Learning from Dissociations*

Choh Man Teng

Institute for Human and Machine Cognition
University of West Florida, Pensacola FL 32501, USA
cmteng@ai.uwf.edu

Abstract. Standard association rules encapsulate the *positive* relationship between two sets of items: the presence of X is a good predictor for the simultaneous presence of Y. We argue that the absence of an association rule conveys valuable information as well. Dissociation rules are rules that capture the *negative* relationship between two sets of items: the presence of X *and* z is *not* a good predictor for the presence of Y. We developed a representation for augmenting standard association rules with dissociation information, and presented some experimental results suggesting that such augmented rules can improve the quality of the associations obtained, both in terms of rule accuracy and in terms of using these rules as a guide to making decisions.

1 Introduction

An association rule [1] is meant to denote a positive relationship between two sets of items. For example, the rule

$$\text{potato chips} \Rightarrow \text{coke} \qquad (^*1)$$

says that people who buy potato chips also tend to buy coke. While this rule provides us with useful information, in many cases it may be beneficial to, in addition, take into account possible exceptions to the association rules we have discovered. For example, we might note that the rule

$$\text{potato chips and pepsi} \Rightarrow \text{coke} \qquad (^*2)$$

is *not* an acceptable association rule. What are the implications of the absence of this rule? Loosely speaking, rule (*1) is concerned with people, or more specifically transactions of people, who buy potato chips in general, and rule (*2) is concerned with a special subsection of this population, namely people who buy pepsi in addition to potato chips. These two groups of people have different buying behaviors with respect to coke: the former group often buy coke, the latter often do not.

Let us denote the absence of rule (*2) by

$$\text{potato chips and pepsi} \not\Rightarrow \text{coke.} \qquad (^*3)$$

* This work was supported by NASA NCC2-1239.

Y. Kambayashi, W. Winiwarter, M. Arikawa (Eds.): DaWaK 2002, LNCS 2454, pp. 11–20, 2002.

Such rules indicate a *negative* relationship between two sets of attributes. We will use the term "dissociations" for this relationship, in contrast to *positive* "associations". We will give a more formal definition of dissociations in Section 2.

Our motivation is closely related to that discussed in [2]. Our approaches differ, however, in that the formulation in the present paper does not rely on the existence of a strict taxonomy, which may be difficult to construct. We make use of the information contained in the "parent" rule in a more general way.

1.1 Subclass Dissociations

Dissociations by themselves can be quite arbitrary:

$$\begin{aligned} \text{potato chips } &\not\Rightarrow \text{cotton swabs;} \\ \text{light bulbs} &\not\Rightarrow \text{wine openers.} \end{aligned}$$

We focus here on a case of particular interest, namely, when a subclass exhibits a dissociation that contradicts an association in the original class. Symbolically, this can be represented as

$$\begin{aligned} X &\Rightarrow Y; \\ XZ &\not\Rightarrow Y. \end{aligned} \qquad (^*4)$$

Rules (^{*1}) and (^{*3}) above are an example of such a situation. The products coke and pepsi are competitors, and people who buy one kind may be disinclined to buy the other kind at the same time.

Subclass dissociations can also indicate a special niche in the market. Consider the two rules

$$\begin{aligned} \text{yeast } &\Rightarrow \text{flour;} \\ \text{yeast and hops } &\not\Rightarrow \text{flour.} \end{aligned} \qquad (^*5)$$

In general people who get yeast also get flour, to bake sumptuous cakes, but if they in addition also get hops, they are more likely to be making beer and have no use for flour. Here the beer-brewers constitutes a niche with a buying pattern that is different from that of the general population.

We can make this even more explicit by eliciting rules such as

$$\begin{aligned} \text{yeast } &\Rightarrow \text{flour;} \\ \text{(yeast and) brewer's yeast } &\not\Rightarrow \text{flour.} \end{aligned} \qquad (^*6)$$

Here there is a strict set-inclusion relationship between yeast and brewer's yeast. Again, brewer's yeast picks out a special subclass of people who are not very interested in flour.

Taxonomically speaking, in the first case (with rules (^{*1}) and (^{*3})), the two items represented by Y (coke) and Z (pepsi) in the pair of rules in (^{*4}) are comparable subclasses of the same superclass "cola". In the rules in (^{*6}), Z is a strict subclass of X. In the rules in (^{*5}), it is intended that "yeast and hops" picks out the same subclass as "brewer's yeast", although the taxonomic relationship between the items are only implicit and approximate.