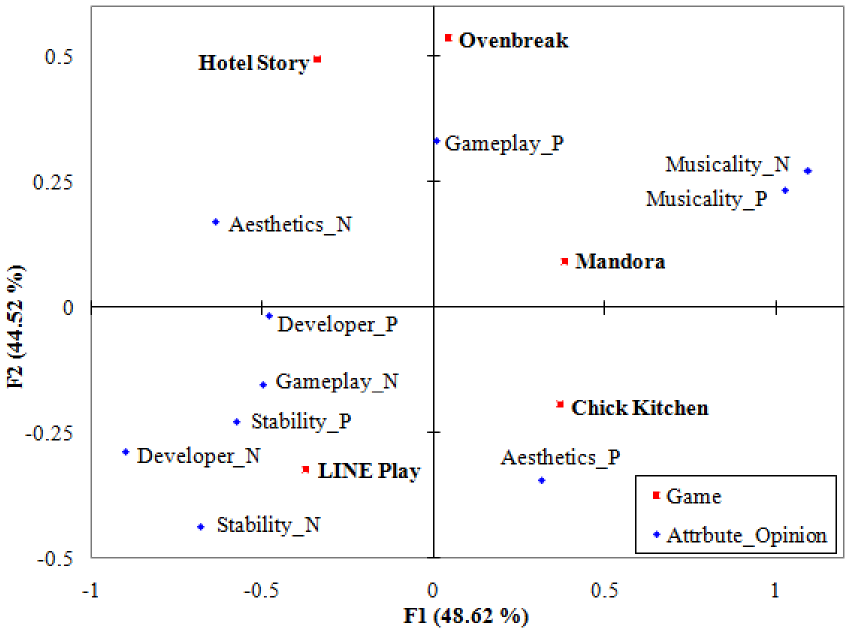


Figure 11. Perceptual Map for Top 5 Games among Females

DISCUSSION

Game apps are facing intense competition due to a fast growing, emerging market, and developers need to be cognizant of how users perceive their products. The interpersonal influence arising from opinion exchange is an important factor influencing users' selection decisions. Users seek quality information from online user reviews prior to purchasing a product. Developers need to put more focus on user perceptions of their games, including what their users want and expect from an app. It is essential that developers constantly monitor and assess online user reviews to identify how users rate the various game attributes.

UGC in Google Play contains substantive information about apps. In this study, the content of users' reviews of free downloadable Android game apps is analyzed. Overall, most reviews focus on arcade & action and casual games game types, as well as two attributes: gameplay and stability. Males' reviews outnumber those by females, particularly for arcade & action, sports, and racing games; the opposite is true for brain & puzzle and casual games. This shows that there are significantly different preferences in terms of game type across genders. Further, the reviews are overwhelmingly positive, particularly for arcade & action and sports games. Our results suggest a J-shaped distribution with mostly 5-star ratings for a series of data sets including games, developers, and reviews by males, females, and all respondents.

With the aid of CA, we can learn more about the relationship between game types and user sentiments regarding particular game attributes, as well as the relationship between star ratings and user sentiments regarding game attributes. Arcade & action and sports games are mostly rated positively in terms of gameplay and musicality, while cards & casino games are rated positively in terms of stability. However, brain & puzzle, casual and racing games are mostly rated negatively in terms of gameplay, aesthetics, musicality, and stability.

Further, 5-star ratings are most often given for gameplay and musicality, while 1- and 2-star ratings are most often given for stability and the developer. These findings suggest that the gameplay and musicality of game apps are generally recognized as positive by users; however, they are less likely to tolerate game instability. Finally, the top 5 popular games for males, females and all respondents are closely associated with users' sentiments regarding game attributes. Our research findings offer critical information associated with users' true experience, which can help developers provide immediate, complete, and accurate improvements to existing products and future designs.

CONCLUSIONS

This study develops an opinion mining approach related to feature-level sentiment classification that extracts online user reviews for free downloadable Android games listed on Google Play. A heuristic n-phrase rule is proposed to extract the attribute-opinion pairs to elicit user opinions about game apps. The combination of content analysis and CA helps to summarize and visualize users' reviews according to different metrics including star ratings, game types (arcade & action, casual, brain & puzzle, cards & casino, sports, racing), and game attributes (gameplay, aesthetics, musicality, stability, developer). Using a sentiment analysis approach can effectively capture positive and negative opinions and distinguish between them. This study provides some insights on users' published reviews, as well as greater clarity on what app attributes and opinions are important to users.

E-commerce is rapidly expanding, thereby facilitating online purchasing. Understanding the relationship between current offerings on the app market and user's experiences and opinions can help developers to better understand their game image and what attributes influence users' selection decisions. In this way, they can ensure the best possible end-user experience. Furthermore, developers should consider how best to exploit social recommendations and tactical sales promotions. They can encourage positive word-of-mouth from existing users as part of their marketing strategy, or provide more information-rich reviews to assist potential users. Developers should also continually track users' opinions to stay cognizant of their weak attributes, such that they can improve in these areas.

However, there are some inherent limitations in contents analytics such as free and short formats of text and the difficulty to generalize findings due to the minority users who will actually express their feeling in text. Thus, the research results can reflect partial users' reviews, instead of for all. Additionally, the complicated Chinese grammar and linguistic structure may result in omission bias or wrong judgment for constructing the attribute–opinion pairs. Future researchers may wish to more fully investigate paid game apps as well as the diversified information sources about them, including game discussion forums. The attribute-opinion pair extraction technique can be improved by including more detailed information on attributes connected to game reviews.

ACKNOWLEDGEMENT

This research was supported by the grants from NSC-102-2410-H-155-035-MY2 of National Science Council, Taiwan.

REFERENCES

[1] Distimo, (2013) '2012 Year in Review', available at http://www.distimo.com/publications (accessed 5 September 2013).

[2] Zagal, J. P., Tomuro, N., & Shepitsen, A. (2012) 'Natural Language Processing in Game Studies Research: An Overview' *Simulation & Gaming*, Vol. 43, No. 3, pp. 356-373.

[3] Liu, B. (2010) 'Sentiment analysis and subjectivity', *Handbook of Natural Language Processing*, 2nd edition, pp.627–666.

[4] Felt, A. P., Ha, E., Egelman, S., Haney, A., Chin, E., & Wagner, D. (2012, July) 'Android permissions: User attention, comprehension, and behavior', *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS 2012), Washington, DC, USA., July 11-13*, pp. 3.

[5] Greenacre, M. J. (1984) 'Theory and applications of correspondence analysis '

[6] Greenacre, M. (2006) 'Tying up the loose ends in simple, multiple, joint correspondence analysis', *Proceedings of Compstat 2006-Proceedings in Computational Statistics*, Physica-Verlag HD, pp. 163-185.

[7] Ha, E., & Wagner, D. (2013) '*Do Android users write about electric sheep? Examining consumer reviews in Google Play*', *Proceedings of* Consumer Communications and Networking Conference (CCNC 2013), pp. 149-157.

[8] Hu, M., & Liu, B. (2004) 'Mining and summarizing customer reviews', *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168-177.

[9] Hu, N., Zhang, J., & Pavlou, P. A. (2009) 'Overcoming the J-shaped distribution of product reviews', *Communications of the ACM*, Vol. 52, No. 10, pp.144-147.

[10] Huang, C. R., Chen, F. Y., Chen, K. J., Gao, Z., and Chen, K. Y. (2000) 'Sinica Treebank: design criteria, annotation guidelines, and on-line interface', *Proceedings of the second workshop on Chinese language processing: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics*, Vol. 12, pp. 29-37.

[11] Huang, C. Y., & Ting, Y. H. (2012) 'Derivations of factors influencing the word-of-mouth marketing strategies for smart phone APPs by using the Fuzzy DEMATEL based Network Process' , *Proceedings of Fuzzy Theory and it's Applications (iFUZZY 2012),International Conference on*, pp. 42-47.

[12] Li, Z., Zhang, M., Ma, S., Zhou, B., & Sun, Y. (2009) 'Automatic extraction for product feature words from comments on the web', *Information Retrieval Technology*, Springer Berlin Heidelberg, pp. 112-123.

[13] Liu, B. (2010) 'Sentiment analysis and subjectivity', *Handbook of natural language processing*, second edition.

[14] Platzer, E. (2011) 'Opportunities of automated motive-based user review analysis in the context of mobile app acceptance', *Proceedings of CECIIS*, pp. 309-316.

[15] Popescu, A. M., & Etzioni, O. (2007) 'Extracting product features and opinions from reviews', *Natural language processing and text mining*, Springer London, pp. 9-28.

[16] Raison, K., Tomuro, N., Lytinen, S., & Zagal, J. P. (2012) 'Extraction of user opinions by adjective-context co-clustering for game review texts', *Proceedings of Advances in Natural Language Processing*, Springer Berlin Heidelberg, pp. 289-299.

[17] Turney, P. D. (2002, July) 'Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews', *Proceedings of the 40th annual meeting on association for computational linguistics*, Association for Computational Linguistics, pp. 417-424.

[18] Vasa, R., Hoon, L., Mouzakis, K., & Noguchi, A. (2012) 'A preliminary analysis of mobile app user reviews', *Proceedings of the 24th Australian Computer-Human Interaction Conference*, pp. 241-244).

[19] Whitlark, D. B., & Smith, S. M. (2001) 'Using correspondence analysis map relationships', *Marketing Research*, Vol. 38, No. 4, pp. 22-27.

[20] Wu, S. H., Shih, C. W., Wu, C. W., Tsai, T. H., & Hsu, W. L. (2005) 'Applying maximum entropy to robust Chinese shallow parsing', *Proceedings of ROCLING*.