

國立政治大學資訊管理學系

碩士學位論文

指導教授:郝方博士

AppCAT: 基於手機應用程式評論情感分析和產品功能發掘 - 以 iOS 為例

AppCAT: Systematic Sentiment Analysis
of Mobile Application Reviews

研究生：黃書韋

中華民國一零五年六月

Abstract

User reviews of mobile apps often contain complaints or suggestions which are valuable for app developers to improve user experience and satisfaction. However, due to the large volume and noisy-nature of those reviews, manually analyzing them for useful opinions is quite challenging. To address this problem, we propose AppCAT, a sentiment and feature mining framework for automated review analysis. AppCAT defines the initial sets of keywords of those comments. And it use word similarity technique to expand the initial sets by grouping other keywords to find out the product features of those apps. Furthermore, AppCAT detects the sentiment and its subject(a product feature) of those reviews and figure out the user attitude towards those product feature of a specific app. AppCAT use those data to plot a bar chart to visualize those feature polarities for users to facilitate if they should consider this app. For the app developers, they can use this system to get the opinion overview of users as a basis of revision.

Contents

1	Introduction	1
2	Related Work	3
2.1	Research on Mobile Application Reviews	3
2.2	Identifying Product Feature	3
2.3	Sentiment Analysis	5
3	Methodology	5
3.1	Review preprocess	6
3.1.1	Download review data	7
3.1.2	Natural Language Processing	7
3.1.3	Filter and Add Frequent Words	7
3.2	Fetch Product Feature	8
3.2.1	Define Initial Product Feature Keywords	8
3.2.2	Transform tokens into vector	10
3.2.3	Form feature keyword sets	10
3.3	Sentiment Analysis on Product Feature	10
3.3.1	Find Review Subject	10
3.3.2	Analyze Sentiment	11
3.3.3	Visualization	12
4	Experiment	12
4.1	Some Apps with privacy issue	13
4.2	Apps in Finance Category	13
4.3	Apps in Weather Category	18
5	Validation with AppReco	21
6	Conclusion	22

1 Introduction

The ownership of mobile devices such as smartphones, smartwatches and tablets, has grown significantly. According to the recent Gartner report [3], these mobile devices will soon exceed 120 million in 2015. To serve mobile device users, mobile applications like iOS and Android Apps have been increasing remarkably in the past decade and become one of the most dominant software applications in history. Both Google Play and Apple's App Store have over 1.4 million apps available for public download [2, 10]. In the recent annual Apple conference, WWDC 2015, Tim Cook announced that the number of apps download in Apple's iTunes Store has crossed 100 billions.

User reviews of mobile apps often contain negative or positive experiences with attitude from user, When a user wants to download an app, he will check the rating and the count of ratings to decide whether it's download-worthy. Furthermore, users can check through the reviews with respect to this app to decide whether it is worthwhile to download.

Unfortunately, mobile app store like Google play [5] or iTunes Store [8]do not provide statistic or quick summary of those reviews. When a customer wants to download applications into their smart devices, they have to scroll down through the reviews to decide whether the app is download-worthy. Besides, in the view of developers, developing app is a highly competitive business, as millions and counting apps of different categories are made available on app markets. Since the revenue and profit of a mobile app is often proportional to the size of its userbase, improving user experience and satisfaction to retain existing users and attract new ones is of important to its developers. User opinions like complaints or suggestions would be valuable for that task.

As mobile app markets typically provide rating and reviewing mechanisms, reviews from users of apps purchased on those markets provide an important source of user opinions. However, analyzing those reviews to get useful opinions would be inherently challenging. First, a popular app with millions of users often gets thousands of reviews each day. Reading all of those reviews would be very time-consuming. In addition, user reviews of mobile apps are often noisy. They can have typos, acronyms, abbreviations, etc. Even

worse, prior research reports that more than 60% of user reviews do not contain useful opinions.

Most reviews are short and lots of apps receive under 50 reviews in the first year. And most app receive more shorter reviews than longer reviews. [14] As a result, manually scan through app reviews is not a effective action for user to do. And if we can quickly provide a summary of automatically analysis of app reviews for user to choose from those similar apps. It would be a valuable work to do. In this paper, an automated framework called AppCAT(Application Comment Analysis Tool) for mining user opinions from user reviews of mobile applications is presented. It combined product feature identification using sentence comparing by word2vec model to find out the comment subject topics and the keywords from review analysis. For each sentence towards a specific product feature of an app. Sentiment analysis techniques are performed to defined users' attitude to this feature.

Product feature is a key factor for user as they provide them with hints about how well a product or service will deliver its benefits [9]. We define five main product features of mobile application regarding to different aspects(UI/UX, system performance, privacy, price, advertisement) A initial set of application product feature is defined above and the keywords in this features are extend by using google's word2vec tool.

And use cluster technique to group those keywords to find out the product features of those apps.

Furthermore, AppCAT detects the sentiment and its subject(a product feature) of those reviews and figure out the user attitude towards those product feature of the specific app. AppCAT use those data to plot a bar chart to visualize those feature polarities for users to facilitate if they should consider this app.

As a result, we also provides some experiments and comparison to other app recommendation system, AppReco, to validate the result of sentiment analysis.

2 Related Work

The research of AppCat involves three parts. First parts are number of empirical and exploratory studies on the importance of app's reviews in app development process. Second part is some research on how to identify the product feature among those reviews to a specific product. It is important that if we can find the review subject, then we can figure out the sentiment of this user towards the product feature. Final part the key part of the study, sentiment analysis. There are lots of research studied on how to find out the sentiment of a sentence.

2.1 Research on Mobile Application Reviews

Phong et al. [25] develop a keyword-based approach to mining user opinions in app reviews. It applies word2vec technique to cluster the keywords appeared in mobile reviews. And it can find problems of a specific mobile application regard to its reviews.

Hoon and Leonard et at. [14] performed a survey on mobile application review and some statistical research on those data. They find out that the rate of review growth and profile of reviews changes between different apps and categories.

2.2 Identifying Product Feature

It is import to identify the comment subject by extracting the product feature from reviews. And there are several of studies mentioned different approaches.

Khairullah et al. [17] use hybrid patterns to identifying product features from customer reviews. It minds the customer reviews on websites for product reputation and sales forecasting.

Su et al. [16] propose a method of extracting product features and opinion words using pattern knowledge in customer reviews.

Fang and Zhan [13] tackle the problem of sentiment polarity categorization, which is the one of the fundamental problems of sentiment analysis and give insight to future analysis.

Leonard, Rajesh, Jean-Guy and John [15] perform an statistic analysis on mobile application review and found that most reviews are short and the majority of apps receive under 50 reviews in their first year. And most of them are short.

Poescu and Ana et al. [23] introduce a system with unsupervised method to mind the review and finds the important product features. The product feature is automatically learn by mind the reviews.

Table 1 illustrates examples and keywords from WordNet [20] projects of emotions. WordNet is a lexical database for English which linked English words together by their semantic relationships.

Popescu and Ana-Maria et al. [24] identify the review mining task as four steps. Including identify product features, identify opinions regarding product features, determine the polarity of opinions and finally, rank opinions based on their strength.

Table 1: Example of Emotion classes and words

Emotion classes	Sentiment Polarity	Emotion word examples
Happiness	Positive	happy, joyful, glad
Pleasantness	Positive	pleasant, enjoy, nice
Relief	Positive	relief, comfort, solace
Fear	Negative	scare, fear, frightening
Sadness	Negative	sad, grief, mourn
Disappointment	Negative	bummer, regret, dejected
Unpleasantness	Negative	dislike, yuck, irksome
Loneliness	Negative	lonely, withdrawn
Anxiety	Negative	anxiety, worry, distress
Anger	Negative	anger, furious, antagonize

AppCAT applies Word2vec by Google, first presented in 2013, Word2Vec is a distributed, data-driven, vector-based representation of words. This technique represents each keyword as a high dimensional vector which it learns from input data.

There are two models behind Word2vec, continuous bag-of-word model(CBOW) and skip-gram(SG), describe below.

The input to the CBOW model could be $\omega_{i-2}, \omega_{i-1}, \omega_{i+1}, \omega_{i+2}$, the preceding and following words of the current word we are at. The output of the neural network will be ω_i . Hence we can think of the task as "predicting the word given its context".

The input to the model is ω_i , and the output could be $\omega_{i-2}, \omega_{i-1}, \omega_{i+1}, \omega_{i+2}$. So the task here is "predicting the context given a word". Also, the context is not limited to its immediate context, training instances can be created by skipping a constant number of words in its context.

2.3 Sentiment Analysis

Sentiment analysis, also known as opinion mining, refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. It is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service.

Bo and Lillian [12] make a summary and an introduction on current trends of opinion mining and sentiment analysis. Also, it covers techniques and approaches that promise to directly enable opinion-oriented information seeking systems. They introduce the new challenges raised by sentiment-aware applications, too. Liu and Bing [19] provides some important research and survey on Sentiment analysis. Also, it introduce some structured method the perform on unstructured data(opinions). Mukherjee and Bhattacharyya [21] advocates an approach to identify feature specific expressions of opinion in product reviews. The system learns the set of significant relations to be used by dependency parsing and a threshold parameter to merge the opinion expressions. Kouloumpis and Efthymios et al. [18] investigate the utility of linguistic features for detecting the sentiment of Twitter messages. the messages in the micro-blog and its tags are captured and analysed by supervised approach. Passonneau and Rebecca [22] use POS-specific prior polarity features and use a tree kernel to obviate the need for tedious feature engineering.

3 Methodology

The methodology of AppCAT described below, the architecture overview is shown in Figure 1 and source codes of AppCAT are available on author's Github [4].

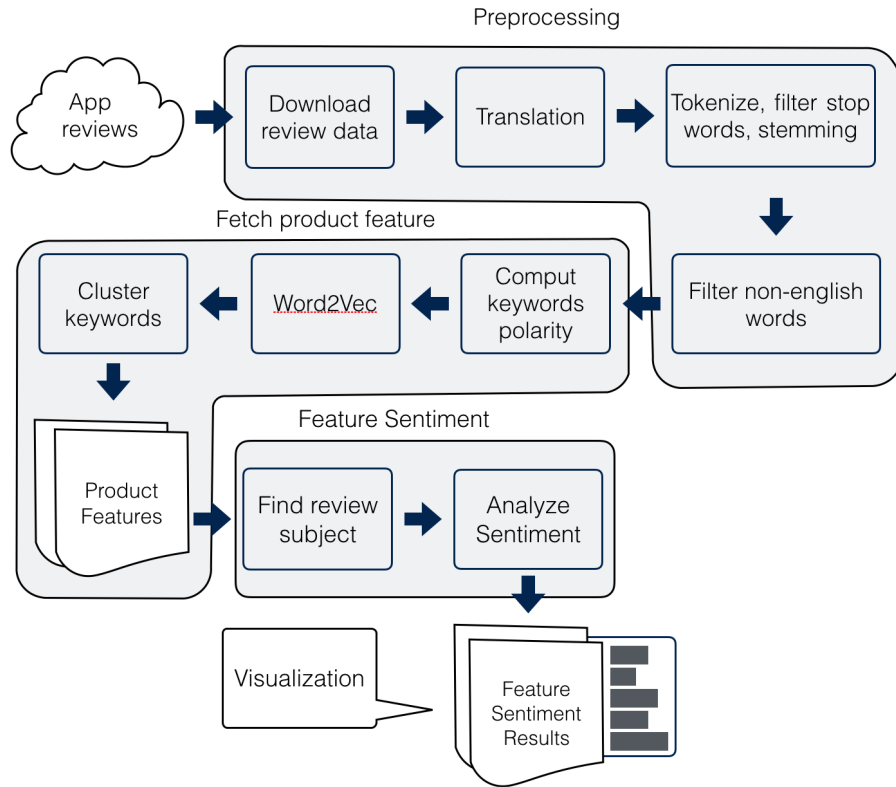


Figure 1: AppCAT architecture

The first part of AppCAT architecture is review preprocess including download, translation, tokenize, filter stop words, stemming and construct dictionary. Second part is fetch product feature including define initial product feature, and form product feature set. Third part is sentiment analysis on product feature including finding review subject, analyze sentiment. The final part is calculating the sentiment of each product feature to this app, provides a bar chart for visualization.

3.1 Review preprocess

Here we introduce how to download the reviews of apps and do preprocess including translation, filter stop words, tokenize, stemming and dictionary construction.

3.1.1 Download review data

App reviews data are collected from App Store api [6], the content of api includes app track id and list of reviews. In each review contains a user id and the user's rating to this app(from 1 star to 5 star), the review title and the content of the review. An example of review is below.

After download is complete, AppCAT then performs natural language preprocess. Described next section.

3.1.2 Natural Language Processing

Reviews in App Store often presented in various languages like English, Japanese, Spanish or French, we need to translate all the reviews to one single language to facilitate the word tokenizing, stemming and product feature extraction afterwards. Here we choose English as our translation target because most comments are originally from English. Google translate api were applied here to do the trick. After all the descriptions are translated to English, word tokenizing is applied to breaking sentences into words and elements. For each token in the list, we use NLTK in python to do word stemming, reducing inflected words to it word stem. From example, "argue", "argued", "arguing", and "argus" reduce to the stem "argu". After stemming process, for each token, we filter the stop words. Stop words are words commonly appear in documents like "the", "is", "at", "which", and need to be filtered out before we start processing. After process progress completed, reviews are stored into mongodb, a no-sql database, for further process and analysis.

3.1.3 Filter and Add Frequent Words

AppCAT maintains a dictionary to filter tokens. tokens that is not in dictionary would be filtered out. The default dictionary is English words. Some words appear frequently in reviews are not English, like facebook, UI, wma, etc. If the word is frequently appear in reviews but not in English, AppCAT mark it as frequent words and store it into dictionary. After that, tokens that are not in dictionary are filtered out.

Table 2: Examples of frequent words that is not in English dictionary

Frequent Words
app
iphone
ipad
ipod
wifi
ok
internet
google
online
facebook
pc
itunes
offline
ui

3.2 Fetch Product Feature

Like all products and services, applications are also products that have various product feature. For example, price, energy consumption, efficiency, user interface, etc. User may leave comments to those product features. Currently, AppCAT defines some product features that user often leave their comments and care about. And by using Word2vec that transform keywords to vector and find similar keywords to initial product features. Finally we can form product feature sets.

3.2.1 Define Initial Product Feature Keywords

According to a recent survey [11], what mobile application user really care about is the privacy and application performance. So currently APPCAT take these into product feature concerns. According to AppDynamic reports [1], 86% of users have uninstalled apps after only using them once due to the poor performance of those apps. A test conduct by JWPlayer [7] suggests that 70% of app users will leave within 11 seconds of waiting for content (like waiting internet connection within an app) to load. According to above survey and research, AppCAT define some initial keywords as those domain specific features, by use some keywords like "performance", "ui", "memory". As a result,

we define four types of product feature and its initial keyword, as shown in table 3.

Table 3: Product feature and keywords

Keyword name	Type	Keywords
Performance	Performance	execution, usability, stability, responsiveness, workmanship
Lag	Performance	laggy, glitchy, choppy, slower, freeze, jumpy
Crash	Performance	freeze, shuts, startup, crashed, crashing, quits
Memory	Performance	space, storage, hogging, ram, hog, battery
Network	Performance	wifi, router, server, lte, connected, wireless, connection
Advertisement	Advertisement	advertising, ad, barrage, popups, commercial, banner
UI	UI/UX	interface, layout, design, gui, sleek, elegant
Setting	UI/UX	default, enable, turned, set, privacy, override, preference
Price	Price	cost, priced, hefty, pricing, reasonable, bargain
Buy	Price	buying, pay, purchase, sell, splurge, purchasing, bought
Privacy	Privacy	administrator, setting, employer, authenticate, enabled, protect
Permission	Privacy	denied, consent, disallow, authenticate, administrator, unauthorized, insisted

First category of performance category, user cares about the speed of the app. They ask the question like "Would the app be crash when starting?", "Is the network connection stable in this app?". AppCAT take those into account and define the initial keyword as "performance", "lag", "crash", "memory", "network". The second category is advertisement, as we noticed, people usually complains pop-ups in apps, so we take this feature into account. Besides, user experience(UX) and user interface(UI) are also the main concern of users when downloading apps, so is the price, so these are third and fourth feature. At last, as mentioned before, privacy is the main concern of user. So we add privacy as the fifth feature. Afterwards, AppCAT use word2vec to form sets of feature topics. Then, by training a *word2vec* model, AppCAT can expand initial set of feature topic with only

some elements in it to a larger group where contains many similar keywords. To do this, first we define some initial keywords that user care about, and compare their similarities to other keywords. Grouping them base on the vector similarities. Detail describe below.

3.2.2 Transform tokens into vector

First presented in 2013, Word2Vec is a distributed, data-driven, vector-based representation of words. This technique represents each keyword as a high dimensional vector which it learns from input data. We use Word2vec here, using the comment corpus as the training data. Transform all words appeared in review dictionary into vector representation.

3.2.3 Form feature keyword sets

Based on above procedure, words in vector can be calculate the vector similarity. We use the initial feature name to find the similar keywords and form the product feature keyword sets, result are presented in table 3.

After this process, each cluster contains some related keywords imply one topic.

3.3 Sentiment Analysis on Product Feature

After we get the product feature of apps we can find the sentiment (positive or negative) of user to a specific feature and do visualization to display it to user. detail described below.

3.3.1 Find Review Subject

Finding the comment subject of a review is important, a user could comment on battery consumption or user interface. By the step above, AppCAT detects the keywords in review to find out which topic this review is about. For each app. we tear the comments down to sentences. And we assume each sentences describe no more than one subject(product feature). AppCAT will not perform analysis on an app if this app has less than 100 sentences. For each sentences, our system compare if the product feature keywords match

any words in the sentence. If so, the sentence is identified as comment toward that product feature. For example, consider the following review:

All the best in easy to access information packaged with great design!

According the keywords we found via Word2vec, word "design" is belong to "UI" product feature and appeared in the above comment. AppCAT will classify this comment as a comment to "UI" topic.

3.3.2 Analyze Sentiment

Currently, AppCAT applies WordNet [20] sentiment analysis tool to identify the sentiment of a sentence. In python, a library called Textblob use WordNet to do sentiment analysis.

All the best in easy to access information packaged with great design!

Textblob gives 0.81 score to this sentence, the higher the score, the better sentiment this review is.

Another example,

The negative: It is a battery vampire; the radar map and animation is not very detailed.

Textblob gives a negative score(-0.23) to this sentence. Because it has words like "negative", "not".

The Textblob sentiment tool can find the polarity and subjectivity of a sentence. Each word in the lexicon has scores for polarity(negative or positive), subjectivity(objective or subjective) and intensity(modifies next word or not).

If it is a modifier, the total score would be x0.5 else would be x2.0. For example, the word "great" has score of 0.8, "not great" uses "not" negative modifier on "good". So the final score of "not great" is -0.4. For each feature and sentences related, AppCAT average the score of those sentence as the feature score of this app. We can see more empirical cases in section 4.

To a specific app, we gather the reviews belong to a specific features. Count the sentiment for each review, and sum up to form the feature score. We sum the negative and positive score separately. As a result, each app has five feature score containing negative and positive scores.

3.3.3 Visualization

AppCAT use bar chat to represent all the product feature of a specific app. As figure 4 shows, the light green bar is positive feature score, while dark green bar is negative feature score.

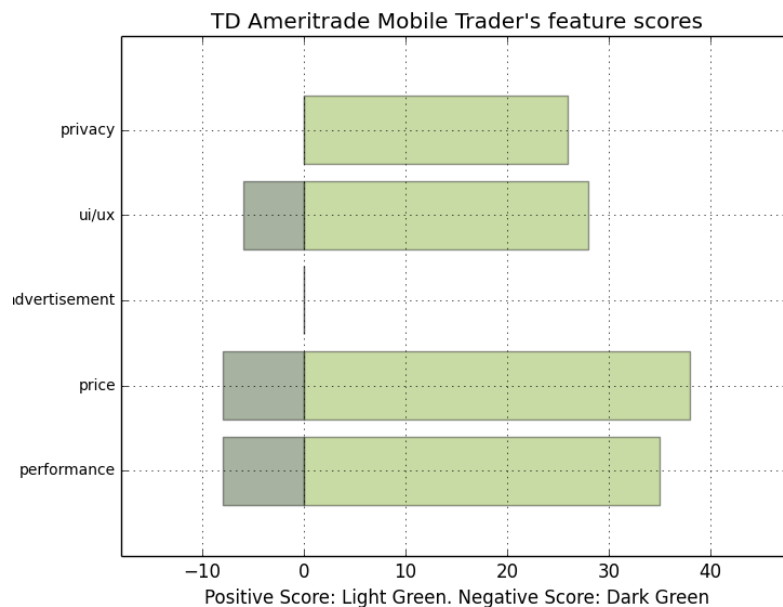


Figure 2: TD Ameritrade Mobile Trader comment bar chart

It more result and bar charts are displayed in section 4.

4 Experiment

In this section we do the AppCAT comment analysis on several apps and interpret the result. Totally AppCAT has analyze about 4384 apps. Among those 4384 apps. 1430 of them are analyzed because of sufficient reviews. In the following section We examine

four apps belong to two category of finance and weather and visualize the results. Detail described below.

4.1 Some Apps with privacy issue

Google Translate is an app developed by Google for user to translate sentences from one language to another. The result of Google translate is shown in table 4. Privacy issue are complained by some of the users. It is reported that there are some users feel unsatisfied with the privacy feature. Some users complain user interface while others like it.

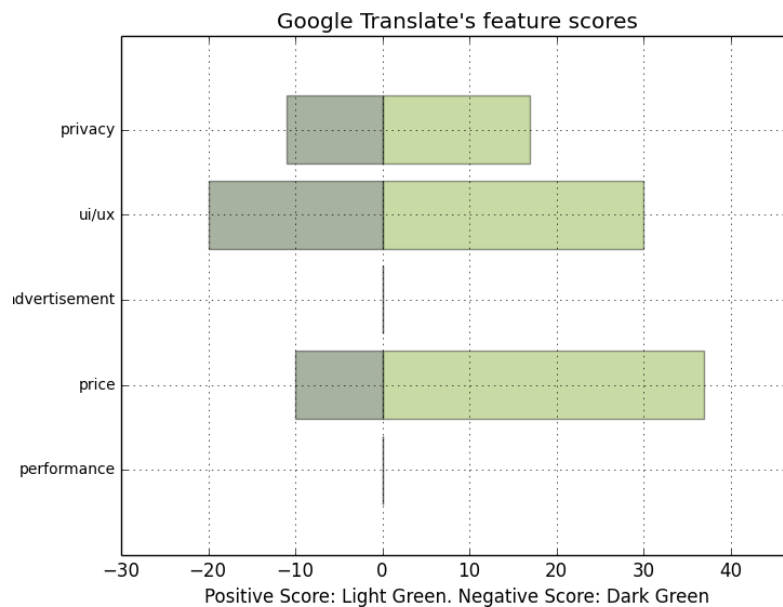


Figure 3: Google translate comment bar chart

4.2 Apps in Finance Category

One of the app analyzed by AppCAT is *TD Ameritrade Mobile Trader*. It's an app that users can trade their stock, options or futures on their mobile devices. The review analysis result is shown in table 5 AppCAT finds out that this app has some problem of crash and freeze, but not too severe. And this app has a high score on performance.

Another app in finance category is *PocketMoney LITE*. It is an app that enable user to track their finance status and keep accounting on the mobile device. The review

Table 4: AppCAT result of Google Translate(track id 414706506)

Feature name	Original rating	Sentiment	Sentence
PERFORMANCE	2	-0.31	Since the last update, it will not Allow you to enter a word or phrase to translate before getting a window fur "network fail"Please fix!
PERFORMANCE	1	0.00	When I type something in, it says network error.
PERFORMANCE	2	0.00	But now theres a bug that makes the app crash, and won't reopen unless I restart the phone, and even then it still crashes after one or two searches
PRICE	5	0.00	One day I'll probably discover they've been using this App to track me in some manner and I'll flip out, but until then, it works and the price works for me.
UI/UX	5	0.16	My suggestion is to have an option of setting a few permanent languages which are used the most, instead of the current setup of having "most recent languages."
UI/UX	5	0.38	I had a Spanish assignment and I could figure it out and the typing took forever, but the picture setting is amazing!
UI/UX	5	0.28	This is a great app, such a powerful engine and simple and functional design, that just works!
PRIVACY	1	0.17	Anyone can access the Internet through older versions of this app by going to the Google privacy policy which brings up a browser.
PRIVACY	3	0.00	The microphone button and the handwritten button will not appear when granted permission (version 4.0.0) iPhone 6 iOS 8.4.
PRIVACY	1	-0.22	Please respect privacy of end users as well without requiring unnecessary access to smartphones that is not required for basic functions that used to work well in the past.

Table 5: AppCAT result of TD Ameritrade Mobile Trader(track id 375660117)

Feature name	Original rating	Sentiment	Sentence
PERFORMANCE	5	0.50	This is a great platform and doesn't use that much memory
PERFORMANCE	5	0.50	Virtually no lag I am even able to scalp.
PERFORMANCE	2	0.40	This application was great but after a recent update the app now crashes every 5 seconds.
PRICE	5	0.50	I'd have to buy software that cost a couple of grand to do what this does.
PRICE	5	0.45	Not as good for research but orders of magnitude better for live time prices and charts.
PRICE	2	-0.80	The app has the annoying, and costly, habit of switching a trade from sell to buy when adjusting quantity.
UI/UX	5	0.25	I love it!On my wish list, I do hope the user study settings can be exported/imported or sync-able via iCloud Drive.A second wish is that users can take notes on stocks or options.
UI/UX	4	0.00	One feature I would like is ability to keep study settings across time frames.
UI/UX	2	-0.10	However, you need to be careful when setting up trades.
PRIVACY	5	0.50	Would like to see more emphasis placed on security.
PRIVACY	4	0.01	It annoying to read news only based on a security one is looking at, the general news tab will be great.
PRIVACY	5	0.43	It's also great for having quick access to news on the particular security you're watching.

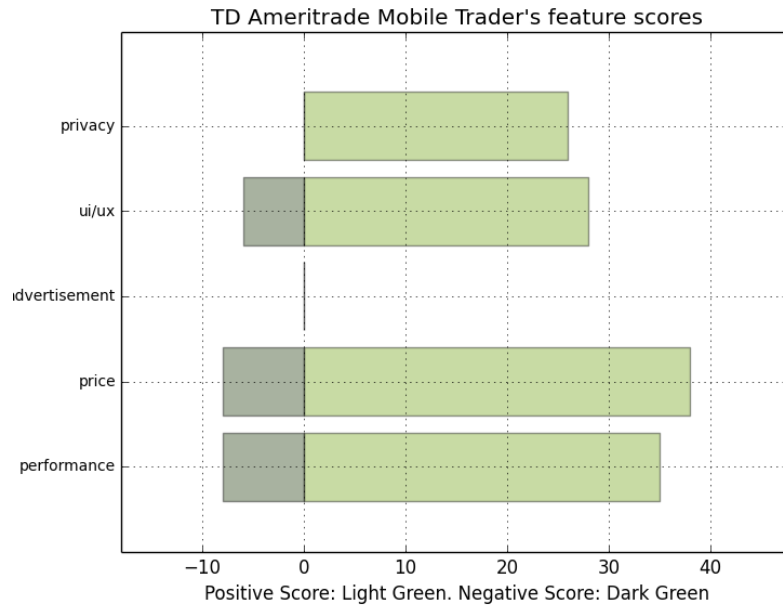


Figure 4: TD Ameritrade Mobile Trader comment bar chart

analysis result is shown in table 6 AppCAT finds out that this app has serious problem of advertisement which is not satisfying users. Also, this app may have crash and freeze issue and thus have a high score (about 3) in crash feature.

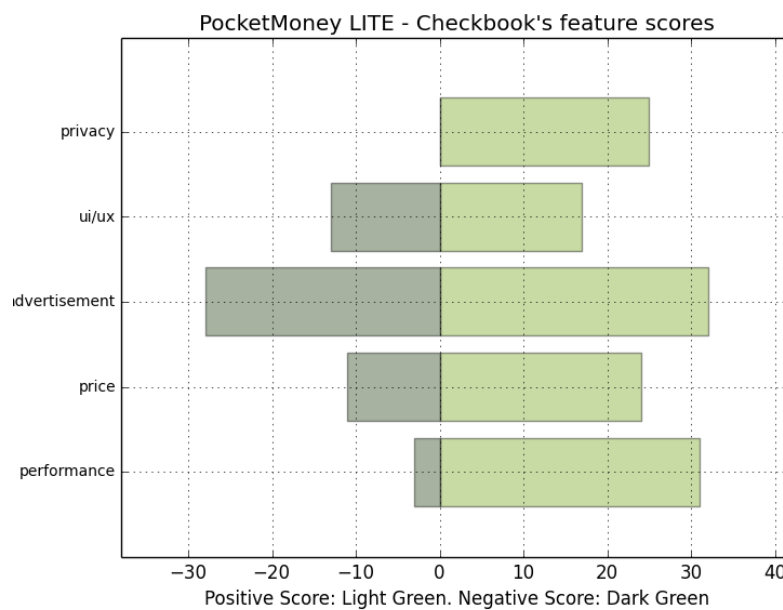


Figure 5: PocketMoney LITE comment analysis result

AppCAT visualize the score of those apps to a bar chart. For app user to choose better

Table 6: AppCAT result of PocketMoney LITE - Checkbook(track id 300715810)

Feature name	Original rating	Sentiment	Sentence
PERFORMANCE	1	0.00	Crashes constantly, adjust balance function won't allow data to be entered.
PERFORMANCE	5	0.19	Yes occasionally it might have a crash issue but its still worth it!
PERFORMANCE	1	0.21	This is by far my favorite app for tracking my bank accounts, but since the latest update it usually crashes before it even opens.
PRICE	1	0.20	Don't see a way to not have to buy the whole thing over again, since it crashed and had to be deleted it because Pocket Money screwed up the update.
PRICE	5	0.05	Will buy in the near future!
PRICE	5	0.00	I want to buy the upgrade but I'm worried it will erase my data!
UI/UX	1	-0.40	Poor design in iOS 7.
UI/UX	2	0.00	Its got all the settings that I need.
UI/UX	3	0.00	I just wish that there was a setting to allow rollover in your budgets from one period to another.
ADVERTISEMENT	1	0.00	The pop up ads have made this app a living nightmare.
ADVERTISEMENT	3	0.15	I almost never write reviews, but I must say the feature set of PocketMoney is impressive as I would expect of a mature software package, but its very difficult to evaluate when ads keep popping up over data entry fields.
ADVERTISEMENT	1	0.29	Thenn they added pop up ads which will come right when you're entering a transaction.
PRIVACY	1	0.25	No talk of encryption security; Beware what you enter; my whole iPhone was stalled out; needed 3-4 pushes to exit & enter anywhere !

app easily.

4.3 Apps in Weather Category

One of the app in "Weather" category analyzed by AppCAT is *Clear Day - Weather HD* which provides pretty videos depicting weather conditions, and a quick view for multiple cities weather. The review analysis result is shown in table 7 AppCAT finds out that this app has some issue of advertisement. And the user interface of this app, is good for users (about 0.44 points). but with some complains (-0.06 points). We can also figure out from the bar chart 6. The app have great design and good performance but also with advertisement issue. The advertisement issue is complained by users. It is still a highly-recommended app for user to download.

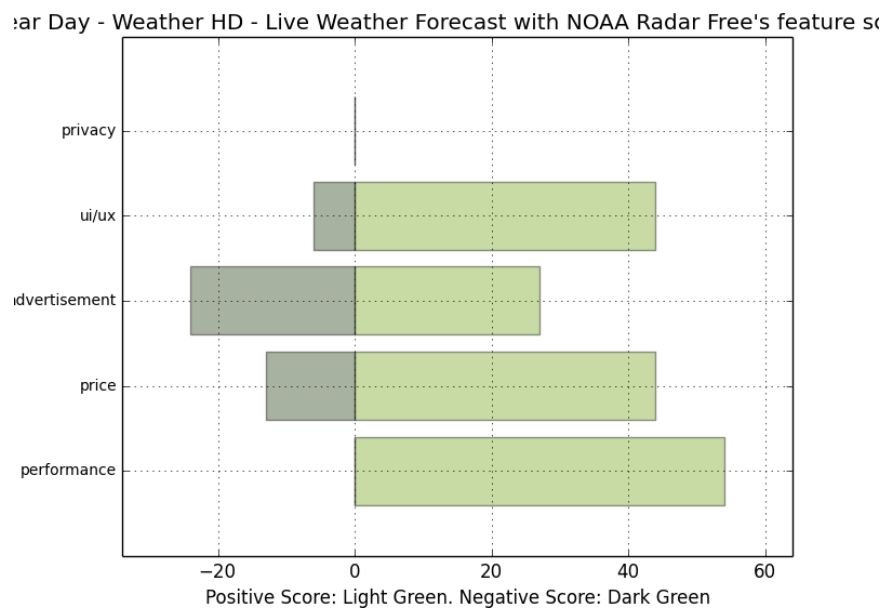


Figure 6: Clear Day - Weather HD comment analysis result

Another app in weather category is *Weatherwise*, it is an app with creative animated weather scenes, photographers. The following review:

Cute and simple and no ads

It implies that the app has no advertisements. AppCAT also identifies that it has little performance issue. So it is a good app to download.

Table 7: AppCAT result of Clear Day - Weather HD - Live Weather Forecast with NOAA Radar Free(track id 412489722)

Feature name	Original rating	Sentiment	Sentence
PERFORMANCE	3	0.54	Good performance, good data, and appealing presentation... at first.
PERFORMANCE	2	0.25	Also, ever since the update, the program crashes upon first opening.
PRICE	2	0.55	I'm happy to buy stuff but this doesn't seem to have functionality worth paying for.
PRICE	5	0.22	It is good because it tells you how the weather is going to be but bad because you haft to buy to get more weather news at the starting you only can have 1 news and that is new your and then you can look for 1 more than you have to buy the rest well it is good for..... NEW YOURK
PRICE	3	0.42	I love this app, and want to buy the full version, but it crashes on some of the maps - it always crashes when I try to play the animation on the precipitation map.
UI/UX	5	0.81	All the best in easy to access information packaged with great design!
UI/UX	5	-0.12	You can change your settings & add other locations to see what the temperature is.
UI/UX	5	0.18	Clean and simple design.
ADVERTISEMENT	1	0.00	I didn't like the ads either.
ADVERTISEMENT	4	0.12	I recommend it...wish there were less ads and pop-ups, but it's free.
ADVERTISEMENT	3	0.20	No radar, no multiple cities in free version,you do get two plus a lot of ads.

Table 8: AppCAT result of Weatherwise(track id 420954273)

Feature name	Original rating	Sentiment	Sentence
PERFORMANCE	4	0.40	Good detail and more accurate than local network forecasts.
PERFORMANCE	5	0.35	Continue the good work as always..Weatherwise crash when I open my iPhone please fix it ..
PERFORMANCE	2	0.20	glad that the icon has changed to the old one but the app crash when i open it !!
PRICE	5	0.30	I might buy a theme pack or something it's pretty cool
PRICE	2	-0.12	You will have to buy the themes again from the other device.
PRICE	2	-0.12	Not like the other apps where if you buy any additional stuff from the app you can apply everything to the other device.
UI/UX	5	0.80	Great design
UI/UX	5	0.33	Simply stated, this is a stunning app that combines thoughtful design with accurate weather and a variety of themes (the default is still my favourite) which reflect the changing weather conditions and time of day.
UI/UX	5	0.62	Love the artwork and interface design!
ADVERTISEMENT	5	0.25	Cute and simple and no ads!!
ADVERTISEMENT	5	0.00	There are no ads in the basic version.
ADVERTISEMENT	5	0.00	Nothing drives me more angry than ads when all I want is the temp today.

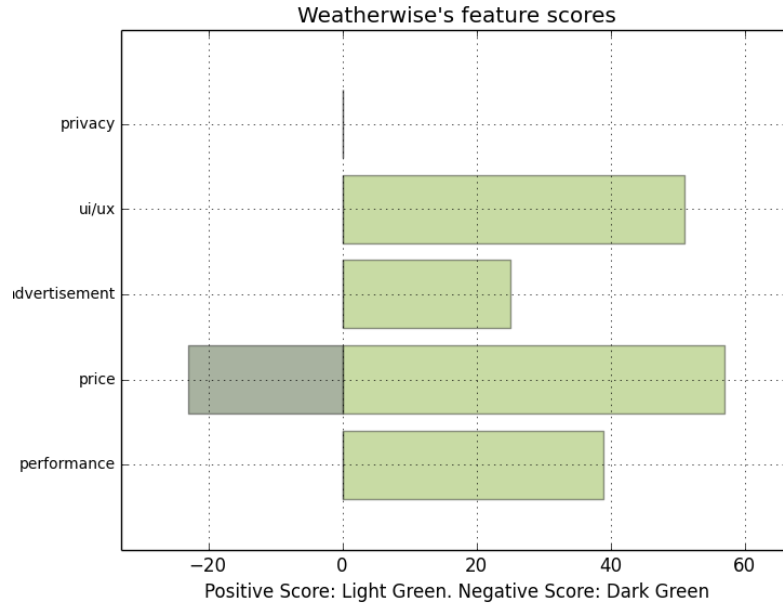


Figure 7: Weatherwise comment analysis result

5 Validation with AppReco

In this section we compare the result of AppCAT review analysis to AppReco to validate the experiment process. AppReco is a system that through static analysis to find out the hidden system call and analyze the behavior of mobile applications. We can compare the result of AppReco and AppCAT with respect to same app.

Take an analysis result in AppReco as an example: "Clear Day - Weather HD", it has 2.0 score in privacy. While AppCAT didn't report any privacy issue, it means that this app has some hidden privacy concerns that user do not notice. And the advertisement score output by AppReco is 0.67. It is reported by AppCAT that Advertisement score is also low due to the serious advertisement issue.

It is interesting that most app in AppCAT do not have a privacy score. AppReco can be a good approach to identify the hidden threat in apps while user doesn't know what api the app is calling.

The second case is comparison of two photo application, "Pic Stitch" and "Pic Collage". Result displayed in table 9 and table 10.

As table shown that the higher the privacy risk in AppReco, the more danger the

app is. In this aspect, "Pic Collage" has a 0.75 in privacy. While user give this app a lower rate on privacy in AppCAT, it is not surprising that two result points to the same conclusion.

Table 9: Comparison of "Pic Stitch" and "Pic Collage" in AppReco

App Name	Privacy	Third Party SDK	Advertisement	Total Risk
Pic Stitch	0.26	0.7	0	0.96
Pic Collage	0.75	0.7	0	1.45

Table 10: Comparison of "Pic Stitch" and "Pic Collage" in AppCAT

App Name	Performance	Price	UI/UX	Advertisement	Privacy
Pic Stitch	0.43	0.54	0	0.17	0.43
Pic Collage	0.42	0.44	0.55	-1.0	0

6 Conclusion

We present AppCAT, a systematic way of analyzing mobile application reviews. It can automatically analyze reviews in App Store and based on the feature defined and extended by word2vec then classify reviews and perform sentiment analysis on each comment sentences. As a result we can plot radar charts to each application for user reference when downloading.

There're some sentences that contains negation like "This app doesn't care about user experience." Such sentence sometimes can't be recognized by our methodology. Unlike blogs and movie review posts, reviews to mobile applications are usually brief in content. Thus, it contains less information. And sometimes the reviews don't have a comment subject. For future work, the sentiment analysis part can be tuned by other methods like deep learning to improve the accuracy of sentiment polarity.

References

- [1] app-attention-span-research-report-1.pdf. <https://www.appdynamics.com/media/uploaded-files/1425406960/app-attention-span-research-report-1.pdf>.

(Accessed on 03/03/2016).

- [2] AppBrainStat. <http://www.appbrain.com/stats>. Accessed: 2015-06-12.
- [3] Gartner Says Worldwide Sales of Cellular-Embedded Mobile PCs, Tablets and Mobile Hot Spot Devices Will Exceed 112 Million in 2015. <http://www.gartner.com/newsroom/id/3064718>. Accessed: 2015-06-12.
- [4] Github - vacuumv/appcat: Mobile application in ios review sentiment analysis. <https://github.com/vacuumv/AppCat>. (Visited on 01/27/2016).
- [5] Google play android. https://play.google.com/store/apps?utm_source=apac_med&utm_medium=hasem&utm_content=May2516&utm_campaign=Evergreen&pcampaignid=MKT-DR-apac-tw-all-med-hasem-py-Evergreen-May2516-1-BKWS%7c%2EHASEM_kwid_43700011378110954&gclid=Cj0KEQjwhZm7BRCUyfs6ho2VjOEBEiQAumpGMt367Hff9cpZOExgclsrc=aw.ds#/now. (Accessed on 06/20/2016).
- [6] <https://itunes.apple.com/rss/customerreviews/id=340233007/sortby=mostrecent/page=1/json>. <https://itunes.apple.com/rss/customerreviews/id=340233007/sortby=mostrecent/page=1/json>. (Accessed on 06/20/2016).
- [7] Improving hls on android - jw player sdk for android 1.2. <https://www.jwplayer.com/blog/improving-hls-android-sdk-1-2/>. (Accessed on 03/03/2016).
- [8] itunes apple. <http://www.apple.com/tw/itunes/charts/free-apps/>. (Accessed on 06/20/2016).
- [9] Product/service features and benefits - entrepreneurship.org. <http://www.entrepreneurship.org/resource-center/productservice-features-and-benefits.aspx>. (Accessed on 06/20/2016).
- [10] The Statistics Portal. <http://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores>. Accessed: 2015-06-12.

- [11] What app users care about when sharing personal data: Permissions – martindale.com. http://www.martindale.com/intellectual-property-law/article_Mintz-Levin-Cohn-Ferris-Glovsky-Popeo-PC_2220436.htm. (Accessed on 03/03/2016).
- [12] Lillian Lee Bo Pang. Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval*, 2(1):1–135, 2008.
- [13] Xing Fang and Justin Zhan. Sentiment analysis using product review data. *Journal of Big Data*, 2(1):5, 2015.
- [14] Leonard Hoon, Rajesh Vasa, Jean-guy Schneider, and John Grundy. An Analysis of the Mobile App Review Landscape : Trends and Implications. pages 1–23, 2013.
- [15] Leonard Hoon, Rajesh Vasa, Jean-Guy Schneider, and John Grundy. An Analysis of the Mobile App Review Landscape: Trends and Implications. Technical report, Department of Computer Science and Software Engineering, Swinburne University of Technology, July 2013.
- [16] Su Su Htay and Khin Thidar Lynn. Extracting product features and opinion words using pattern knowledge in customer reviews. *TheScientificWorldJournal*, 2013:394758, 2013.
- [17] Khairullah Khan, Baharum Baharudin, and Aurangzeb Khan. Identifying product features from customer reviews using hybrid patterns. *International Arab Journal of Information Technology*, 11(3):281–286, 2014.
- [18] Efthymios Kouloumpis, Theresa Wilson, and Johanna Moore. Twitter Sentiment Analysis : The Good the Bad and the OMG ! pages 538–541, 2011.
- [19] Bing Liu. Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167, 2012.

- [20] George A. Miller. Wordnet: A lexical database for english. *COMMUNICATIONS OF THE ACM*, 38:39–41, 1995.
- [21] Subhabrata Mukherjee and Pushpak Bhattacharyya. Feature specific sentiment analysis for product reviews. *Computational Linguistics and Intelligent Text Processing*, pages 1–12, 2012.
- [22] Rebecca Passonneau. Sentiment Analysis of Twitter Data.
- [23] Ana Maria Popescu and Orena Etzioni. Extracting product features and opinions from reviews. *Natural Language Processing and Text Mining*, (October):9–28, 2007.
- [24] Ana-Maria Popescu and Orena Etzioni. *Extracting Product Features and Opinions from Reviews*, pages 9–28. Springer London, London, 2007.
- [25] Phong Minh Vu, Tam The Nguyen, Hung Viet Pham, and Tung Thanh Nguyen. Mining User Opinions in Mobile App Reviews : A Keyword-based Approach. 2015.