Chapter 12

Strong Symmetric Association Rules and Interestingness Measures

Agathe Merceron
University of Applied Sciences TFH Berlin, Germany

ABSTRACT

Strong symmetric association rules are defined as follows. Strong means that the association rule has a strong support and a strong confidence, well above the minimum thresholds. Symmetric means that \(X \rightarrow Y\) and \(Y \rightarrow X\) are both association rules. Common objective interestingness measures such as lift, correlation, conviction or Chi-square tend to rate this kind of rule poorly. By contrast, cosine is high for such rules. However, depending on the application domain, these rules may be interesting regarding criteria such as unexpectedness or actionability. In this chapter, the authors investigate why the above-mentioned measures, except cosine, rate strong symmetric association rules poorly, and show that the underlying data might take a quite special shape. This kind of rule can be qualified as rare, as they would be pruned by many objective interestingness measures. Then the authors present lift and cosine in depth, giving their intuitive meaning, their definition and typical values. Because lift has its roots in probability and cosine in geometry, these two interestingness measures give different information on the rules they rate. Furthermore they are fairly easy to interpret by domain experts, who are not necessarily data mining experts. They round off our investigation with a discussion on contrast rules and show that strong symmetric association rules give a hint to mine further rare rules, rare in the sense of a low support but a high confidence. Finally they present case studies from the field of education and discuss challenges.

INTRODUCTION

Association rules are very useful and used in Educational Data Mining, which is a particular application domain: data used to extract association rules come from learning systems. Examples in this application domain include Merceron & Yacef (2004) in which association rules are used to find mistakes often made together while students solve
exercises in propositional logic. Wang (2006) and Wang & Shao (2004) used association rules, combined with other methods, to personalize students’ recommendation while browsing the web. Minaei-Bidgoli et al. (2003) used them to find various associations of student’s behavior in their Web-based educational system LON-CAPA. Lu (2004) used fuzzy rules in a personalized e-learning material recommender system to discover associations between students’ requirements and learning materials. Romero et al. (2002) combined them with genetic programming to discover relations between knowledge levels, times and scores that help teachers modify the course’s original structure and content. A more extensive overview is given in Romero & Ventura (2007).

We extracted association rules from the data stored by the Logic-ITA, an intelligent tutoring system for formal proof in propositional logic Merceron & Yacef (2003). Our aim was to know whether there were mistakes that often occurred together while students are training. This information could be used to act on the course or on the tutoring system itself. The results gave strong symmetric associations between three mistakes. Strong means that all associations had a strong support and a strong confidence. Symmetric means that $X \rightarrow Y$ and $Y \rightarrow X$ were both extracted association rules.

It is well known that even rules with a strong support and confidence may in fact be uninteresting (Han & Kamber, 2006). This is why, once the association rule $X \rightarrow Y$ has been extracted, it is wise to double-check how much $X$ and $Y$ are related. About 20 objective measures have been proposed in the literature to do so. We explore in this paper a few measures in the context of our data. We have observed that common objective interestingness measures such as lift, correlation, conviction or Chi-square tend to rate our strong symmetric association rules poorly. By contrast, cosine is high. However, the extracted rules are actionable in the sense that the information they give can be used to act on the course and on the tutoring system. Thus this kind of rule can be qualified as rare as they are actionable but would be pruned by a number of objective interestingness measures.

In this chapter, we investigate various interestingness measures for the strong symmetric association rules we have obtained and show that the underlying data have a quite special shape. Further, we look for restricting the number of interestingness measures a teacher, who is not necessarily a data mining expert, has to consider while picking out meaningful rules from all the extracted rules. Therefore we focus on two measures, lift and cosine, and investigate them in more depth. The reason to focus on these two measures is twofold. First, because lift has its roots in probability and cosine in geometry, these two interestingness measures give different information on the rules they rate. Second, they are fairly easy to interpret by domain experts, who are not necessarily data mining experts. We round off our investigation of strong symmetric association rules with a section on contrast rules. We investigate whether strong symmetric association rules can give a hint to mine other rare association rules, rare having the more classical meaning of rules with a low support but a strong confidence. This is indeed the case, however lift and cosine rate these rare rules poorly. Finally we present case studies that use cosine and lift to prune the extracted association rules, the case study of the Logic-ITA already mentioned and a case study with the Learning Management System Moodle. This chapter merges and deepens results presented in Merceron & Yacef (2007) and Merceron & Yacef (2008).

**BACKGROUND**

As has been said above, it is well known that, once association rules have been extracted, a prune step is necessary. This step involves using further interestingness measures to double check how much $X$ and $Y$ are related. Unfortunately, no
17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

www.igi-global.com/chapter/strong-symmetric-association-rules-interestingness/36907?camid=4v1


www.igi-global.com/e-resources/library-recommendation/?id=7

Related Content

A Framework to Detect Disguised Missing Data
www.igi-global.com/chapter/framework-detect-disguised-missing-data/46888?camid=4v1a

Learning SKOS Relations for Terminological Ontologies from Text
Wei Wang, Payam M. Barnaghi and Andrzej Bargiela (2011). Ontology Learning and Knowledge Discovery Using the Web: Challenges and Recent Advances (pp. 129-152).
www.igi-global.com/chapter/learning-skos-relations-terminological-ontologies/53884?camid=4v1a

A Successive Decision Tree Approach to Mining Remotely Sensed Image Data
www.igi-global.com/chapter/successive-decision-tree-approach-mining/24903?camid=4v1a

Developing an Online Course On-the-Fly with an IR State of Mind
Kevin Gosselin and Hansel Burley (2012). Teaching Cases Collection (pp. 174-193).
www.igi-global.com/chapter/developing-online-course-fly-state/60847?camid=4v1a