

# Risk Mediation in Association Rules

## The Case of Decision Support in Medication Review

Michiel C. Meulendijk<sup>1</sup>(✉), Marco R. Spruit<sup>2</sup>,  
and Sjaak Brinkkemper<sup>2</sup>

<sup>1</sup> Department of Public Health and Primary Care,  
Leiden University Medical Center, Leiden, The Netherlands  
m. c. meulendijk@lumc.nl

<sup>2</sup> Department of Information and Computing Sciences, Utrecht University,  
Utrecht, The Netherlands  
{m. r. spruit, s. brinkkemper}@uu.nl

**Abstract.** We propose a model for the incorporation of risk in association rule application. We validate this model using data gathered in a randomized controlled trial from a recommender system for medication reviews in primary care. The model's outcomes are found to have predictive value when tested against decisions made by physicians on 261 patients' health records.

**Keywords:** Association rules · Decision support · Risk management

## 1 Introduction

Association rule mining remains one of the most prominent knowledge discovery methods in existence. Association rules have been implemented in countless software applications in a wide variety of domains, including the domain of primary care [1].

The typical example of association rule implementation, the 'market basket', reveals one of its weaknesses: grocery items that are frequently purchased together form an association rule, which is used for personalized adverts. In a safe, insensitive domain such as shopping this poses no problems, but in a precarious domain with vulnerable datasets, where association rules can have potentially far-reaching implications, these kinds of associative suggestions are risky.

Therefore, in this paper we propose a model for the incorporation of risk in association rules. We implement and validate it in a primary care setting.

## 2 Background

### 2.1 Recommender Systems in Precarious Domains

The impact association rules may have depends on the sensitivity of the dataset on which they are applied. This danger involved in applying discovered association rules to sensitive datasets in precarious domains has not received substantial attention. For the remainder of this paper, we define *precarious domains* as 'domains in which

association rules' consequences have potentially major impact on its datasets'. An example of a precarious domain is that of primary care.

### 2.2 Risk as a Post-mining Metric

Risk management involves a set of coordinated activities to direct and control an organization with regard to risk. Risk is typically understood as the probability of loss in any given situation, or, in a more generic definition, the effect of uncertainty on objectives. In risk management, risk is usually incorporated as a function of these uncertain effects' impact and their likelihood of occurring.

Association rules and risk share the concept of probability; association rules are expressed with a degree of confidence, while risk incorporates the likelihood of dangerous consequences. While association rules' impact is not a standardized concept, the implications their consequences have imply its risk's severity (Table 1).

**Table 1.** Relations between association rules' characteristics and risk.

		<i>Risk Probability</i>			
		High	Low		
<i>Rule Impact</i>	High	✘	?	High	<i>Risk Severity</i>
	Low	?	✓	Low	
		Low	High		
		<i>Rule Confidence</i>			

### 3 Risk Model Formulation

Recommender systems, whether powered by explicit knowledge bases or implicit content-based or collaborative-based filtering, depend on inference rules. Inference rules are logical functions which analyze premises and, based on their syntax, return conclusions. Recommender systems depend on datasets containing all relevant items to trigger their inference rules. In propositional logic inference rules can be written as  $x \rightarrow y$ , with a dataset  $D = \{d_1, \dots, d_n\}$  and  $x \in D$ . Thus, for a specific rule dataset  $D$  contains its premises, along with other items, but never its consequence.

The risk associated with a rule is a function of its unwanted consequences and their likelihood of occurring. The formula to determine the risk of an inference rule  $x \rightarrow y$  reads:

$$risk(x \rightarrow y) = (1 - probability(x \rightarrow y)) * \sum_{i=D,y} severity(i) \tag{1}$$

The probability of the consequences being unacceptable is one minus the confidence with which the rule is accepted. In the case of association rules that have been discovered using common algorithms, the probability in the formula is equal to its confidence.

The severity of a rule's risk is the sum of the impact of the objects associated with it. This comprises not just the danger associated with the inference's consequence  $y$ , but also detrimental characteristics of items in dataset  $D$ ; it may be riskier to perform a certain action on a vulnerable dataset than on a safer one. The danger associated with these items can be estimated through risk formulas as well. As such, when implemented in a domain, the final formula results in the sum of its associated objects' risks, multiplied with its rule's inverse probability.

## 4 Implementation

The above-mentioned formulae are demonstrated and validated by applying them to the STRIP Assistant (STRIPA), a recommender system for medication reviews in primary care [1]. STRIPA's hybrid rule base consists of guidelines and inference rules acquired through association rule mining, and has been shown to be effective [2].

### 4.1 Health Records' Risk

The dataset in STRIPA consists of a patient's health record. The items in this dataset, such as diseases or drugs, serve as premises for the system's inference rules. All inference rules modify one or more drugs by prescribing new medicines, removing existing ones, or adjusting their dosages. A dataset  $D$  comprising a certain patient's diseases, drugs, contra-indications, allergies, and measurements can be described as a set:

$$D = \left\{ \begin{array}{l} disease_1, \dots, disease_k; drug_1, \dots, drug_l; \\ contraindication_1, \dots, contraindication_m; \\ measurement_1, \dots, measurement_n; \\ allergy_1, \dots, allergy_p \end{array} \right\} \quad (2)$$

Following  $x \in D$  and the fact that all inference rules adjust drugs, the implemented risk formula reads:

$$risk(x \rightarrow drug) = (1 - probability(x \rightarrow drug)) * (severity(D) + severity(drug)) \quad (3)$$

### 4.2 Health Records' Severity

Determining the risk associated with a patient's health record, or dataset, involves taking into account a multitude of domain-dependent variables, such as his or her age, frailty, physical properties, and cognitive state. This results in the severity of a dataset  $D$  being the sum of its domain-relevant, patient-specific, risk factors:

$$severity(D) = \sum_{riskFactor \in D} riskFactor \quad (4)$$

The Dutch multidisciplinary guideline for polypharmacy in elderly people<sup>1</sup> proposes seven factors that increase patients' risk of harm due to inappropriate drug use: age (over 65 years old), polypharmacy (using four or more drugs simultaneously), impaired renal function, impaired cognition, frequent falling, decreased therapy adherence, and living in a nursing home. In this study's dataset, four of these risk factors were available: age, polypharmacy, impaired renal function, and impaired cognition.

### 4.3 Drug Severity

A drug's severity can be expressed as a function of its dose, or toxicity, and its adverse effects, or harm:

$$severity(drug) = toxicity(drug) * harm(drug) \quad (5)$$

An adverse effect is a response to a drug that is noxious and unintended and occurs at normal doses. These effects are usually classified in terms of their likelihood of occurring. Their frequency and potential impact are used to determine drugs' safety for prescribing. As such, the function of the number of adverse effects a drug has and their frequency can be useful to determine a substance's potential harm:

$$harm(drug) = \sum_{e \in E} e.frequency \quad (6)$$

The sets of adverse effects, defined per active substance, were retrieved from a database maintained by the Royal Dutch Pharmacists Association<sup>2</sup>. Adverse effects are classified due to their likelihood of occurring: often (over 30% of patients or more), sometimes (10–30% of patients), rarely (1–10% of patients), and very rarely (less than 1% of patients). As such, each drug's adverse effects can be described as a set  $E = \{e_1, \dots, e_n\}$ , where  $e_i = (id, frequency)$ .

The definition provided for adverse effects also takes into account the prescribed dosage, something not accounted for in the formula above. Research has shown that the probability of adverse effects generally increases with higher dosages being used. For each drug, the World Health Organization has defined an average strength with which it is typically prescribed. This Defined Daily Dose<sup>3</sup> (DDD) is the assumed average maintenance dose per day for a drug used for its main indication in adults. Dividing a patient's actual daily dosage of a drug by the DDD provides a factor that can be used to relativize the drug's risk. Its toxicity can thus be calculated as such:

$$toxicity(drug) = \frac{prescribedDailyDose(drug)}{definedDailyDose(drug)} \quad (7)$$

<sup>1</sup> <https://www.nhg.org/themas/publicaties/multidisciplinaire-richtlijn-polyfarmacie-bij-ouderen>.

<sup>2</sup> <http://www.apotheek.nl/medicijnen>.

<sup>3</sup> [http://www.whooc.no/ddd/definition\\_and\\_general\\_considera/](http://www.whooc.no/ddd/definition_and_general_considera/).

## 5 Validation

The risk model was validated by comparing its predictions with actual actions taken by experts. STRIPA was used on real patient cases by dedicated teams of GPs and pharmacists for the duration of a year as part of a randomized controlled trial, which was performed in 25 general practices in Amsterdam, the Netherlands and included 500 patients [3]. For the 261 of these patients that were placed in the intervention arm, four teams consisting of one GP and one pharmacist each used the software to optimize their medical records. The users would respond to patient-specific advice, recommending them to prescribe new drugs for particular diseases. Their responses to advices were gathered; each time a suggestion was heeded or ignored, the instance, along with relevant patient case information, was logged. A total of 776 responses to advices, of which 311 were heeded, has been gathered and will be used to validate the risk model.

Assuming that users will strive for the minimization of risk, we hypothesize that users will have chosen the least risky option whenever possible. Based on the assumption that riskier patients – i.e. patients who have multiple risk factors in their dataset (or health record) – are best served with as little change to their drug regimen as possible, it was hypothesized that the higher an action's risk was, the least likely it was to be performed by users. To test this, the risk factors of each generated recommendation were calculated; its proposed drug's risk and the relevant patient's risk factors were summed according to the introduced model. An independent t-test affirmed the hypothesis, showing a statistical difference in the risk associated with proposed actions which were followed ( $M = 2.42$ ,  $SD = 0.57$ ) and the risk of proposed actions which were not followed ( $M = 2.57$ ,  $SD = 0.60$ );  $t(623) = 3.040$ ,  $p = .002$ .

## 6 Conclusion

In this study, we explored the potential usefulness of the concept of risk in association rules. Validation shows that the risk model has predictive power in the domain of medication reviews. Application examples can be found in a technical report [4].

We would like to thank Pieter Meulendijk for his contributions to the conceptual model following his expertise in risk management.

## References

1. Meulendijk, M., et al.: STRIPA: A Rule-Based Decision Support System for Medication Reviews in Primary Care. In: Proceedings of ECIS 2015, Münster, Germany (2015)
2. Meulendijk, M., et al.: Computerized decision support improves medication review effectiveness: an experiment evaluating the STRIP Assistant's usability. *Drugs Aging* **32**(6), 495–503 (2015)
3. Meulendijk, M., et al.: Efficiency of clinical decision support systems improves with experience. *J. Med. Syst.* **40**(4), 76–82 (2016)
4. Meulendijk, M., Spruit, M., Brinkkemper, S.: Risk Mediation in Association Rules: Application Examples, Utrecht University, UU-CS-2017-004 (2017)