Optimizing Stratified Datalog with Count

Fernando Hechavarrría Fajardo, Heba Aamer, and Bas Ketsman

Vrije Universiteit Brussel, Belgium
{Fernando.Hechavarria.Fajardo,Heba.Mohamed,Bas.Ketsman}@vub.be

Ensuring consistent results in rule-based systems is critical for multi-core and distributed environments. In contrast to single-core systems, which operate sequentially, such distributed systems face coordination and synchronization challenges due to the asynchronous nature of execution. In such distributed settings, coordinating the execution order and synchronization of rules across different nodes is challenging, as delays and disorder are inevitable. Positive Datalog, which is known for its confluence property, offers stability but falls short in expressive power, making it not suitable for many practical applications involving constructs like negation or aggregation. To address these limitations, stratified Datalog organizes rules into strata to handle such constructs. However, this approach introduces synchronization requirements between strata, leading to increased latency in distributed systems [1].

In this work, we present a variant of stratified Datalog that incorporates count aggregation and investigate whether adding this feature can reduce the number of required strata, thereby minimizing synchronization overhead [2, 3]. We discover that while the expressive power of positive Datalog with count remains the same as regular positive Datalog, in multi-strata programs, stratified Datalog with count can express programs using fewer (or an equal number of) strata compared to standard stratified Datalog. Thus, this reduction has the potential to enhance performance in distributed systems by minimizing synchronization.

References

[1] Alvaro, P., Marczak, W. R., Conway, N., and Hellerstein, J. M. (2011). *Consistency analysis in Bloom: A CALM and collected approach.* In Fifth Biennial Conference on Innovative Data Systems Research, CIDR 2011, Asilomar, CA, USA, Online Proceedings, pages 249 - 260. www.cidrdb.org,2011. URL: http://cidrdb.org/cidr2011/Papers/CIDR11_Paper35.pdf.

[2] Koch, C., and Lindner, P. (2024). *Query optimization by quantifier elimination.* Proc. ACM Manag. Data, 2(2), 106. doi:10.1145/3651607.

[3] Hechavarría Fajardo, F., Aamer, H., and Ketsman, B. (2024). *Optimizing Stratified Datalog with Count.* Unpublished manuscript.