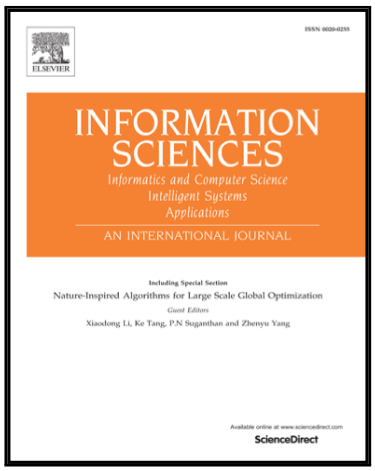

A Latent-class Model for Estimating Product-choice Probabilities from Clickstream Data

Naoki Nishimura, Noriyoshi Sukegawa, Yuichi Takano, Jiro Iwanaga

PII: S0020-0255(16)32134-X
DOI: 10.1016/j.ins.2017.11.014
Reference: INS 13242

To appear in: *Information Sciences*

Received date: 21 December 2016
Revised date: 17 June 2017
Accepted date: 8 November 2017

Please cite this article as: Naoki Nishimura, Noriyoshi Sukegawa, Yuichi Takano, Jiro Iwanaga, A Latent-class Model for Estimating Product-choice Probabilities from Clickstream Data, *Information Sciences* (2017), doi: 10.1016/j.ins.2017.11.014

A Latent-class Model for Estimating Product-choice Probabilities from Clickstream Data

Naoki Nishimura[a], Noriyoshi Sukegawa[b,*], Yuichi Takano[c], Jiro Iwanaga[d]

[a] *Product Management Unit, Internet Business Development Division, Recruit Lifestyle Co., Ltd. GranTokyo SOUTHTOWER, 1-9-2 Marunouchi, Chiyoda-ku, Tokyo 100-6640, Japan*
[b] *Department of Information and System Engineering, Faculty of Science and Engineering, Chuo University 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan*
[c] *School of Network and Information, Senshu University 2-1-1 Higashimita, Tama-ku, Kawasaki-shi, Kanagawa 214-8580, Japan*
[d] *Retty, Inc. Takanawa Park Tower, 3-20-14 Higashi-Gotanda, Shinagawa-ku, Tokyo 141-0022, Japan*

Abstract

This paper analyzes customer product-choice behavior based on the recency and frequency of each customer's page views on e-commerce sites. Recently, we devised an optimization model for estimating product-choice probabilities that satisfy monotonicity, convexity, and concavity constraints with respect to recency and frequency. This shape-restricted model delivered high predictive performance even when there were few training samples. However, typical e-commerce sites deal in many different varieties of product, so the predictive performance of the model can be further improved by integrating such product heterogeneity. For this purpose, we develop a novel latent-class shape-restricted model for estimating product-choice probabilities for each latent class of products. We also give a tailored expectation-maximization algorithm for parameter estimation. Computational results demonstrate that higher predictive performance is achieved with our latent-class model than with the previous shape-restricted model or latent-class logistic regression.

Keywords: Product choice, Latent class, EM algorithm, Optimization, E-commerce, Clickstream data

1. Introduction

Nowadays, a wide variety of products are viewed and purchased on e-commerce sites [43]. This enables the automated collection of clickstream data, which is a record of a visitor's page view (PV) history. Consequently, the analysis of clickstream data has been drawing intense research interest with respect to various topics, such as website browsing and navigation, Internet advertising, and online purchase behavior on e-commerce sites [5]. This paper presents an analysis of customer product-choice behavior based on clickstream data. The results of this research could be used to help visitors go to a target page on an e-commerce site and find the products they want. It could also be useful in demand forecasting for inventory management [20].

The recency and frequency of a customer's previous purchases have been shown to be key indicators for forecasting repeat purchases [10, 11, 22, 38, 39]. In light of this fact, Iwanaga et al. [21] recently devised optimization models for estimating product-choice probabilities from the recency and frequency of each customer's previous PVs. These models exploit the properties of recency and frequency of PVs to enhance their predictive performance. In particular, the monotonicity–convexity–concavity (MCC) model estimates

*Corresponding author
Email address: sukegawa@ise.chuo-u.ac.jp (Noriyoshi Sukegawa)