Using Machine Learning to Facilitate the Delivery of Person Centered Care in Nursing Homes

Gerald C. Gannod,1 Katherine M. Abbott,2 Kimberly Van Haitsma,3 Nathan Martindale,1 Rachel A. Jennings,1 Chelsey N. Long1

1Department of Computer Science, Tennessee Technological University
2Department of Sociology and Gerontology, Miami University-Oxford
3College of Nursing, Pennsylvania State University, Polisher Research Institute, Abramson Center for Jewish Life

Abstract
Nursing home providers are moving towards a model of care that is person centered in order to improve the quality of care and quality of life for individuals residing in their communities. The State of Ohio has mandated that providers use the Preferences for Everyday Living Inventory (PELI) to assess resident preferences. This paper and pencil assessment adds to an increasing data management barrier to efficiently incorporate preferences into care. We are in the process of developing the Care Preference Assessment of Satisfaction or ComPASS system which supports data collection and reporting in order to better integrate preferences into the everyday care of residents. With this platform we are exploring how machine learning can be used to provide more personalized care in nursing homes by providing insights and recommendations based on resident preferences while lessening the data collection and management burden. In this paper, we describe ComPASS, discuss our initial investigations into using machine learning for long-term care, present initial findings, and suggest future directions for this research.

Introduction
The Ohio Department of Medicaid adopted a pay for performance initiative in 2015 whereby provider communities are incentivized to assess nursing home (NH) resident preferences via the Preferences for Everyday Living Inventory (PELI) to integrate preferences into the care planning process (Carpenter et al. 2000; Van Haitsma et al. 2013). Pay for performance programs are innovative initiatives aimed at influencing systemic change among large structural aspects of health care. As a team of researchers, we began working with the nursing home communities of Ohio to: 1) guide them on ways to integrate the PELI assessment into daily operational practices; 2) provide education/training about using the PELI data to inform care; and 3) to evaluate the barriers to preference-based, person-centered care implementation and provide solutions for long-term sustainability.

In addition we are developing tools to address key concerns needed to facilitate preference-based care in provider communities. In particular, our investigations explore three areas: 1) easing the collection of data that captures resident preferences and satisfaction with preferences being fulfilled, 2) creation of convenient reporting mechanisms to support incorporation of preferences into individual care conferences, and 3) use of machine learning to provide insight and personalization into the care of residents through the use of their preferences. Long-term care providers need a scalable and efficient way to collect data associated with the PELI. In addition, as preference and satisfaction data is collected, practitioners seek to use PELI data to inform quarterly resident care conferences and quality improvement efforts.

In this paper, we describe our efforts to support collection, reporting, and personalization through our work on the creation of the Care Preference Assessment of Satisfaction or ComPASS system. Of special interest is our application of machine learning approaches to ease the effort required to collect data through the creation of a recommender system that identifies association rules using rule-based collaborative filtering.

The remainder of this paper is organized as follows. In the next section, we provide background information on person-centered care, recommender systems, and review related work. We then present the ComPASS system, an application that has been created to support data collection, reporting, and analysis. We then discusses the recommendation system that we have built followed by a brief evaluation of the recommender. Finally, we draw conclusions and describe future investigations.

Background
In this section we provide background information in the areas of person-centered care and recommender systems, and review relevant related work to provide context for the research described in the remainder of this paper.

The culture change movement in long-term care began in the early 1980s as a widespread effort led by consumer advocacy groups, policy makers, and health care providers to improve the quality of care and quality of life for individuals residing in nursing homes. The overarching goal of the culture change movement is to transform care delivery from a “medical model” to a more comprehensive, holistic model of care that recognizes all aspects of the person beyond their disease or disability. Following the Omnibus Budget Reconciliation Act of 1987, nursing home providers were required by law to provide, “services sufficient to attain and maintain his or her highest practicable physical, mental, and psy-
- (F0400B) How important is it to take care of your personal belongings?
- (F0400C) How important is it to choose what clothes to wear?
- (F0400D) How important is it to choose what clothes to wear?
- (F0400E) How important is it to choose your own bedtime?
- (F0400F) How important is it to choose who you would like involved in discussions about your care?
- (F0400G) How important is it to go outside to get fresh air when the weather is good?
- (F0500B) How important is it for you to listen to the music you like?
- (F0500C) How important is it for you to choose between a tub bath, shower, bed bath, or sponge bath?
- (F0500D) How important is it for you to listen to the music you like?
- (F0500E) How important is it to choose your own bedtime?
- (F0500F) How important is it to choose who you would like involved in discussions about your care?
- (F0500G) How important is it to go outside to get fresh air when the weather is good?

Figure 1: Subset of the PELI-NH Minimum Data Set 3.0 Questions (with MDS reference indices in parentheses)

As a result, providers began incorporating more individualized approaches to care delivery and the concept of person-centered care emerged. Identifying and documenting residents' preferences is an important first step towards providing individualized, person-centered care. The Preferences for Everyday Living Inventory (PELI-NH) (Van Haitsma et al. 2013) was developed as a standardized assessment of psychosocial preferences to be used with nursing home residents.

Several items from the PELI instrument informed the development of the Centers for Medicare and Medicaid Services Minimum Data Set (MDS) 3.0 Section F “Preferences for Customary and Routine Activities”; a required assessment of all residents in certified nursing facilities (Saliba et al. 2008). A subset of the questions from the PELI-NH and MDS is shown in Figure 1.

The PELI-NH consists of 72-items asking nursing home residents to rate preferences on a four-point rating scale: 1 = “very important”; 2 = “somewhat important”; 3 = “a little important”; 4 = “not at all important”; and 9 = “important but can’t do/no choice.” The questions fall under major groupings including activities, privacy, food and dining, socializing, medical care, and personal care. One example preference question posed to residents in this context that falls in the personal care category is “How important is it for you to choose what clothes to wear?”

Recommender systems (Aggarwal 2016) have gained in popularity as a practical application of machine learning. Consumer websites have long seen demonstration of this technology in areas such as e-commerce (Amazon), movies (Netflix), and music (Pandora). Different strategies for implementing recommender systems have been suggested, including the use of association rules (Cakir and Aras 2012; Lin 2000; Bendakir and Ameur 2006). In our work, we apply the use of a combination of association rule mining using the Apriori algorithm (Agrawal and Srikant 1994) and linear regression using a generalized linear model (Nelder and Wedderburn 1972).

In the healthcare domain, the idea of using data mining is not a new one (Patel et al. 2009; Simovici 2012). Applications have included treatment effectiveness and condition identification (Koh and Tan 2005). Hu, for instance, has suggested using Apriori for mining medical data but as a means for diagnosis of conditions (Hu 2010). In the area of long-term care, we are unaware of work being done to understand resident preferences.

Care Preference Assessment of Satisfaction

The Care Preference Assessment of Satisfaction System, or ComPASS is a web-based implementation of an Excel tool created to support the use of the PELI-NH with a quality improvement. ComPASS was developed using the Ruby on Rails framework and seeks to provide a scalable, user-friendly system to facilitate data collection and data analytics of long-term care residents’ preferences.

At the highest level, the system is organized at the facility or community level, which represents a community of residents at an individual nursing home. An administrative user adds employees (also known as mentors) to the system; mentors ask the PELI questions through interviews with residents. ComPASS supports the interview process by providing an interface similar to the one shown in Figure 2. In our initial offering of the system, each mentor is required to ask the 16 MDS 3.0 subset of questions.

When a resident is admitted into a community on ComPASS, they are given a first name, last name, unique id, an assigned mentor, and a neighborhood. Neighborhoods are a sub-group of residents typically organized by a geographic location, such as a particular hallway or floor. Neighborhoods allow for aggregated reporting and analytics. All of the settings for a resident can be edited at a later date.

In our work, we have also added additional nested questions that explore specifics regarding the top-level preference question. For instance, for the question posed above regarding clothing, nested questions explore topics such as “What do you like to wear to sleep?” or “What jewelry do you like to wear?” These additional questions are open ended and provide a way for a care provider to capture details about preferences. After a set of questions have been answered, the interviewer can proceed to the next question by clicking a button at the bottom of the interview. Interview
progress is visually marked by a progress bar and is automatically saved at each step allowing interviewers to pause an interview for any reason and resume it at a later point.

New residents can be led through an initial interview by a variety of personnel, including social workers, activity therapists, or volunteers. An initial interview includes only preference questions without any of the related satisfaction questions as resident are unable to report on their satisfaction level of their preferences until they have spent some time in the community. After completing the preference questions, care providers will have the option of doing a follow-up interview that asks residents how satisfied they are with their previously stated preferences. Satisfaction questions are associated with each preference question in the form of “How well do you feel this preference has been satisfied?” Responses to this question include very satisfied, somewhat satisfied, not satisfied, and no response or N/A. After completing both an initial interview and a follow-up interview, all future interviews include both sections as described above.

ComPASS supports display of reports of resident satisfaction in order to provide input and insight into the care of the resident during individual quarterly care conferences. These reports show the breakdown of each question marked with importance as well as how satisfied the resident is with the care they have received with respect to that preference. In the report, this is displayed as a pie chart with 4 sections: unsatisfied, needs improvement, satisfied, and important but can’t do. The report also provides detailed information regarding of each of these sections including questions that fall within each as well as feedback and notes as recorded by the direct care worker. At the community level, we display a dashboard of information that provides an overview of several metrics that are important to nursing home communities. These metrics are an aggregation of data collected for an entire nursing home community, an example of which is shown in Figure 3.

**ComPASS Recommendation Engine**

One of the goals of our work is to use ComPASS as a mechanism for conducting research in long-term care. While the system provides the means for collecting, analyzing, and reporting on important resident preference data, we also believe that the use of data mining and machine learning creates an opportunity for providing insight into potential preferences and diagnosis of conditions.

The MDS 3.0 is a subset of the PELI-NH and consists of the 16 core questions, a subset of which is shown in Figure 1. While it is a much more accessible set of questions for conducting preference and satisfaction surveys when compared to the 72 questions of the PELI-NH, there is a desire to explore the topics and concerns of the full 72 question PELI-NH. In particular, it is desirable to be able to use the PELI-NH for both personalized care and research purposes, but conducting 72 question surveys for the PELI-NH can often lead to survey fatigue (Hess, Hensher, and Daly 2012). As such, it is very difficult to perform statistical research on long-term care based on the use of the PELI-NH as it can be challenging to conduct the surveys.

To assist in the reduction of fatigue while also providing personalization of resident experience we have been developing a recommender as a way to provide guidance to direct care workers as they explore the next best set of preference questions to ask a nursing home resident. Our initial investigations have focused on using a rule-based collaborative filtering approach, with responses to preference questions used in a manner analogous to selection of products in a market basket.

The process of generating a recommendation occurs following an interview. Once a resident has provided responses to the initial 16 MDS questions, the list of recommended questions and categories from the rest of the 72 - 16 = 56 preferences are generated and displayed, and the interviewer can select which questions to add to the rest of the interview. As shown in Figure 4, top-3 items are displayed followed by a ranking of questions identified during the logistic regression phase of our approach. The bar to the right of the questions provides a visual representation of the score used during application of logistic regression to rank the quality of a rule. In this context, the score metric is obtained via the use of support and confidence.

Our approach uses the basic Apriori algorithm (Agrawal and Srikant 1994) with no fine-tuning of the algorithm beyond parameterization. In addition to fine-tuning and using other similar algorithms, our future investigations will explore the use of neighborhood-based recommendation methods that look at both user-based and item-based models. In our approach, we map the importance levels of “very important” and “somewhat important” response categories to true or 1 and the response categories of “not very important” and “not important” to false or 0 in order to create a unary ratings matrix. In our investigations, we have found that the choice between resident responses of “very important” and “somewhat important” to be indistinguishable (Van Haitsma et al. 2014). In addition, the existence of the “not very important” and “not important” response categories allow us to treat the data as a binary matrix.

The recommender is built from a collection of rule sets generated via the Apriori algorithm (Agrawal and Srikant 1994) from the arules R package (Hahsler et al. 2017). These rules are structured as \( \{ A \} \rightarrow B \), where the preference for all of the preferences in the antecedent set \( A \) suggests the preference for a consequent \( B \). While the antecedent and consequent pairs may refer to multiple preferences, they are only fired if a given resident indicates a preference for all of the preferences in \( A \). For example, one such rule is: “F0400B, F0400F \rightarrow How important is it to choose your own care professional?” where the left hand side is an antecedent set of questions from the MDS (see Figure 1), and the right hand side is a question from the remaining 56 PELI-NH questions.

Rules are generated using minimum support, confidence, and lift values, and are filtered to only contain the 16 MDS items in the antecedents and the remaining 56 preferences in the consequents. In order to generate recommendations, the initial 16 MDS preference responses collected from a particular resident in an interview are passed into the recommender using the mapping described above.
The aggregated statistics for each consequent collected from the matching rules are then run through a logistic regression model, where each consequent is classified as either a good or a bad recommendation.

In addition to making recommendations on specific preferences, we also support recommendations at the category level. Specifically, rule sets for recommending categories of questions are incorporated into the recommendation process, as shown in Figure 5. In this case, all questions contained in a given category are displayed, with questions contained in the recommendation consequent set highlighted (including the bar representing the ranking of a given question). The same 16 MDS preference responses are fed into the categorical rule set, where the consequents contain categories of preferences the resident is likely to consider important. The resulting categories from the rules that fired are then expanded into their constituent individual preferences, and these are added to the list of consequents prior to consequent aggregation.

**Evaluation**

To evaluate the ComPASS recommender, we are using a dataset that was collected through a convenience sample of 28 nursing home facilities located in the Greater Philadelphia, Pennsylvania area. For a description of this dataset and other findings regarding preferences, see (Abbott et al. 2018). We use this dataset as the basis for exploring the viability of using machine learning to provide personalization and insights into the care of residents in nursing homes.

The dataset contains 255 de-identified residents, along with the results of two separate interviews. The interviews consist of responses to the importance of preferences via the PELI-NH questions at two points in time, three months apart. The dataset also contains almost 1000 columns of additional resident information collected from the two interviews. This information holds data such as medical conditions, behavioral evaluations, and psychological symptoms. Although these columns were not used in current experiments, they may prove to be helpful in future work.

Experiments were run both with and without different lo-
<table>
<thead>
<tr>
<th>Metric</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>recall (TPR)</td>
<td>( \frac{TP}{TP + FN} )</td>
</tr>
<tr>
<td>specificity (SPC)</td>
<td>( \frac{TN}{TN + FP} )</td>
</tr>
<tr>
<td>precision (PPV)</td>
<td>( \frac{TP}{TP + FP} )</td>
</tr>
<tr>
<td>accuracy (ACC)</td>
<td>( \frac{(TP + TN)}{(TP + TN + FP + FN)} ) (UNIQUE/(PELI - MDS))</td>
</tr>
<tr>
<td>scope</td>
<td>( \frac{2 \cdot (TPR + PPV)}{(TPR + PPV)} )</td>
</tr>
</tbody>
</table>

where \( TP \) is number of true positives, \( FP \) is number of false positives, \( TN \) is the number of true negatives, \( FN \) is the number of false negatives, \( UNIQUE \) is the number of unique results in the consequents of a rule, \( PELI \) is the total number of PELI questions (72), and \( MDS \) is the total number of MDS questions (16).

Table 1: Evaluation Metrics

<table>
<thead>
<tr>
<th>Model</th>
<th>TPR</th>
<th>SPC</th>
<th>PPV</th>
<th>ACC</th>
<th>( f_1 )</th>
<th>scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>glm.1</td>
<td>0.8135</td>
<td>0.4021</td>
<td>0.7903</td>
<td>0.6985</td>
<td>0.7968</td>
<td>0.7968</td>
</tr>
<tr>
<td>glm.2</td>
<td>0.8319</td>
<td>0.3700</td>
<td>0.7845</td>
<td>0.7047</td>
<td>0.8044</td>
<td>0.9107</td>
</tr>
<tr>
<td>glm.3</td>
<td>0.8134</td>
<td>0.4022</td>
<td>0.7903</td>
<td>0.6985</td>
<td>0.7968</td>
<td>0.9107</td>
</tr>
<tr>
<td>glm.4</td>
<td>0.8021</td>
<td>0.4189</td>
<td>0.7919</td>
<td>0.6979</td>
<td>0.7953</td>
<td>0.9107</td>
</tr>
<tr>
<td>nomodel</td>
<td>0.9349</td>
<td>0.3385</td>
<td>0.7500</td>
<td>0.7212</td>
<td>0.8301</td>
<td>0.9107</td>
</tr>
</tbody>
</table>

Table 2: Evaluation Results by varying parameters: Confidence = 0.75, Support = 0.01, Lift = 1.1

Table 3: Evaluation Results for Generalized Linear Regression Model: varied size and confidence

<table>
<thead>
<tr>
<th>Model</th>
<th>TPR</th>
<th>SPC</th>
<th>PPV</th>
<th>ACC</th>
<th>( f_1 )</th>
<th>scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>glm.1</td>
<td>0.8135</td>
<td>0.4021</td>
<td>0.7903</td>
<td>0.6985</td>
<td>0.7968</td>
<td>0.9107</td>
</tr>
<tr>
<td>glm.2</td>
<td>0.7632</td>
<td>0.4202</td>
<td>0.7834</td>
<td>0.6087</td>
<td>0.7705</td>
<td>0.8929</td>
</tr>
<tr>
<td>glm.3</td>
<td>0.9007</td>
<td>0.2458</td>
<td>0.7666</td>
<td>0.7260</td>
<td>0.8280</td>
<td>0.8929</td>
</tr>
<tr>
<td>glm.4</td>
<td>0.7854</td>
<td>0.3846</td>
<td>0.7783</td>
<td>0.6758</td>
<td>0.7706</td>
<td>0.9107</td>
</tr>
<tr>
<td>glm.5</td>
<td>0.8319</td>
<td>0.3700</td>
<td>0.7845</td>
<td>0.7047</td>
<td>0.8044</td>
<td>0.9107</td>
</tr>
<tr>
<td>glm.6</td>
<td>0.7662</td>
<td>0.4169</td>
<td>0.7831</td>
<td>0.6699</td>
<td>0.7718</td>
<td>0.8929</td>
</tr>
<tr>
<td>glm.7</td>
<td>0.8886</td>
<td>0.2682</td>
<td>0.7696</td>
<td>0.7223</td>
<td>0.8241</td>
<td>0.8929</td>
</tr>
<tr>
<td>glm.8</td>
<td>0.8219</td>
<td>0.3195</td>
<td>0.7696</td>
<td>0.6872</td>
<td>0.7928</td>
<td>0.9107</td>
</tr>
<tr>
<td>glm.9</td>
<td>0.8134</td>
<td>0.4022</td>
<td>0.7903</td>
<td>0.6985</td>
<td>0.7968</td>
<td>0.9107</td>
</tr>
<tr>
<td>glm.10</td>
<td>0.7870</td>
<td>0.3919</td>
<td>0.7809</td>
<td>0.6797</td>
<td>0.7811</td>
<td>0.8929</td>
</tr>
<tr>
<td>glm.11</td>
<td>0.9005</td>
<td>0.2459</td>
<td>0.7665</td>
<td>0.7259</td>
<td>0.8279</td>
<td>0.8929</td>
</tr>
<tr>
<td>glm.12</td>
<td>0.8157</td>
<td>0.3368</td>
<td>0.7725</td>
<td>0.6867</td>
<td>0.7914</td>
<td>0.9107</td>
</tr>
<tr>
<td>glm.13</td>
<td>0.8021</td>
<td>0.4189</td>
<td>0.7919</td>
<td>0.6979</td>
<td>0.7953</td>
<td>0.9107</td>
</tr>
<tr>
<td>glm.14</td>
<td>0.7777</td>
<td>0.4126</td>
<td>0.7844</td>
<td>0.6786</td>
<td>0.7786</td>
<td>0.8929</td>
</tr>
<tr>
<td>glm.15</td>
<td>0.9014</td>
<td>0.2448</td>
<td>0.7665</td>
<td>0.7261</td>
<td>0.8282</td>
<td>0.8929</td>
</tr>
<tr>
<td>glm.16</td>
<td>0.8106</td>
<td>0.4314</td>
<td>0.7968</td>
<td>0.7071</td>
<td>0.8019</td>
<td>0.9107</td>
</tr>
</tbody>
</table>

gistic regression models and with different combinations of rule sets. Specifically, we experimented with a Generalized Linear Regression (glm) and Stochastic Gradient Descent (sgd) models. Due to space constraints, we only report on glm here. Each experiment was evaluated with 5-fold cross-validation with the best results obtained using a logistic regression model. In addition to measuring recall (i.e., true positive rate or TPR), specificity (SPC), precision (i.e., positive predictive value PPV), and accuracy (ACC), we measured scope, or the percentage of the preferences that the rule sets had the capability to predict. In this context we define each of these measures according to the definitions found in Table 1.

A sampling of the results of these experiments are shown in Tables 2 and 3. Table 2 shows five variants of applying our combined Apriori-Linear Regression approach where confidence was set to 0.75, support to 0.01, and lift to 1.1. While we have not fine-tuned the approach beyond four different parameterizations of the generalized linear regression (glm) model, we have been able to achieve an average recall (TPR) of about 81% and an average precision (PPV) of about 79%. Our overall accuracy (ACC), however, is closer to about 70% for the evaluations shown in the table. Also contained in the table are our F1 measures for each version, which show that in general we have a relatively balanced set of recommendations for preferences.

Table 3 shows a comparison of four variations of the glm model that vary a number of parameters including confidence, support, and lift. The numbered samples were configured with the following confidence, support, lift, size1, and size2 vectors, respectively: Sample 1: (0.75, 0.01, 1.1, 4977, 676), Sample 2: (0.8, 0.01, 1.1, 3276, 440), Sample

7: (0.8, 0.01, 1.1, 3276, 783), Sample 10: (0.75, 0.01, 1.1, 4977, 404). In addition, we vary the size of the rules in the associations between the MDS items and individual preferences as well as the category-based rule set sizes. In the table, “size1” represents the number of rules in the set of associations between MDS items and individual preferences, and “size2” represents the number of rules in the set of associations between MDS items and preference categories. In these executions, the average precision (PPV) is about 78% while the average recall (TPR) comes in at about 82%. Overall accuracy (ACC) remains about 70% with the F-measures ranging from .7705 to .8280.

The results shown in the tables reflect a relatively high scope (i.e., the ability of the approach to identify relevant rules). In developing the recommender, we have focused on positive results from precision and recall for crafting a recommendation set, as is consistent with the work of Lin (Lin 2000). As such, we are not interested in maximizing specificity (SPC), or rather, predicting lack of preference. Overall, we expect that as we continue to develop ComPASS that we will be able to continue to improve the rule-based recommender. In addition, we plan on exploring other potential methods for suggesting preferences.

Threats to validity. We evaluated the learning approach using a relatively small data set (n = 255) with an assumption that responses of individuals are independent and are not impacted by test-retest consistency. We also assumed that the responses of very important and somewhat important indicate “preference”, while other responses indicated “non-preference”. These assumptions were consistent with the findings of Van Haisma et al. (Van Haisma et al. 2014).

309
Conclusions and Future Work

The PELI-NH is increasingly being used as a guide for providing preference-based person-centered care in nursing home communities. The amount of time needed to complete the PELI-NH and the difficulty in managing data on paper are considerable barriers. Our machine learning mechanism has the potential to reduce the time needed for completing the PELI-NH interview while still incorporating important resident preferences. Our initial work has shown that we can achieve a reasonable rate of accuracy in providing recommendations on potential preferences for a resident with an acceptable rate of precision, a fruitful potential for providers.

As we work to deploy the ComPASS system to nursing home communities, our goal is to use the system as an active learning system, with the hope of gaining the benefits of collaborative filtering. In addition, we are interested in applying other recommendation strategies, including item-based collaborative filtering (Sarwar et al. 2001). Finally, we are investigating how other forms of machine learning, including classification (Zhang et al. 2017), can allow us to combine other information contained in electronic medical records with preferences to enable activities such as diagnosis.

References


Hahsler, M.; Buchta, C.; Gruen, B.; and Hornik, K. 2017. Title Mining Association Rules and Frequent Itemsets.


Simovici, D. A. 2012. Data Mining of Medical Data: Opportunities and Challenges in Mining Association Rules. In IALS.

