

## Association Rules – Review

I have annotated the document where I have spotted typos or formalities that are wrong. Apart from that, I still have some general remarks:

To me, it is not entirely clear how the transactions refer to the association rules in the formal definition. First, it appears as if a transaction simply consists of all the items in the basket, as seen in Table 1. However, you define a transaction as only a subset of items from the basket  $I$ . Also, do we even need transactions to define association rules? Perhaps it would be better to first define association rules and then define transactions as a separate concept. Then you could proceed to explain how association rules can be obtained from a list of transactions.

You mention the support constraint of the association rule mining process twice: First you mention it in the definition and then again under additional constraints. Also, you introduce the syntactic constraints but you don't explain how one could fulfill them during mining.

What exactly is the use of the lattice structure? You mention it and show a figure, but you don't really explain what it is or what it does. Why do we need a lattice structure instead of simply computing the subsets of  $I$ ?

As I have outlined in one annotation, all figures as well as algorithms should have a caption that describes what their purpose is. For example, you can consider putting Table II-V in one figure with a caption that states what we can see from these tables. Also you might consider not placing the figures and algorithms in an inline fashion as this is very unconventional.

I think you should describe the difference between the SETM and AIS algorithm in more detail as I haven't really understood what the difference is from reading your paper. For example, it is not clear how the SETM algorithm addresses the drawbacks of the AIS algorithm.

What is the advantage of Apriori-TID in contrast to the standard Apriori algorithm? I suggest you briefly compare them or describe why one would choose one algorithm over the other. Besides that, I have to admit that I couldn't quite follow your explanation of the Apriori-TID algorithm. What exactly are the sets  $\overline{C_k}$  and what is their purpose?

There is a mistake in the proof for the theorem for confidence-based pruning. You state that "Since  $X'$  is a subset of  $X$ ,  $|X'| \geq |X|$ ". However, the cardinality of  $X'$  is less (or equal) if  $X'$  is subset and can never be greater. Maybe you mean the number of transactions containing  $X'$  instead of the cardinality?

Despite those points of criticism, I like the general structure of your paper. The sections are arranged in a well-structured way and it is clear what you want to explain in each section. However, as I have outlined above, the content of the sections themselves is sometimes a bit unclear and I suggest you resolve all formal problems with your paper (maybe also take a look at the IEEE guidelines).