

Concordia University
Department of Computer Science and Software Engineering
COMP 6521 - Advanced Database Technology and Applications.
Fall 2018

Algorithm Description:

- First we read the entire input file and extract important information like number of baskets, minimum support, items in each basket and also the count of each item in the entire data set and store them in hashmaps and variables.
- To get the solution of frequent itemset and that satisfies given minimum support, we use PCY instead of normal Apriori to improve performance. It stands for Park Chen Yu Algorithm and uses hash function to reduce the number of pairs and triplets.
- For the first pass that is singleton we check if the support of individual item greater than threshold then that item is included in the frequent item set.
- We use the items in the first frequent item set are used to create pairs. Along with it side by side we also hash the buckets using a hash function as $(i*j)\%11$. Which will hash the pairs to its respective buckets according to the hash value.
- There are multiple conditions before forming triplets, quadruplets, etc. to optimize code and help by eliminating unnecessary combinations before getting into huge loops.
- After formation of combinations again the support of each combination is checked with the threshold if it satisfies the threshold then it is added to next frequent item set.
- Along with this all the combinations which qualifies as a frequent item set are written into an output file.

- Experimental Results

Baskets	Minimum Support	Time Taken(ms)
10	1	145
100	30	56323
1000	300	39794

Team Contribution

Yash Sheth - Coding + Algorithm

Samip Thakkar - Coding + Algorithm

Divyansh Thakar - File Reader

Harsh Vaghani - Optimizations

References:

<http://stackoverflow.com/>
<https://medium.com/weekly-data-science/the-pcy-algorithm-and-its-friends-ecba67216190>
<http://infolab.stanford.edu/~ullman/mining/pdf/assoc-rules2.pdf>
<http://cs.brown.edu/courses/cs195w/slides/freqitems.pdf>
<https://en.wikipedia.org/wiki/PCY>
http://data.science.uoit.ca/wp-content/uploads/2016/01/BDA_06-associrules-2.pdf

"We certify that this submission is our original work and meets the Faculty's Expectations of Originality"

Y.S.

S.T.

D.T.

H.V.