IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 12, December 2015.
www.ijiset.com
ISSN 2348 – 7968

Heuristics Motorized Taxonomy of Mobile Apps

A.Murali Yadav[1] M.Sreenivasulu [2]

[1] Department of CSE, J.N.T.U Anantapur, Tirupathi, Andhra Pradesh, India, Email- yadav100.murali@gmail.com

[2]Departments of CSE, Sathyabama University, Chennai, Tamilnadu, India, Email- srm200546@gmail.com

Abstract

Mobile applications and their evolutions square measure quickly growing; during this association finding the users' expectations and their interests could be a difficult issue. One attainable resolution is live the usage of the applications through log messages and categories to predefined classes and their options. Victimization touch of environmental info regarding the Mobile applications it's tough to categories, as a result of such info largely doesn't carry the captions even someday conflicts. To handle this drawback we have a tendency to introduce a replacement prediction model that extracts the appropriate info from the computer program results and also the log messages that helps to predict the appliance class, and later users' expectations and their interests. We have a tendency to deploy the rear propagation algorithmic rule to boost potency.

Keywords: *Mobile Apps, App Taxonomy, Context Log, Dataset*

1. Introduction

With the wide unfold use of mobile devices in recent years, a large range of mobile Apps are developed for mobile users. For instance, as of the top of July 2013, there are a unit quite one.9 million Apps and one hundred billion additives download at Apple's App store and Google Play. Indeed, mobile Apps play a crucial role in the daily lives of mobile users. Intuitively, the study of the use of mobile Apps will facilitate to grasp the user preferences, and therefore motivates several intelligent personalized services, like App recommendation, user segmentation and target advertising. However, the knowledge directly from mobile Apps is usually terribly restricted and ambiguous. For instance, a user's preference model might not totally perceive the knowledge "the user sometimes plays Angry Birds" unless the mobile App "Angry Birds" is recognized as a predefined App class "Game/Strategy Game". Indeed, owing to the massive range and high increasing speed of mobile Apps, it's expected to have an efficient and automatic approach for mobile App classification. All the same, one might argue that some mobile Apps area unit related to predefined tags or descriptions as information within the App delivery platform (e.g., App Stores) and this knowledge are often directly used for recognizing the latent linguistics meanings. However, this knowledge might be tough to get by the third party services, especially in the case that there exist multiple App delivering channels and it's ineffective to trace the supply of a mobile App, such as the sensible state of affairs of the robot scheme.

1.1 Existing System

Many existing algorithms were introduces to extract the hidden info required for categorization ex: Phan et al. Sahami et al. Cao et al. projected ways, on the other hand think about solely dataset, however each application access history and classes don't seem to be concern. Indeed, mobile App classification isn't a trivial task which continues to be under-development. The most important challenge is that there aren't several effective and specific options on the market for classification models as a result of the restricted discourse information of Apps on the market for the analysis. Specifically, there is restricted discourse data regarding mobile Apps in their names, and also the solely on the market specific options of mobile Apps are the linguistics words contained in their names. However, these words are sometimes too short and sparse to mirror the connectedness between mobile Apps and particular classes. As an example, Fig. 1 shows the distribution of the amount of mobile Apps with relevance the name length in our real-world knowledge set. During this figure, we can observe that the distribution roughly follows the ability law, and most Apps solely contain but three words in their names.

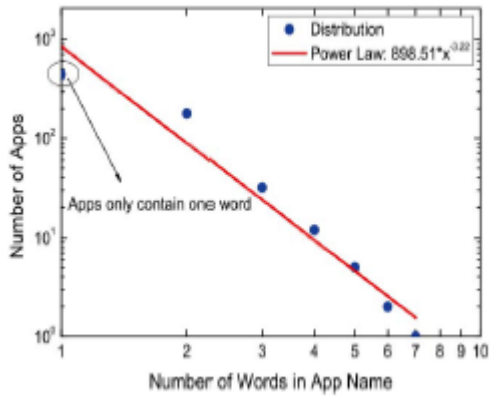


Fig. 1 Distribution of the amount of mobile Apps with regard to the name length in our real-world date set.

1.2 Disadvantages of Existing System

- Uses only pre-defined data sets.
- Results are not effective.
- Only Categorization but not access history.

2. Proposed System

The environmental info of the mobile applications is force out program results, conjointly force from access logs known as heuristics, the ultimate section combine the properties retrieved from each the environments, which is able to be helpful for the preparation the prediction model & categorizations. Coaching section of back propagation needs attribute alternatives to equip the categorization Engine to predict the classes of a mobile application, it doable that on application belongs to multiple categories. Is associate degree optimized methodology with sturdy and fruitful? In our experiments, Google program is employed for extracting to get the connected odds and ends of Applications. The rear propagation is heuristic primarily based prediction mechanism that improves the standard of prediction.

We propose to leverage each Web information and real-world contexts for enriching the contextual info of Apps, therefore will improve the performance of mobile App classification. in step with some state-of-the-art works on short text classification [9], [1], [3], [5], [7], a good approach for enriching the first few and thin matter options is investment net knowledge. Impressed with those works, we have a tendency to propose to require advantage of an online computer program to get some snippets to describe a given mobile App for enriching the matter options of the App. The leveraged net computer program will be a general computer program like Google or the vertical App search engine provided by associate degree App store. However, sometimes it may be troublesome to get comfortable net information for new or seldom used mobile Apps. During this case, the relevant real-world contexts of mobile Apps could also be helpful. Some observations from the recent studies [4], [7], [9], [8], [7], [1] indicate that the App usage of a mobile user is usually context-aware. as an example, business Apps square measure likely used beneath the context like "Location: Work Place", "Profile: Meeting", whereas games square measure typically contend beneath the context like "Location: Home", "Is a holiday?: Yes". Compared with net information, the relevant real-world contexts of new or seldom used mobile Apps could also be a lot of accessible since they will be obtained from the context-rich device logs of the users WHO used them in mobile devices. Therefore, we additionally propose to leverage the relevant real-world contexts of mobile Apps to enhance the performance of App classification.

2.1 Advantages of Proposed System

- ✓ Uses predefined data sets, search results and usage history.
- ✓ Efficient results obtained.
- ✓ Latest updates also considered in predictions.

2.2 Contributions

To be specific, the contributions of this paper are summarized as follows.

1. First, automatic mobile App classification could be a novel problem that remains under-development. To the most effective of our information, we tend to area unit one among the primary tries to review this problem. Moreover, we tend to area unit the primary to leverage both Web knowledge and relevant real-world contexts to counterpoint the limited discourse data of mobile Apps for finding this drawback.
2. Second, we tend to study and extract many effective options from each internet information and real-world contexts through the progressive data processing technologies. Then, we propose to exploit the utmost Entropy model (MaxEnt) [7], [2] for combining the effective options to coach a really effective and economical App classifier.
3. Finally, to judge the planned approach, we conduct extensive experiments on the context-rich mobile device logs collected from 443 mobile users, that contain 680 distinctive mobile Apps and over eight.8 million App usage records. The experimental results clearly show that our approach outperforms two progressive benchmark approaches with a big margin.

3. Related Work

Automatic mobile App classification could be a novel application problem; however, it can also be considered the matter of classifying short & thin texts. Short & thin texts square measure very common in real-world services, like question terms and SMS, which regularly contain restricted and thin matter information for utilizing. Compared with ancient text classification tasks, classifying short & thin text is extremely challenging and therefore attracts several researchers' attention. For

example, Phan et al. [3] planned to leverage hidden topics to boost the illustration of short & thin text for classification. The hidden topics square measure learnt from external data set with seeds choice to avoid noise, such as Web data. Sahami et al. [5] planned a completely unique similarity measuring approach for brief text snippets, which can also be established by a kernel, operate. Specifically, this approach utilizes a Web computer program to counterpoint original matter information, which may be leveraged for brief & thin text classification. Moreover, Yih et al. [3] improved the measuring approach by exploiting an extra learning process to form the activity additional economical. Broder et al. [9] planned to extract info from the highest connected search results of the question from an internet computer program, and Shen et al. [7] studied employing a internet directory to classify queries. Cao et al. [11] planned to use internet data for enriching each the discourse options and native options of Web queries for question classification.

Indeed, a number of on top of techniques will be leveraged for our App classification task. For instance, recently, according to Cao's work [11], Ma et al. [2] planned associate degree automatic approach for normalizing user App usage records, which can leverage search snippets to create vector house for each App usages and classes, and classify App usage records according to the circular function house distance. Compared with these works, the work reported during this paper doesn't solely comprehensively cash in of additional internet data based options however conjointly leverages the relevant contexts of mobile Apps that mirror their usage patterns from user perspective.

In recent years, with fast development of mobile devices, several researchers studied investment real-world contexts to boost ancient services, like customized context-aware recommendation [7], [9], [6], [4], context-aware user segmentation [2] and user context aware tour guide[10], [8]. As a result, researchers have found several user behaviors square measure typically context-aware, that is, some user behaviors square measure additional probably to look beneath a particular context. As an example, Cao et al. [10] projected an economical approach for mining associations between user App usage and real-world contexts. A unique metric to count support is developed for addressing the unbalanced occurrences of App usage records and context knowledge. Bao et al. [6] studied investment unattended approaches for modeling user context and App usage. During this work, the raw context knowledge and App usage records are initial divided and then sculptured by topic models. Ma et al. [2] studied the way to leverage the associations between contexts and user activities for discovering similar users with reference to habit by addressing the sparsity of such associations. Impressed by these works, we have a tendency to argue that the categories of mobile Apps that a user can use could also be relevant to his (or her) contexts. Thus, in this paper we have a tendency to propose to leverage the relevant discourse information of mobile Apps for raising the performance of App classification.

4. Overview

Here, we have a tendency to introduce many connected notions and provides an overview of our mobile App classification approach.

4.1 Preliminary

- App Taxonomy. To acknowledge the linguistics meanings of Apps, we will classify every App into one or additional classes according a predefined App taxonomy. Specifically, an App taxonomy ϒ may be a tree of classes wherever every node corresponds to a predefined App class. The linguistics meaning of every App is often outlined by the class labels along the trail from the foundation to the corresponding nodes. Fig.2 a pair of shows an area of the App taxonomy employed in our experiments.

Fig.2 Example of the mobile App taxonomy

- Search Snippets. In our approach, we tend to propose to leverage the Web information to counterpoint the matter info of Apps. To be specific, we tend to initial submit every App name to an internet program (e.g., Google or alternative App search engines), then acquire the search snippets because the extra textual info of the corresponding App. A search snipping is that the abstract of the net page that square measure returned as relevant to the submitted search question. The textual info in search snippets is transient however will effectively summarize the corresponding web content. Thus, they are wide used for enriching the first matter info in the short text classification downside. Fig. 3 shows some samples of search snippets for the App "Plant vs. Zombies" from Google.

Fig.3 Snippets in the result pages from Google.

- Context Log. Sensible mobile devices will capture the historical context information and therefore the corresponding App usage records of users through context-rich device logs, or context logs for short. for instance, Table one shows associate example of context log that contains many context records, and each context record consists of a timestamp, the foremost elaborate contextual info at that point, and therefore the corresponding App usage record captured by the mobile device. The contextual info at a time purpose is pictured by several discourse options (e.g., Day name, Time range, and Location) and their corresponding values (e.g., Saturday, AM8:00-9:00, and Home), which may be annotated as discourse feature-value pairs. Moreover, App usages records will be empty (denoted as "Null") as a result of users don't invariably use Apps. In Table 1, location connected data within the context logs, like GPS coordinates or cell IDs, have been transformed into linguistics locations like "Home" and "Work Place" by a location mining approach [20]. The basic idea of such approach is to search out the clusters of user positions and acknowledge their linguistics meanings through the time pattern analysis.

Timestamp	Context	App usage records
t_1	{(Day name: Monday),(Time range: AM8:00-9:00),(Profile: General),(Battery Level: High),(Location: Home)}	Angry Birds
t_2	{(Day name: Monday),(Time range: AM8:00-9:00),(Profile: General),(Battery Level: High),(Location: On the Way)}	Null
t_3	{(Day name: Monday),(Time range: AM8:00-9:00),(Profile: General),(Battery Level: High),(Location: On the Way)}	Twitter
.....		
t_{55}	{(Day name: Monday),(Time range: AM8:00-9:00),(Profile: General),(Battery Level: High),(Location: On the Way)}	UC Web
t_{56}	{(Day name: Monday),(Time range: AM8:00-9:00),(Profile: General),(Battery Level: High),(Location: On the Way)}	Null
t_{57}	{(Day name: Monday),(Time range: AM8:00-9:00),(Profile: General),(Battery Level: High),(Location: On the Way)}	Music Player
.....		
t_{359}	{(Day name: Monday),(Time range: AM10:00-11:00),(Profile: Meeting),(Battery Level: High),(Location: Work Place)}	Null
t_{360}	{(Day name: Monday),(Time range: AM10:00-11:00),(Profile: Meeting),(Battery Level: High),(Location: Work Place)}	SMS

TABLE 1 Example of Context Log from a Mobile User in Our Real-World Data Set

4.2 System Architecture

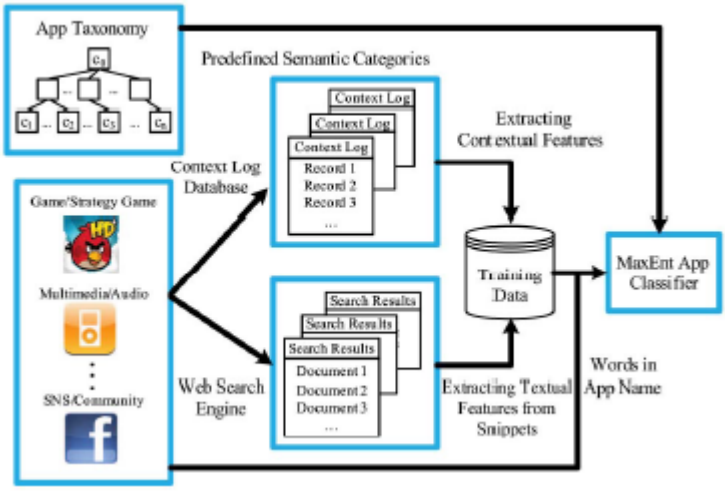


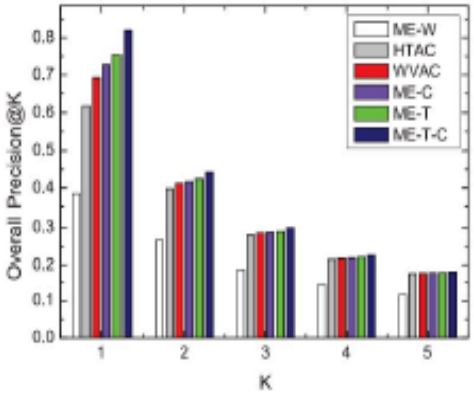
Fig.4 Framework of our App classification approach.

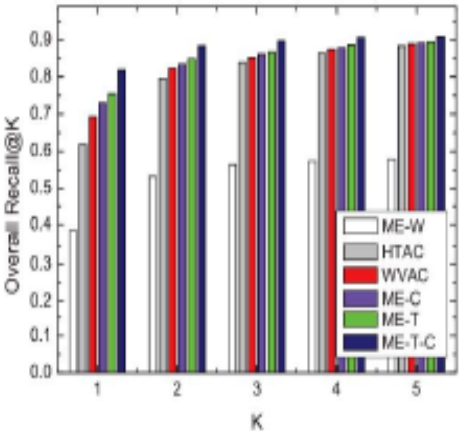
The projected approach for mobile App classification consists of two main stages. First, we tend to collect several context logs from mobile users, and extract each net information primarily based features and real-world discourse options for the Apps appearing in these logs. Second, we tend to cash in of the machine learning model for coaching an App classifier. Fig. 4 illustrates the most framework of the projected approach.

5. Simulated Result

Fig. 9(a) compares the typical overall Precision of 2 baseline strategies WVAC, HTAC and our approach with different options, namely, ME-W, ME-T, ME-C and ME-TC in the 10 rounds of tests. First, from the figure we are able to observe that the classification performance of solely leverage the short & thin texts in App names (i.e.,

ME-W) are very restricted. Second, compared with the 2 baselines WVAC and HTAC, the typical overall Precision of our approaches ME-T, ME-C and ME-T-C is improved systematically. To be specific, for the highest one results (i.e., given K = 1), the development is over Sept. 11 (ME-T), 6% (ME-C) and nineteen (ME-T-C) with reference to WVAC, and 22%, nineteen and thirty fourth with reference to HTAC. Third, comparing ME-T and ME-C, we are able to observe that the online information based matter options area unit slightly simpler than discourse features tho' each of them effectively improve the performance of App classification than ME-W, which only leverages the words in App names. Last, ME-T-C outperforms all different approaches in terms of average overall Precision. the typical improvement than ME-W across different K is over seventieth (the improvement exceeds 110% given K = 1), that clearly validates our motivation of leverage each net information primarily based matter features and real-world discourse options for rising the performance of App classification.





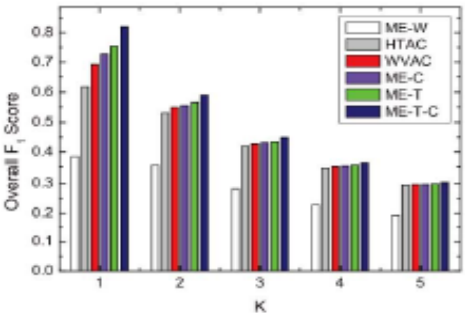


Fig. 9 Overall performance of each classification approach with different evaluation metrics in the cross validation. (a) Overall Precision@K. (b)Overall Recall@K. (c) Overall F1 Score.

Similarly, Fig. 9(b) compares the typical overall Recall@K of ME-W, ME-T, ME-C, ME-T-C and 2 baselines with reference to completely different K within the 10 rounds of tests. From this figure we are able to observe that our approaches outstrip the baselines and ME-T-C has the simplest performance. Another observation is that the typical overall Recall@K of each check approach will increase with the rise of K, which is affordable as a result of the likelihood that the bottom truth class label is roofed by the expected results can increase with a lot of expected class labels. Moreover, Fig. nine (c) compares the typical overall F1 score of all test approaches within the 10 rounds of take a look ats. From this figure we can observe that ME-T-C systematically outperforms other approaches and ME-W has the worst classification performance in terms of F1 score.

6. Conclusion

Although our current approach is each economical and effective for determination the matter of automatic App classification, it is still associate degree open downside regarding the way to plant this approach into mobile devices. Since mobile devices have terribly restricted computing resources, it's necessary to design a simpler service framework. Moreover, different users could have totally different App usage behaviors, thus how to integrate such personal preferences into discourse feature extraction are a motivating analysis direction. Finally, in our future analysis, we tend to conjointly attempt to mix our classification approach with alternative context-aware services, such as context-aware App recommender system, to enhance user experiences.

References

[1] G. Heinrich, "Parameter estimation for text analysis," University of Lipzig, Leipzig, Germany, Tech. Rep. 2008.
[2] M. Kahng, S. Lee, and S.-G. Lee, "Ranking in context-aware recommender systems," in *Proc. WWW*, Hyderabad, India, 2011, pp. 65–66.
[3] X. Li, H. Cao, H. Xiong, E. Chen, and J. Tian, "BP-growth: Searching strategies for efficient behavior pattern mining," in *Proc. MDM*, Bengaluru, India, 2012, pp. 238–247.
[4] Q. Liu, Y. Ge, Z. Li, E. Chen, and H. Xiong, "Personalized travel package recommendation," in *Proc. ICDM*, Vancouver, BC, Canada, 2011, pp. 407–416.
[5] H. Ma, H. Cao, Q. Yang, E. Chen, and J. Tian, "A habit mining approach for discovering similar mobile users," in *Proc. WWW*, Lyon, France, 2012, pp. 231–240.
[6] R. Malouf, "A comparison of algorithms for maximum entropy parameter estimation," in *Proc. COLING*, Stroudsburg, PA, USA, 2002, pp. 1–7.
[7] K. Nigam, "Using maximum entropy for text classification," in *Proc. IJCAI Workshop Machine Learning for Information Filtering*, 1999, pp. 61–67.
[23] X.-H. Phan *et al.*, "A hidden topic-based framework toward building applications with short web documents," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 7, pp. 961–976, Jul. 2010.
[8] X.-H. Phan, L.-M. Nguyen, and S. Horiguchi, "Learning to classify short and sparse text & web with hidden topics from large scale data collections," in *Proc. WWW*, Beijing, China, 2008, pp. 91–100.
[9] M. Sahami and T. D. Heilman, "A web-based kernel function for measuring the similarity of short text snippets," in *Proc. WWW*, Edinburgh, U.K., 2006, pp. 377–386.
[10G. Salton, A. Wong, and C. S. Yang, "A vector space model for Automatic indexing," *Common. ACM*, vol. 18, no. 11, pp. 613–620, Nov. 1975.

A.Murali Yadav the B.Tech Degree in Computer Science and Engineering from Shree Institute of Technology & Sciences college of Engineering, University of JNTUA in 2012.He is currently working towards the Master's Degree in Computer Science and Engineering, in Shree Institute of Technology & Sciences University of JNTUA. He interest lies in the areas of Web Development Platforms, SQL, and Cloud Computing Technology.

M.Sreenivasulu Received M.Tech in Computer Science and Engineering from Sathyabama University. Currently he is an Assistant Professor in the Department of Computer Science and Engineering at Shree Institute of Technology & Sciences-Tirupati.