Building Near Realtime Contextual Recommendations for Active Communities on LinkedIn

Hema Raghavan
LinkedIn Corp
Mountain View, CA, USA
hraghavan@linkedin.com

ABSTRACT

At LinkedIn our mission is to build active communities for all of our members such that members are able to disseminate or seek professional content at the right time on the right channel. We mine a variety of data sources including LinkedIn's Economic Graph and member activities on the site and use large scale machine learning algorithms to recommend members to connect to people they might know to build active communities. We build real-time recommendations to disseminate information so that members never miss a relevant conversation that is going on in any of the communities they are part of. Through this talk we will showcase how we are trying to solve some of the most challenging problems on internet-scale social network analysis, streaming algorithms, and multi-objective optimization.

KEYWORDS

social networks, machine learning, distributed systems

1 INTRODUCTION

At LinkedIn the social network of a member is the core asset that drives the other value propositions that LinkedIn has to offer. A good network is not only critical to finding a job, but also helps professionals get and seek advice to make them more productive and successful in their current roles.

At the core of our strategy is building a network for a member. The People You May Know product is a connection recommendation product that drives more than half of the connections formed on LinkedIn. In the first half of this talk I will focus on technologies we have built to power PYMK. Connected professionals can then form a community that (a) shares content eg., articles/videos or slides that are professionally relevant, (b) gives and seeks professional advice. PYMK on LinkedIn is powered by an engine that allows for triangle closing and other graph walk algorithms in real time. It also allows models to consider near real-time features based on a member's context. We will demonstrate through A/B tests, significant improvements we have observed from adding additional capability in PYMK that allows for near realtime triangle closing over simply using batch offline recommendations.

We will then move on to show that treating every new connection equally could potentially lead to a suboptimal experience for certain members in the ecosystem. Every new connection brings in a different value to the LinkedIn ecosystem as well as to the members whose network this connection is being added to. There is a need to drive network growth in a direction which is not just optimal for increasing the total number of connections but also directs it towards providing value enhancement to all members. We will show that how a member's network evolves plays an important role in their downstream engagement. The problem is formulated as a multi-objective optimization problem

Finally, we will present our work on near real time optimization of activity-based notifications. Through near realtime systems we ensure that our members never miss a conversation that matters. The right notification has the power to delight members. However, notifications and emails when irrelevant can result in app disablements and/or higher unsubscribe rates. We formulate this problem also as a multi-objective optimization problem that trades off different utilities like engagement and volume. We will show through A/B tests that delivering the right information to the right user (through better content targeting) at the right time (through delivery time optimization and message spacing) is critical to building actively engaged communities.

The machine learning systems serving the above use cases are served by large scale distributed infrastructure that we have built to do real time computations. For example our graph computing platform can do random walks on a graph composed of tens of billions of directed edges. The realtime decision engine that makes decisions on whether and when to send a notification is built on top of Samza, an open source stream processing framework originally developed at LinkedIn. The talk will expose the listener to the underlying technology we use to power our AI products.

2 BIOGRAPHY

Hema Raghavan heads the team that builds AI and ML at LinkedIn solutions for fueling the professional social network's growth. Prior to that she was a Research Staff Member at IBM T.J Watson. Hema started her career in industry in Yahoo Labs. Her interests span the broad area of applications of AI and her experience spans a spectrum of products she has built in the areas of Search, Advertising, Question Answering and Recommendations. She has published in several conferences like WWW, SIGIR, ACL and COLING.