Clinical decision support : distance-based, and subgroup-discovery methods in intensive care
Nannings, B.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 7. CONCLUSION AND DISCUSSION
7.1. Principal Findings

The main contribution of the research presented in this thesis is to provide a better understanding of Clinical Decision Support Systems (CDSSs) from two angles. The first part of this thesis focuses on the application of the subgroup discovery algorithm named Patient Rule Induction Method (PRIM) [1] for answering medically relevant questions. Not only are the results of these studies important, also the investigation of the possibilities and limitations of the PRIM method is a valuable contribution from a medical informatics perspective.

In the second part of the paper, we look at CDSSs in a telemedicine context from a bird’s eye view. We propose valuable definitions, conceptual models and a tool for categorizing such systems which together form a framework for understanding them.

In the first chapter of this thesis, the general introduction, we stated a number of objectives. In the paragraph below we reiterate the objectives and discuss how they were reached.

- To assess the value of PRIM subgroups, and compare them to ones obtained from logistic regression, in predicting the mortality in the population of very elderly intensive care patients.

In Chapter 2 we used PRIM on a dataset of 12993 consecutive admissions of elderly (80+) patients to a number of intensive care units (ICUs) in the Netherlands. The goal was to determine subgroups of patients with a very high mean mortality. We compared the characteristics of these subgroups to those of subgroups found using a conventionally used prognostic model (SAPS II [2]).

Using PRIM almost 10% of elderly ICU patients were identified as having a risk greater than 85% to die before hospital discharge. The subgroups are defined as conjunctions of simple conditions (patient characteristics) based on data which are routinely collected in the first 24 hours after ICU admission. Examples of patient characteristics used to define the subgroups are urine production, whether patients required mechanical ventilation, and what the lowest systolic blood pressure was, generally conditions that medical professionals associate with high risk of mortality.

The quality of the subgroups obtained with these methods were comparable, but using PRIM as opposed to conventional prognostic models also carries some additional benefits: PRIM requires less data to collect as subgroup definitions we found are based on only few input attributes while prognostic models such as SAPS II requires many input variables to calculate the patient’s probability of mortality. To obtain SAPS II patient subgroups we consider patients that share the fact that they have a (similar) predicted high probability of mortality. PRIM subgroups are more homogenous than subgroups of SAPS II patients as SAPS II mortality is calculated using a score that consists of many components; two patients having the same score may still differ greatly in their input variables. For the same reasons, the PRIM group may also be easier to understand; it is more clear which are the common (harmful) conditions of the patients within a subgroup. For these reasons, PRIM subgroups may be more useful for decision makers.
To analyze the ability of PRIM to find subgroups of hyperglycemic intensive care patients, as a first step to improve blood glucose control.

In Chapter 3 we used PRIM on a dataset of glucose measurements taken during the stay of patients in the intensive care unit. Although the patients were treated according to a blood glucose regulation protocol, hyperglycemia was still very common. By identifying subgroups of high (hyperglycemic) glucose measurements and correlating these outcomes with available explanatory variables we were able to identify patient characteristics that possibly may cause patients to be unresponsive to the glucose control treatment.

Most of the patient characteristics (possible determinants of hyperglycemia) that were used to define the subgroups were known to have a relation with hyperglycemia, e.g. the relation between a glucose measurement and its previous value, body temperature and bicarbonate concentration are all well-known. Two attributes for which no known relation exists to hyperglycemia are albumin serum levels and the admission type. It was also the case that some patient characteristics for which their relation to hyperglycemia is known, were not found in our subgroups.

The attributes we found are only possible determinants of hyperglycemia, further research that refines the treatment protocols according to our results can verify whether this leads to a reduction of hyperglycemic patients in the intensive care unit.

To analyze the weaknesses/strengths of PRIM by comparing it with the CART methodology and applying the methods to a large medical dataset to find subgroups of high mortality patients in a population of intensive care patients.

In Chapter 4 we apply PRIM to a large medical dataset to find subgroups of patients having a high risk of mortality and compare the resulting subgroups with those discovered by CART (classification and regression trees) [3]. In our dataset CART generally outperformed PRIM because of PRIM’s inability to find a large contiguous group that was found by CART. This subgroup was defined as all patients having a Glasgow Coma Score of 4 or lower.

We conclude that PRIM has problems with peeling data at the mode of an ordinal attribute (e.g. the Glasgow Coma Scale). This can be especially problematic if this mode is located near the variable’s minimum or maximum value. As ordinal scores are used frequently in the medical domain, this is an important fact to consider when using PRIM. We propose suggestions for improving PRIM such as implementing a form of backtracking (e.g. beam search), and making use of global information to choose variables for peeling.

To provide a single definition of Decision Support Telemedicine Systems (DSTS) and to propose a framework of properties helpful to characterize such a system.
In Chapter 5, we propose a Characterizing Property Set (CPS) consisting of 14 properties based on a literature study. We grouped these properties in 3 categories: “Problem”, “Process” and “System”, containing respectively 5, 3 and 6 properties. Properties of the “Problem” category are related to the medical problem and the environment in which a DSTS is introduced, such as purpose of a DSTS and in which medical domain the system is used. Properties of the “Process” category are related to the behavior and dynamic aspects of the DSTS such as the synchronicity of the communication and the passivity of the decision support component of a DSTS. Properties of the “System” category are related to the system and data that the system uses, such as what reasoning process the decision support component of a DSTS uses; if it contains a knowledge-base; and how this knowledge is structured. Unfortunately we did not find emergent properties unique to DSTSs. The CPS can be used to describe, compare, classify and cluster DSTSs by making their types explicit. We exemplify its use by applying it to two different types of DSTSs.

- To provide a conceptual model of DSTSs for its application in different forms of healthcare provision.

In Chapter 6, based on literature search, we propose definitions and conceptual models that are useful for understanding DSTSs. This may help different parties such as physicians, CDSS developers and telemedicine specialists in understanding and developing future DSTSs.

The conceptual models are expressed by Unified Modeling Language (UML) Class models, showing the relation of different components within a DSTS. We provide a single general model that should be useable for most DSTSs, and provide a number of template models which are aimed at specific types of DSTS, e.g. diagnosis or monitoring DSTSs. In both the general model and the template models we encapsulate properties (as class attributes) from the CPS that was described in Chapter 5.

In the following paragraphs we describe strengths and weaknesses of our approach, the implications of the work, related research, future work and concluding remarks.

7.2. Strengths and Weaknesses of our Approach

In the first part of this thesis we presented a significant effort to investigate PRIM and its possible usefulness for medical informatics research. Although PRIM has been applied in the medical domain before [4], our work distinguishes itself by using a large clinical database of high dimensionality, by comparing PRIM to parametric (logistic regression) and non-parametric methods (CART) and by relying on bootstrap techniques for evaluating subgroup performance.

The specific subgroups identified by PRIM have to be considered as validated subgroups with a markedly higher average outcome than the global average. These subgroups are not necessarily the best subgroups possible because PRIM is not an exhaustive search algorithm but in essence a hill-climbing search algorithm. In addition,
altering the meta-parameters of the PRIM algorithm may lead to different (possibly better) subgroups. It should also be noted that the subgroup descriptions might not necessarily generalize to external settings although a multicenter database was used for the mortality prediction problem.

A limitation of the comparison of PRIM with logistic regression and CART is that our evaluation was purely based on performance measures. We did not formally consider the complexity of the obtained rules and the usefulness of applying the knowledge obtained from discovered subgroups in practice.

In addition, our analytical scenarios for comparing PRIM to CART could not possibly mimic the flexibility of a human performing the analysis with PRIM. However, our choices for the scenarios were motivated by the idea to cover the general analytical goals an analyst might have in mind. Since the comparison of the results of the algorithms is difficult, because the subgroups resulting from the application of both methods may not be the same, we used the principle of matching the two algorithms on support and/or target mean. The relative rigidity in performing the analysis has the advantage that our scenarios can be completely automated and hence the analyses are reproducible.

The first part of this thesis can be seen as exploring “does PRIM work (and when not)?”. We did not try to answer the “does it help?” question by actually using the knowledge to influence (treatment) decisions. In future investigations this question should be addressed.

A weakness of the second part of our thesis is that we do not give much attention to semantic interoperability. This becomes an increasingly important issue when one aims to reuse the same CDSS for various databases and “clients” such as different types of Electronic Patient Records. Most of these systems will store information in their own way and a mapping is needed between the concepts used by the CDSS and the various systems to which it is connected. We did not give much attention to this problem because it is a problem that occurs for CDSSs irrespective of whether they are used as a telemedicine application or not. A solution to the problem for guideline-based decision support may be the vMR (virtual medical record) [5]. Johnson et al suggest a vMR that supports (1) a structured data model for representing information related to individual patients, (2) domains for values of attributes in the data model and (3) queries through which guideline decision-support systems can test the states of the patient. The vMR allows guideline authors for example to encode clinical guidelines using a rich and well-defined model of patient data. The vMR does not contain a data model that replicates everything that an EPR holds, but only those distinctions necessary for modeling guidelines and protocols. The authors suggest to use the HL7 Reference Information Model [6] as the basis for a standardized virtual medical record.

The properties we found to be relevant in DSTSs are not unique for those types of systems. In other words, a CDSS with or without a telemedicine component has the same properties. However, we believe our framework is useful for describing DSTSs. In the literature we found a number of DSTSs that were not clearly described because
some properties were not mentioned. E.g. in [7] the type of knowledge representation and the reasoning process of the CDSS component of the DSTS were not described. Our framework can in such cases be used as a checklist for determining whether relevant issues are discussed in a system description.

The fact that we did not find emerging properties has to do with the granularity of the conceptual model. The network part is introduced almost as a black box. For the description of many DSTSs that is not a problem: the only thing that counts is the connection with a specific CDSS. But the availability of a network also makes it possible for a client to choose various services. Both the CORBA (Common Object Request Broker Architecture) standard and web services can now be used for communication between clients and servers. Using the CORBA standard one can select even in runtime a certain service. Our conceptual model could be extended by characteristics that describe these approaches, like the type of object request broker, the presence of a name server, etc. In the case of computer-interpretable guidelines (CIGs) the OpenClinical Group [8] suggested a model for publishing CIGs on the web. In this model, executable guidelines are published as Web-accessible services.

We have stated that our characterizing property set and conceptual models may help in developing a DSTS. However, we do not support the development directly as our work does not contain guidelines about how to develop such systems. Our framework does however provide a ‘language’ to facilitate communicating about these systems and thus indirectly supports the development of these systems.

7.3. Implications

In this thesis we address two forms of decision support: decision support related to subgroup discovery, and decision support systems embedded in a telemedicine setting.

The idea behind PRIM is attractive and it also provides a battery of diagnostics to guide the analyst in performing his or her task. Hence we encourage researchers to explore PRIM in more depth. Analysts should however be aware of the limitations we discovered when using PRIM. We suggest that researchers and analysts complement PRIM with the use of other algorithms or incorporating a suitable backtracking mechanism.

Some of the subgroups we found agree with the literature and seem plausible. For example, the relation between Glasgow Coma Score and mortality that we found in elderly patients is well documented. Of others, we are not sure of their exact meaning (what the underlying cause for a high value of the outcome is). However, these subgroups may prompt other investigators to investigate these subgroups and report about their statistical properties.

Our subgroup discovery related work could also have implications for clinicians, as the results described in this research may eventually lead to improvement of clinical practice guidelines (e.g. ICU blood glucose management guidelines), of course this necessitates additional research to be carried out.
In the second part of this thesis we harmonize the work in two fields based on an extensive literature search: clinical decision support systems and telemedicine. Both disciplines have literature dedicated to it, but literature about DSTSs is scarce, while such systems are becoming increasingly more important with the advent of the Internet and Information Communication Technology (ICT) in general.

We provide a framework that will help parties involved in requirements analysis processes and the development of DSTSs. It focuses on important concepts and their relations from a DSTS perspective. At the start of the requirements analysis the framework may help stakeholders to identify important questions to ask, and aids them in designing a high-level architecture of the DSTS.

Aside from providing support during the analysis and development of a DSTS, the framework provides a means for describing, comparing and clustering DSTSs. While description is important from a research point of view, comparing and clustering DSTSs can be important when carrying out systematic reviews of such systems or evaluation studies.

7.4. Related Research
The first part of this thesis applies PRIM to different purposes: comparing it with other algorithms and evaluating its performance. Although PRIM has been applied before, it has not been applied to a real-world large high dimensionality dataset such as ours.

In this section we contrast PRIM with other algorithms/methods that can be used for subgroup discovery and note the main differences between PRIM and related algorithms. PRIM is a non-parametric, patient, subgroup discovery hill-climbing algorithm without a backtracking mechanism (aside from pasting which however has a very local nature).

The first method that we compared to PRIM was the Simplified Acute Physiology Score (SAPS) II model. SAPS II is used to score the severity of illness of ICU patients, and the model allows us to compare the quality of care of different ICUs. Unlike PRIM, SAPS II is a global parametric model based on logistic regression. It is global because it can predict the probability of the outcome for any subject in the population. It is parametric because it pre-supposes the form of the model. Strictly speaking, it cannot be considered a subgroup discovery algorithm, but it can be used to rank subjects based on their probability of showing an event. An example of a subgroup obtained using SAPS II is all the patients that have a predicted mortality > 90%. Limitations of using a logistic regression model for subgroup discovery are: a) the coefficients of the variables are determined by maximizing the likelihood of the model taking into account all observations, not just those in a subgroup, b) all variables should be known and used in order to determine the probability of an event while a subgroup description on PRIM may use fewer variables (in its application, for subgroup definition generation it does need all the variables), c) the subgroups do not tend to be contiguous in the input variable space, they include all those with a very high (or very low) risk of the event, d) the outcome of
the model is more difficult to interpret than the symbolic representation of outcomes in PRIM.

The second algorithm that we compared to PRIM was CART (Classification and Regression Trees). Like PRIM, CART is a non-parametric hill-climbing model without backtracking. In contrast to PRIM, CART is a global model and is greedier. Using CART for subgroup discovery shares some of the limitations noted above of a global model, as it is not optimized on subgroups but rather on splits in the data. The greedier character of CART means that once a split (a constraint on a variable's values) is determined this split is permanent since there is no backtracking mechanism. If the split, in retrospect, turned out to be a bad one CART would not recover from this sub-optimal choice. PRIM is patient and hence attempts to save enough data for future decisions. However, as we showed in Chapter 4, PRIM's insistence on patience without allowing for backtracking makes it vulnerable too. The adoption of the penalty function in f-PRIM [9] may allow PRIM to make different peeling decisions but without backtracking this does not solve PRIM's vulnerability.

CN2-SD [10], APRIORI-SD [11] and Data Surveyor [12] are all subgroup discovery algorithms. They show two main differences with PRIM. First they are greedy (with DataSurveyor being the most greedy) but they do provide backtracking by applying beam-search. However, it is unclear which beam width one should select. In addition there is a risk that the beam is filled with various constraints of just one dominant variable (e.g. age > 32, age > 41, age> 45) hence defeating the idea of keeping track of truly alternative candidates.

An important related work for part two of this thesis is [13]. This paper describes a service-oriented architecture for distributed clinical decision support. The architecture aims to leverage information exchange between health information systems. Although web service oriented architectures (Web services are a W3C standard) are used in many domains, it is not very prevalent in the domain of medicine. The architecture specifies a series of protocols/communication standards such as HL7 [6], SNOMED [14], the National Council for Prescription Drug Programs (NCPDP) SCRIPT [15], RxNorm [16], and National Drug Codes [17], and Service Oriented Architecture related standards such as Simple Object Access Protocol (SOAP), Extensible Markup Language (XML), Universal Description, Discovery and Integration language (UDDI) and the Web Service Definition Language (WSDL). It aims to promote modularity (services are provided in reusable components) and abstraction (it is not necessary to know how a system works, but only how to use its services) and sets the agenda for future decision support research and development. This research differs from our work since it focuses on the integration task and specifies standards to facilitate this integration whereas our work focuses on properties of the internals and externals of such systems. Common Object Request Broker Architecture (CORBA) [18], Microsoft's Distributed Component Object Model (DCOM) [19] and SUN's Java Remote Method Invocation (RMI) [20] are all standards/approaches highly similar to web services, which will help to promote development and use of DSTSs.
7.5. Recommendations for future research

Future research is mostly related to the weaknesses we mentioned earlier. While we compared PRIM to CART and SAPS, comparisons with other subgroup discovery algorithms still have to be carried out. It would also be interesting to make a comparison between PRIM and a version of PRIM that implements the changes that were suggested in Chapter 4.

It is also important to apply PRIM to other (large) medical datasets, perhaps PRIM has other weaknesses or strengths that did not show in our research because of the specific dataset we used.

The user interactivity that is part of PRIM will be most advantageous when the analysis is carried out by an analyst who has expert knowledge of the relevant medical domain. Perhaps having an expert analyst performing subgroup discovery with PRIM will reveal subgroups far superior to the ones we found using our ‘algorithmic approach’. However, this could pose a problem for studies evaluating PRIM, as it will be unclear which part of the subgroup discovery process can be attributed to the algorithm and which part is attributed by the analyst’s (analytical and domain) knowledge.

It is also worth to investigate the usefulness of the discovered subgroups for clinical practice. Perhaps subgroups can be used to adjust clinical practice guidelines. This however needs rigorous evaluation in carefully designed clinical trials.

The characterizing property set and UML models that we provide for DSTSs need to be applied in practice (e.g. using them to perform requirements analysis for a DSTS) to learn more of their applicability and get feedback to improve them.

Future work related to the DSTS framework should focus also on integrating CDSSs and clinical data sources through e.g. web services. Increase of standards have made web service oriented architectures very common in general ICT. Applying this technology in medicine will increase interoperability of systems and will help to bring together medical data and CDSSs, which has great potential in terms of improvement of quality and efficiency of care.

7.6. Concluding Remarks

In this thesis we have investigated two forms of CDSSs. The ever increasing amount of medical data, and the wish to improve healthcare by applying medical informatics methods will likely boost the development and use of the types of CDSS that we have described in this thesis. Our analysis of PRIM on a large medical dataset revealed both good and poor qualities, and we provided suggestions on how to improve the PRIM algorithm. For DSTSs we provided a characterizing property set and conceptual models that we hope will help future DSTS stakeholders to get acquainted with the basics and will enable them to focus on the essentials of these systems.
7.7. References


