**IGB: A New Informative Generic Base of Association Rules**

Gh. Gasmi¹, S. Ben Yahia¹,², E. Mephu Nguifo², and Y. Slimani¹

¹ Département des Sciences de l’Informatique, Faculté des Sciences de Tunis, Campus Universitaire, 1060 Tunis, Tunisie
{sadok.benyahia, yahya.slimani}@fst.rnu.tn
² Centre de Recherche en Informatique de Lens-IUT de Lens, Rue de l’Université SP 16, 62307 Lens Cedex mephu@cril.univ-artois.fr

**Abstract.** The problem of the relevance and the usefulness of extracted association rules is becoming paramount, since an overwhelming number of association rules may be derived from even reasonably sized real-life databases. A possible solution consists in using results of Formal Concept Analysis to generate a generic base of association rules. This set, of reduced size, makes it possible to derive all the association rules via an adequate axiomatic system. In this paper, we introduce a novel generic and informative base of association rules, conveying two types of knowledge: "factual" and "implicative". We present also a valid and complete axiomatic system allowing to derive the set of all association rules. Results of the experiments carried out on real-life databases showed important profits in terms of compactness of the introduced generic base.

**Keywords:** Association rules, Generic base, Galois connection, Axiomatic system.

1 Introduction

The problem of the relevance and usefulness of extracted association rules is of primary importance. Indeed, in most real life databases, thousands and even millions of high-confidence rules are generated, among which many are redundant. This problem encouraged the development of tools for rule classification, according to their properties, for rule selection according to user-defined criteria, and for rule visualization. The Selection without loss of information is mainly based on the extraction of a generic subset of all association rules, called generic base, from which the remaining (redundant) association rules are generated.

In this paper, we introduce a new generic base of association rules called **IGB**. Through **IGB**, we introduce a novel characterization of generic association rules instead of the classical one, *i.e.*, exact and approximative. In fact, we shall distinguish between the "factual" and the "implicative" generic association rule. Indeed, a factual generic association rule, fulfilling the premise part emptiness,
permits to highlight item correlation without any conditionality. However, for an
implicative generic association rule, where the premise part is not empty, item
correlation is conditioned by the existence of premise items. The introduced
generic base $\mathcal{IGB}$ fulfills the ”informativeness property”, i.e., the support and
the confidence of the derived association rules can be exactly retrieved. In order
to derive valid association rules from the $\mathcal{IGB}$ base, we introduce an axiomatic
system, which it is shown to be valid and complete.

We conducted several experiments on typical benchmarking datasets to assess
the $\mathcal{IGB}$ compactness. The introduced generic rule characterization permitted
to explain the ”atypical” behavior of the variation of the reported generic associa-
tion rules number versus the variation of the $\text{minconf}$ value, i.e., the number
of the reported rules does not necessarily decrease with the augmentation of the
$\text{minconf}$ value.

The remainder of the paper is organized as follows. Section 2 presents the
basic mathematical foundations for the derivation of generic bases of association
rules. We devote section 3 to a review of the literature relating to the extraction
of the generic bases. Section 4 introduces a novel informative base of generic
association rules and the associated axiomatic system. Results of the experiments
carried out on real-life databases are reported in section 5. The conclusion and
future work are presented in section 6.

2 Mathematical Background

Due to lack of available space, interested reader for key results from the Galois
lattice-based paradigm in FCA is referred to [1].

**Frequent Closed Itemset:** An itemset $I \subseteq \mathcal{I}$ is said to be closed if $I = \omega(I)$,
and is said to be frequent with respect to the $\text{minsup}$ threshold if $\text{support}(I) = \frac{|\mathcal{P}(I)|}{|\mathcal{O}|} \geq \text{minsup}$ [2].

**Minimal Generator:** An itemset $g \subseteq I$ is said to be minimal generator of a
closed itemset $I$, if and only if $\omega(g) = I$ and does not exist $g' \subseteq g$ such that
$\omega(g') = I$ [3, 4].

**Iceberg Galois Lattice:** When only frequent closed itemsets are considered
with set inclusion, the resulting structure $(\hat{\mathcal{L}}, \subseteq)$ only preserves the join oper-
ator [1]. This is called a join semi-lattice or upper semi-lattice. In the remaining of
the paper, such structure is referred to as ”Iceberg Galois Lattice” [5]. Therefore,
given an Iceberg Galois lattice in which each closed itemset is ”decorated” with
its associated list of minimal generators, generic bases of association rules can be
derived in a straightforward manner. Indeed, generic approximative rules repre-
sent ”inter-node” implications, assorted with the confidence measure, between
two comparable equivalence relation classes, i.e., from a sub-closed-itemset to a
super-closed-itemset when starting from a given node in the partially ordered
structure [6].