Social Media Based Personalized Advertisement Engine

Hansi De Silva, Poorna Jayasinghe, Ashen Perera, Sithira Pramudith, Dharshana Kasthurirathna
Department of Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka

Abstract— **Online advertising has become a global phenomenon that affects the retail market substantially. Advertisements engines are an effective solution to the mobile application market to push advertisements. This paper reports evidence that AdSeeker, User Preference Based Advertisement Engine Based on Social Media is an effective solution to improve the business value of the marketing and advertising. Since the internet is used by vast number of people, it essentially needs a comprehensive method to push personalized advertisements to the right people. Adseeker is a system built using ontological mapping and social media content based semantic analysis to direct personalized. Identifying personal relationship hierarchy, and ontological approach for advertisement classification helps to identify the most appropriate advertisement for each user. AdSeeker uses the tweets posted by users to capture the preference of each and every user. Each user pushed advertisements based on their individual preferences. Based on the social experiments done using Adseeker, we could demonstrate that the social media profile based advertising is effective in providing highly relevant advertisements.**

Keywords— personalized advertising; ontology; machine learning; semantic analysis;

I. INTRODUCTION

Social media are computer-mediated tools that allow people and organizations to create, share, or exchange information, career interests, ideas, pictures/videos in virtual communities and networks [1]. It facilitates the development of online social networks by connecting a user's profile with other individuals and/or groups.

Since people post unstructured content in social media, it is relatively difficult to extract product related information from such content. Using a users' profile it is easy to gather thinking patterns such as interests, likes, hobbies, thoughts and so on. However, social media content has a huge potential in delivering more personalized advertisements. Most of the commercial advertisements and web applications are not based on users' actual preferences. They do not have social media related advertisements classification system, self-updating or a continuously updating user character profile more personalization and preference based advertisements pushing mechanism. Therefore users, advertisers, and corporations are not often harnessing the full potential of social media based advertising. In this paper, we present an efficient way to provide personalized advertisements to the right people at the right time as well as advertisements pushing mechanism based on user preferences. We used semantic analysis and ontology mapping, advertisements classification mechanism to achieve this target in our AdSeeker advertisement engine.

The following section of the paper explains about background of the proposed system using literature survey. The next section explains the methodology of the proposed system. Finally this paper explains how results are carried out using survey and future works.

II. BACKGROUND

A. User Influence in Twitter

Twitter is a popular social media around the world. Users in Twitter have varying influence level on the users about various topics. Normally Twitter users follow each other to getting information of what they posted on their account. When it comes to the highly followed user's profiles and things they post about the effect the massive influence on their followers. There is a significant advantage that can be gained through social media-based advertising. A trend can be initiated by anyone, and if the environment is right it will spread. Following three activities represent the different types of an influences of people.

1. In-degree influence, the number of followers of the user. Using this you can directly identify the audience of the particular user.
2. Retweet influence, which we measure through the number of retweets containing one's name.
3. Mention influence, which we measure through the number of mentions containing one's name, indicates the ability of that user to engage others in a conversation [2] [3] [4].

B. Relavance of Tweet and Sentiment Analysis

It is important to identify the relevance of tweets by analyzing the different types of features and understanding the semantic meaning of the tweets. According to the tweets manually labeled as relevant or not relevant by using logistic regression [2].

To identify the relevance of tweets, one of the most appropriate way is sentiment analysis. Sentiment analysis can be used to classify tweets into negative, positive and neutral

978-1-5386-4602-1/17/$31.00 ©2017 IEEE

classes. Previously there are several models used for classify tweets [5]. In our proposed system also used sentiment analysis to identify the relevance of data.

C. Create and Maintain Ontology to Build User Profile

Ontology is a set of concepts and categories in a subject area or domain that shows their properties and the relations between them [6]. When it comes to advertising, the system will need domain knowledge (Technology, Automobile, and Fashion) of ontology to identify the contexts of the social media updates. Ontology is very useful to identify different categories of a domain knowledge. Previously Ontology used for improve the quality and quantity of available knowledge by extracting, analyzing existing data. We can use open source ontology building software like Protégé to build massive ontologies with a relationship that is suitable to implement the analyzing system and context identifying modules. After building ontologies when analyzing social media updates the system can correctly identify the contexts of the social media updates and build the user character (preference) profile according to those social media updates. User character profile means a profile that contains the products and the brands that user has explicitly mentioned in their tweets. We build this after analyzing all the tweets [7] [8].

D. Existing Advertisement Engines and Web Clients

At present, there are several advertisement engines and web applications exists including Google AdMob, eBay, Ikman.lk, etc.

- Google AdMob

Google AdMob is one of the popular advertisement pushing engines. It's an advertising platform which is designed for mobile apps. AdMob works with search queries and push relevant advertisements. It has a mechanism to location based advertisements pushing to guest users and preference based advertisements pushing mechanism as well. AdMob has no mechanism to social media related advertisements classification and self-updating character profile [9].

- eBay.com

eBay is e-commerce company which do sales via internet. It mainly works using web application. Advertisers can provide advertisements to eBay as well. eBay uses machine learning mechanism to identify user's search queries and push advertisements according to previous search and purchase history [10].

- Ikman.lk

Ikman.lk is one of the popular Sri Lankan mobile advertising web application. Users can search for advertisements and users can post advertisements freely on the website. Classification mechanism is based on the district wise of the country. User preference identification mechanism not included [11].

III. METHODOLOGY

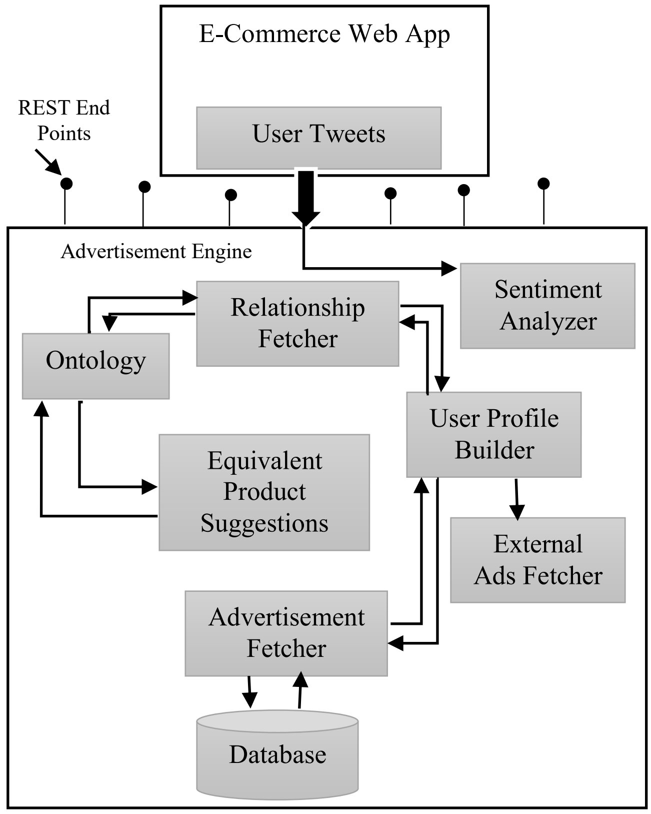


Fig. 1. Overall Architecture

Our proposed advertisement engine built using JAVA JAX-RX Restful API web services, Apache Jena which is used to connect and query with an ontology, hibernate to deal with database, MySQL Server used as the database, sentiment analysis [5] [12] [13] using Naïve Bayers [14] and used NLTK libraries. In order to demonstrate advertisement engine we used PHP web application with Codeigniter framework. Fig. 1 describes the overall architecture of proposed system. According to the overall architecture following steps explains the methodology of proposed system.

A. Ontology (Constructing Domain Knowledge)

Since this research is based on user preference based advertisement pushing mechanism, we had to develop ontology domains using Protégé software. In this research, we mainly deal with Technology and Automobile domains. An ontology consist with "onto-classes" and "individuals". Our proposed system ontology contains four main levels called Top-Category, Product Type, Brand and Product. As a class hierarchy, added accurate details to the ontology.

Ex: - technology->computer-> VGA-> titan x

As the above example added individual "Titan x" details to the relevant ontology hierarchy. Likewise by adding relevant classes, subclasses we built the 2 domains of ontology. Web Ontology Language used to process information in ontology and used the format as RDF/XML. Our ontology contains two properties called object properties and data properties. In data properties each and every class and individual contain an annotation called “isTypeof” and it has four types such as a product, brand, product types and top category.

Example:-

Technology → Top Category

Mobile Phone → Product Type

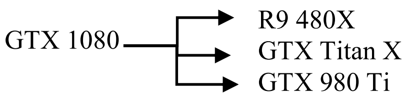
Apple → Brand

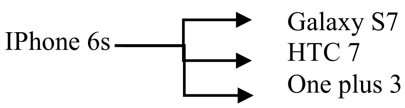
IPhone 6s plus → Product

B. Equivalent Product Suggestions

In our system we used object properties to map same range objects. It used to map individuals with same price range and performance range. All these details can be derived from the ontology.

As an example, For GTX 1080 and IPhone 6s can be mapped similar price range and similar performance categories.





Suppose a user has tweeted about GTX 1080, our system can identify GTX 1080 is a VGA card, identify its part of a computer category and its part of the technology section. Then it can get related product to the GTX 1080 like R9 490X, Titan X which are similar to GTX 1080. All these details can be derived from the ontology.

C. Sentiment Analyzer

Social media is used by a vast number of people in the world. Out of the many social media platforms Twitter can be considered as one of the most popular used across the globe. In order to obtain twitter data, user should have been logging with twitter. Using twitter id of a user gather twitter to the JAX-RX (Java API for RESTful Web Services) RESTful endpoints. Then after that check whether user existence in the database.

If the user is a new user (do not exist in the database), get all the tweets from users account and send those tweets to the sentiment analysis web services to get the positive tweets to the system.

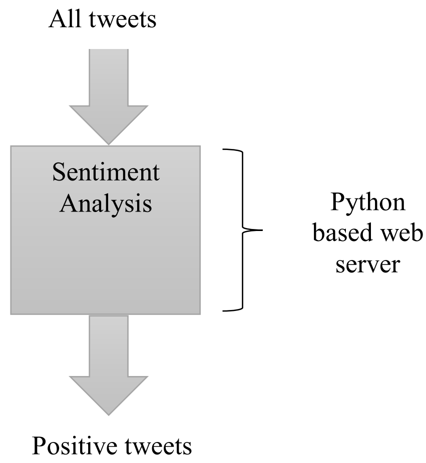
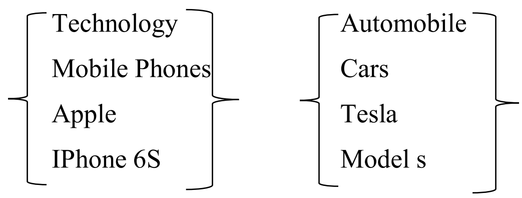


Fig. 2. Sentiment Analysis

D. Relationship Fetcher

After fetching all the positive tweets as in Fig. 2, send them to “FilterWordList” to filter words. Then send all remaining tweet variants to the ontology to get the relationship hierarchies. Then send all the relationship hierarchies to get user preference mention profile algorithm as shown in step E.

Sample Relationship Hierarchy,



E. Users' Character Profile Builder (Product Amount Mentioned Algorithm)

In order to build the product amount mentioned algorithm following steps were taken. Refer to the section A, it needs to create annotations for every class and individual. Then analyzing positive tweets and according to the section C. Imagine a sample tweet as “I really want to buy the new iPhone 6s and new model S”. After analyzing we can get a profile mention as follows.

TABLE I. USER MENTION PROFILE

		No of Occurrences
Product	iPhone 6s	1
	Model S	1
Brand	Apple	1
	Tesla	1
Product Type	Mobile Phone	1
	Car	1
Top Category	Technology	1
	Automobile	1

In our system, we maintain user profiles with user mentions as above Table 1 for all the tweets. Imagine if one person has tweeted twice that he wants to buy Model S. Then Model S count increases and all the hierarchy related to the Model S also increases. According to user's mention profile, we can obtain preferred products and hierarchies.

TABLE II. SAMPLE PRODUCT MENTION PROFILE

		No of Occurrences
Product	iPhone 6s	1
	iPhone 6s plus	2
	Model S	5
	GTX 1080	1
Brand	Apple	3
	Tesla	5
	NVidia	1
	Mobile Phone	3
Product Type	Car	5
	Computer	1
	Technology	14
	Automobile	5
Top Category	iPhone 6s	1
	iPhone 6s plus	2
	Model S	5
	GTX 1080	1

F. *Advertisement Fetcher Algorithm*

Relationship hierarchy object which is saved from above step D, E is needed to send to the advertisement fetcher algorithm in order to fetch how many advertisements we need. Using this algorithm it is going to fetch advertisements according to the product, brand, product type, and top category. The below Fig. 3 represent the basics of advertisement fetcher algorithm.

Advertisements fetcher
({Relationship object}, {Total advertisements count})

Fig. 3. Advertisement Fetcher Algorithm

After analyzing the profile (as an example Table II) identify how many advertisements we want to fetch from the database in order to display to the user. The top to bottom categories which mentioned in Table II, we assign more priority to the Products section and then Brand, Product Type AND Main Category respectively. The reason of giving highest priority to the product section is the user explicitly mentioning that he/she likes those products.

Imagine the situation the home feed has requested 20 advertisements. If the product category doesn't contain at least 10% of the mentions from the advertisements have requested (if the system has requested 20 advertisements at least 2 products should be there (10%) to get advertisements related to the product) next it goes to the brand's section and fetches advertisements from it.

Example 1:-

Product → iPhone 6s → 1

Brand → Apple → 1

Tesla → 3

Now the system fetches 50% of the requested advertisements which are related to the product and 50% related to the brands. According to the above scenario system will fetch 10 advertisements related to iPhone 6s and 10 advertisements related to the tesla products. If the brand section didn't contain 10% mentions of the requested advertisements it moves to the product type category and fetches 33% from the product section, 33% from the brand section and 33% from the product type section. If product type section didn't contain 10% mentions of the advertisements count it fetches 25% percentage of advertisements from each section.

Example 2:-

TABLE III. EXAMPLE 2 - SAMPLE PRODUCT MENTION PROFILE

		No of Occurrences
Product	GTX 1080	2
	Model S	2

The above mention profile is 100% priority because the client asked for 20 advertisements.

4/20*100=20% (ratio)

20% >= 10% (true)

So it gives 100% priority to the product section.

20 ads → 2 advertisements/4 total advertisements * 100

= 50% → for GTX 1080

20 ads → 2 advertisements /4 total advertisements * 100

= 50% → for Model S

According to the algorithm, system should fetch 10 advertisements from the database belong to GTX 1080 and 10 advertisements related to Model S.

G. *Advertisements Classifier Algorithm*

Advertisement classifier algorithm gets relevant advertisements from the database according to the country and according to an internal advertisement distribution function. Advertisement distribution function means distribution of advertisements according to the country. For example if you live in Sri Lanka you tend to prefer Sri Lankan advertisements. Therefore, marketing and business value will improve if we push advertisements to the right person at the right time. Thus Sri Lankans getting advertisements from Sri Lankans would be more profitable.

After fetching related advertisements from the database push those from the end point to the relevant user. The above steps can continue if the user exists in the database (Step D, E and F).

In our comprehensive system, we implemented admin panel from our web application. Using this it is easy to add individuals and classes to the ontology. Every 7 days user mention profile automatically updates. Old profile is moved to a different section of the database with a timestamp. Therefore we can get reports regarding a user's change of preferences.

Each search query the user searched (advertisements) is saved in the database along with the user profile and they also get archived every 7 days same as user character profile. This procedure helps to push advertisements to the home feed of the user.

IV. RESULTS

Compare with the existing systems and proposed system, it has new features for users to get preferable advertisements to their home feed. The below Table IV, V explains the new features of proposed system.

TABLE IV. COMPARISON WITH ADMOB ENGINE

	Google AdMob	Proposed System
Location based advertisement pushing to guest users	Yes	Yes
Preference based advertisement pushing mechanism	Yes	Yes
Social media related advertisement classification	No	Yes
Self-updating character profile	No	Yes

TABLE V. COMPARISON WITH WEB APPLICATIONS

	Ikman.lk	eBay.com	Proposed System
Advertisers can provide advertisements	Yes	Yes	Yes
Advertisers can categorize the advertisement	No	No	Yes
Categorization according to the person	No	No	Yes
Personalized advertisements	No	No	Yes

To test the proposed system with the relevance of advertisements, we have done the experiment. Basically the survey launched for check the accuracy of our advertisement engine with the other existing advertisement engines. Therefore we compared the number of relevant advertisements on eBay home feed and number of relevant advertisements on proposed system home feed. We developed the web client, which can push advertisements according to the user by calculating the number of advertisements using our advertisement engine.

The users that participate for the experiment have been using their eBay accounts actively to buy goods online. From the proposed system, to calculate the no of advertisements of specific user we have collected tweets with their permission, and by analyzing them got the advertisements count from different categories. The below chart illustrate how many relevant advertisements the home feed of the particular user's eBay and AdSeeker engine has suggested.

Therefore an experiment was done with 20 participants and the results showed that our home feed provides around 90% relevant advertisements to the user. The participants of the experiment had been using eBay and Twitter accounts for more than 3 years. Table VI shows the average of relevant advertisements between eBay and AdSeeker based of number of clicks of the relevant advertisements.

TABLE VI. RESULTS WITH EBAY AND ADSEEKER COMPARISON

	eBay		AdSeeker	
	A	B	C	D
Person 1	5	20.83%	18	90%
Person 2	4	16.66%	19	95%
Person 3	5	20.83%	18	90%
Person 4	6	25%	16	80%
Person 5	6	25%	17	85%
Total	26		88	
Average	26/5= 5.2	108.32/5 **=21.66%**	88/5= 17.6	440/5 **=88%**

A – No of relevant advertisements in eBay home feed (out of 24)

B - Percentage of relevant advertisements in eBay home feed

C – No of relevant advertisements in AdSeeker home feed (out of 20)

D - Percentage of relevant advertisements in AdSeeker home feed

V. DISCUSSION AND FUTURE WORK

This paper discusses the influence of social media to improve the accuracy of the Adseeker advertisement engine. This advertisement engine proves that user preference based advertisement engine is more efficient way to improve the relevance of the advertisements in mobile applications. Users, advertisers as well as corporations have many advantages from this advertisement engine.

Currently used commercial web applications and advertisement engines provide results based on previously searched queries most of the time. They do not consider the social media content generated by user. Therefore it may not be the actual preference of the user. But social media based system may provide more accurate results.

There are some limitations of our system as well. We collect user preferences through social media. Therefore we aware of what users published in there profiles not what they actually thinks. As a future work, it may be interesting to see whether the relevance of the advertisements improve when both the social media content and the search queries of the users are considered. As well as it can incorporate with the content generated by other social media outlets like Facebook and Instagram Preference results are further fine-tuned in order to increase the accuracy level of the system. Therefore it becomes a more usable and profitable product.

REFERENCES

[1] "Social Media," [Online]. Available: https://en.wikipedia.org/wiki/Social_media.

[2] Ke Tao, Fabian Abel, Claudia Hauff, Greet-Jan Houben, “What makes a tweet relevant for a topic”, Web Information Systems, TU Delft PO Box 5031, 2600 GA Delft, the Netherlands, 12 June 2015, pp 1-9.

[3] Meeyoung Cha, Hamed Haddadiy, Fabr´ıcio, Benevenutoz Krishna, P. Gummadi, “Measuring User Influence in Twitter: The Million Follower Fallacy”, Max Planck Institute for Software Systems (MPI-SWS), Germany, 2010, pp 1-8.

[4] Lalindra De Silva and Ellen Riloff, “User Type Classification of Tweets with Implications for Event Recognition”, School of Computing University of Utah, 2014, pp 1-11.

[5] Apoorv Agarwal Boyi Xie Ilia Vovsha Owen Rambow Rebecca Passonneau, “Sentiment Analysis of Twitter Data”, Department of Computer Science Columbia University New York, NY 10027 USA, 2009, pp 1-9.

[6] "Ontology," [Online]. Available: https://en.wikipedia.org/wiki/Ontology.

[7] Kavi Mahesh and Sergei Nirenburg, “A situated Ontology for practical NLP”, Computing research laboratory, USA, 19 August 1995, pp 1-10.

[8] Natural Language Interface for Querying Ontologies. https://www.academia.edu/22509718/NLP-Reduce_A_na_%C4%B1ve_but_Domain-independent_Natural_Language_Interface_for_Querying_Ontologies.

[9] "Google AdMob," [Online]. Available: https://www.google.com/admob/.

[10] "EBay," [Online]. Available: http://www.ebay.com/

[11] "ikman.lk," [Online]. Available: http://ikman.lk/.

[12] "Sentiment Analysis," Wikipedia, [Online]. Available:https://en.wikipedia.org/wiki/Sentiment_analysis.

[13] Mariam Adedoyin-Olowe, Mohamed Medhat Gaber and Frederic Stahl, “A Survey of Data Mining Techniques for Social Network Analysis”, School of Computing Science and Digital Media, Robert Gordon University, 2013, pp 1-25.

[14] "Naïve Bayes," [Online]. Available: https://en.wikipedia.org/wiki/Naive_Bayes_classifier.