

Large-Scale Analytics of Dynamics of Choice Among Discrete Alternatives

Andrew Tomkins
Google, Inc.
Mountain View, CA
atomkins@gmail.com

ABSTRACT

This talk will discuss the theory of discrete choice with a particular focus on aspects that are of interest to practitioners of large-scale data mining and analysis. We'll look at some example types of choice problems, including geographic choice as in restaurant selection, repeated sequential choice as in music listening, and the induction of nested models of choice.

CCS Concepts

•Computing methodologies → Supervised learning by regression;

Keywords

Discrete Choice; Repeat Consumption; IIA; Geographic Choice

1. INTRODUCTION

Jean-Paul Sartre writes, “We are our choices.” Every day we are faced with a stream of these choices demanding our attention: when to leave home, how to travel, what to listen to, where to eat, whom to meet, and so forth. In many of these cases we select a single element from a slate of alternatives: which restaurant, which song, or which movie. This is the setting of *discrete choice*, and is the subject of this talk.

The models commonly studied for discrete choice may be traced back to the 1920s, but significant activity began in the 1950s. A commonly-referenced starting point is the model of Bradley and Terry [3], which formalizes the probability that a user will select one option rather than another.

During the same decade, R. Duncan Luce extended this model, and introduced *Luce's axiom of choice* [5], which states axiomatically that a user's relative likelihood of choosing option A versus B is unaffected by the presence of option C in the slate of alternatives. It is simple to construct examples in which this axiom is violated. However, Luce showed that adopting the axiom led to a particularly simply algebraic form of likelihood of choice, and that the mathematical

problem of learning the parameters for this simple form is equivalent to the well-known and scalable multinomial logistic regression.

In this talk we'll begin with a summary of discrete choice models, and we'll present some recent results studying to what extent Luce's axiom holds in real large-scale online datasets. We'll also consider what can be done when the axiom of choice does not hold [2].

With this background in mind, we'll then apply the tools of discrete choice to some large datasets.

We'll begin with a study of geographic choice [4], in which users select a restaurant from a number of nearby alternatives. We'll show how discrete choice may be applied with features to capture some of the subtleties of proximity and density of alternatives. We'll further show how multinomial regression allows us to build a more detailed understanding of the particular form by which distance and related features influence likelihood of choice.

Finally, we'll consider some recent work [1, 2] on repeat consumption. In this setting, the user consumes a sequence of items, for example, repeatedly selecting a restaurant for dinner, or a TV series or youtube video channel to watch, or a particular song to listen to. In some cases the user chooses to listen to a new song, but in others, the user chooses to “reconsume” an already-familiar song. We study the dynamics of this behavior pattern, and develop models that can be used to predict when a particular song is likely to be reconsumed.

The work described in this talk is partly due to other researchers, and partly joint with various colleagues including Ashton Anderson, Austin Benson, Ravi Kumar, Mohammad Mahdian, Bo Pang, Sergei Vassilvitskii and Erik Vee.

2. REFERENCES

- [1] A. Anderson, R. Kumar, A. Tomkins, and S. Vassilvitskii. The dynamics of repeat consumption. In *WWW 2014*, pages 419–430.
- [2] A. R. Benson, R. Kumar, and A. Tomkins. On the relevance of irrelevant alternatives. In *WWW 2015*.
- [3] R. A. Bradley and M. E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39 (3/4)(324), 1952.
- [4] R. Kumar, M. Mahdian, B. Pang, A. Tomkins, and S. Vassilvitskii. Driven by food: Modeling geographic choice. In *WSDM 2015*, pages 213–222.
- [5] R. D. Luce. *Individual Choice Behavior a Theoretical Analysis*. John Wiley and sons, 1959.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

WebSci '16 May 22–25, 2016, Hannover, Germany

© 2016 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4208-7/16/05.

DOI: <http://dx.doi.org/10.1145/2908131.2908133>