Redefining Data Dependencies: Enhancing Attribute Relationships for Today's Data Challenges

Nishttha Sharma
McMaster University
sharmn99@mcmaster.ca

1 Problem

Data is not static, and attribute value changes often trigger changes in another set of attributes. Functional Dependencies (FDs) [1] and Order Dependencies (ODs) [2] do not capture attribute value changes in both antecedent and consequent attributes. While Sequential Dependencies (SDs) [3] and Time Series Data Dependencies (TSDDs) [4] account for the value of antecedents influencing the range of change in consequent attribute values, they overlook the amount of change in the antecedents. Differential Dependencies (DDs) [5] do capture how changes in antecedents impact the changes in consequents but do not order the tuples, which limits their ability to model sequential relationships.

2 Motivating Example

Table 1 shows a sample instance of daily stock prices and trading volumes for two companies, A and B. We aim to explore how changes in stock prices affect their corresponding trading volumes, such as "a change of S_A% in Stock Price A, leads to a V_A% change in Volume A."

Stock Price A	Volume A	Stock Price B	Volume B
100	2000	150	3000
105	2300	145	3200
110	2400	140	2900
115	2600	150	3300
120	2700	155	3400
125	2800	165	3600
130	3000	160	3500

Table 1: Sample data for Company A and Company B.

Let us explore what some of the existing data dependencies would capture from this dataset and their limitations. ODs state that when tuples are ordered in ascending order along Stock Price A, then Volume A is also in ascending order. However, the semantics of ODs do not declaratively specify the change in attribute values between two records for both the antecedent and consequent attributes.

SDs state that when tuples are ordered in ascending order along Stock Price A, then Volume A would always increase, within a range of (100, 300), in consecutive tuples. Unlike ODs, SDs do declaratively specify the change in consequent attributes. However, they fail to capture the change in the antecedent and they also assume that changes are tracked between adjacent tuples.

DDs capture the change in attribute values for both the antecedent and consequent attributes. Here, DDs state that when the difference in Stock Price A between any two tuples is within a range (5, 10), then the difference in Volume A between the same tuples will be within a range (100, 400). The issue with DDs is that they assume that changes are tracked within a set (with no inherent ordering). By not capturing order, DDs miss critical contextual information like trends or patterns across consecutive tuples, which can be vital in sequential data.

Given these limitations, we aim to define a new data dependency that captures changes in both the antecedent and consequent attributes in an ordered relational instance. Our goal is to capture how changes in antecedent attributes (X), within the context of a window (defined over a set of X values), cause changes in consequent attributes (Y), in a relational dataset. We propose a framework to model and capture these attribute relationships.

References

[1] Catharine Wyss, Chris Giannella, and Edward Robertson. Fastfds: A heuristic-driven, depth-first algorithm for mining functional dependencies from relation instances extended abstract. In *Data Warehousing and Knowledge Discovery: Third International Conference, DaWaK 2001 Munich, Germany, September 5–7, 2001 Proceedings 3*, pages 101–110. Springer, 2001.

[2] Jaroslaw Szlichta, Parke Godfrey, Lukasz Golab, Mehdi Kargar, and Divesh Srivastava. Effective and complete discovery of order dependencies via set-based axiomatization. *arXiv preprint arXiv:1608.06169*, 2016.

[3] Lukasz Golab, Howard Karloff, Flip Korn, Avishek Saha, and Divesh Srivastava. Sequential dependencies. *Proceedings of the VLDB Endowment*, 2(1):574–585, 2009.

[4] Xiaoou Ding, Yingze Li, Hongzhi Wang, Chen Wang, Yida Liu, and Jianmin Wang. Tsddiscover: Discovering data dependency for time series data. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, pages 3668–3681. IEEE, 2024.

[5] Shaoxu Song and Lei Chen. Differential dependencies: Reasoning and discovery. *ACM Transactions on Database Systems (TODS)*, 36(3):1–41, 2011.