THE USE OF SOCIAL NETWORKS IN AID OF PERSONAL HEALTH MONITORING

Castilo, Mark Anthony, Go, Matthew Phillip, Langilao,Cathy, Sanhi, Christelle Mae, Cheng, Danny
mark_anthony_castilo@dlsu.edu.ph, matthew_go@dlsu.edu.ph, cathy_langilao@dlsu.edu.ph,
christelle_sanhi@dlsu.edu.ph , danny.cheng@delasalle.ph

College of Computer Studies
De La Salle University
* Cheng, Danny : danny.cheng@delasalle.ph

Abstract: Personal health is becoming a growing field with the major platform and device vendors providing alternative solutions that address anything from manual software tracking to automated device based sensor tracking. Although there has been several existing personal health monitoring and management systems, personal health records (PHRs) monitoring applications are being limited to only a relatively small amount of users that uses them. A major reason is that people see healthcare as only needed during times of sickness and is about doctors, clinics, and hospitals integrating and updating information rather than a personal responsibility that has to be monitored and maintained regularly and personally. This leads them not to update their health information in the PHR/s. Our attempt to address the lack of information and the irregular and inconsistent updating of personal health records is to make use of major social media websites that have already attracted millions of users that share their information, whereby some are health related, on these sites. This paper discusses HealthGem, a cloud-based PHR system that focuses on prevention by monitoring and maintaining a healthy lifestyle rather than looking at the cure. The system incorporates an automated mining and retrieval of health related posts from Facebook using data mining approaches and crowd sourced data to increase the amount of health information it will contain to have a more comprehensive view of a person's overall health. By using the social media as a source of data, contributions from the community can also be included in the mined data thereby increasing the volume and granularity of data captured. This aims to reduce the amount of manual input to be performed by the user as data is already retrieved automatically. Currently, the system was able to mine and classify data with an accuracy and precision rate of greater than 90 percent with a sampling population of 500 to 1000 posts per user within a span of 1 year. This shows that social media can assist in populating health related information in aid of personal health monitoring.

Key Words: Personal Health Record, Social Media, Data Mining

1. INTRODUCTION

Personal health records (PHRs) are developed for users to let them input and keep track of their health information. Despite having many PHRs available for free having good functionalities and ease of use, there are only relatively few people that seems to use them. A reason for this is that people are lazy in taking preventive measures of their health. They see healthcare as "having someone else to make it better, and not about personal responsibility" (Cerrato, P. (2012, January 12)).

In contrast to this, social media has gained millions of online users posting information about the things they do, eat, and others (Social Media Report 2012). The information that are present in their profiles contain details that are related to them and can be used in the field of data mining. Data mining is "the process of analyzing data from different perspectives and summarizing it into useful information" (Data Mining (n.d.). Retrieved June 06, 2014). Data mining is currently applied to fields such as sentiment analysis, product analytics and others. Some of the information in the profile of the users are directly related to health such as the food they eat, and the activities they do. HealthGem will be able to automatically retrieve food and activity related posts from the user's Facebook account and add them as an entry to the user's PHR. Additionally, users can link their Instagram account to their Facebook account to allow more information to flow in. This allow users to view more of their health information based from the things they have posted on their social media accounts.

To make sure that the automated retrieval is kept up to date with the new trends in foods and activities, HealthGem has a crowdsourcing mechanism that collects new data from the users. Crowdsourcing is "the practice of outsourcing tasks to a broad, loosely defined external group of people" (Rouse, M. (n.d.) Crowdsourcing 2014). Allowing users to add new food and activities entries that are currently not in the HealthGem reference list and using these new inputs in the automated retrieval allows the app to be up-to-date.

HealthGem is a cloud-based PHR system that has an automated retrieval of health-related posts from Facebook using data mining approaches and crowdsourced data to increase the amount of health-related information in the user's PHR.

3. RELATED LITERATURE
2.1 Personal Health Records

There are many different PHRs available free for use, Microsoft Healthvault, Apple HealthKit, Samsung S Health, Capzule PHR, and others. Some basic information contained in these PHRs are personal information, list of illnesses/surgeries, progress notes, exercise routines, diet plans, and others. A commonality found among these systems are their keyboard input methods. However, HealthVault and Capzule PHR allow users to input through taking pictures. This feature enable users to store documents which are not electronically available. Another method of input is through importing files. For instance, digital copies of lab results could be added into the person's health.

2.2 Personal Health Records

In Culotta's research (2010) (Culotta, A. (2010, July 28)), he tried to detect influenza outbreaks from analyzing Twitter messages. Using a small number of flu-related keywords such as 'flu', 'cough', and others, the algorithm checks if a message contains such keywords and increases the influenza-illness statistics. It produced up to 95% correlation rate however, false correlations were detected such as when a message talks about 'flu shot' and not 'flu'. To solve this, he trained a bag-of-words classifier and manually labeling the data set whether the keywords were used in the right context or not. This generated an accuracy of 84.29%, 92.8% precision, and 88.1% recall.

2.3 N-gram Based Text Classification

One of the problems in text classification of the Chinese language is that it does not have natural delimiter/s unlike other languages like English that has white spaces separating the words. In the research by Luo, et.al (n.d.) (Luo, X., Ohyama, et al), they used n-gram based approach where the text is split into several n-grams. The n-grams created are then used for text classification and other algorithms needed to achieve their desired results. After performing feature selection, principal component analysis and support vector machines algorithm, it generated an 88.65% f-measure in one of their experiments.

3. METHODOLOGY

HealthGem is a cloud-based PHR app that operates on Android devices and a Java web server that acts as the host/cloud. It has multiple features along with the key feature which is automatically retrieving user's health information from their Facebook profiles using data mining algorithms. The app is targeted to users who are interested in tracking their health and regularly posts on social media.

Fig. 1. Tracker Post Screen

HealthGem has three main modules: trackers, display, and verification. The tracker module consists of eight different trackers that allow users to input and track their records: food, activity, weight, blood pressure, blood sugar, checkups, and notes. The eighth tracker is the calorie counter which computes the user's daily calorie intake from the food and activity trackers and informs the user whether they have met their respective recommended calorie intake for the day based on their weight and height. The food tracker allow users to record their food intake by choosing for a food from a predefined and crowdsourced list of foods. FatSecret, a food-calorie service, was also included into the system to increase the number of food choices for the users. The predefined food list were gathered from the Munchpunch website, one of the leading restaurant menu websites in the Philippines. Figure 1 shows a sample of the post screen of the mobile application. The activity tracker allow users to record their activities performed by also choosing from a list available. The list was retrieved from the FatSecret website that contains the list of activity names with their corresponding metabolic equivalent of task (MET). Crowdsourced entries are gathered from new entries made by the users if they have not found the entry they wanted in the list. The other five trackers allow users to input either numeric or text values for recording purposes. Adding status message and images along with the input values in the trackers are encouraged to give the users a sense of personalization in each of their entry. Additionally, the users may also choose to publish their entries to their Facebook account for information sharing.

The display module mainly consist of the timeline, home screen, and tracker display. Figure 2, is the timeline screen which is a compilation of all the tracker entries of the user arranged by time. This allow the user to track their updates, status, and images in chronological order. The home screen, as seen in Figure 3, is the first screen that the user will see once he/she opens the app. It contains a visual summary of the user's health for the day based on their latest updates. It evaluates the user's health information and informs the user correspondingly. Standard health metrics for height and weight, glucose level, blood pressure and recommended daily intake for calories, and nutrition (protein, fats and carbohydrates) were used for evaluation. For example, based on the user's weight, it can determine whether the user is normal, overweight, or underweight.

This information is displayed in the home screen along with other evaluations from the other trackers.
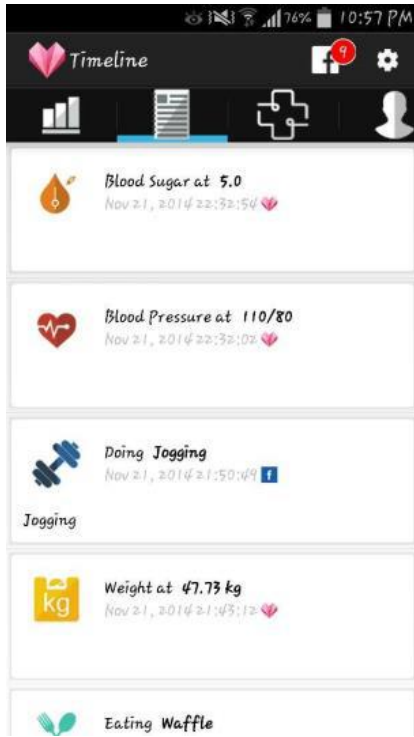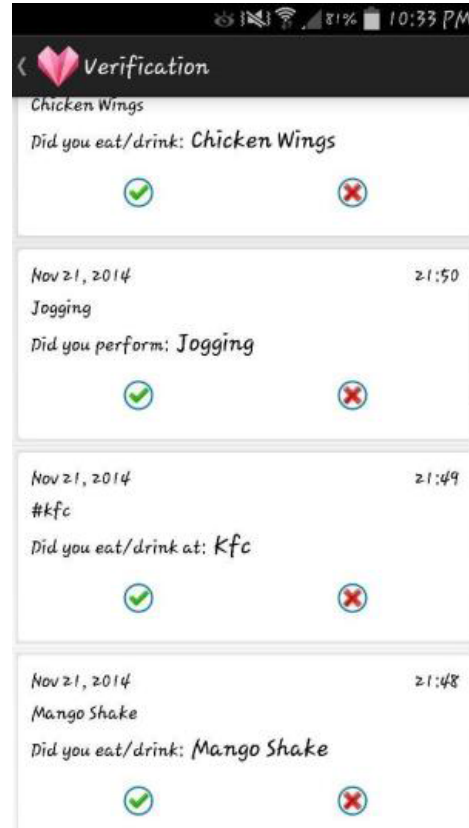
Fig. 2. Timeline Screen
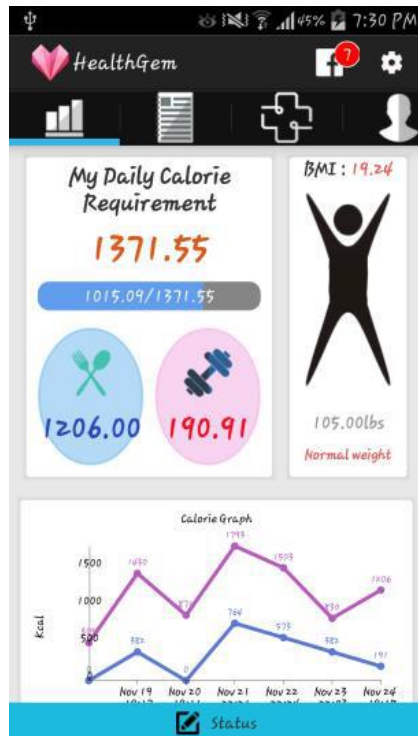

Fig. 3. Main Status Screen

Fig. 4. Main Status Screen

Figure 4 shows the verification screen of the mobile application. The verification module asks the user whether he/she eat/done the health information retrieved by the automated filtering algorithm. The algorithm automatically associates the respective calorie and/or nutritional information to the food/s or activity/ies identified from the user's Facebook account. The system also looks for restaurant and sports establishment names in the posts if neither food nor activity was found. If a restaurant or sports establishment name was found in a user's post, the verification module asks the user what he/she ate or did in that location. It presents a list of choices available either the list of food available in that

restaurant or the list of activities that can be performed in the sports establishment. To ensure that the user still has the control over his/her PHR, all these entries must still be verified by the user whether he/she wants to add these information to his/her PHR.

The filtering algorithm uses a large word corpus stored in the web server. The word corpus is divided into four different categories: food, activity, restaurant, and sport establishments. The food corpus is the list of food names available for the food tracker module. These words are retrieved from Munchpunch, a leading restaurant menu provider, FatSecret, a leading food-calorie service, and the list of foods and activities crowdsourced from the HealthGem users

3. RESULTS AND DISCUSSION

The algorithm was tested on four users, the first two actively posting their foods and activities and the other two are completely random users that share different things related to academics, songs, television shows, and others.

The accuracy measures the performance of how the algorithm classifies all the posts, both health-related and unrelated. Whereas, the precision and recall measures the performance of how the algorithm classifies the health-related posts.

Table 1. Performance Summary of Test User(s)

Criteria	Sample1	Sample2	Sample3	Sample4
Accuracy	97.62% 493/505	94.8% 566/597	96.67% 145/150	96.01% 939/978
Precision	94.16% 129/137	88.74% 197/222	-	49.35% 38/77
Recall	97% 129/133	97.04% 197/203	-	100% 38/38

As seen in Tables 1, the accuracy are relatively close at around 98% and recall at around 96%. This is mainly due to the ability of the filtering algorithm to filter out unrelated posts with relatively

few errors. The performance in recall mainly depends on how many words are added into the corpus list to increase the number of posts filtered in. However, this may lower the precision as more words may be present in posts that would be used in a different sense. As for precision, for the four users it had performances of 94.16%, 88.74%, 0%, and 49.35% respectively. This is mainly due to how the users used the corpus words in their posts. Because the algorithm only checks if corpus words are present in posts, it also retrieves posts even if it contains corpus words used in a different sense. For example, a post with the status "Had an apple for breakfast" has 'apple' as the food word which the user has eaten. However, a status containing this phrase "To all Apple iPhone users…" pertains to the Apple brand. These two posts would be filtered in by the algorithm because it both contained the food word 'apple'. This resulted for the first two users, who mainly posts about their foods and activity receiving high performance rates, while the last two users had performance rates below 50%. Depending on the kinds of posts users make in their social media account, the performance of the algorithm may vary significantly.

It is also observed that some filtering had minor errors wherein instead of referring a food as one word, say "strawberry cupcake", it will be referred as two foods: "strawberry" and "cupcake". This is mainly because the corpus contain the two individual words but not the compound word.

4. CONCLUSIONS

The filtering algorithm works well on the type of users that always posts about their foods and activities. It is also sufficient to conclude that the filtering algorithm's performance vary depending on the type of users and how these users have used the words in the corpus. This is because the algorithm has the capability and retrieving the posts containing corpus words as long as they were used in the right sense.

As seen in the results, the flaws of the algorithm are due to the lack of corpus words in the list and because the algorithm did not have a word-sense disambiguation. By adding more corpus words in the posts, it implies that there would be more posts to be filtered in. Although this solution may

reduce the precision and may retrieve more posts that contain words in a different sense, this can be solved by the existing functionality in the app wherein users can decide whether to add or ignore the posts and add it to their trackers. Future implementation may also include word-sense disambiguation techniques wherein other algorithms will be added in to allow the system to decide whether a corpus word was used in a proper sense before it is filtered in to the user's list of health-related posts.

4. REFERENCES

Cerrato, P. (2012, January 12). Why Personal Health Records Have Flopped. Retrieved March 6, 2014, from Information Week HealthCare: http://www.informationweek.com/healthcare/patient-tools/why-personal-health-records-have-flopped/d/d-id/1102247?

Culotta, A. (2010, July 28). Detecting Influenza Outbreaks by Analyzing Twitter Messages. Hammond, Louisiana.

Data Mining (n.d.). Retrieved June 06, 2014, from http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm

Luo, X., Ohyama, W., Wakabayashi, T., & Kimura, F. (n.d.). Approach, Automatic Chinese Text Classification Using Character-based and Word-based. Tsu, Mie, Japan

Rouse, M. (n.d.) Crowdsourcing. Retrieved from http://searchcio.techtarget.com/definition/crowdsourcing

Social Media Report 2012: Social Media Comes of Age. (2012, December 03). Retreived from http://www.nielsen.com/us/en/newswire/2012/social-media-report-2012-social-media-comes-of-age.html