Analysis of User-Generated Multimedia Data on Medication Management and Consumption Behavior Using Data Mining Techniques

Chaiwoo Lee, Lisa A. D'Ambrosio, Richard Myrick, Joseph F. Coughlin, and Olivier L. de Weck

AgeLab, Engineering Systems Division, Massachusetts Institute of Technology, Cambridge, MA 02139, United States
{chaiwoo,dambrosi,dmyrick,coughlin,deweck}@mit.edu

Abstract. Technology-enabled tools have been suggested as a solution to assist older adults in the management and consumption of medications. However, existing systems and studies are often limited by incomplete understanding of the potential users' behaviors. This study uses a web-based survey and photo submission system to collect and analyze user profiles and behavioral characteristics. Various data mining techniques, including association rules, clustering and classification, are used on quantified data to find important behavioral patterns, group users with similar characteristics, and discern factors related to risky medication management behaviors. This paper presents the process and results of analysis, including a detailed description of coding scheme and model development. Practical and methodological implications are also discussed.

Keywords: Medication compliance, assistive technology, user observation, survey research, design for aging population.

1 Introduction

An average older American takes 3 to 5 medications [1-2]. Of all older adults seeing a physician, 61% are prescribed at least one medication [3]. However, older adults often experience limitations in terms of physical and cognitive capabilities, making it challenging for them to manage and consume their medications as directed. As a result, noncompliance, which can cause serious problems in terms of health outcomes and healthcare costs, is reported to be prevalent in up to 59% of older adults [4].

As a solution, technology-enabled tools for medication management and their potential benefits have been discussed and explored in various academic disciplines. For example, studies in artificial intelligence mainly looked at sensor technologies and algorithms to track information related to medication management and consumption [5]. In engineering design, the configuration of system architecture and the design of devices and interfaces have been the major topics [6]. In addition, relevant research from human factors perspectives has focused on the interactions between users and the tools [7].

C. Stephanidis and M. Antona (Eds.): UAHCI/HCII 2013, Part III, LNCS 8011, pp. 490–499, 2013.

While current products and systems are developed to aid older adults' management and consumption of medication, many are based on insufficient understanding of user behaviors and use cases. For example, systems were often developed with a focus on performance and reliability, while the match among the interface design, information flow, and actual user behaviors has not yet been fully discussed [8].

Although medication compliance is an important topic and a complex concept, there exists little information on the ways that people, especially older adults, manage their medications. This study aims to observe and analyze older adults' behaviors in storing, managing and consuming their medications. A novel web-based photo collection survey was conducted for a remote yet detailed observation of medication management and user characteristics. A data mining approach was used to discern behavioral characteristics and patterns from the survey data and photo submissions. To find behavioral patterns, describe characteristics related to risky behaviors, and build taxonomy of users, various data mining techniques were used.

This paper describes the process of data collection, the characteristics of data and its variables, and the methods used for data mining analysis. The results and their implications are also presented. With the comprehensive investigation of user characteristics and behavior in managing medications, this study is expected to contribute by providing implications to the design of medication management systems as well as suggesting an effective method of participatory user study.

2 Data Collection

The purpose of this study is to find important characteristics and patterns that can inform the design and delivery of medication management systems. For the study, a survey method was used to ask people about their medication management behaviors and characteristics. Also, for a remote observation, a photo submissions system was added to the survey. The collected data were tabulated and coded for quantitative analysis. This section describes how the data were collected and processed, as well as what fields and values the dataset included.

2.1 User Survey and Photo Collection

A web-based user survey was conducted for collection of data. Individuals in the MIT AgeLab participant database, who have agreed to be contacted with studies they may be eligible to participate in, were emailed with information about the survey. The email included a link to the survey web page that included the instructions, the survey, and the photo uploading tool. Figure 1 shows a portion of the survey web page.

The survey included a brief questionnaire about the participants, including demographics and living situation. They were asked to answer questions on their consumption and management of medications, including the types of medications that they take, the amount of time they spend on managing and taking medications, and their perceived difficulty of managing.

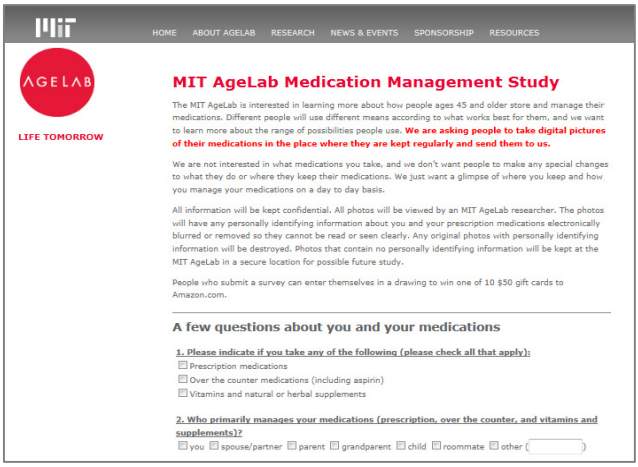

Fig. 1. Survey web page

In order for the survey to be submitted, participants were required to upload at least one photo that describes where and how they store their medications. They were asked to take one or more photos that show the medications and their surroundings including the room and the containers. Figure 2 shows a sample of collected photos.

Fig. 2. Photo collection sample

The survey was distributed only to people who are 45 or older to focus the interest of the study on the older adult population. As they are a significant population of users who consume various types of medications, they, and their caregivers, face a complex task of managing multiple medications. Thus, the topic of medication management applies more importantly to this population.

2.2 Coding and Pre-processing

Data were first pre-processed to remove any personal identifying information such as names and addresses in the collected photos. Any irrelevant data, such as photos without any medication information in them, were removed prior to analysis.

The information in the collected photos was coded quantitatively according to a comprehensive set of descriptor variables, including the room that medications are stored in, whether or not they are in an enclosed space, characteristics of packaging, display of risky storage behaviors, and more. The complete set of variables for metadata coding is summarized in Table 1.

Table 1. Coding scheme for photo information quantification

Category	Description	Binary (yes or no) variables
Location	The type of the room where medications are being stored as shown in the photo	— Kitchen — Bathroom — Bedroom
Enclosure	The type of space or container where medications are stored or enclosed as shown in the photo	— Enclosed cabinet, drawer or box — Open container or shelf — Tabletop — Bags or purses
Combination	The method of which medications are kept individually or mixed with other drugs as shown in the photo	— Individually kept — Mixed in bulk — Mixed by dose
Packaging	The method of which medications are kept in their original packaging or repackaged by the user as shown in the photo	— Kept in original packaging — Kept in a pill minder — Kept in some other bottle or box
Space sharing	Items or products that are kept in the same space as the medications as shown in the photo	— Medications alone — Kept with food — Kept with bath/ beauty products — Kept with healthcare products — Kept with office supplies — Kept with flatware and/or utensil
Risky behaviors	Whether or not any behaviors undesirable for appropriate management of medications are shown in the photo	— Medications kept near heat — Medications kept near water — Medications kept with the lid open

2.3 The Dataset

The data includes 112 observations, or responses from 112 individuals. As some people submitted more than one photo, the total number of photos collected was 213. For those who submitted multiple photos, the information from the photos were collectively merged into one row. For example, if a person submitted a photo showing medications in a bathroom and another photo showing medications in a kitchen, the row corresponding to the individual was marked with the value 1 under both variables of *Kitchen* and *Bathroom*.

A total of 36 variables, including participant ID and basic demographics, are included in the data. Out of the 36, 21 binary variables, as described in Table 1, are used

for quantitative description of photos collected. More binary variables were also used for describing if the respondents take prescription medications, over-the-counter medications and/or supplements. Categorical variables were used for marital status, living situation, and other people that a person manages medications for. Ordered categorical variables were used for age, time spent managing medications and perceived difficulty. A ratio scale was used for describing the number of people in household.

3 Analysis and Results

A number of different data mining techniques were used to investigate find behavioral patterns, group users, and to describe predictors of risky behavior. XLMiner, a data mining program operated on Microsoft Excel, was used for applying the techniques. This section describes the specific goals, methods and results of the analysis.

3.1 Medication Management Behavior Patterns

For the objective of finding important and common behavioral patterns in management and consumption of medications, the associations between variables were investigated. Association rules analysis was used on the binary variables to see if some behavioral and user characteristics are associated with others.

Variables with binary values were used for the association rules analysis. Some relevant categorical and continuous variables were transformed into binary format. For example, the variable *Number of People Living With* was transformed into *Live with Others*, with value 0 for living alone and 1 for living with someone. Also, one observation that had missing values was removed prior to analysis.

Using XLMiner, association rules with the minimum support of 20 and minimum confidence of 80% were generated. The initial result generated many rules were redundant and didn't contain meaningful information. For example, one rule stated that "90.91% of females who keeps prescription medications individually and mixed by dose also consume supplements and use both the original packaging and pill minders." This rule is meaningful in illustrating that in some cases people employ multiple methods and use multiple containers to manage their medications. However, the specific causal order may not have a significant meaning, especially since the related variables were found in different orders in other rule statements. Due to the limitations of directly interpreting the raw results, this part of the study focused more on identifying and examining the co-occurring relationships that were common in many rules. A subset of the result is summarized in Table 2.

Table 2 lists several rules that were found by examining the common co-occurring associations. The first rule suggests that when prescription medications are organized by dose, pill minders are used 97% of the time. However, only 75% people who use pill minders organized their prescription medications by dose. This rule indicates that while pill minders are the primary strategy in organizing medications to keep the regimen, they are not always correctly used. The second rule states that 71% of the time when pill minders are used, they are kept in an enclosed space. Since pill minders are

intended to be used as an accessible, simple and often portable means for complying with medication regimen, one can expect to see pill minders in a more open space, such as tabletops, or near a purse. However, pill minders were very often kept deep inside big boxes and cabinets, which, along with the first rule, suggests that people who have pill minders may not be using them as intended. The third rule states that 66% of people who use pill minders and keep their medications mixed by dose are females, which is higher than the overall proportion of females in the sample. This indicates that people who employ more strategized and planned method of organizing medications are largely females. The fourth rule states that 80% people who manage medications for other people in the household are females living with others. Although it was not statistically tested, rules 3 and 4 may be related in how females are often faced with more complex tasks and may seek ways for easier management. The fifth rule states that 69% of people who live with others and take both prescriptions and supplements keep their medications in individual, original packaging in enclosed spaces. No gender association is stated in this rule, but it may have implications similar to rule 4 in that it suggests the relationships between complexity of task and simplicity of management methods.

Table 2. Association rules (a subset of secondary analysis results)

A	B	Confidence[1]		Lift ratio[2]
		A→B	B→A	
Dose-mixed, Prescription	Pill minder	97%	75%	2.25
Pill minder	Enclosed	71%	50%	1.16
Female	Pill minder, Dose-mixed	44%	66%	1.18
Female, live with others	Manage for others	28%	80%	2.07
Live with others, Prescription, Supplement	Enclosed, Individual, Original packaging	69%	44%	1.16

3.2 User Groups

The second part of the data mining analysis concerns the variety in user characteristics and types of user behaviors related to medication management. This aims to categorize and group users to build a taxonomy of users according to their behavioral characteristics. For this part of the study, clustering analysis was used.

Prior to the analysis, variables that were thought not to contribute directly to medication management behavior were excluded. These include *ID*, *Marital Status*, *Pets*, *State* and *Race*. Hierarchical clustering was first applied to see the overall structure in which the observations can be clustered or set apart. Based on the hierarchical clustering result, it was determined that five distinct clusters can be formed with the data.

[1] The confidence of rule A→B shows the proportion of events A that also meet consequent B.
[2] The lift shows the performance of a rule compared to random, independent occurrences. A lift higher than 1 indicates positive dependence between events.

Accordingly, a k-means clustering with k=5 was conducted. The result from the k-means clustering analysis is summarized in Table 3.

From the k-means clustering results, with k=5, it can be seen that respondents can be grouped based mainly on the types of medications, their age, the number of people in household, whether or not they manage medications for someone in the household, the primary storage location, their use of pill minders and reorganization of medications, and the things that they store their medications with.

Table 3. K-means clustering result (K=5)

Cluster no.	No. of observations	Distinguishing characteristics based on cluster centers
1	4	— Spend the most time managing medications — Live alone and among the oldest of the sample — Store medications in kitchen — Mixes by dose and don't keep original packaging
2	51	— Don't reorganize medications in any way — Manage only for self
3	24	— Take both prescriptions and supplements — Mostly female — Store medications in kitchen and bathroom — Manage for others in household
4	15	— Take over-the-counter medications — Keep original packaging
5	17	— Take both prescriptions and supplements — Use pill minders to organize medications

3.3 Behavior Classification

Another objective of this study is to identify and describe characteristics related to undesirable, or risky, behaviors such as storing medications near heat, near water, or with the lids open, as exemplified in Figure 3. With the identification of common characteristics, safety measures and tools can be targeted to help people with the identified characteristics. For this part of the study, classification methods were used to develop models that classify risky and non-risky users based on the data.

The response variable was newly defined to integrate the variables describing risky behaviors - *Near Heat*, *Near Water*, and *Container Open*. If a person had 1 under one or more of the three variables, the person was marked with a 1 under the new variable *Risky Behavior*. That is, the new variable *Risky Behavior* had two values - 0 for no sign of risky behavior and 1 for display of some risky behavior. Variables that were hypothesized to be irrelevant, or found to have little variance in its values were excluded from the analysis. For example, only 4 observations had value 0 for the variable *Original Packaging*, and the difference in the value of this variable was found not to be associated with risky behaviors. Thus the variable *Original Packaging* was

removed before classification. In the same manner, variables *Individual*, *Bags* and *Bulk-mixed* were excluded from classification analysis.

Fig. 3. Examples of risky medication management behaviors

For this supervised approach, the data was partitioned into a training set and a validation set. Prior to the classification analysis, 60% of observations were randomly labeled as training data, and the other 40% were grouped into the validation set. The training set was used for developing classification models, and the trained model was applied to the validation data for evaluation.

A total of five different methods - naïve Bayes, k-nearest neighbors, logistic regression, discriminant analysis and neural networks - were used to develop classification models. For the k-nearest neighbors model, k=3 was used for analysis as it was found to be the best k with the lowest validation error. For logistic regression, a best subset selection was done using an exhaustive search, and a model with 11 predictor variables was selected. During the neural networks analysis, 100 networks were first generated and a simple model with the lowest error was selected. Classification tree method was also used, but was not included in the final comparison as it generated a model where all observations were predicted to belong to the majority and didn't base the decision on any of the behavior variables. The error results from the five classification models are summarized in Table 4. In Table 4, class 1 error refers to the probability of incorrectly classifying risky individuals in the validation set as non-risky.

Table 4. Classification model comparison

Classification model	Training error	Validation error	Class 1 error
Naïve Bayes	8.96%	22.73%	100.00%
k-nearest neighbors	14.93%	18.18%	100.00%
Logistic regression	10.45%	22.73%	62.50%
Discriminant analysis	11.94%	20.45%	75.00%
Neural network	4.48%	20.45%	87.50%

While the k-nearest neighbors model showed the lowest overall validation error, it classified all risky individuals in the validation set as non-risky. If higher costs are associated with the class 1 error, this result can be very costly. On the other hand, the

logistic regression model showed the best performance in correctly identifying individuals with risky behavior, although it showed a somewhat higher overall validation error. If one was to develop a product or service targeted at individuals who are more likely to show risky behavior, the logistic regression model may be chosen to best inform related design decisions. The logit model below, as developed from with the logistic regression analysis, shows that individuals who consume supplements, store medications in bathroom or open tabletops with various other products, have a pill minder, and live with other people are more likely to show risky behaviors. On the other hand, it shows that people who keep medications in enclosed spaces are less likely to show risky behaviors.

$$\begin{aligned}\text{Logit} = &-4.23 + 1.80\times\textit{Supplements} + 0.87\times\textit{No. of people living with} + 1.81\times\textit{Bathroom} \\ &- 4.68\times\textit{Enclosed} - 1.79\times\textit{Open box} + 1.99\times\textit{Tabletop} + 2.07\times\textit{Pill minder} + \\ &2.25\times\textit{With food} + 2.77\times\textit{With bath or beauty} + 2.82\times\textit{Flatware utensil}\end{aligned}$$

4 Discussion and Conclusion

In this study, user-generated data were analyzed using various data mining techniques to find results that can inform the design of medication management systems. The goal of this study was to find important behavioral characteristics and patterns that people, especially older adults, show as they manage and consume medications. Along with the survey data, information coded from a collection of photos was used for a comprehensive analysis. To find patterns, investigate user characteristics, and determine what leads to risky behaviors, association rules analysis, clustering analysis, and various methods of classification, respectively, were applied.

Through association rules analysis, this study found co-occurring associations among variables. The results can be expected to inform designers as they analyze use cases, investigate possible errors, and set specific development goals. For example, medication management tools that can be used by a whole family can be targeted a female adult user, since they are more likely to be actually in charge of managing medications. This study then took a step in building a taxonomy of users and their behavioral characteristics using hierarchical and k-means clustering. The result of clustering can be useful as designers write scenarios and detailed use cases around system interactions and interfaces. The clusters can inform practitioners if they need to target certain behaviors or specific groups of people. The third part of the study concerned classifying users into risky and non-risky groups based on their behavioral characteristics. Five different methods were used to find classification models, which were analyzed comparatively based on their performance. In order to make medication support tools more beneficial and effective, targeting the behavioral characteristics associated with risky behaviors can be helpful. For example, systems can be designed with an enclosed storage space without requiring users to organize medications into pill minders to lower the likelihood of showing risky behaviors.

This study has implications not only for developers and designers of medication management tools, but also for practitioners in other fields. For example, pharmaceutical companies can use the result of clustering analysis for targeting their products at a more specific segment of customers. Healthcare professionals such as physicians and

nurses can better advise their patients to store and manage their prescription medications in a less risky manner. Also, home care personnel can be informed to assist their customers and patients in managing their medications in a safer way.

A novel web-based research approach was used for data collection. Previous studies aimed at studying medication consumption behaviors were based on home visits and direct observations [8-9]. While such methods allow a close analysis, they may not be ideal for studying the heterogeneity and variance of behavioral characteristics, and can be costly for studying a larger sample. As the use of information and communications technology become prevalent among older adults, the user-driven remote observation method can be effectively and efficiently utilized. Also, the process of metadata coding can inform development of algorithms for analyzing graphic media. The definition of variables and attributes, as well as the process of quantification, can inform ways in which features are extracted and analyzed in multimedia data mining.

In short, this study extends the current understanding of potential users of medication management systems, especially older adults, by investigating the variance and patterns in their behavioral characteristics. The use of a new user-driven approach, application of quantitative data mining techniques, and the findings from the study can be expected to inform system designers, healthcare providers and caregivers, as well as researchers and scholars in the disciplines of user studies and data mining.

References

1. Giron, M.S.T., Wang, H.X., Bernsten, C., Thorslund, M., Winblad, B., Fastbom, J.: The Appropriateness of Drug Use in an Older Nondemented and Demented Population. J. Am. Geriatr. Soc. 49(3), 277–283 (2003)
2. American Society of Health-System Pharmacists (ASHP): Snapshot of Medication Use in the U.S., ASHP Research Report (2000), `http://www.ashp.org/s_ashp/docs/files/PR_snapshot.pdf`
3. Rathore, S.S., Mehta, S.S., Boyko, W.L., Schulman, K.A.: Prescription Medication Use in Older Americans: A National Report Card on Prescribing. Fam. Med. J. 30(10), 733–739 (1998)
4. Malhotra, S., Karan, R.S., Pandhi, P., Jain, S.: Drug Related Medical Emergencies in the Elderly: Role of Adverse Drug Reactions and Non-Compliance. Postgrad. Med. J. 77(913), 703–707 (2001)
5. Sadri, F.: Multi - Agent Ambient Intelligence for Elderly Care and Assistance. In: Proc. Internat. Electron. Conf. Comput. Sci., pp. 117–120 (2007)
6. Silva, J.M., Mouttham, A., El Saddik, A.: UbiMeds: A Mobile Application to Improve Accessibility and Support Medication Adherence. In: Proc. ACM SIGMM Internat. Workshop Media Stud. Implement. Help Improv. Access Disabl. Users, pp. 71–78 (2009)
7. McGee-Lennon, M.R., Wolters, M.K., Brewster, S.: User-Centred Multimodal Reminders for Assistive Living. In: Proc. ACM SIGCHI Conf. Hum. Factors Comput. Syst., pp. 2105–2114 (2011)
8. Geetanjali, N.: Prospective Randomized Survey Study on Assessment and Education of Home Medicine Cabinet among General Population of Community, Doctoral Thesis. Rajiv Gandhi University of Health Sciences, Bangalore, India (2010)
9. Palen, L., Aaløkke, S.: Of Pill Boxes and Piano Benches: “Home-Made” Methods for Managing Medication. In: Proc. ACM SIGCHI Conf. Comput. Support. Coop. Work, pp. 79–88 (2006)