Summary of Research Paper: Development of a New Metric to Identify Rare Patterns in Association Analysis: The Case of Analyzing Diabetes Complications

(Saeed Piri , Dursun Delen , Tieming Liu , William Paiva)

Abstract:

- Proposed system aims to address the *rare item detection* problem in association rule mining.
- A new assessment metric, called adjusted_support, is proposed for rare items detection.
- A large size dataset with the data of about 600,000 patients is used to test the proposed metric.
- Adjusted_support is applied to discover rare association rules for diabetes complications.
- Comorbidity index of diabetic patients in various demographic groups is analyzed.

Drawback of existing system:

One of the most critical problems in association rule mining is the identification/discovery of rare item/patterns. This problem emerges when there are items that occur rarely compared to more frequent items.Setting a high threshold for support may lead to losing rare association rules. On the other hand, specifying a low threshold for support will lead to over generating association rules. Hence in this paper they proposed a new metric for assessing association rules known as adjusted_support. By using adjusted_support, the problem of rare items would be addressed by extracting rare rules without over-generating association rules.

Experiment results:

All of the generated rules in the analysis were assessed by both objective and subjective metrics. For objective assessment, proposed metric-adjusted_support was used beside support and lift. In addition, for subjective assessment consultation from medical advisors were taken into account.

Limitations:

First, the scope of the research was limited to the complications of diabetes that are specified in ICD 9 and ICD 10. Therefore,other diseases that patients may have been diagnosed with, were not considered in the study. Another limitation was related to the nature of EHR data. Because these types of dataset are collected for reasons other than the purpose of research, they may lack some degree of accuracy, for instance, some of the complications of a patient may not be recorded in their visit. However, the large amount of the data that was available in the study could compensate for this limitation.