Privacy Friendly Mobile Intelligent Advertising Framework

Preethi Satishchandra, Sanjay Addicam, Kalpana Algotar
Retail Services Division
Intel Corporation
Chandler, USA
emails: {preethi.satishchandra, addicam.v.sanjay, kalpana.a.algotar}@intel.com

***Abstract*— This paper explains a framework known as "Mobile Intelligent Advertising Framework" which provides personalized recommendations to the user in a privacy friendly manner. Mobile devices, such as smartphones, and tablets are everywhere now. So, retailers and advertisers want to rely more on mobile data to recommend products to their consumers and most importantly they want to understand consumer interests to recommend meaningful ones. Our framework includes an Android app, which tries to understand the user passively using the available mobile data that the user gives access to, such as browsing history, accelerometer data, call log history, etc., and recommends a product after data analysis. The app can also communicate with any digital sign nearby which triggers the sign to play targeted ads to the user viewing the sign instead of random ones. This paper explains the important components of the framework briefly and presents an overview of current state of the art in capturing user's interests through mobile data for providing relevant recommendations.**

Keywords-Mobile; Intelligent Advertising Framework; Digital Signs; Data Mining; Lifestyle Analysis; Privacy.

I. INTRODUCTION

In recent years, mobile devices such as smartphones and tablets have grown significantly and they present new opportunities and challenges. This applies to the field of retail too where retailers and advertisers now not only have a new medium to showcase their products and ads to their consumers but also should give importance to new and important fields such as mobile data to show relevant stuffs to the consumers. Mobile data is a good source of data for providing personalized recommendations. Unlike random ones, these personalized targeted recommendations such as ads can provide more value to the consumers thereby benefitting both retailers and advertisers too. There are few applications out there with context-aware platforms, such as Qualcomm's Gimbal [7] and Google Now [6]. But, these do not address privacy as our framework attempts to do by handling all the analysis locally within the device.

Our paper is organized in 3 sections. The first section is Introduction; the second section explains the entire framework, which is further divided into many subsections. The last section is the conclusion and future work.

Section 2 begins by explaining the entire framework briefly. This is followed by describing the data needed for analysis, various data sources and its collection. This is followed by the description of the lifestyle analysis and the lifestyle model. Using this model, lifestyle analysis is done based upon which relevant recommendations are provided to the user. This is followed by the subsection, describing the various tools that were used to build the framework - RapidMiner, Lingpipe. The following sub section explains the communication between the digital sign and the mobile phone, to play targeted ads on the sign instead of random ones which is followed by the explanation of the privacy importance emphasized by our framework which distinguishes it from many other existing context-aware platforms and finally the subsection presenting experiments and results.

II. MOBILE INTELLIGENT ADVERTISING FRAMEWORK (MIAF)

Big data in terabytes and petabytes is floating around from various sources, which can provide valuable insights on current and future trends of an industry. This applies to many industries, such as healthcare, transportation, retail, finance, and many more. Data have proved to be important in mobile too, which can be utilized to provide many recommendations, such as ads, fitness updates, etc. These mobile data can come from different sources and sensors. Some of them are call log data, browsing history, accelerometer data, battery usage, ambient light, app usage, location data and much more. Developing an intelligent framework which fetches the right data, integrates data from various sources, does analysis and provides relevant recommendations in a privacy friendly manner is not trivial. Our framework approaches to solve this problem.

As essential part of MIAF, an Android app was created to facilitate data analysis within the mobile. When a user installs this app, he or she has to provide some minimum information to get recommendations which is age and gender bracket, e.g., Adult Male. The app can access only the part of mobile data to which the user gives permission to.

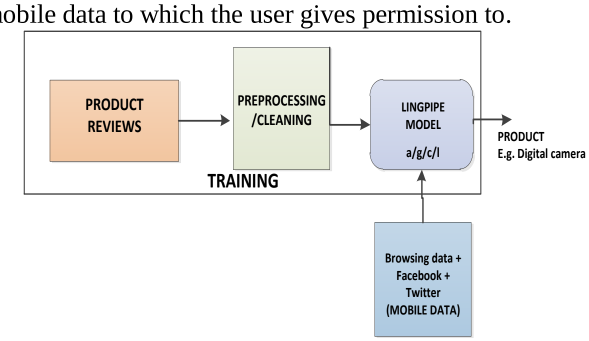

Figure 1. MIAF Architecture

 ISBN: 978-1-61208-364-3

Fig. 1 explains the overall MIAF architecture briefly. We are collecting data, such as product reviews, cleaning the data and training a classification engine based on Lingpipe [3], a text processing tool kit suitable for mobiles. The products selected are grouped into different age, gender, city, lifestyle groups (a/g/c/l, respectively). Lingpipe classification models for different a/g/c/l groups are built by training the model with respective products. Currently, we have around 80 a/g/c/l profiles. So, there are 4 age groups (Child, Young Adult, Adult, and Senior), 2 genders, 2 cities and 5 lifestyles. For example, adult male, from New York with social lifestyle. So, when a user installs the app, information such as age, gender, location is easily captured and further lifestyle analysis of the user is done. Based on the user's profile, the respective trained Lingpipe model is imported to the mobile. The new real-time mobile data, such as browsing history, Facebook and Twitter data are processed and passed into this model to get relevant product recommendations. So, the entire mining and analysis happens locally without the data leaving the device. Besides providing product recommendations, the app can also communicate to any digital sign nearby to play relevant ads and also this app can be integrated behind any retailer's app who is interested in providing personalized recommendations to their loyal consumers. This framework also includes many other features, such as targeted advertising based on facial detection [1][9]; this paper mainly concentrates on the mobile part of the framework.

A. Data

The Lingpipe classification engine requires training data to provide product and lifestyle recommendations. Product reviews fetched from various websites served as a good source of training data using which valuable product insights were extracted for each of the product category. Reviews were preferred instead of product description as they can be correlated with the user's data more and also reviews have become more important as consumers now trust reviews more than description before purchasing a product. A web crawler developed using PHP [14] was used to fetch the reviews for a set of preselected products. Further data based on different demographics (a/g/c) were obtained from Neilsen Company [5]. These data comprised of different kinds of mobile data, such as app usage data, browsing history, battery usage, call log data, etc.

B. Lifestyle Analysis

Knowing the lifestyle of a user helps to recommend better products as consumers usually buy products based on their lifestyles, e.g., a fitness cautious person may prefer product, such as running shoes, sports good, etc. Determining the lifestyle of a user after doing analysis of mobile data is not a trivial problem. MIAF architecture can recommend different lifestyles to a user which are social, talks a lot, travels a lot, fitness cautious and busy. The varied data sources used to determine lifestyle are call log (duration over some period), battery usage (percentage over some period), accelerometer (jogging, running or walking) and number of locations (e.g., number of distinct locations visited in a day) data. Table 1 depicts how data for recommending lifestyle looks for particular demographics after collecting and doing some analysis in a day.

TABLE I. TABLE FOR LIFESTYLE ANALYSIS

Data sources	**Young Adult/Male/New York**		
Call log (min)	23	100	130
Battery usage (percentage)	50	40	80
Accelerometer	Jogging	Walking	Walking
App Usage (Social)(percent)	3	10	4
App Usage (Talks a lot)	4	10	20
App Usage (Productivity)	3	10	2
App Usage (Fitness)	5	12	4
App Usage (Travels)	4	15	10

C. Lifestyle Model

A lifestyle model was built using the historical data obtained from Neilsen [5] company. The varied data sources used to compute lifestyle were obtained based on different demographics, as shown in Table 1. To build this model, a weighted table was constructed by clustering each of the data sources into 5 groups (since 5 lifestyles) for every a/g/c group. K means clustering algorithm [10] was used and the 5 centroids of the 5 clusters were assigned as weights for each of the data sources. Table 2 shows the weights computed for some of the data sources after clustering the historical data for a particular a/g/c group. Such weighted tables are created for all the data sources for every a/g/c profile.

TABLE II. SAMPLE TABLE FOR LIFESTYLE MODEL

Data sources	**Social**	**Talks a Lot**	**Travels**	**Fitness**	**Busy**
Call log (min)	0.3	0.2	0.4	0.1	0.5
Battery usage (percentage)	0.4	0.8	0.3	0.5	0.5

These weights are used as multipliers; so, when the new user mobile data comes (all in numerical values), they are multiplied with these multipliers and finally summed up to determine one lifestyle. The lifestyle with the greatest weight will be selected. This is explained in Table 3; if Sum1 is the greatest, that particular user is tagged with social lifestyle. The weights are updated as and when, new data is collected and added to the historical data collection.

TABLE III. SAMPLE LIFESTYLE TABLE WITH WEIGHTS

Data sources	**Social**	**Talks a Lot**	**Travels**	**Fitness**	**Busy**
Call log – **50min**	50*w1	50*w2	50*w3	50*w4	50*w5
Battery usage**40%**	40*w1	40*w2	40*w3	40*w4	40*w5
App Usage (Social) **30%**	30*w1	30*w2	30*w3	30*w4	30*w5

 ISBN: 978-1-61208-364-3

App Usage (Talks a lot) **20%**	20*w1	20*w2	20*w3	20*w4	20*w5
App Usage (Productivity)**10%**	10*w1	10*w2	10* w3	10*w4	10*w5
App Usage (Fitness) **20%**	20*w1	20*w2	20*w3	20*w4	20*w5
App Usage (Travels) **20%**	20*w1	20*w2	20*w3	20*w4	20*w5
Sum	Sum1	Sum2	Sum3	Sum4	Sum5

D. Rapidminer

Rapidminer [2], an open source data mining tool, served as a scratchpad to understand and build some data mining models such as classification and clustering. This understanding eased the development process of MIAF app. Product reviews fetched from various websites were processed and a classification model based on Support Vector Machine (SVM) [11] was developed in Rapidminer. The measure of accuracy of the model gave some confidence to build the same on the Android Mobile platform. SVM outperformed other classification engines, such as K- Nearest Neighbor (KNN) [13] and Naïve Bayes [12].

Rapidminer was also used for implementing clustering algorithm for detection of lifestyle as discussed above. K – Means clustering algorithm served well as we wanted to divide the data into 5 (k = 5) lifestyles to find 5 centroids.

E. Lingpipe

Lingpipe [3] is a Java-based framework text processing tool kit which served well for Android data mining. Its Software Development Kit (SDK) is easy to integrate with the mobile platform. Lifestyle models were built based on different demographics as discussed above. Further, for various a/g/c/l profiles, Lingpipe product classification models are built for product recommendation and these are built offline in the cloud. Depending upon the user profile respective trained model is imported to the mobile through MIAF. Cloud is used to build all the pre-trained data models and storage of the same. None of the user information is pushed into the cloud as it is handled within the device itself. By having all the training handled in the cloud there is no overload on the mobile Central Processing Unit (CPU) and the battery, as training of the data model requires more CPU overhead. User privacy is also preserved by utilizing the cloud only for the required and by not pushing everything to it.

TF-IDF (Term Frequency – Inverse Document Frequency) classifier [15] was used as it had better classification accuracy then other classifiers, such as Naïve Bayes [12]. SVM was not supported by Lingpipe. Tokenization, transformation of tokens, stop word removal, stemming and filtering of tokens was used for cleaning of text. Depending on the user profile (a/g/c/l), the respective model is imported from the cloud to the mobile to provide product recommendations. This model keeps updating, too, periodically, if needed. For example, if a user previously was in New York and recently moved to San Francisco, a new product model will be imported to the user mobile to provide relevant product recommendations.

MIAF app receives mobile data which are made up of browsing history keywords, twitter keywords and Facebook keywords. Browsing pages are parsed and top n words based on TF-IDF score are extracted, same with Facebook page and Twitter page. All these top TF-IDF words extracted are concatenated and passed into the model imported into the mobile to recommend products. All the word vectors are stored in SQLite database [16] within the device. So, no user mobile data leaves the mobile. The data are cleaned from SQLite every week, so that there is no overload on the database and not cleaned very frequently too for the creation of meaningful user profiles. Currently, in a device, products can be recommended from 100 product lists.

F. Talking of Mobile and Digital Sign

Digital signs [4] are part of every industry where they are present in airports, shopping malls and retailer shops. MIAF app can make the ads played on the digital sign more valuable. To bring more value for ads, we wanted our app to talk with the digital sign. Extensible Messaging and Presence Protocol (XMPP) [17], a messaging protocol was used to establish this communication channel. So, when a user with MIAF app walks near any digital sign communication channel between them is built and the ad played on the digital sign will be related to the ad recommended on the user's mobile. This is explained in Fig. 2, where the communication between the mobile and the digital sign is established using XMPP, iPhone ad is played on the sign after the word "iPhone" is transferred from mobile to the sign.

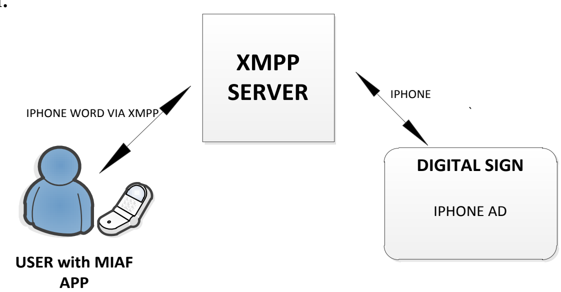

Figure 2. Communication between Phone and Digital Sign

G. Privacy

Most of the applications that exist now do not handle user's data securely, as most of the mobile data analysis are pushed into the cloud due to limited mobile computing resources. Our app handles all the required analysis locally in the mobile with minimal effect on the battery usage. Only a single word vector such as iPad, sun glass, etc., goes out of the device. No user data are withheld. The user can opt in and opt out of the service anytime. The products or ads are targeted in a more generalized way, where the targeting is done based on demographics, but not for any particular individual. We are looking into adding more levels of security and privacy by encrypting the data and also investigating on integrating with McAfee [8] software on Android.

 ISBN: 978-1-61208-364-3

H. Experiments and Results

MIAF app was shared among a few employees within the company for feedback and also among few customers. The results have been positive. Due to time constraints and other issues we have not documented the results for publishing purposes. But, we have conducted an experiment to prove targeted advertising is more relevant then non – targeted advertising based on AVA [1] technology, which is an Intel based technology used to target ads, based on demographics. The experiment was conducted in a supermarket over a period of 9 months. There was monthly sale increase of 16% and 3% monthly viewership time increase with targeted ads compared to non- targeted ones and the results are published [9]. By adding the mobile part, we believe sales and viewership can increase more bringing in more value to the consumers and the retailers too. Further, we compared MIAF to other similar applications out there. Google Now [6] has the proximity beacon missing and with Gimbal [7], the analytics apps need to be created by the customer and most importantly both the applications do not handle the analysis locally within the device like MIAF.

III. CONCLUSION AND FUTURE WORK

In this paper, we have attempted to explain a software framework for mobiles and digital signage using which relevant and personalized recommendations can be reached to the right consumer at the right time in a privacy friendly manner. With the limited data available, user's interest is captured passively by analyzing the lifestyle of the user and all this analysis happens locally within the device to protect the privacy. Further, based on lifestyle relevant products are recommended. In the future, we want to make this application more efficient and scalable by adding more features such as Graph databases and also by expanding the product lists from 100's to 1000's. We are also extending the framework to not only recommend ads or products, but also predict next user's activity, e.g., a user usually calls at noon every day; so, it is noon now and he/she may want to call. In the future, we want to extend MIAF to other platforms than Android.

REFERENCES

[1] P. Tian, S.V. Addicam, K. Chiranjeevi and S. Malik, "Intelligent Advertising Framework for Digital signage," in Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 2012, pp. 1532–1535.

[2] I. Mierswa, M. Wurst, R. Klinkenberg, M. Scholz, and T. Euler, "YALE: Rapid Prototyping for Complex Data Mining Tasks," in Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-06), August 2006, pp. 935–940.

[3] Alias-i. 2008, LingPipe 4.1.0 - http://alias-i.com/lingpipe [retrieved: May, 2014]

[4] Digital Signage - http://en.wikipedia.org/wiki/Digital_signage [retrieved: May, 2014]

[5] Neilsen - http://www.nielsen.com/us/en.html [retrieved: June, 2014]

[6] Google Now - http://www.google.com/landing/now/# [retrieved: May, 2014]

[7] Qualcomm Gimbal - https://www.gimbal.com/ [retrieved: May, 2014]

[8] McAfee - http://www.mcafee.com/us/ [retrieved: May, 2014]

[9] K. Algotar, S.V. Addicam and P. Satishchandra, "Case Study: Using Video Analytic Data to Target Advertisements," in proceedings of Journal of Emerging Trends in Computing and Information Technology, vol. 5, no. 3, March 2014, pp. 206–209.

[10] K-means Clustering - http://en.wikipedia.org/wiki/K-means_clustering [retrieved: June, 2014]

[11] Support Vector Machine - http://en.wikipedia.org/wiki/Support_vector_machine [retrieved: June, 2014]

[12] Naïve Bayes Classifier - http://en.wikipedia.org/wiki/Naive_Bayes_classifier [retrieved: June, 2014]

[13] KNN - http://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm [retrieved: June, 2014]

[14] PHP - http://www.php.net/ [retrieved:May, 2014]

[15] TFIDF Classifier - http://alias-i.com/lingpipe/docs/api/com/aliasi/classify/TfIdfClassifierTrainer.html [retrieved: May, 2014]

[16] SQLite database - http://www.sqlite.org/ [retrieved: June, 2014]

[17] XMPP - http://xmpp.org/ [retrieved: May, 2014]