Analysis of guideline compliance – a data mining approach

Vojtěch SVÁTEK[1,2], Antonín ŘÍHA[1,3], Jan PELEŠKA[1,4], Jan RAUCH[1,2]
[1]*European Centre for Medical Informatics, Statistics and Epidemiology - Cardio*
[2]*University of Economics, Prague, Czech Republic*
[3]*Institute of Computer Science, Academy of Sciences, Prague, Czech Republic*
[4]*General University Hospital, Prague, Czech Republic*
E–mail: {svatek, rauch} @vse.cz, {riha, peleska}@euromise.cz

Abstract. While guideline–based decision support is safety–critical and typically requires human interaction, offline analysis of guideline compliance can be performed to large extent automatically. We examine the possibility of automatic detection of potential non–compliance followed up with (statistical) association mining. Only frequent associations of non–compliance patterns with various patient data are submitted to medical expert for interpretation. The initial experiment was carried out in the domain of hypertension management.

1 Introduction

The most frequently discussed role for guideline–based software is interactive decision support, with focus on decisions associated with an individual patient. In this process, formalised statements from the guidelines are complemented and possibly amended with situated human judgement, rather than being obeyed literally. Furthermore, since even well–designed generic guidelines are not applicable on all individual patients and under all type of local/institutional conditions, adaptations are carried out prior to their deployment in formal representation.

This use case of computerised guidelines is however not the only one. Here, we concentrate on the comparison between the *literal* content of the guidelines and a larger collection of patient records, carried out outside the clinical environment. The ultimate addressee of such analysis – possibly viewed as analysis of guideline *compliance* – is a guideline authoring body rather than just a field clinician.

We base our approach on two crucial assumptions:

- Since compliance analysis is not safety–critical, errors concerning individual cases can be tolerated. The analytical system thus can be granted more autonomy and ideally, can be run *offline*. This would result in fast performance desirable for operation on larger collections of patient records.

- Information on frequently occurring non–compliance patterns may be useful for guideline authors and/or promoters, in particular, when *associated* with other patterns in patient data. Clinical expert then may start to form initial hypotheses on the causes of non–compliance, based on associations, even before examining the individual cases.

It is important to note that, although we technically speak about *non–compliance* patterns, these would often be indicators of other phenomena than trivial non–compliance (e.g. by ignorance). The guidelines may be, for example:

- outdated, or with errors in text (non–compliance is then highly desirable)
- too generic, requiring adaptation to local conditions
- inadequate with respect to data chosen for analysis (cf. section 2)
- their formalisation may have been erroneous (cf. section 3).

Our method consists in a two–phase process: detection of potential non–compliance in individual patient records, followed up with mining in patient records enriched with non–compliance patterns. In section 2 we briefly introduce the guideline document, data, and tools used in our experiment. In section 3 we characterise the process of rendering the guidelines into a formal representation. In sections 4 and 5 we describe both phases of analysis and their results. In section 6 we survey some related research. Finally, in section 7 we outline directions for future research.

2 Experiment setup

For our experiments, we chose the domain of *hypertension*, which is an example of diagnosis with long–term outpatient follow–up. It is frequently mentioned in connection with computerised guidelines. As underlying *guideline document*, we selected the 1999 WHO/ISH hypertension guidelines [1]. The document is relatively comprehensible for non–experts, and fairly generic. The *data* were collected in the hypertension clinic of the 2nd Department of Medicine, General University Hospital in Prague, by a hypertension specialist (co–author of this paper); all the data describe patients with essential hypertension. The main motivation for choosing this dataset for the initial experiment thus was availability of deep insight into the individual records by the physician – member of the research team. On the other hand, the data reflected specialist care, following up with initial diagnosis established and (usually unsuccessful) treatment applied by GPs. Therefore, they were not fully adequate with respect to the guideline document (which is intended for primary care starting from the first contact). Furthermore, since only a subset of the fields was recorded in structured form, manual pre–processing had first to be applied. Therefore, only a fraction of the database was available for experiments to date, namely, the records of 48 patients (of the total of approx. 200).

The software used for the first phase of compliance analysis—processing individual patient records—was developed in the *OCML* language [5]. OCML is a powerful knowledge representation and reasoning language previously used for construction of ontologies and other knowledge models. It combines Prolog–like backward chaining with inheritance in class hierarchies and calls of the underlying CommonLisp engine. In the second phase of analysis—association discovery—we used an existing data mining tool called *LISp–Miner*[1] [8]. LISp–Miner is a modular system that enables to apply various (statistically–inspired) mining techniques on tabular data; it scales well to vast quantities of data thanks to numerous optimisations. In this project, we only used a small fraction of its capabilities, to date.

3 Guideline formalisation

In order to proceed fast, we did not adopt one of existing generic guideline models and tools. Instead, the given guideline document was manually transformed into a (declarative) OCML program, explicitly describing different patient states and treatment strategies considered in

[1] Surprisingly, the system has nothing to do with the Lisp language. LISp stands for ”Laboratory for Intelligent Systems, Prague”, where it was originally developed.

the document[2]. Clearly, this solution has limited reusability, and should be substituted by a generic model[3] if the whole methodology proves viable.

Although the informaticians who developed the program consulted most unclear points with clinical experts, there was no systematic introduction of *background knowledge*[4]. This simplification relied on the assumption that missing background knowledge will, during compliance analysis, result in *errors of commission* rather than those of omission[5]. In other words, non–compliance will often be indicated incorrectly, but true non–compliance will rarely remain unveiled. A frequent but fictive non–compliance pattern (as artefact of 'sloppy' formalisation) causes extra workload for the expert who interprets the results; nevertheless, this workload probably pays off compared to preventive addition of all conceivable background knowledge in the phase of initial model building. The model (OCML program, in our case) can naturally be amended with knowledge *a posteriori* identified as missing.

4 Detection of potential non–compliance

The OCML program was run against the time–stamped patient records, and all unexpected findings were semi–manually assigned to generic non–compliance patterns. We ended up with ten patterns, such as 'non–administration of indicated drug', 'administration of contraindicated drug', 'therapy change despite good response', 'long pause between visits' etc. Since the data were small, we did not take into account the period of follow–up in which the event occurred, and even abandoned the distinction of drug classes and factors influencing drug choice when shifting to the second phase of analysis. In this way, we obtained patterns that were quite coarse but their frequencies allowed application of a data mining tool.

5 Mining for associations

The first step of analysis yielded, thanks to pattern confluence along the time axis, a single table with one row per patient and ten binary columns corresponding to non–compliance patterns (NCPs). In the *pre–processing* phase of the second step, we glued this table with other patient data: both timeless (incl. anamnestic) data, and aggregations of time–stamped data such as maximal/minimal values of BP. This second part of the table contained 39 attributes: 4 related to BP, 6 to presence of risk factors for coronary heart disease, 5 to associated clinical conditions, 3 to target organ damage, 14 to drug treatment, the remaining ones to patient's history, treatment duration and frequency of visits. Using LISp–Miner's own pre–processing tool, we also converted nominal and numerical attributes to binary ones.

Associations with other data were expected to act as clue for guessing the nature of non–compliance prior to examination of individual cases. We thus first ran the mining task relating observations from the one part of the table (NCPs) to observations from the other part of the table, possibly conditioned by further observations (see [8] for explanation of LISp–Miner hypothesis types). Given the small size of data, the run times were negligible (fractions of a second). The extent of results was appropriate for submission to experts, e.g. for the default setting of parameters, we obtained eight unconditional associations. Fig. 1 shows statistical

[2]More details on OCML formalisation of 1999 WHO/ISH hypertension guidelines are in [9].

[3]In this respect, *Prodigy* [3] looks as most promising solution, since its modular scenarios could probably be mapped on non–compliance patterns.

[4]Except for operationalisation of intentionally vague statements (such as 'good response' in terms of BP lowering). In such situations, we preferred the most benevolent (but still consensual) value among those suggested by co–operating physicians.

[5]This assumption seems to be consistent with the study done by Patel et al. [6].

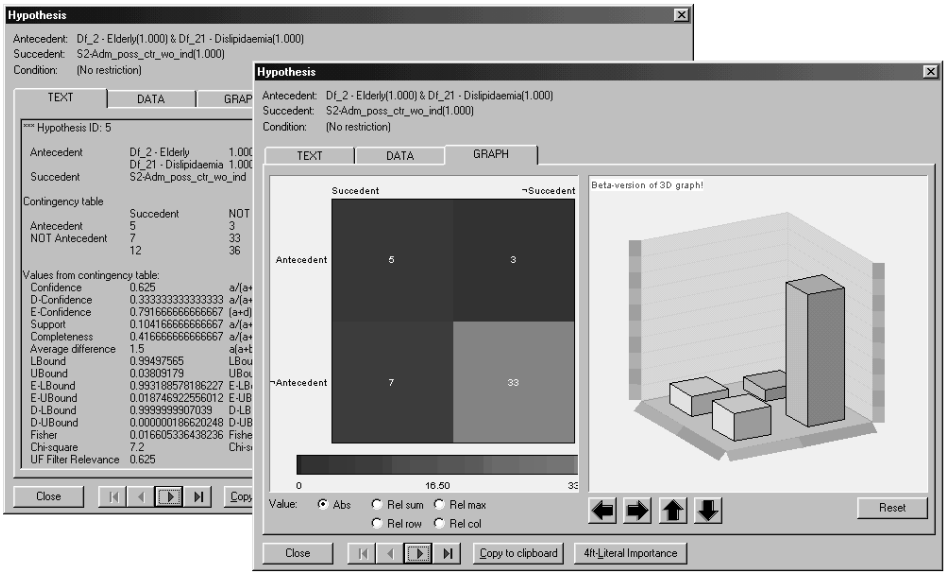

Figure 1: Details of an association in textual and graphical form

details (computed from the contingency table) of one of strong associations—"elderly patients with dyslipidaemia were often treated with possibly contraindicated drugs"[6]—in textual and graphical form. We also experimented with associations *among the NCPs themselves*. Since NCPs have relatively low frequencies, their co-occurrence was always supported by a few cases only. The most promising rule (still far below statistical plausibility) associated patients with unusually sparse visits with patients treated with possibly contraindicated drugs.

6 Related work

Although (the increase of) compliance is declared as important goal associated with guideline computerisation, few systematic attempts have been undertaken so far to determine it via automated analysis (data mining) in larger collections of patients. Let us mention four partially related projects; interestingly, the first two also deal with hypertension management. Persson et al. [7] examined more than 300 hypertensive patients by means of a rule–based decision–support system, and identified several interesting non–compliance patterns. The study only covered the drug selection problem, and ignored temporal aspects of treatment (which we, in turn, embedded into the OCML–based software). Advani et al. [2] proposed a complex language (QUIL) for evaluation of quality of clinician actions. Their approach assumes augmentation of computerised guidelines with information not explicitly stated in the original text, in particular with importance of individual statements and underlying intentions. In contrast, we only consider the literal content of the guidelines, treat all non–compliance patterns as equally important, and only filter them by frequency of occurrence. The interpretation of severity of non–compliance (and whether it is non–compliance at all) is left to the final phase of analysis in our approach. In this way, it could be seen as complementary to Advani's, as fast, initial examination of the literal guideline, imposing minimal requirements on expert–supplied background knowledge, and possibly yielding such knowledge indirectly. Marcos et al. [4] carried out a small–scale but thorough study on compliance with short–term protocols in the domain of neonatal jaundice. Experts were asked to provide solutions for a set of cases, and their suggestions were manually compared with those provided by the formalised proto-

[6] We identified these drugs as beta–blockers. Their adverse effect on lipid level has recently been assessed as transient, and the possible contraindication was withdrawn in 2003 ESH/ESC Hypertension Guidelines.

col. The interpreted outcomes of the project were in accordance with our own typology: some 'non–compliance patterns' were identified as local deviations, some as potential gaps in the protocols, and some as artefacts of imperfect protocol formalisation. The interpretation was however carried out for individual cases rather than for frequent patterns. Finally, in the experiment described by Seroussi et al. [10], knowledge extracted from the (cancer–treatment) guidelines was presented to the physicians in the form of a decision tree, together with patient data. The degree of compliance with the guideline knowledge before and after such confrontation was statistically measured. This approach however lacks direct comparison of guidelines with treatment outcomes reflected in patient data.

7 Conclusions

In this paper we examined the way data mining tools could assist in guideline compliance analysis. Association rule mining looks promising as source of hypotheses relating (previously identified) non–compliance patterns among themselves or to other patient data. The initial experiment was carried out on 48 data records of patients treated for hypertension.

Since the sample was small and not fully adequate with respect to the guideline document, we do not make any claims about the medical validity of results. Expected benefits of offline compliance analysis for guideline providers/promoters will only materialise when there are large numbers of patient data available, ideally from different points of care. Statistically grounded association hypotheses could then (also) be generated at lower degree of abstraction, thus shedding more light on the nature of non–compliance. Key bottleneck in the data acquisition process is the necessity to convert textual patient records into structured form in a semi–manual manner. Hopefully, ongoing efforts to deploy structured EHR systems into clinical practice will change this state of affairs in a not too distant future.

The research has been partially supported by the project LN00B107 of the Ministry of Education of the Czech Republic.

References

[1] WHO/ISH Hypertension Guidelines for the Management of Hypertension. *Journal of Hypertension*, 17, 1999, 151–183.

[2] A. Advani, Y. Shahar, and M. A. Musen, Medical Quality Assessment by Scoring Adherence to Guideline Intentions, 2001 Annual AMIA Conference, Washington, Hanley and Belfus 2001.

[3] P. D. Johnson, S. Tu, N. Booth, B. Sugden, and I. N. Purves, Using Scenarios in Chronic Disease Management Guidelines for Primary Care. AMIA Annual Symp. 2000, 389–393.

[4] M. Marcos, M. Balser, A. ten Teije, and F. van Harmelen, From informal knowledge to formal logic: a realistic case study in medical protocols, in: Proc. EKAW'02, Springer 2002, 49-64.

[5] E. Motta, Reusable Components for Knowledge Modelling. IOS Press, 1999.

[6] V. L. Patel, V. G. Allen, J. F. Arocha, and E. H. Shortliffe, Representing Clinical Guidelines in GLIF: Individual and Collaborative Expertise. *JAMIA* 5(5):467–83, 1998.

[7] M. Persson, T. Mjorndal, B. Carlberg, J. Bohlin, and L. H. Lindholm, Evaluation of a computer-based decision support system for treatment of hypertension with drugs: retrospective, nonintervention testing of cost and guideline adherence. *J.Intern Med.*, 2000 Jan, 247(1), 87-9.

[8] J. Rauch and M. Šimůnek, Mining for 4ft Association Rules. In: (Arikawa S., Morishita S., eds.) Discovery Science 2000. Springer Verlag 2000, 268-272.

[9] A. Říha, V. Svátek, P. Němec, and J. Zvárová, Medical guideline as prior knowledge in electronic healthcare record mining, in: Data Mining III. (Eds. Zanasi A., Brebbia C.A., Ebecken N.F.F.E.,Melli P.), WIT Press, Southampton, 2002, pp.809-818.

[10] B. Seroussi, J. Bouaud, E.-C. Antoine, L. Zelek L. and M. Spielmann, An experiment in sharing and reusing OncoDoc's breast cancer guideline knowledge. Computer-Based Support for Clinical Guidelines and Protocols, Studies in Health Technology and Informatics, IOS Press, 2001.