Applied Machine Learning in Care for Initial Orientation for Family Caregivers

Paul Wunderlich
inIT - Institute Industrial IT
OWL University of
Applied Sciences and Arts
Lemgo, Germany
paul.wunderlich@th-owl.de

Frauke Wiegräbe
inIT - Institute Industrial IT
OWL University of
Applied Sciences and Arts
Lemgo, Germany
frauke.wiegraebe@th-owl.de

Helene Dörksen
inIT - Institute Industrial IT
OWL University of
Applied Sciences and Arts
Lemgo, Germany
helene.doerksen@th-owl.de

I. Motivation

For 2019, the German Federal Statistical Office calculated 4.13 million people in need of care as defined by SGB XI. Of these, around 80% are cared for in the home environment [1]. Two-thirds of family caregivers are employed, primarily immediate family members [2]. Due to the demographic aging of society, the number of people in need of care will continue to rise. The German Federal Ministry of Health (BMG) anticipates 6.5 million people in need of long-term care in Germany by 2050 [3]. However, there is already a shortage of skilled nursing staff today. And future scenarios indicate that, due to the shortage of skilled workers in all forms of nursing care, people in need of help and care can no longer be adequately cared for by professional caregivers. Accordingly, support for family caregivers is needed to ease the burden and improve the balance between work and care.

II. Approach

Therefore, we have developed the Digital Case Manager to serve as an initial orientation support for family caregivers. The approach to achieve this objective is illustrated in Fig. 1.

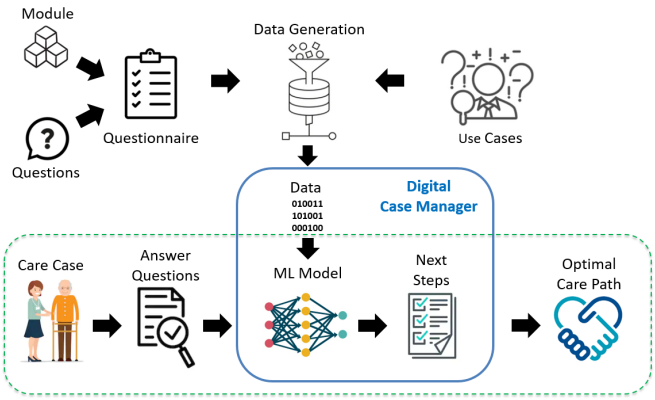

Fig. 1. Approach Digital Case Manager

This work was developed in the project work & care, which is funded by the Ministry of Economic Affairs, Innovation, Digitalization and Energy of the State of North Rhine-Westphalia (MWIDE NRW) under grant number 34.EFRE-0300198.

The core of the approach is the Machine Learning Model (ML Model), which learns the relationship between the care situation and the next steps. Learning the ML Model requires data, which is very hard to obtain in the world of healthcare. Therefore, it was necessary to generate artificial data.

Once the ML Model has learned, it can enter the application case, which is outlined in green dashed lines in Fig. 1. The family caregiver can easily fill out a questionnaire in case of a new care case. The ML Model recommends next steps to the family caregiver based on the responses, which thus contributes to an optimal care path.

A. Data Generation

The generated data should be as close as possible to reality in order to capture a care situation also by means of data. For this purpose a questionnaire has been developed. This questionnaire is composed of seven modules that are already commonly used by the Medizinischer Dienst der Krankenkassen (MDK) [4]. The seven modules are Mobility, Cognitive and Communication Skills, Behavioral and Psychological Problems, Self-Care, Dealing with Demands Caused by Illness or Therapy, Organizing Daily Life and Social Contacts, and Organizing Care. There are four to eleven questions per module.

In addition to the questionnaires, a variety of use cases in care were also created by a healthcare expert. There are currently about 16 use cases and more are being worked on. So far the use cases are restricted to the field of stroke, care giving and dementia. These use cases serve as the foundation for random-based data generation. For each of these use cases, the possible answer options and corresponding next steps were defined. Subsequently, an algorithm was developed, which randomly selects one of the use cases. Based on the selected use case, the questionnaire is filled out randomly with the allowed answer choices for that use case. It also assigns the appropriate next steps to the questionnaire.

The developed algorithm for data generation can produce a number of n answered questionnaires based on the aforementioned use cases. For the training of the ML Model, a data set with 10000 answered questionnaires and the corresponding next steps was generated to have sufficient data.

B. Learning the Model

The ML Model is the core of the Digital Case Manager. It learns the correlations between the answers of the questionnaire and the next steps to be recommended. This is a multi-label classification problem, since none, one or more steps can be recommended per answered questionnaire [5].

For the multi-label classification problem, a variety of classification algorithms such as k-nearest neighbor, decision tree or random forest can be considered. For the ML Model in the demonstrator a random forest classifier was applied.

Input data for the ML Model are 10000 answered questionnaires, as well as the corresponding next steps. First, the categorical answers were transformed into integers. Furthermore, as usual, the entire data set was split into a training data set (75%) and test data set (25%). Using the training data set, the random forest classifier is trained. There is no restriction on the number of branches or the number of cases for a leaf. Only the amount of decision trees for the random forest classifier was restricted to 20 in order to reduce the training time.

When evaluated on the test data, the ML Model achieves an accuracy of 100%. Likewise, the Hamming loss value [5] is 0. However, the still limited number of use cases in a data set of 10,000 questionnaires must be taken into account. Presumably, this is an over-adaptation of the ML Model to the existing use cases. Therefore, in the upcoming steps we will try to generalize the ML Model further.

III. Demonstrator

As a proof of concept a demonstrator was developed. A part of this demonstrator can be see in Fig 2.

Fig. 2. Demonstrator of the Digital Case Manager

The demonstrator consists of three parts introduction, modules and analysis. The introduction is explaining for new users what the Digital Case Manager is and how to use it. Further information about the modules and the questions can be obtained by opening an expander via the plus symbol.

After the introduction part comes the modules. Each module is an independent expander and can be opened or closed according to the user's preference. If you open the expander, you will see the questions associated with the module.

The final part is the analysis. After all questions were answered, the user can press a button called "Start Analysis" to begin the inference of the ML Model. When the inference is finished, a list of recommended next steps appears. In theory this next steps could be linked to offered services by health professionals to support the family caregivers further.

IV. Outlook

As mentioned in Subsection II-B, the overfitting of the ML Model should be reduced to allow a better generalization to other care situations. In addition to the already mentioned data augmentation, methods such as classical cross-validation, feature reduction and regularization will be applied. For data augmentation, the previous use cases are slightly modified to obtain further variations of them. This leads to a greater variety in the data set. In addition, a feature reduction by a feature selection or a dimensionality reduction via PCA is considered. In feature selection, individual questions could be combined or eliminated altogether. Likewise, pruning is intended to be applied, e.g., to limit the number of branches for each decision tree in the random forest.

In general, the Digital Case Manager can also be extended with a feedback system in order to improve independently. This means after someone has used the Digital Case Manager and thus received recommendations for next steps, an evaluation of the recommendations is made by the user or by professional health experts. They correct the recommendations and this new record could then be stored in a database. After a large enough amount of new data is received, the model would be relearned based on the previous training data and the newly added data records. As a result, it would also allow learning the correct recommendations for next steps for other care situations that have not yet been included in the training data.

In addition, further development of the tool as support for professional case managers would also be conceivable.

References

[1] Destatis - Statistisches Bundesamt, "Pflegestatistik - Pflege im Rahmen der Pflegeversicherung - Deutschlandergebnisse - 2019," 2020.
[2] U. Schneekloth, S. Geiss, and M. Pupeter, "Abschlussbericht - Studie zur Wirkung des PflegeNeuausrichtungs-Gesetzes (PNG) und des ersten Pflegestärkungsgesetzes (PSG I)." 2017.
[3] Bundesministerium für Gesundheit (BMG), "Zahlen und Fakten zur Pflegeversicherung," 2022.
[4] Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e.V., "Richtlinien des GKV-Spitzenverbandes zur Feststellung der Pflegebedürftigkeit nach dem XI. Buch des Sozialgesetzbuches," 2021.
[5] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," *International Journal of Data Warehousing and Mining (IJDWM)*, vol. 3, no. 3, pp. 1–13, 2007.