

Combining Fuzzy Information: an Overview

Roanld Fagin

Abhishek Srivastava

Student ID: 861307778

February 28, 2017

CS 236, Winter 2017

The problem:

The paper states that accessing data in a sorted manner based on fuzzy information is also necessary for certain applications such as multimedia data and we want to choose top-k elements based on their fuzzy value of attributes in database systems. Previously stated algorithms were not optimal for these applications.

The contribution:

The authors defines fuzzing property of a data as a value of an attribute between 0 and 1 instead of *True* or *Not True*. Author presents a **Threshold** Algorithm, which is an improved version of Fagin's Algorithm. Fagin Algorithm was probabilistically optimal but under certain cases it did not performed well.

The method:

The Authors presents Threshold algorithm which was improvement over Naive and Fagin's algorithm. Naive algorithm idea was to calculate the grade of each data using an aggregate function and then get the top-k elements.

Fagin's algorithm was based on to do sorting access in parallel at all the lists, randomly select lists to get k objects appear in all the lists, calculate the grade or score of the objects and compute top-k objects.

Authors stated following steps in the Threshold algorithm:

- Do sorting access in the parallel at all the lists which includes for each object R that has been seen at least once in any of the list.
- Do random access to get the attribute value of R from the list where objects has not been seen yet.
- Compute the grade of selected object and repeat until threshold value is less than the grade value.
- If threshold is less than lowest aggregated grade of top-k set of objects then halt.

Threshold Algo is more optimal than Fagin's Algo but there are conditions where data can not be able to access in sorted manner or random manner. For those *Restricting Sorted Access* and *Restricting Random Access* algorithms are presented.

Comments:

The paper improves upon existing methods to find top-k elements based on their fuzzy properties over multiple attributes. Threshold Algorithm performs in general better if there is not certain restrictions in access patterns.

However, Some of drawbacks of the algorithms were:

- If the number of attributes and number of objects are very high in the, sorting cost can also become prevalent.
- It does not handle the case when all the stored data can be same which can be the worst case scenario where it may have to access all the stored objects.