FRAMEWORK FOR COMPUTER AIDED ANALYSIS OF MEDICAL PROTOCOLS IN A HOSPITAL

Rene Schult, Pawel Matuszyk and Myra Spiliopoulou
Otto-von-Guericke-University, Universitätsplatz 2, D-39106 Magdeburg, Germany

Keywords: Knowledge discovery from medical protocols, Knowledge discovery from anesthesia protocols, Knowledge discovery, Data mining, Healthcare information systems.

Abstract: We study the potential of analyzing medical protocols with data mining methods for resource planing.
Background. Medical protocols can be exploited in several resource planing applications, such as optimizing occupancy of surgery rooms or scheduling teams for surgery operations. Literature has identified many variables that can be used to predict resource demand; some of them can be extracted from medical protocols.
Contribution. We propose a high-level framework for knowledge discovery from medical protocols, and present a first instantiation in a German hospital. We report on the findings of this instantiation for the task of predicting surgical room occupancy time.

SCIENCE AND TECHNOLOGY PUBLICATIONS

1 INTRODUCTION

Hospitals are increasingly facing the demand for efficient resource planing, not least in response to economic recession and demographic evolution. Of particular interest is the efficient management of resources needed for expensive types of treatment, such as intensive care, and of resources with high demand, such as surgical rooms. Eijkemans et al. point out that more than 60% of hospital patients undergo some surgical treatment (Eijkemans et al., 2010). So, there is need for methods for predicting and optimizing occupancy of surgical rooms and intensive care units.

Medical protocols encompass information that can be used to improve room planing. Eijkemans et al. have identified several predictive variables for ”operation time” (Eijkemans et al., 2010); some of these variables are routinely recorded in anesthesia protocols. In this study, we consider these protocols for the prediction of *surgery room occupancy time (”SRO time”)*, which we define as the elapsed time between the entry of the patient to the operation room until the exit moment. This is equivalent to ”operation time” in (Eijkemans et al., 2010), but we pertain to the more explicit ”SRO time”, because in some hospitals (including the one we studied) patients occupy the surgery room until they wake up from anesthesia.

Predictive variables can serve as aid to resource planers. However, the variables recorded vary among hospitals. Recording all desirable variables may require process redesign and thus incur additional costs. Hence, it is necessary to exploit the predictive power of available variables to the largest possible extent.

To this purpose, we propose a lightweight framework for knowledge discovery from medical protocols and report on its use for prediction of resource demand. The overarching idea is that the framework should allow a reporting or prediction task to be *plugged* into existing processes, without requiring process modifications nor additional activities from the medical staff. We report on a first instantiation of our framework in a hospital for the analysis of intensive care unit protocols and anesthesia protocols. We show how knowledge discovery from anesthesia protocols can lead to better prediction of SRO time; the full report is in (Schult et al., 2011).

In section 2 we discuss related work. In section 3, we describe our framework at an abstract level; in section 4 we present its instantiation in a hospital on two types of medical protocols. Findings on the prediction SRO time from anesthesia protocols are summarized in section 5. Section 6 concludes our study with lessons learned and planed next steps.

2 RELATED WORK

The importance of information technology in the health care industry is reflected in increasing investments in appropriate IT systems (Wilson and Tulu,

Schult R., Matuszyk P. and Spiliopoulou M..
FRAMEWORK FOR COMPUTER AIDED ANALYSIS OF MEDICAL PROTOCOLS IN A HOSPITAL.
DOI: 10.5220/0003776702250230
In *Proceedings of the International Conference on Health Informatics* (HEALTHINF-2012), pages 225-230
ISBN: 978-989-8425-88-1
Copyright © 2012 SCITEPRESS (Science and Technology Publications, Lda.)

2010). Avison and Young point out that decision support systems are one important application in health care information systems (Avison and Young, 2007), while Combi et al. stress the importance of timestamped data for reasoning, e.g. for clinical diagnosis and for devising care plans (Combi et al., 2010).

Reasoning, prediction and other forms of decision support require generic frameworks that allow for the particularities of each hospital. For example, consider prediction of surgery room occupancy: Dexter et al. report that the average duration of a given surgery between the second-fastest and the second-slowest clinic they investigated may differ by up to 50% (Dexter et al., 2006). This implies that predictive models must be learned for each hospital, on the data recorded in this hospital. Accordingly, we propose a framework at a high level of abstraction, and we show how its instantiation in a German hospital lead to the exploitation of medical protocols for prediction.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

3 FRAMEWORK FOR MEDICAL PROTOCOLS IN DECISION SUPPORT

We propose a framework for decision support on the basis of medical protocols. The main purpose of these protocols is to store all medical activities that refer to a patient of a hospital. They are essential for the patient's treatment, but also for accounting and billing, for resource management and planing, and for auditing. They can also be used for scientific research, studies on new treatments and medication, and for the analysis and optimization of internal processes.

In our framework, we consider medical protocols for decision support in resource management, and anticipate two core tasks - reporting on resource use, and predicting resource use. For prediction we focus on (supervised) data mining methods. All tasks we anticipate are depicted in Figure 1 and described below.

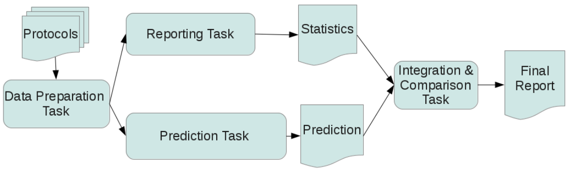

Figure 1: Framework for reporting and analyzing medical protocols for decision support.

There are different types of medical protocols that can be used as input to our framework, such as anesthesia protocols or intensive care unit protocols; they are compiled by different members of the medical staff, are incorporated in different processes, and intended for different recipients. This affects the contents of the protocols, the number of variables recorded and the format used. Some protocols have the form of a single record per patient, while others adhere to the entity-attribute-value model (Stead et al., 1983) and consist of several records per patient – one per variable of interest for this patient. These differences influence the prediction task, and must be dealt with during data preparation.

Data Preparation Task. This task involves algorithms that prepare the data for reporting and for data mining. Conventional data preparation tasks include finding and handling missing values and errors in the data, detecting correlations between variables, and determining the target variable for the subsequent prediction task, depending on the problem at hand. There are many statistical tools and also mining algorithms available for such purposes; in our experiment (Section 5) we report on those we used for our instantiation to predict SRO time.

Less conventional data preparation is needed to transfer data from the entity-attribute-value model to a format that can be used by mining algorithms. We elaborate more on this issue in Section 4.

Reporting Task. This task involves utilities for data querying and summarization, as provided conventionally with a database management system or data warehouse. The concrete information to be reported depends on the objective of decision support. For example, optimizing surgery room occupancy (SRO) requires an overview of SRO times for different variables, such as type of surgical treatment, patient age etc. Relevant variables are listed in (Eijkemans et al., 2010), but the contents of the output report *Statistics* depend obviously on the variables recorded in the protocols used. In our instantiation (Section 4), we used anesthesia protocols.

Prediction Task. This task involves data mining algorithms, as provided by commercial suites or open source tools (free for research purposes). Prediction can run independently of the reporting task, but often reporting precedes prediction: some reports can provide insights on predictive variables.

The algorithms used depend obviously on the concrete objective for decision support. In our instantiation, we wanted to predict discrete SRO time slots ("bins") rather than exact SRO time, since surgery rooms are rather occupied in time slots than up to the minute. is more interessting as the prediction of the concrete time. For the prediction of values of a continuous variable, regression methods should be used, while the prediction of discrete values requires classification algorithms.

The prediction task involves specification of the target variable and of appropriate evaluation criteria. Prediction requires training, tuning and comparing several learners, before a learner (or an ensemble) is chosen to be used for the prediction over unknown data. In our experiment (cf. Section 5), we compared several classifiers, but we also compared data preparation algorithms, because they turned to influence classifier performance.

Integration and Comparison. This task involves placing the report (from the reporting task) and the results of the prediction task together, including visualizations. The tools needed here are usually part of the suites appropriate for the Prediction, resp. the Reporting task. Integration serves foremostly the juxtaposition of findings acquired from reports via simple statistics and querying, and those acquired by machine learning. The final report serves as basis for a human decision maker for planing, or as input for a simulation tool that may consider different resource planing scenaria. In the long term, comparison also concerns the juxtaposition of a predictor learned some time back with newer data; changes in healthcare processes or external factors may require re-learning of predictive models.

In the next two sections, we discuss an instantiation of our framework in a German hospital for intensive care unit protocols and anesthesia protocols, and summarize our insights from applying our framework to predict SRO time.

4 INSTANTIATION OF THE FRAMEWORK IN A GERMAN HOSPITAL

We present an instantiation of our framework for knowledge discovery in a German hospital. We study two types of medical protocols recorded from 2007 till 2009, namely intensive care unit protocols and anesthesia protocols. As pointed out in section 3, such medical protocols are typical inputs to our framework. We first describe briefly the challenges and potentials of using intensive care unit protocols for knowledge discovery. Then, we focus on anesthesia protocols, which we use for the prediction of surgery room occupancy time (SRO time, cf. Section 1). Our experiment on these protocols is described section 5.

4.1 Intensive Care Unit Protocols

An intensive care unit protocol contains data on the treatment of a patient in an intensive care unit. Such data include diagnosis and medication. Special emphasis is put on the patient's vital signs, e.g. body temperature, pulse and blood pressure, which have to be monitored constantly.

The data stored in intensive care unit protocols are very complex: depending on the patient and the treatment, different variables must be stored. Given the limitations of database systems with respect to the maximum number of columns in a table, Stead et al. proposed to use the *entity-attribute-value* model (Stead et al., 1983), where the *entity* is the patient identifier, and *(attribute, value)*-pairs contain the specific variables and values to be stored for this patient. This model was used for the intensive care unit protocols in our instantiation, whereby an identifier and a timestamp was added to each record.

The entity-attribute-value model is not appropriate for data mining. The reason is that a data mining algorithm requires that all values belonging to one instance (here: one patient of the intensive care unit) be in one record, so that the algorithm accesses and analyses all attributes of the record together. In contrast, under the entity-attribute-value model the values belonging to one patient are spread over time and stored separately, as if they belonged to different stays and treatments of the patient. Hence, the data of each intensive care unit protocol must be collected and integrated into a single record. However, since there is a large number of possible attributes but only a few are recorded for each specific patient, the density of information for each patient could be too low for learning.

One solution to this problem could be the following: group protocols for which the same attributes have been recorded (preparation task), perform reporting and knowledge discovery on resource demand, such as bed occupation or drug utilization, *for each group* separately (reporting task / prediction task), and then integrate the findings of the groups into a report (integration and comparison task).

4.2 Anesthesia Protocols

An anesthesia protocol contains an exact description of all anesthetic activities performed during a surgical treatment. Among the data contained in such a protocol are involved personnel, important time points (e.g. time point of the incision and of the end of a surgery), and data about medication.

In the hospital of our study, these data were recorded by an anesthetist during the surgery, using the Anesthesia Information Management System (AIMS) NarkoData. NarkoData contains all data associated to anesthesia during the whole anesthesia process, including drugs, laboratory results, relevant vital signs, as well as data on the attributes specified by the German Society of Anesthesiology and Intensive Care Medicine [1](DGAI, 1993). NarkoData also contains data from the hospital information system, including data on patients and medical staff. Patient's attributes are age, weight and body size, disease according to the ICD Classification (WHO, 2011), physical status according to the ASA-Classification, and type of anesthesia. Information on medical staff is limited to the identifiers of surgeons and anesthesists. This allows us to distinguish among staff members without disclosing personal information.

The time points recorded in the anesthesia protocols are very important: they can be used to predict the duration of future, similar surgical treatments. In the next section, we present the findings of the framework's instantiation in the German hospital for the prediction of SRO time using anesthesia protocols. We discuss the concrete activities of

data preparation (Section 5.1) and learning (Section 5.2), including results and lessons learned. Section 5 is a summary of (Schult et al., 2011), where all details can be found.

5 KNOWLEDGE DISCOVERY EXPERIMENT ON ANESTHESIA PROTOCOLS

The goal of knowledge discovery from anesthesia protocols in the hospital under study was to learn a model that predicts the SRO time of future surgery treatments better than the current baseline. This involved an instantiation of the data preparation task and of the prediction task (cf. Figure 1). In our experiment, we consider three discretization methods for data preparation, and four classification algorithms for prediction, and we compare the quality of the models learned by the twelve combinations.

Our target variable is a discretized version of SRO time. As described in section 1, we define "SRO time" as the elapsed time between entry and exit of the patient to/from the surgery room. We thus cover hospitals that do not have a separate room where patients stay after surgery until they wake up.

We discretize SRO time for learning, because room occupancy plans deliver *time slots* (equiv. *bins*) rather than exact values. The size of the bin affects prediction, so we experimented with different binning methods in the data preparation task described below.

For the evaluation, we compare our models to a baseline predictor ICDavg: for each class of surgery according to the ICD classification (WHO, 2011), ICDavg finds all protocols refering to treatments of this class, adds their SRO times and computes the average. Then, for each anesthesia protocol in the testset, ICDavg identifies the ICD class of the treatment and returns the corresponding SRO time average.

To compare our models to ICDavg, we map back the predicted bin of SRO time to the mid value of the bin (e.g. a bin of 90 min is mapped to 45 min). Then, we define a function that computes the *Cumulation of Absolute Differences between true and predicted SRO time* (SROCD) for the whole period of study. The lower the SROCD, the higher is the model's quality.

5.1 Data Preparation

The data preparation task in our instantiation for SRO time prediction involved following activities: (a) incorporation of the surgeonID in each record, (b) computation of the SRO time per record and (c) discretization of the SRO time into a fixed set of intervals/bins, so that the SRO time bin becomes the label to be predicted in the prediction task.

The incorporation of the surgeonID into the anesthesia protocols is important because this variable is predictive (Eijkemans et al., 2010). However, some surgical treatments involve more than one surgeon, so that the incorporation of the identifier transformed one protocol (from the originally 33,862 anesthesia protocols) into multiple records, whenever multiple surgeons were participating. This was not a problem for our experiment, because the duplicates were considered both by the baseline and by each learner. However, in a real scenario, a more elaborate approach is needed for the incorporation of multiple identifiers of surgeons into a single record.

The computation of the SRO time of each protocol is performed by using the timestamps and is fairly straightforward. For discretization, one may provide the target number of bins as input, or consider methods that both estimate this target number and do the binning. Since there are cases where the latter type of methods is not of advantage (see section 5.2), we propose following approach, under the assumption of representative data, to specify the target number of bins: generate bins for different input numbers, learn a classifier for each number of bins, compute the SROCD, and identify the moment of saturation of the SROCD curve.

To test this approach we have experimented in (Schult et al., 2011) with the discretization methods (i) *Equal Width Interval Binning* (EWIB) that parti-

[1] DGAI: Deutsche Gesellschaft für Anästhesiologie und Intensivmedizin

tions the SRO times in the anesthesia protocols into bins of equal size, and (ii) K-Means that groups similar SRO times into K clusters, whereby each group becomes a bin. Unlike EWIB, K-Means builds bins that are not necessarily of equal width.

Figure 2 depicts the SROCD curve of a J4.8 decision tree classifier upon bins computed with EWIB: the curve does converge. The saturation is on 50 bins, the same value was found when using K-Means instead of EWIB. Hence, we can use this experimental approach to determine the number of bins, provided that the data set is representative. Anesthesia protocols are recorded anyway for each surgical treatment, so representative samples can be drawn from them.

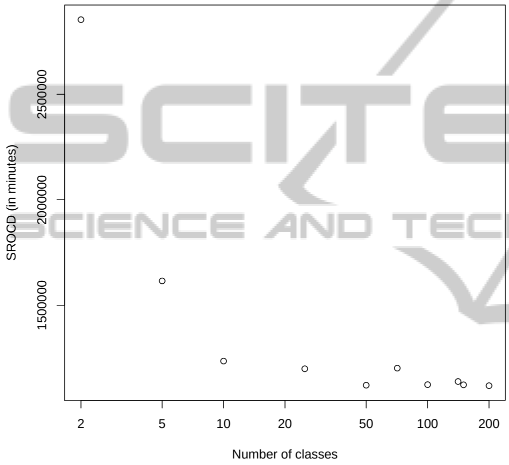

Figure 2: SROCD (in minutes) of a decision tree learner (J4.8) for $K = 2, \ldots, 200$ bins; the contents of a bin for each value of K was computed with EWIB.

5.2 Prediction

In our instantiation, prediction translates into a classification task, because the target variable (SRO time) has been discretized. It is essential to define a baseline and to study how different learners behave in comparison to this baseline. Then, for the operative task of prediction on new, unlabeled data, the best model or an ensemble of learned models should be used.

As baseline we used the ICDavg described at the beginning of this section: it computed an expected accumulated SRO time of 1,279,567 minutes. We compared it to Naïve Bayes, to the ID3 decision tree classifier of (Quinlan, 1986), to the J4.8 Java implementation (Witten and Eibe, 2005) of C4.5 (a successor of ID3), and to a random forest (Ho, 1995) - an ensemble of decision trees. For all learners we performed 10-fold cross validation. For binning, we used K-Means and EWIB ($K = 50$ bins), and the *Tree-Based Unsupervised Bin Estimator* TUBE of (Schmidberger and Eibe, 2005), which estimates the number of bins. Parameter settings can be found in (Schult et al., 2011).

The lowest SROCD value (882,513 minutes) was achieved by ID3 after binning with K-Means. However, ID3 classifiers abstain from classifying some records in the test set, hence it is inferior to the learners build by the other algorithms, which assigns labels to all records. The overall best performance is achieved by Random Forest with 15 trees (910,383 minutes), and the second best by J4.8 (1,073,231 minutes), in both cases after binning with K-Means. The single tree of J4.8 improves the baseline by 16.2%.

An overview of the results is given in Figure 3. The horizontal line is the SROCD value achieved by the baseline ICDavg; of interest are only predictors that achieve lower SROCD, i.e. improve the baseline. For each learner, we depict the performance achieved by each binning method. We see that TUBE leads to worst performance, K-Means to best performance, EWIB being only slightly inferior to it.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

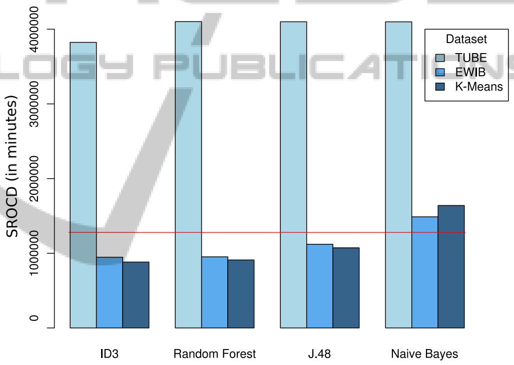

Figure 3: SROCD values for each classification algorithm (X-axis) combined with each binning method (legend); the baseline ICDavg is depicted as horizontal line. Lower values are better, and only values below the baseline correspond to an improvement in predictive power.

In classification tasks, it is usual to evaluate on accuracy. In Table 1, we juxtapose accuracy to SROCD for Naïve Bayes (worst SROCD) and J4.8 (second best SROCD), considering each binning method. The juxtaposition shows that accuracy is inappropriate for the task at hand, as it behaves contrary to SROCD. In particular, we see that both learners achieve best accuracy under TUBE, but much poorer SROCD values than under EWIB or K-Means. In contrast, the accuracy under EWIB and K-Means is very low.

This is an artifact that may lead the decision maker to wrong conclusions, so we explain it here: EWIB and K-Means produced 50 bins, distributing records evenly among them. TUBE produced 99 bins, but most of the records were placed in only 3 of them. Hence, under TUBE, the classifiers essentially learned to distinguish among three labels/bins. The

Table 1: Impact of binning method on the performance of the classifiers learned by Naïve Bayes (NB) and J4.8: performance is measured as accuracy (higher values are better) and SROCD (lower values are better). Accuracy is an ill choice for the task at hand, because it shows bias to the number of labels and to the distribution of the data among the labels.

Binning method	# bins	Accuracy (in%)		SROCD (in min)	
		NB	J4.8	NB	J4.8
EWIB	50	21.23	41.11	1,488,230	1,120,729
K-Means	50	14.39	35.41	1,639,846	1,073,231
TUBE	3+96	82.12	88.01	4,099,064	4,111,336

prior probability of a miss (wrong label assignment) is higher if there are 50 labels than if there are only three. Accuracy is sensitive to the number of labels, so it is an ill choice if classifiers are learned with different numbers of labels or with a strong bias towards only a few labels.

Summarizing, the instantiation of the prediction task for the hospital resulted in predictors that improved the baseline. Among the lessons learned are the impact of discretization on the learners and the importance of selecting a proper evaluation measure.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

6 CONCLUSIONS

We presented a high-level framework for knowledge discovery from medical protocols, and its instantiation in a German hospital for the prediction of surgical room occupancy time (SRO time). Such data are primarily recorded for medical purposes , but can be used to support planing decisions, too, provided they are appropriately prepared and analyzed.

In the instantiation of our framework in a German hospital we studied intensive care unit protocols and anesthesia protocols. Instantiation on the former is still under data preparation, since the intensive care units' data were in a format not yet appropriate for data mining. Anesthesia protocols have been successfully analyzed after a preprocessing task that involved computation and discretization of the target variable (SRO time). We reported on what steps should take place during preprocessing and analysis, how different algorithms can affect the predicting power of the learned models, and how they should be compared.

Next steps include the refinement of our framework towards specific activities for decision support tasks, and instantiations for knowledge discovery from other types of medical protocols, foremostly from intensive care unit protocols.

REFERENCES

Avison, D. and Young, T. (2007). Time to rethink health care and ICT? *Communications of the ACM*, 50(6):69–74.

Combi, C., Keravnou-Papailiou, E., and Shahar, Y. (2010). *Temporal Information Systems in Medicine*. Springer.

Dexter, F., Davis, M., Halbeis, C. E., Marjamaa, R., Marty, J., McIntosh, C., Nakata, Y., Thenuwara, K. N., Sawa, T., and Vigoda, M. (2006). Mean operating room times differ by 50% among hospitals in different countries for laparoscopic cholecystectomy and lung lobectomy. *Journal of Anesthesia (2006) 20:319–322*.

DGAI (1993). Qualitätssicherung und Datenverarbeitung in der Anästhesie. *Kerndatensatz Qualitätssicherung in der Anästhesie. Anästh Intensivmed*, 34:331–335.

Eijkemans, M. J. C., van Houdenhoven, M., Nguyen, T., Boersma, E., Steyerberg, E. W., and Kazemier, G. (2010). Predicting the unpredictable: A new prediction model for operating room times using individual characteristics and the surgeon's estimate. *Anesthesiology 2010; 112:41–9*.

Ho, T. K. (1995). Random decision forests. *3rd Int'l Conf. on Document Analysis and Recognition*.

Quinlan, J. (1986). Induction of decision trees. *Machine Learning 1: 81-106, 1986*.

Schmidberger, G. and Eibe, F. (2005). Unsupervised discretization using tree-based density estimation. *Lecture Notes in Computer Science, Volume 3721/2005, 240-251*.

Schult, R., Matuszyk, P., and Spiliopoulou, M. (2011). Prediction of surgery duration using empirical anesthesia protocols. In *The First International Workshop on Knowledge Discovery in Health Care and Medicine (KDHCM 2011)*, pages 66 – 77.

Stead, W., Hammond, W., and Straube, M. (1983). A chartless record - is it adequate? *Journal of Medicine Systems*, 7:103 – 109.

WHO (2011). World health organization: International classification of diseases (ICD). http://www.who.int/classifications/icd/en/.

Wilson, E. V. and Tulu, B. (2010). The Rise of a Health-IT Academic Focus. *Communications of the ACM*, 53(5):147–150.

Witten, I. H. and Eibe, F. (2005). *Data mining : practical machine learning tools and techniques*. Amsterdam: Elsevier; San Francisco, CA: Morgan Kaufmann.