

Constructing Clinical Decision Support Systems for Adverse Drug Event Prevention: A Knowledge-based Approach

Vassilis Koutkias, PhD¹, Vassilis Kilintzis MSc¹, George Stalidis, PhD¹, Katerina Lazou, MSc¹, Chrysa Collyda, PhD¹, Emmanuel Chazard, MD, MSc², Peter McNair, MD, Dr.Med.Sci³, Regis Beuscart, MD, PhD², the PSIP Consortium, Nicos Maglaveras, PhD¹
¹Aristotle University, Medical School, Lab of Medical Informatics, Thessaloniki, Greece; ²Lille University Hospital, EA2694, France; ³Region Hovedstaden & Kennedy Center, Copenhagen, Denmark

Abstract

A knowledge-based approach is proposed that is employed for the construction of a framework suitable for the management and effective use of knowledge on Adverse Drug Event (ADE) prevention. The framework has as its core part a Knowledge Base (KB) comprised of rule-based knowledge sources, that is accompanied by the necessary inference and query mechanisms to provide healthcare professionals and patients with decision support services in clinical practice, in terms of alerts and recommendations on preventable ADEs. The relevant Knowledge Based System (KBS) is developed in the context of the EU-funded research project PSIP (Patient Safety through Intelligent Procedures in Medication). In the current paper, we present the foundations of the framework, its knowledge model and KB structure, as well as recent progress as regards the population of the KB, the implementation of the KBS, and results on the KBS verification in decision support operation.

Introduction

The potential of Information Technology (IT) tools and techniques towards Adverse Drug Event (ADE) prevention has been highlighted in various studies¹. Major focus of IT-based research on ADEs has been automatic or semi-automatic identification of ADEs by employing machine learning and statistical inference techniques applied to patient data repositories²⁻⁴, e.g. Electronic Health Records (EHRs). In this regard, studies have been concentrated on the development of IT tools capable of providing evidence on the origin of ADEs, following typically experts review evaluation of the obtained results. These outcomes were foreseen to constitute the basis for advancing the decision support functionalities on ADEs offered by clinical information systems, such as Computerized Physician Order Entry (CPOE) systems. However, the majority of the proposed approaches did not elaborate further

towards the incorporation of the ADE signals identified into actual Clinical Decision Support Systems (CDSSs) capable of interoperating with clinical information systems, e.g. CPOEs and EHRs.

As more mature evidence on ADE prevalence is gained, the focus of IT research has been also attracted by the incorporation of the identified ADE patterns or signals into sophisticated knowledge-based models. For example, a data mining tool has been proposed aiming to improve signal detection algorithms by performing terminological reasoning on MedDRA codes described in description logic⁵. In addition, an integrated approach for the identification of ADE signals has been proposed that is based on Semantic Web technologies⁶; as the information required for identifying signals during drug design and development resides in heterogeneous, distributed data repositories, the Semantic Web paradigm offers new capabilities for data integration that exploits explicit semantics and well-defined ontologies. Following also an ontology-based approach, a prototype medical intelligent assistant has been proposed aiming to improve patient safety by reducing medical errors within hospitals⁷. The ontology encapsulates hospital care concepts including activities, procedures and policies, as well as medical knowledge, and is particularly designed to track the implications of medical decisions taken by health professionals within the context of guidelines/regulations of the medical environment, and the established medical knowledge.

Aligned with these knowledge-based efforts, in the scope of the European project PSIP (Patient Safety through Intelligent Procedures in Medication), a knowledge engineering framework is constructed aiming to systematically represent and manage identified and validated ADE signals following a knowledge discovery and elicitation phase⁸. The core of this framework consists of a Knowledge Base (KB) encapsulating the abovementioned signals that are provided in the form of rules. This framework

constitutes the basis for the construction of contextualized CDSSs for ADE prevention. In this paper, we extend our preliminary work concerning the initial establishment of this framework⁹, by presenting the knowledge model elaborated and its foundations, the underlying formalism adopted, the query and inference mechanisms employed, focusing also on KBS implementation and verification aspects.

Methods

The current work focuses on the construction of a rule-based knowledge framework which, in addition to ADE identification, is designed to support ADE prevention through effective decision support that is delivered via alerts and recommendations to the clinical personnel. It incorporates a context-sensitive, meta-rule level, which is used to address rule ranking and decide on the applicability of ADE signals per case, in order to eliminate over-alerting. The proposed knowledge model is mapped to a data model specifically designed within the PSIP project for querying the KB with actual patient data¹⁰. In the following, we present the major parts elaborated to synthesize the proposed knowledge framework.

Rules Structure: ADE signals elaborated in the PSIP knowledge model constitute production rules¹¹. These rules follow the general form:

$$p_1 \text{ AND } p_2 \text{ AND } \dots \text{ AND } p_n \rightarrow h, \quad (1)$$

where p_1, p_2, \dots, p_n are atomic formulae of some accepted language (e.g., propositional logic, attributive logic, first order logic, etc.) and h is the conclusion, action or decision. In PSIP, p_i are pseudo-variables which correspond to groups of a) drug codes in ATC, b) lab results in C-NPU/IUPAC, or c) diagnosis codes in ICD-10 classification; the h part constitutes the effect of the rule, i.e. the actual ADE, which typically corresponds to a diagnostic pseudo-variable.

Since these rules are statistically inferred by applying data mining techniques to diverse EHRs⁸, the importance and applicability of each rule is determined based on its statistical significance in the local context that is being triggered, i.e. hospital/department. Thus, statistical features such as the confidence (probability of having the effect knowing that the conditions are met), the support (probability of having the effect and matching the conditions at the same time), the Fisher test p -value and so forth constitute rule meta-data that may be particularly used to address over-alerting.

Besides data mining originated rules, a commercial knowledge source capturing drug to drug interactions,

drug contraindications, drug to allergy class associations, as well as drug to lab value or medical parameter associations, is made available in the project by the partner VIDAL (<http://www.vidal.fr/>), following also a rule-based formalism and based on the abovementioned standard terminologies.

Knowledge Base Structure: The knowledge employed in PSIP belongs in three categories: a) *domain knowledge*, defining types and facts, which are generally static and structured via concepts (i.e., classes), relations-associations, attributes, and rule types (expressions); b) *task knowledge*, in terms of functional decomposition, and control; in this regard, knowledge is elaborated with respect to combination of tasks to reach a goal/workflow, or oppositely, decomposition of complex tasks into separate processes; c) *inference knowledge*, corresponding to the basic reasoning steps that can be followed in the domain and are applied by tasks.

According to the above, the PSIP KB comprises of a set of ontology-based structures, either PSIP-specific or standard classifications. In addition, a rule-based component is included that is defined via a set of classes and populated with ADE rules. The ontology-based structures and the rule-based component constitute the fundamental elements to define complex procedural logic in terms of protocols and guidelines, according to the computer interpretable guidelines formalism¹², which constitutes the core knowledge engineering methodology employed in this work. This formalism enables the unification of the former knowledge components, so as to provide a common knowledge framework based on which the CDSS will offer its services.

Query and Inference Mechanisms: As the KB developed constitutes the core part of CDSS modules for ADE prevention, an appropriate interface for querying the KB with patient records has been defined that is based on a common data model¹⁰, so as to test the case(s) of interest against ADE signals incorporated in the KB. This interface relies on the mapping of each concept/attribute defined in the knowledge model (presented in the following) with the relevant fields/tables defined in the data model.

Considering a tuple $\langle Dr, Di, Bi \rangle$, where Dr corresponds to drug values, Di to diagnosis values and Bi to lab results of a patient stored in his/her EHR as the input of the CDSS, and a set of rules R incorporated in the KB, an inference mechanism f is introduced, in order to match the above tuple in R , i.e.:

$$f: \langle Dr, Di, Bi \rangle \rightarrow R. \quad (2)$$

The outcome of this procedure is a new set $R_A \subseteq R$, resulting in potential ADE signal(s). In case of multiple applicable ADE signals, i.e. multiple rules triggered, a major issue in the application of discourse constitutes the elimination of over-alerting, i.e. multiple rules firing. In this regard, the CDSS is fine-tuned (in both construction and runtime mode) following a context-sensitive strategy that applies meta-rules in terms of thresholds concerning the statistical significance of the corresponding triggered rules, to determine the most significant alerts or recommendations that will reach the CDSS end-user. Thus, a new mechanism g is introduced that maps R_A into a set $R_B \subseteq R_A$ according to the context criteria C_X :

$$g: R_A / C_X \rightarrow R_B. \quad (3)$$

It has to be noted that there are cases where it is necessary to preserve a part of R_A as an outcome of (2), independently of the meta-rules defined in C_X . Such a potential is taken into account in the mechanism g . Finally, the CDSS outcome is a list of effects that is associated with R_B , along with appropriate explanations of the respective rules, the importance of the potential ADEs, the data that made the rules fire, as well as recommendations for actions.

Knowledge Model: The core component of the proposed model is in the form of rules associating a number of conditions to an effect. A set of additional components are used as terminology for efficiently expressing these rules. Moreover, a separate rule-based component is included, which encapsulates knowledge on the applicability of basic rules to various clinical practice circumstances. Thus, three levels/types of rules have been defined, *main rules*, *intermediate rules* and *meta-rules*.

Figure 1 illustrates the conceptual schema of the PSIP knowledge model. *Rule* constitutes the primary concept corresponding to the main rules on predicting ADEs and is linked with *Conditions* and *Effect*. The cardinality of the relevant relations are in line with the basic form shown in (1), according to which a rule comprises of an arbitrary number of conditions and generates only one effect. Note that all possible instances of *Effect* are enlisted, i.e. each effect identified in PSIP is coded and given a name in the form of pseudo-variable. It is also expected that more than one rule may be linked to the same effect. This property is considered for rule grouping and inference optimization.

Each *Rule* in the KB is uniquely indexed and the corresponding source is recorded. Attributes related to rule meta-data are also defined. These parameters are initially evaluated for each rule during the data

mining phase for rule generation. However, the statistical features may also be evaluated in the specific clinical epidemiology in which the KB is to be used. Hence, the thresholds for rule application may be adjusted dynamically to the hospital/department where the KB is used. The value of such statistical parameters is an indication of how likely it is that a rule will fire (sensitivity), and with what confidence the predicted event will actually happen (predictive value). Thus, their usefulness is two-fold: (a) filtering the rules according to the desirable sensitivity and predictability, in order to avoid over-alerting and (b) adjusting the KB to a specific country, hospital and department by evaluating the parameters locally.

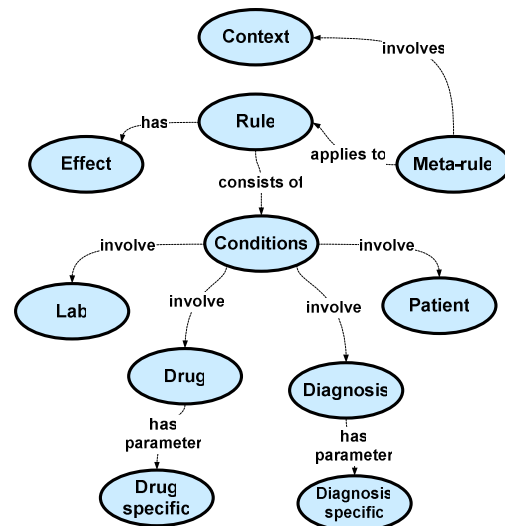


Figure 1. Conceptual schema of the PSIP knowledge model.

The *Condition* consists of the value of one variable, a comparison operator and one reference value, while a TRUE/FALSE value is returned according to the result of the comparison. The variable checked by each condition may be either a prescribed or suppressed drug, a diagnosis (acute or chronic patient condition), and a patient characteristic (e.g., age, weight) or a lab value (e.g. the result of a biology test) compared to a threshold. Each of these variables is defined as a concept, since it corresponds to a set of attributes.

The approach adopted for expressing diagnoses and medication was to define a set of pseudo-variables. These express groups of conditions or drugs, which for the purposes of ADEs have the same effect. *Diagnosis* and *Drug* correspond to such PSIP-defined pseudo-variables and are implemented as intermediate rules. Each instance of *Diagnosis* or

Drug is associated with a set of specific diagnoses or specific drugs, respectively. The population of *Diagnosis* and *Drugs* with groups of specific diagnoses and types of medications with similar effects has resulted by the knowledge discovery phase⁸. It is considered as part of the domain knowledge and is subject to refinements and updates. The *Patient* concept is associated with patient data that participate in *Condition*.

The mechanism that controls the applicability of rules to a particular context and the subset of alerts to be forwarded as decision support output, as formulated in (3), is modeled via meta-rules. For this purpose, the concepts *Meta-rule* and *Context* are defined. Each instance of *Meta-rule* is a filtering mechanism which controls the subset of basic rules that will be visible to the decision support service. The parameters considered by each meta-rule are defined in *Context* (i.e. whatever is related to the environment and the needs of the targeted user).

Results

The presented knowledge framework has been implemented in GASTON (<http://www.medecs.nl/>). The core of GASTON consists of a guideline representation formalism¹³, relying on a combination of knowledge representation approaches and concepts, i.e., primitives, problem-solving methods and ontologies, upon which the conceptual schema illustrated in Figure 1 has been implemented. Currently, the KB is populated with 245 ADE rules originated from data mining (that correspond to about 400 intermediate rules for pseudo-variables concerning drugs, diagnosis, patient data and lab results) and stored as Protégé frames (<http://protege.stanford.edu/>) as well as in an XML-based representation (both supported/offered by GASTON). Including standard classifications, the current KB encapsulates 46,702 classes and 47,810 instances.

Furthermore, an SQL-based, parameterized interface has been developed to interface the KB with the VIDAL drug knowledge source, as an additional source for ADE prevention that the CDSS modules can exploit. A mapping of each concept/attribute defined in the knowledge model with the relevant fields/tables defined in the data model¹⁰ enabled the development of an XML-based request-response interface for querying the KB. In addition, a basic terminological reasoning mechanism has been developed to appropriately expand/narrow terms contained in the data requests (queries), according to

the semantics of the intermediate rules as defined/implied in source knowledge.

Figure 2 illustrates an example rule, implemented as a guideline in GASTON. The rule is: “*renal failure & NO Suspension of antithrombotic & high weight heparin & NO Suspension of proton pump inhibitor & NO sympathomimetic & NO Suspension of high weight heparin* → *Appearance of hyperkalemia ($K^+ > 5.3$)*”. This rule consists of six conditions, one related to diagnosis and the rest to drugs. The initial conditions checked via intermediate rules and the procedural logic according to which the rule is implemented are depicted.

Several options are explored for contextualizing the KB, according to the inference mechanism described. As knowledge sharing between the elicitation and engineering phases in PSIP is based on custom XML schemas, mechanisms for automatically importing source knowledge into the KB have been developed, performing in advance syntactic verification. Knowledge maintenance mechanisms have also been elaborated, with particular emphasis on the capability of defining/altering rules’ metadata in the runtime mode of the CDSS.

Verification of the populated KB was also performed, aiming to identify whether the implemented rules fire on the same cases as identified by the data mining techniques. For this purpose, the CDSS responses for the same clinical database used in the data mining, as well as a list of hospital stays (i.e. a hospitalization period for a patient) fitting each rule conditions were used as input. It has to be noted that each stay corresponds to many cases of KB request and inference. The outcome of this procedure was assessed via contingency tables¹⁴, comparing the identities of the KB rules fired on each hospital stay with its counterpart from the data mining. The verification was based on a selected set of ADEs, i.e. hyponatremia, hyperkalemia, coagulation problems and renal insufficiency. In this procedure, a total of 21,331 hospital stays comprising 3,796,918 classification tasks were analyzed. Overall, the verification process, which was iteratively performed as consecutive versions of the KB were released, identified about 5-10% of the hospital stays in which potential errors in rules implementation had to be analyzed and corrected. This was primarily due to development errors and ambiguities in the description of source knowledge.

Conclusion

In this paper, we presented a knowledge-based approach for constructing CDSSs for ADE

prevention. The selection of the knowledge engineering approach and the design of the proposed knowledge framework have been primarily driven by the knowledge sources characteristics and the problem to be solved. The verification of the KB content as well as the inference and query mechanisms indicated that the approach followed is effective and technically sound. A clinical validation phase of the decision support services offered by the proposed knowledge framework is currently conducted to assess the added value that may be introduced in actual clinical settings.

Acknowledgement

The research leading to these results has received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under Grant Agreement n° 216130 – the PSIP project.

References

1. Bates DW, et al. Detecting adverse events using Information Technology. J Am Med Inform Assoc 2003;10(2):115–28.
2. Morimoto T, Gandhi TK, Seger AC, Hsieh TC, Bates DW. Adverse drug events and medication errors: detection and classification methods. Qual Saf Health Care 2004;13(4):306-14.
3. Murff HJ, Patel VL, Hripcsak G, Bates DW. Detecting adverse events for patient safety research: a review of current methodologies. J Biomed Inform 2003;36(1-2):131-43.
4. Estacio-Moreno A, Toussaint Y, Bousquet C. Mining for adverse drug events with formal concept analysis. Stud Health Technol Inform 2008;136:803-8.
5. Bousquet C, Henegar C, Lillo-Le Louët A, Degoulet P, Jaulent M-C. Implementation of automated signal generation in pharmacovigilance using a knowledge-based approach. Int J Med Inform 2005;74(7-8):563-71.
6. Stephens S, Morales A, Quinlan M. Applying Semantic Web technologies to drug safety determination. IEEE Intell Syst 2006;21(1):82-6.
7. Payne VL, Metzler DP. Hospital Care Watch (HCW): an ontology and rule-based intelligent patient management assistant. IEEE Symp Proc on CBMS 2005;2005:479-84.
8. Chazard E, et al. Data-mining-based detection of adverse drug events. Stud Health Technol Inform 2009;150:552-6.
9. Koutkias V, et al. A knowledge engineering framework towards clinical support for adverse drug event prevention: The PSIP approach. Stud Health Technol Inform 2009;148:131-41.
10. Chazard E, et al. Detection of adverse drug events: proposal of a data model. Stud Health Technol Inform 2009;148:63-74.
11. Ligêza A. Logical foundations for rule-based systems. Springer-Verlag Berlin Heidelberg 2006.
12. De Clercq P, Kaiser K, Hasman A. Computer-interpretable guideline formalisms. Stud Health Technol Inform 2008;139:22-43.
13. de Clercq PA, Hasman A, Blom JA, Korsten HHM. Design and implementation of a framework to support the development of clinical guidelines. Int J Med Inform 2001;64(2-3):285-318.
14. Brender J, Talmon J, Egmont-Petersen M, McNair P. Measuring quality of medical knowledge. Proc of MIE 1994:69-74.

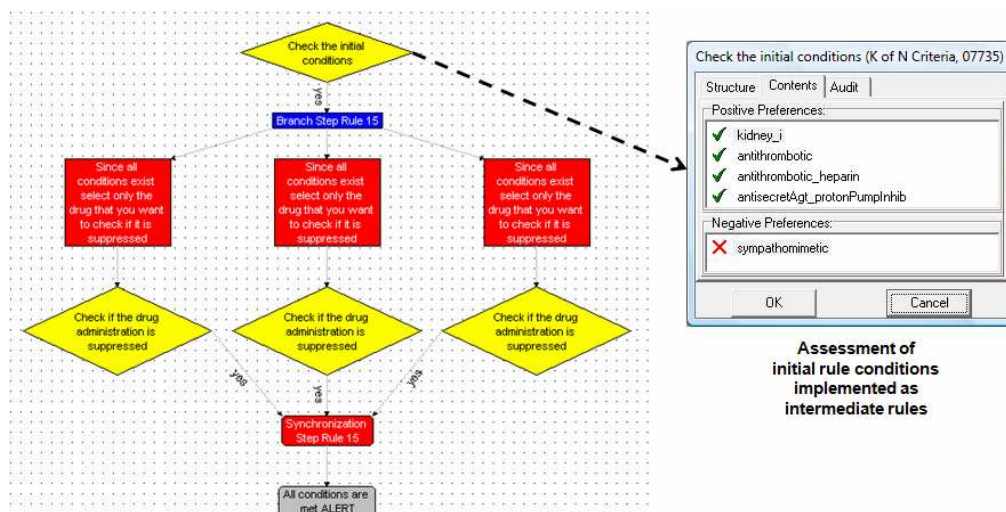


Figure 2. Example rule implementation as guideline illustrated in the knowledge authoring tool employed.