Profile Based Mobile App Classification On Analyzing The App Reviews By Using Contextual Information Of App

Ashwini Sanap[1], *M.B.Vaidya.*[2]

[1] *Student, Computer Department, Amruthvahini College Of Engineering, Maharashtra, India.*
[2] *Assistance Professor, Computer Department, Amruthvahini College Of Engineering, Maharashtra, India.*

ABSTRACT

Today, use of mobile devices is increasing rapidly day by day; huge numbers of mobile apps are coming into the market. Mobile apps are coming from different vendors. These apps are useful and effective for various purposes like making it easy and time efficient for the user to select the required app, understanding the user preferences that motivates an intelligent personalized services, etc. Website and apps can be managed through the smart phones. Mobile website display similarly as other websites those are browser-based HTML pages are linked together and accessed via the Internet. Effective classification of mobile apps is required for which there is a need to have detailed information about the app. However this is non-incidental task as limited contextual information is available. As the information available about the apps is very less and unused, classification of the apps can also be considered as coming into the category of classification of less and unused text. Different methods are available to classify the less and unused text are present, which can be used for classification of the mobile apps that are not effectively worked to gain the proper contextual information of mobile apps.

Keywords: *Mobile app classification, Risk, Malware, Web knowledge, enriched contextual information, data mining, android app.*

1.Introduction

Recently, with the growing reputation of mobile devices large numbers of mobile apps are developed. Mobile apps plays important role for mobile users Mobile App classification is not a small task various issues and ways are revealed while working on it. Limited contextual information of Apps is the major challenge to classification models due for the analysis. Specifically, there is limited contextual information about mobile Apps in their names, and the only available certain features of mobile Apps are the semantic words contained in their names. However, these words are usually too short and sparse to reflect the relevance between mobile Apps and particular categories. However, it is difficult to obtain this information for the third party services. Usually, those tags are not very accurate [4]. Indeed, classification of apps is in trivial task and it is still in development process. Due to the limited contextual information of Apps explicit information is available for mobile apps are not effective for classification model. As the information available about the apps is very less and unused, classification of the apps can also be considered as coming into the category of classification of less and unused text. Different methods are available to classify the less and unused text are present, which can be used for classification of the mobile apps that are not effectively worked to gain the proper contextual information of mobile apps. Thus, to improve the performance of mobile App classification there must be researches that propose Web knowledge and *real*-world contexts for enriching the contextual information of mobile Apps. For averaging Web knowledge state-of-the-art works on short text classification [10], [12], is an effective approach which is useful to us for enriching mobile App. For better the performance of App classification this paper propose real-world contexts of mobile Apps.

2 Literature Survey:

Hengshu Zhu, Enhong Chen, Hui Xiong, Huanhuan Cao, and Jilei Tian[1] extract 443 mobile users' device logs to validate the proposed approach of enriching the contextual information of mobile Apps. For averaging Web knowledge state-of-the-art works on short text classification is an effective approach which is useful to us for enriching mobile App. For better the performance of App classification propose real-world contexts of mobile Apps.

M. Kahng, S. Lee, and S.-G. Lee [2] proposed a method for context-aware recommendation and ranking features for various contextual effects. Rank framework combines these features. This technique adopt ranking framework called as learning to rank. It is used to combine the features that reflect various contextual effects. This approach improves the quality of recommendation by simplifying the appropriate features into the ranking model. In this ranking model helps to produced optimal function for the application.

Q. Liu, Y. Ge, Z. Li, E. Chen, and H. Xiong[3] described BP-Growth algorithm. It used to combine behavior patter matching. To obtain the behavior pattern mining in a practical mobile computing environment this system systematic evaluate the several algorithms. This method introduces GCPM algorithm. This helps to reduced redundant data to minimize the running time for behavior pattern mining

Robert Malouf,[4] describe the comparisons of different algorithm by evaluating the parameters of Conditional maximum entropy (ME) model. ME model uses machine learning technique. It includes Generalized Iterative Scaling and Improved Iterative Scaling. In the implementation stage they used PETSc system library to solve the large system problems. This approach suggests the more comprehensive results as compare to the convergence rate as well as accuracy of various algorithms on a wider range of problems.

Kamal Nigam knigam, John Lafferty, Andrew McCallum [5] uses maximum entropy for text classification. It used to estimates the conditional distribution of the class variable given the document. Over fitting is reduced, and performance improves when maximum entropy is used with a Gaussian prior.

X.-H. Phan, L.-M. Nguyen, and S. Horiguchi [6] represents general framework to build classifiers those are required for short and sparse text. It discovered most hidden topics of Web data from huge text & Web collections. For processing short and sparse documents a hidden topic-based framework is introduced. This framework mainly solves data sparseness and synonyms/homonyms and also evaluates Web search result classification and matching/ranking for contextual advertising having large-scale data sets.

M. Sahami and T. D. Heilman [7] describes kernel function in a large-scale system. It is required to suggest related queries for search engine users. This function measures the semantic similarity between pairs of short text snippets. Kernel-based machine learning method determines its ability for improvement classification and clustering of text. Latent Semantic Indexing is used to retrieval of information in which kernel matrix is computed over text documents.

Dou Shen, Jian-Tao Sun, Qiang Yang, Zheng Chen [8] provides solution for generating a ranked list of target categories. Here an intermediate taxonomy is used. It helps to train classifiers bridging the queries and target categories. Maximum Entropy model combine all the enriched contextual information. This proposed approach outperforms into two state-of-the-art benchmark methods which having a significant margin.

K. Yu, B. Zhang, H. Zhu, H. Cao, and J. Tian [9] proposed a sequence-based semi-supervised approach.This approach is used to model the personalized context for mobile users. In this approach, to model the context in the form of probabilistic distributions and transitions of raw context data (BHMM) modeling is used. To measure the relevancies between categories this system firstly represents categories as topic distributions. Authority ranking, to collectively exploit the information in both target and relevant categories Topical Random Surfer model is used. Gibbs sampling method is used to estimate the conditional probabilistic distribution. Kullback Leibler (KL) divergence measures asymmetric, category relevancies.

H. Zhu, H. Cao, E. Chen, H. Xiong, and J. Tian [10] exploits the additional Web knowledge of mobile apps contextual information from internet. In this MaxEnt model outperforms the two baselines WVAC and HTAC, to indicate the combination of multiple Web knowledge based textual features and basic app name. This method

primarily learns the common context-aware preferences from the logs of many users. Also individual user can be represented as a distribution of common context-aware preferences.

H. Zhu, H. Cao, H. Xiong, E. Chen, and J. Tian [11] represents crowd wisdom based approach is proposed to mine personal context-aware preferences from context logs of mobile users. In this individual user can be represented as a distribution of common context-aware preferences. Two different assumptions are used for mining common context-aware preferences. This system has relatively higher computational cost.

H. Zhu et al [12] incorporate context information into the problem of query classification. For this classification CRF model is used. The CRF model does not consider different languages. To model the query context basic Linear-chain CRF is used. CRF model is applied on a commercial search engine log to extract the real data set.

A. Z. Broder et al. [13] proposed Query classification for information retrieval. This methodology uses search results as a source of external knowledge to address the problem of short queries which contains the insufficient information with them. This system produces a query classifier directly for the target taxonomy without using auxiliary taxonomy.

T. Bao, H. Cao, E. Chen, J. Tian, and H. Xiong [14 proposed an unsupervised approach. This approach is used for modeling of personalized contexts of mobile users. Two mobile contexts modeling are used 1) MU and 2) LDA. This model assumes every document is generated by a prior distribution of topics instead of a single topic. Context model based on LDA outperforms the MU based context model and hence it improves the effectiveness of mobile context modeling.

A. L. Berger, V. J. D. Pietra, and S. A. D. Pietra [15] maximum entropy model. It is used to build real value feature and constraint from the set of features. To built real-valued features and constraints maximum entropy model building blocks are introduces in this paper. Further, this paper describes the various applications such as, automatic machine translation system which demonstrates the efficiency of the maximum entropy.

3. CONCLUSIONS

In this review paper, study some previous Mobile App Classification systems. Effective classification of the mobile apps is important as everyday there are number of similar kind of apps coming in the market. In the study of literature survey observed that various classification techniques are presented for classifying the short and sparse data, which can be adapted for classifying the mobile apps. But the result obtained from these techniques does not give us the effective classification of the apps, as they take into consideration only single factor for classification i.e. web knowledge or contextual information. In base paper, Maximum Entropy model (MaxEnt), is used to combine effective information which is extracted from web knowledge and relevant real-world contexts. According to study of literature survey as well base paper we conclude that, combining personal preferences into contextual feature extraction is an interesting investigation area in classification of mobile apps.

4. REFERENCES

[1] Hengshu Zhu, Enhong Chen, Senior Member, IEEE, Hui Xiong, Senior Member, IEEE,Huanhuan Cao, and Jilei Tian," Mobile App Classification with Enriched Contextual Information,".

[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] R. Malouf, "A comparison of algorithms for maximum entropy parameter estimation," in Proc. COLING, Stroudsburg, PA, USA, 2002, pp. 1–7.

[5] K. Nigam, "Using maximum entropy for text classification," in Proc. IJCAI Workshop Machine Learning for Information Filtering, 1999, pp. 61–67.

[6] 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.

[7] 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.

[8] D. Shen, J.-T. Sun, Q. Yang, and Z. Chen, "Building bridges for web query classification," in Proc. SIGIR, Seattle, WA, USA, 2006, pp. 131–138.

[9] K. Yu, B. Zhang, H. Zhu, H. Cao, and J. Tian, "Exploiting enriched contextual information for mobile app classification," in Proc. CIKM, Maui, HI, USA, 2012, pp. 1617–1621.

[10] H. Zhu, H. Cao, H. Xiong, E. Chen, and J. Tian, "Towards expert finding by leveraging relevant categories in authority ranking," in Proc. CIKM, Glasgow, U.K., 2011, pp. 2221–2224.

[11] H. Zhu et al., "Mining personal context-aware preferences for mobile users," in Proc. ICDM, Brussels, Belgium, 2012, pp. 1212–1217.

[12] H. Cao et al., "Context-aware query classification," in Proc. SIGIR, Boston, MA, USA, 2009, pp. 3–10.

[13] aZ. Broder et al., "Robust classification of rare queries using web knowledge," in Proc. SIGIR, Amsterdam, Netherlands, 2007, pp. 231–238.

[14] T. Bao, H. Cao, E. Chen, J. Tian, and H. Xiong, "An unsupervised approach to modeling personalized contexts of mobile users," in Proc. ICDM, Sydney, NSW, Australia, 2010, pp. 38–47.

[15] A.L. Berger, V. J. D. Pietra, and S. A. D. Pietra, "A maximum entropy approach to natural language processing," Comput.Linguist., vol. 22, no. 1, pp. 39–71, Mar. 1996.