Modeling User Transition Behavior Among Networks With a

Probabilistic Programming Approach

Introduction

Imagine one typical 9-5 worker. She wakes up at 8, goes to work at 9, comes back at 5. Has dinner at 6. Spends two hours surfing the web at night. Now imagine how her daily network usage would look like. She would be on her home network in the morning. She uses a combination of the company and LTE network at work. Sometimes she eats lunch at a coffee shop and uses the network there. She spends some time on her home network at night. Her active network traffic ends at 11, but some passive email checking and other maintenance from her device persist throughout the night. Her daily network usage is very predictable. If researchers can predict her network usage, notifications can be customized to make her life more convenient.

Imagine another person traversing through a college campus while checking his email on the phone. As he moves along, his phone’s network switches from LTE and his college network alternatively. Every time the user switches network, his phone loses some data and has to send extra requests to recover lost data. The LTE and campus wifi, despite being geologically close by, can be vastly distant in network topology. Hence a request may have to go up multiple layers, before finally reaching a common ancestor. This extra path traversed carries inefficiency in the current network.

Multiple questions can be asked in those two scenarios. To what accuracy is human network behavior habitual? To what degree can network behavior be predicted? How should we evaluate network usage behavior? How often do such costly network transitions happen? How do we measure the cost of one such transition? How should we design our networks to avoid this costly transition?

This paper focuses on a quantitative approach to evaluate and predict typical user network behavior. Previously, some works were done in a quantitative approach. In Yang [1], a three state Markov transition model is described evaluate user transitions in networks. In Beverly [2], statistical learning was utilized to predict which bit of the IP address impacts traffic significantly. This paper tries a different approach as it explores more the modeling space, in addition to software to facilitate the modeling programming.

To predict a user behavior, researchers can model users in multiple ways: as one generalized type of user, as multiple representative types of users, or each user of his own. We were more interested in the generalized and representative types of users, as they provide more scalability and generality. A crucial part of modeling involves data refining. The researchers tried their best to extract time periods, sessions, and other important information that can distinguish the type of network access.

[Paragraph on modeling]

Probabilistic programming is a new type of programming that approaches the problem of modeling from a more statistician perspective. PP saves the researcher time and effort to write his own inference methods. With a booming probabilistic programming community, it is worth a while to explore this new technology. In this research, we evaluate the pros and cons of pymc3, pgmpy, and other python-based probabilistic programming packages.

[Paragraph on probabilistic programming]