SCALING CHOICE MODELS OF RELATIONAL SOCIAL DATA

Jan Overgoor, George Pakapol Supaniratisai, and Johan Ugander

SIAM Workshop on Network Science 2020
July 9–10 · Toronto

Summary

Discrete choice models are a natural approach to modeling network formation, but don't scale well. We present two techniques for scaling the conditional logit model to large graphs. First, importance sampling of non-chosen alternatives reduces the choice sets while keeping a balanced distribution of features. Second, De-mixing is a novel model simplification technique that makes it both conceptually and practically feasible to fit mixed logit models to large datasets. We illustrate the benefits of these techniques on synthetic data, as well as on a large graph of 501M transactions on the Venmo platform.

Discrete choice for relational events

Many modern challenges in mining social data (link prediction, anomaly detection, recommendation systems) can be cast as modeling the likelihood of edges or events between nodes. Across these problems there is a common language of *relational events*, which are events involving two or more units, viewed as nodes in a graph where events connect these units by edges. Relational event modeling can thus be applied to a large number of data mining applications. Discrete choice modeling provides a natural framework for modeling relational events. Each event is viewed as a choice made by one node to involve another node, and modeled based on features of all the alternatives (see Fig 1). The conditional logit model of relational events subsumes and extends many existing models of network dynamics, such as preferential attachment and triadic closure [5]. However, the ability to work with large datasets is a major concern, as modeling relational events as choices raises both practical and conceptual issues.

For a graph of n nodes, when the originator of the choice is known then every *event* represents a choice with $O(n)$ alternatives. For large and sparse graphs, the large slates of potential alternatives makes direct inference intractable. Existing frameworks for modeling network formation with a logit model, such as REMs [2], are severely restricted in the size of the data they can directly handle [3]. However, the non-chosen alternatives can be sampled via a procedure commonly called "negative sampling" which produces estimates that are consistent for the estimates on the full data [4].

Availability, mixing, and de-mixing

Another issue with applying the conditional logit model to large graph datasets is the *availability* assumption that the chooser is a rational actor who has complete information about their available options and their features. This assumption is obviously not realistic in large social networks, where nodes are generally not aware of the existence of most others, and act mostly within their local social neighborhood. At the same time, in most social networks *some* edges happen outside the direct social neighborhood. Such a mixture of relational processes happening at different scales can be modeled with a mixed logit model.

However, likelihood maximization for mixed logit models is generically much harder than for the conditional logit, and the use of negative sampling does not cleanly transfer to the mixed logit setting. We therefore introduce a model simplification technique for discrete mixture models that we call "de-mixing" whereby the individual modes are assumed to operate on disjoint sets. When the modes of a mixed logit model operate over disjoint choice sets, the resulting model reduces to a collection of individual conditional logit models. De-mixing circumvents standard challenges with maximizing the likelihoods of mixture mod-

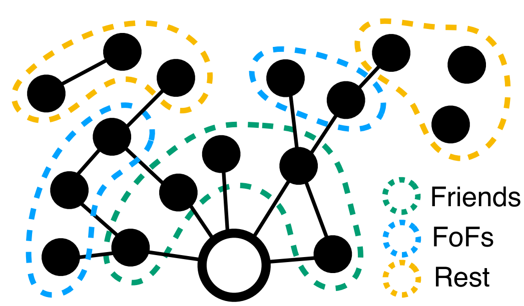

Figure 1: Illustration of a choice process in a small network. The ego chooses who to interact with, where different choice models may apply to friends, FoFs, and others.

els, and also opens the door to straight-forward importance sampling-based approaches to negatives sampling, which we separately find to be an important underutilized tool for choice models of network formation.

Importance sampling

In practice, negative sampling is typically done uniformly, a procedure that can be very inefficient when important features are rare within the population. This inefficiency is particularly pronounced for modeling large-scale social networks, as they are frequently driven by activity within one's local social neighborhood, which covers only a small subset of the full node set (making "is a friend of a friend" a rare feature). This concentration is even more pronounced for relational events, where pairs of individuals can interact repeatedly and often do.

To circumvent these issues and enable estimation of choice models on large graphs, we use importance sampling, a standard technique for approximate inference, to sample non-chosen alternatives non-uniformly. Importance sampling produces consistent estimates when coupled with an adjustment, based on the probability of being sampled, to the likelihood function [1]. As a result, we can feasibly fit models that incorporate activity in the local neighborhood to very large graphs.

Applications

We illustrate the value of negative sampling on synthetic data with a known data generating process. Non-uniform importance sampling is especially effective for rare features that commonly drive network formation, where it can reduce the variance of the estimates by an order of magnitude. As an important inspection, we examine the trade-off between an overall downsampling of the data (n) vs. sampling non-chosen alternatives (s) within data points, and generally find that additional data points (larger n) are much more valuable (in terms of mean squared error) than additional negative samples (larger s) at a fixed ns budget, up to a point (see Fig 2, left). Estimating a conditional logit model for mixed logit data leads to biased and, worse, inconsistent estimates (Fig 2, right). The conditional logit does retrieve the correct estimates for data from a de-mixed model (which fits within the conditional logit model class).

To illustrate the feasibility of fitting discrete choice models to very large graphs, we introduce and analyze

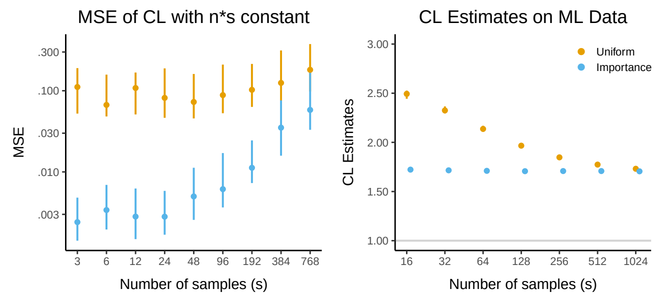

Figure 2: Left: MSE of conditional logit (CL) model on synthetic data, varying the number of negative samples s while keeping ns at constant. Right: CL estimates on misspecified mixed logit (ML) data for different sampling policies.

a large-scale dataset of 501M public transactions on the Venmo platform. A single conditional logit model on this data results in extreme parameter estimates that are hard to interpret. However, a de-mixed mixed logit model shows that for local activity, well-known dynamics like reciprocity take place, while activity outside the local neighborhood is primarily driven by preferential attachment. In this case, the de-mixed model provides significant new insight in the formation dynamics of this real world graph.

Conclusion

By de-mixing models that better approach the availability assumption and by leveraging non-uniform importance sampling, discrete choice models can be scaled to large-scale relational event data with great potential for impact. These advances open up their use to diverse data mining applications, including but not limited to link prediction, anomaly detection, and general recommendation systems.

References

[1] M. E. Ben-Akiva, S. R. Lerman, and S. R. Lerman. *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press, 1985.

[2] C. T. Butts. A Relational Event Framework for Social Action. *Sociological Methodology*, 38(1):155–200, 2008.

[3] J. Lerner and A. Lomi. Reliability of relational event model estimates under sampling: how to fit a relational event model to 360 million dyadic events. 2019.

[4] D. McFadden. Modelling the choice of residential location. Cowles Foundation Discussion Papers 477, Yale University, 1977.

[5] J. Overgoor, A. Benson, and J. Ugander. Choosing to grow a graph: Modeling network formation as discrete choice. In *WWW*, pages 1409–1420. ACM, 2019.