

Review : Explaining Query Answers with Explanation-Ready Databases

Mili Shah(mashah@cs.umass.edu)

April 9, 2018

1 Summary

An explanation system attempts to explain trends and anomalies in data by giving explanations that it deems to be most responsible to cause anomalies or trends. The problem of such an explanation system is important because the number of users working with and analyzing data has been growing, and the backgrounds of users has also been increasingly varying. So, a need for such data analysis tools that can help automate detecting causes of trends has been growing. Such tools previously did not exist. Other proposed solutions are not efficient in performance, and they provide limited types of explanations because of performance issues.

With this work, the authors introduce an approach for an explanation system, consisting of explanation-ready databases and explanation-query. Explanation-ready databases pre-compute interventions. Interventions are obtained by removing tuples corresponding to an explanation from a database and evaluating its effect on the answer to a query. Explanation-query scores all the candidate explanations and gives a ranked list, by employing incremental query computing techniques for interactive speed. The authors also evaluate their approach on some datasets.

2 Strengths

The biggest strengths of the proposed approach are that first, the explanations that this framework provides are richer than those of the previous explanation systems, and second, it gives interactive performance. Computing interventions(making ERD) is the costliest part, and since they are pre-computed, the only computation left is for explanation-query - scoring the candidate explanations, and they are computed simultaneously with a single SQL query. Thus, the performance improves over previous systems remarkably. This boost in performance also enables the system to consider much deeper explanations of varied forms, such as, explanations involving combination of attributes and tables, explanations with aggregates, and explanations having top-k form. This was not done in previous systems due to performance issues.

3 Weaknesses and Extensions

1. There is a huge amount of manual effort involved for the ERD. Domain expertise in the domain of the database is required to define the types of explanations, their associated queries, and the causal dependencies. This is a significant shortcoming, which prevents this framework from being used inexpensively.
2. Pre-computing all interventions also requires a significant amount of extra space.
3. Extensions are possible in the direction of support of more types of inputs, such as null values, support for all orders of operations, etc. Extensions are also possible to improve performance for large datasets.