CCSU
DEPARTMENT OF MATHEMATICAL SCIENCES

COLLOQUIUM
Friday, December 5

1:00 – 1:30 PM

Davidson Hall, Room 207

APPLYING ASSOCIATION RULES TO OPTIMIZE ENTERPRISE SOFTWARE DEVELOPMENT

JEFFREY RICHARDSON

(Data Mining MS Thesis Presentation)

CENTRAL CONNECTICUT STATE UNIVERSITY

Abstract: For most software companies, development consumes a substantial portion of their resources. As the volume of products offered increases, costs rise dramatically. For large development organizations with significant portfolios of product offerings, this can become a limiting factor for future development and growth. We theorize that by applying association rule algorithms to software usage data, we can find patterns and inherent relationships that can be used to streamline development and consolidate products.

We analyzed one year of software usage data from a leading CAD software vendor. The data included daily usage records of software for over 700,000 users. We started with 113,261,451 data points and began a comprehensive analysis of usage patterns. Through exploratory data analysis were able to prune this data set down and remove trial usage, beta programs, and incomplete data. We applied Apriori and CARMA algorithms to various aggregations and transformed versions of this data.

We were able to show that both the Apriori and CARMA algorithms successfully produced interesting and useful rule sets on our data. Apriori was unable to handle the scope of our original data, both in volume and in item set complexity, but was able to work on an aggregated data set. CARMA outperformed Apriori in speed, memory usage, and rule generation. Apriori was able to generate tens of thousands of rules, while CARMA was able to generate hundreds of thousands of rules. Our rules clearly showed numerous relationships between products and usage patterns.

We were able to share this data with decisions makers and product managers at the company, who immediate acted on this information. Development was shifted to combine titles into streamlined offerings and in one example a division was able to combine 25 separate offerings into 5 condensed products. The result was a $2,000,000 savings in annualized development cost through resource reallocation and increased focus on core products. This was done without sacrificing revenue and, notably, without workforce reduction. The methods used are being applied to products throughout the company now and a full consolidation/redesign effort is underway.

For further information:

gotchevi@ccsu.edu 860-832-2839

http://www.math.ccsu.edu/gotchev/colloquium/