Computers in Human Behavior 30 (2014) 698–707

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Towards healthcare business intelligence in long-term care An explorative case study in the Netherlands

Marco Spruit *, Robert Vroon, Ronald Batenburg

Utrecht University, Department of Information and Computing Sciences, Princetonplein 5, 3584 CC Utrecht, The Netherlands

ARTICLE INFO

Article history:
Available online 24 August 2013

Keywords:
Healthcare business intelligence
Knowledge discovery
Data mining
Crisp-dm
Long-term care

ABSTRACT

This research contributes to the domain of long-term care by exploring knowledge discovery techniques based on a large dataset and guided by representative information needs to better manage both quality of care and financial spendings, as a next step towards more mature healthcare business intelligence in long-term care. We structure this exploratory research according to the steps of the CRoss Industry Standard Process for Data Mining (CRISP-DM) process. Firstly, we interview 22 experts to determine the information needs in long-term care which we, secondly, translate into 25 data mining goals. Thirdly, we perform a single case study at a Dutch long-term care institution with around 850 clients in five locations. We analyze the institution's database which contains information from April 2008 to April 2012 to identify patterns in incident information, patterns in risk assessment information, the relationship between risk assessments and incident information, patterns in the average duration of stay, and we identify and predict Care Intensity Package (ZZP) combinations. Fourth and finally, we position all data mining goals in a two-by-two matrix to visualize the relative importance of each goal in relation to both quality of care and financial state of care institutions.

1. Introduction: long-term care in the Netherlands

This research uncovers the relatively unexplored long-term care sector in the Netherlands and the applicability of knowledge discovery techniques as a next step towards the strategic goal of more mature healthcare business intelligence (Mettler & Vimarlund, 2009). During the last fifteen years knowledge discovery has evolved in the healthcare domain from predicting epidemics (Prather et al., 1997) to a broad spectrum of data mining applications (Koh & Tan, 2011). The data that are being captured by healthcare organizations is enormous in size and, therefore, a treasure for data analysts (Lucas, 2004). Koh and Tan (2011) argue that the use of knowledge discovery techniques have become increasingly popular, if not essential for healthcare organizations. However, within the healthcare domain the long-term care sector has been left out in research, so it seems. Up until now.

In 2008 the following general policy for long-term care in the Netherlands was formulated by Mot, Aouragh, Groot, and Mannaerts (2010):

> *"To ensure that for persons with a long-term or chronic disorder of a physical, intellectual or psychological nature, care of good quality is available and that the cost level of this care is acceptable to society."*

Long-term care is care for people with a long-term or chronic disorder, where the chronic disorder can be either of a physical, intellectual or psychological nature. The policy contains two goals for long-term care, which are 'care of good quality' at an 'acceptable cost level'. These goals apply to all care institutions that deliver long-term care. In order to support these goals, care institutions should have insight in the quality and financial state of the internal organization.

Electronic Client Record (ECR) software is used to keep track of the quality and financial state of the internal organization. All the information stored in ECR software is client related, which makes the client the central entity. ECR software contains personal details, medical information, financial information, production information, care plan, incidents, documents, treatment plans, presence and absence. One should note that at least in the Netherlands, ECR software is different than Electronic Patient Record (EPR) software, which is mainly used in hospitals. One could argue that the two are closely related, but one overarching system is unfortunately not yet in place in the Netherlands at this moment. Both ECR and EPR systems are tailor-made for the sector in which they are used.

Long-term care within the Netherlands has become one of the biggest expenses at this moment for the Dutch government, consuming no less than 38% of the total healthcare budget (Schäfer et al., 2010). The expenditures of the Exceptional Medical Expenditures Act (AWBZ) alone have steadily increased from €14 billion in 2000 up to €27 billion (budgeted) for 2012, which is a doubling in just 12 years (Ministerie van Volksgezondheid & Welzijn en Sport,

* Corresponding author. Tel.: +31 30 253 3708.
E-mail address: m.r.spruit@uu.nl (M. Spruit).

0747-5632/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.chb.2013.07.038

2011a, 2011b). According to the population forecast of the Dutch Central Bureau for Statistics (CBS) in 2010, the number of elderly in the Netherlands will increase from 2.4 million to 4.6 million in 2040 (Duin & Garssen, 2010). This implies that a growing number of people will need some form of care, which is also addressed in (Mot et al., 2010). Also, as life expectancy increases, more people will need some form of care for a longer period of time (Duin & Garssen, 2010). From a governmental perspective it is therefore important to increase the efficiency of long-term care in order to avoid that long-term care expenditures will become uncontrollable. The Ministry of Health, Welfare and Sport (2011) also addressed the importance of management information on quality:

> *"In order to supply care of good quality, it is essential that the care institution is managed well and that the institution has permanent information about the quality of the institution."*

From the perspective of long-term care institutions it is important to have a better insight into the internal organization. At this moment ECR software is mainly used to support the core process of care institutions, which is the delivery of care. The ECR systems contain a lot of valuable data that are nowadays used to receive budget from the government and, moreover, to justify their existence. ECR data are also used on an individual basis to support a client in the best possible way. However, at this moment, the data that are collected in ECR software are not fully exploited yet. If only the collected unstructured data could be made explicit to some extent, the information arising therefrom could then be used to improve the efficiency and effectivity of the long-term care processes (Feelders, Daniels, & Holsheimer, 2000).

Therefore, this research aims to firstly discover the information needs of long-term care institutions, and secondly, the extent to which the desired information needs can be made explicit based on a wide range of already proven knowledge discovery techniques, which is captured in the following research question:

> *"How can knowledge discovery techniques support Dutch long-term care institutions to manage their internal organization?"*

Note that we pursue a meta-algorithmic approach in this research. Instead of focusing on developing new algorithms to mine for new insights, we aim to uncover new knowledge by reusing already proven algorithms in new configurations and application domains. In general, a meta-algorithmic knowledge discovery approach provides an informatics perspective onto data analytics research by modelling knowledge discovery *technology* selection to facilitate *process* analysis to improve *people*'s performance.

The remaining part of this paper is structured as follows. Section 2 provides a description of the data material under investigation and the research approach. In Section 3 the Top 5 data mining goals are modelled and their outcomes are elaborated upon. An overall interpretation of these findings is provided in Section 4. Section 5 contains our main conclusions and further discussion of this work.

2. Material and methods: a data-driven case study

Knowledge discovery techniques embed data mining models within an overarching application process to help discover new interesting knowledge from unstructured data. However, Knowledge discovery and Data mining are often used as synonyms by many researchers (Kurgan & Musilek, 2006). In this research data mining is used as one step in an encompassing knowledge discovery process (Cios, Pedrycz, & Swiniarski, 2007). We have structured this exploratory research according to the steps of the CRoss Industry Standard Process for Data Mining (CRISP-DM) process (Chapman et al., 2000) after researching and comparing CRISP-DM with three other knowledge discovery processes: the Knowledge Discovery in Databases (KDD) process by Fayyad, Piatetsky-Shapiro, and Smyth (1996), the Sample Explore Modify Model Assess (SEMMA) method as referred to in (Azevedo & Santos, 2008) and the Three Phases Method (3PM) by Vleugel, Spruit, and van Daal (2010). CRISP-DM is a clearly described process and has been widely used for knowledge discovery processes ever since its inception, making it the 'de facto standard' in the field, for developing data mining and knowledge discovery projects (Giraud-Carrier & Povel, 2001; Harding, Shahbaz, Kusiak, & Srinivas, 2006; Onwubolu, 2009).

The CRISP-DM process consists of the following six phases: business understanding, data understanding, data preparation, modelling, evaluation and deployment. These phases are also employed in this research.

2.1. Business understanding

Multiple unstructured in-depth interviews have been performed to create an understanding of the long-term sector. Unstructured in-depth interviews are appropriate when the richness of detail through clarification of questions and answers is to be ensured. Yin (2009) states that open interviews are the best way to discover explorative information.

Experts from different perspectives and various organizations have been interviewed in order to create a reliable and complete picture of the information needs in the entire long-term care sector. Our experts represent information needs of Nursing homes, Care homes and Home care (VVT), Mental care (GGZ) and Disability care (GZ). In total 22 experts were interviewed in 18 sessions. Table 1 shows that we interviewed eight experts from the board of directors level, seven experts from the management level and seven experts from stakeholder positions. The concept of Valuation in the left-most column will be explained in Section 3.

Information needs are the result of the interviews, which are consequently translated to data mining goals during this phase. The use of information needs is slightly different than the business goals as prescribed in the CRISP-DM method. In our view information needs are more elaborate than business goals. The data mining goals are input for the next phase: data understanding.

2.2. Data understanding

To allow proper modelling, it is important to understand the gathered data. We performed a single case study, which means that the data that has been gathered by one long-term care institution are used to do the modelling. The care institution has currently around 850 clients, both intramural and extramural, and has been working with the ResidentWeb ECD system since 2008. Table 2 shows the codes, colours and number of beds per location of the long-term care institution in our case study.

In order to use the data, all personal information such as names, BSN numbers, and addresses had to be deleted first. This research was commissioned by the software development company behind ResidentWeb, and the care institution under investigation is closely involved in the on-going development of the ECR software.

Table 1
Overview of interviewed experts in the long-term care sector.

Type of interviewee	Number of experts	Number of sessions	Valuation
Board of directors/Director	8	8	10
Management	7	5	6
Stakeholders	7	5	3
Total	22	18	–