Module 9 describes association rules – a.k.a. the study of “what goes with what” – as a means of picking out item clusters in sets of transactions or events (called transactional datasets). Popularized by Amazon in the world of online retail, association rules works to predict users’ interests and buying patterns, helping to suggest additional items for them that they might also be interested in. As Module 9 goes on to note, the first step of association rules is to consider all possible combinations of items in the given dataset of transactions, in order to come up with the rules that delineate possible relationships between items. In practice though, analyzing all possible combinations of items is time intensive – imagine trying to review all possible combinations of items for sale on Amazon, for example. We can take a shortcut around this by looking only at frequent itemsets - item combinations that occur above a certain frequency in the dataset. The level of frequency that determines those frequent itemsets is in turn determined by the support parameter.

Shmueli’s Data mining for Business Analytics refers to support as the “number of transactions that include both the antecedent and consequent itemsets”. In other words, within a given transactional dataset, how many times did two possible combinations of items occur as part of the same transaction? In the context of online retail for example, this could manifest as how many times a certain book A and books B and C were bought together; if there are 8 transactions, of which 2 contain all of books A B and C, that would be a support parameter of 2/8 or 0.25. In a medical context, this concept could show up as how many times a patient presented with both coughing and sneezing as well as heart palpitations – of 250 patient cases, for example, if 25 patients experienced coughing, sneezing, and palpitations, that would be a support parameter of 25/250 or 0.1. This support value in turn affects the number of rules that will be generated by our analysis. A lower support parameter means a lower threshold for determining which instances in the dataset are frequent itemsets – and the more frequent itemsets there are, the more possible rule candidates there will be for potential item associations. A lower support parameter value begets a higher number of rules; a higher support parameter results in a lower number of rules.

How do we decide the value of support? Knowing the domain knowledge for the field we are working in helps here. Using the sales field for example, we may perhaps know that “sales of items beyond a certain proportion tend to have a significant impact” on profits (KDnuggets). So it would make sense to use that proportion as the support parameter in that case. In the medical field, an example of domain knowledge might be knowing that symptoms occurring together above a certain frequency are more likely to be indicative of a particular diagnosis. Having domain knowledge like this is what helps us determine the support parameter value when we are running association rules in the real world.

**References**

Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2020). Data mining for Business Analytics: Concepts, techniques and applications in Python. John Wiley & Sons, Inc.

Association rules and the Apriori Algorithm: A tutorial. KDnuggets. (n.d.). Retrieved December 10, 2022, from https://www.kdnuggets.com/2016/04/association-rules-apriori-algorithm-tutorial.html