

# Objective Novelty of Association Rules: Measuring the Confidence Boost<sup>1</sup>

José L Balcázar\*

\*Departamento de Matemáticas, Estadística y Computación  
Universidad de Cantabria  
Santander, Spain  
joseluis.balcazar@unican.es

**Résumé.** On sait bien que la confiance des règles d'association n'est pas vraiment satisfaisant comme mesure d'intérêt. Nous proposons, au lieu de la substituer par des autres mesures (soit, en l'employant de façon conjointe a des autres mesures), évaluer la nouveauté de chaque règle par comparaison de sa confiance par rapport á des règles plus fortes qu'on trouve au même ensemble de données. C'est á dire, on considère un seuil "relative" de confiance au lieu du seuil absolute habituel. Cette idée se précise avec la magnitude du "confidence boost", mesurant l'increment relative de confiance prés des règles plus fortes. Nous prouvons que notre propose peut remplacer la "confidence width" et le blockage de règles employés a des publications précédentes.

## 1 Introduction

The outcome of a Data Mining project is expected to offer some degree of novelty. However, to formally study the novelty of Data Mining results is far from being a trivial task. Novelty refers to facts that are somehow unexpected, and therefore some expectation, lower than actually found, must exist, due to some alternative facts or prediction mechanisms.

Here we consider the specific task of Association Rule Mining, and follow up a recent proposal that, as a minimum, each rule should be evaluated for novelty according to the rest of the rules mined, treated as "alternative" mechanism [4], [5]. These works propose the *confidence width* as a measure of a relative form of objective novelty or surprisingness of each individual rule with respect to other rules that hold in the same dataset, and propose also to *block* some rules in case they do not bring in enough novelty with respect to a "blocker" itemset. Our contribution here is a new novelty notion, the *confidence boost*, similar in its syntactic definition to confidence width, but very different in its semantics; its main trait is that it encompasses at once both the bound on the confidence width and the ability to detect that a rule would be blocked, so that the confidence boost bound embodies both of the bounds proposed in [5].

The notions of confidence width and rule blocking from [5] are similar to the "pruning" proposal from [14], in that the intuition is the same; also our proposal here follows an analogous intuitive path. Major differences are that, in the proposals we discuss, a large portion of the pruning becomes unnecessary because we work on minimum-size bases (representative rules, proposed in [1], in [13], and in [16]) and, more importantly, that the pruning in [14]