Virtual Coaching Activities for Rehabilitation in Elderly

Call: H2020-SC1-2016-2017

Grant Agreement Number: 769807

D4.1 Clinical Knowledge Representation Framework Design

Extended summary

This project vCare has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 769807.

The deliverable deals with the design decisions for the knowledge representation framework (KRF) of the vCare system. The major outcome is aligning the medical domain ontology specified with knowledge integration and learning requirements the vCare knowledge layer in order to consolidate the medical domain terminology with the technical perspective. This is required in order to get a common understanding between medical- and technical partners, which is essential for developing the baseline ontology. The latter will implement the basic decision-oriented structure we propose in this deliverable, which enables the integration of clinical pathways. The latter reuse modeled concepts of the baseline ontology and use them to structure the patient treatment process. Machine Learning-based tasks can then be triggered within clinical pathways to recommend activities, which are also represented in the resulting domain ontology.

Based on clinical requirements of the vCare solution, use case narratives as well as a regular, structured communication with medical experts, we derive exhaustive competency questions that stipulate what functionality the KRF has to provide. The competency questions thus define the central requirements for the interface between the technical- and medical partners of the vCare project and, among others, cover aspects about the area of expertise of physicians, possible repercussions of medications, available devices in a smart home or available algorithms for automatic activity recommendations.

We focus on two research fields for our survey, namely **structured knowledge representation** and **machine learning (ML)**. A structured knowledge representation is central for defining a language between medical and technical partners as well for integrating heterogeneous data sources of vCare. However, the representation has to be compatible and sufficiently detailed for all learning tasks in WP4, which requires us to find suitable ML paradigms for virtual coaching. The survey is conducted by using the competency questions as baseline for the project's requirements with respect to the KRF and elaborate on the eligibility of technologies to fulfil them.

We first survey **structured knowledge representation frameworks**, such as the Resource Description Framework (RDF) or Labeled Property Graphs (LPGs), and **engage in a critical analysis about their eligibility with respect to the KRF**. RDF – potentially used in combination with relatively heavy-weight modeling constructs based on RDF Schema (RDFS) or the Web ontology language (OWL) – is sufficient for integrating the heterogeneity of information we face in vCare (e.g. clinical- or personal patient information, sensor data or pathway templates), but needs to be paired with a logic-based formalism to enable defining expert-based rules (i.e. medical guidelines by clinicians). The latter are essential for constraining the system based on the medical experience of clinicians and for creating the baseline for learning personalized patient recommendations. We argue to use Notation3 (N3) Logic, which enables to model implications of semantic annotations modeled in RDF. The combination of RDF and N3 Logic is sufficient for the KRF, as all relevant information can be sustainably modeled, easily retrieved and eventually transformed into suitable representations for learning-related tasks.

To additionally ensure that the KRF is sufficiently rich for subsequent learning tasks (e.g. personalizing the activities for a patient or generalizing among patients) we also survey and assess suitable **ML paradigms**. We therefore compare the supervised ML setting, where all data points (i.e. individual situations within the clinical pathway) are fully labeled (i.e. the optimal personalized activity for a patient is always known) before deriving a predictive model, with more restrictive settings such as Reinforcement Learning (RL), where only a subset of possible outcomes (i.e. only the quality of the recommended activity) are available for the learning algorithm. While standard ML settings such as **Active Learning (AL) are useful to interactively enhance the training process of the vCare system, RL is especially suitable for medical pathway recommendations**, where dependent activities (i.e. choosing an activity influences the subsequent one) are gradually adapted to an individual patient's preferences and medical records. We additionally investigate possibilities to deal with the initial starting period of the KRF, where no available data is available for learning (i.e. no activities have been recommended to patients), and argue to use modeled expert rules in order to teach the agent a global (i.e. not yet personalized) internal model to deal with the patient. As a long-term objective of WP4 is to **derive insights into how to generalize among different patients with varying degrees of similarity**, we further explore advances in the fields of ML with missing features, which are suitable to reuse and extend learned behavior of already past patients. The experience can then – in addition to the simulated global baseline – be used to accelerate the personalization of the system to patients.

Based on our survey about knowledge modeling approaches and ML paradigms (conducted on the grounds of the derived competency questions for the vCare KRF), **we articulate the need for an abstract, structured representation of states the virtual coach is acting in, which can be directly grounded with respect to a medical pathway** (i.e. the initial vCare schema). The abstract state representation integrates heterogeneous information about the patient and its sensor-enriched home, as well as actions the virtual coach can perform, which are related to activities of the pathway modeled in different abstraction degrees. We also elaborate on how to add necessary information about the resulting quality of an action (i.e. eventual impact of an activity on the patient's life), which is essential for gradually improving the virtual coach's decisions with RL. We finally present first schema of the baseline ontology based on the stipulated requirements. The central next step comprises implementing an initial proof-of-concept given the baseline ontology, where an exemplary (and potentially fictive) patient is (1) modelled with concepts and properties of the ontology, (2) the semantically patient is situated in an exemplary clinical pathway, (3) structured rules collaboratively derived with physicians are applied in respective states of the pathway and (4) the latter's appropriateness is assessed and augmented with initial personalization approaches with ML.