Author: Veronika Gondová
Supervisor: Prof. Mária Bieliková
Slovak University of Technology

User Modeling in the Domain of E-commerce

How to work with e-commerce data to get additional information for the recommendation?

Motivation

The recommendation performance is influenced by available data and the chosen algorithm. However, current approaches use only a small part of the available data in the form of a user model. Usage of the full stack of available information (or their combinations) can significantly improve the results of recommendation.

Main contributions

DATASET
- preprocessed dataset from real
e-commerce available for future research

RESEARCH PAPERS
- 2 journals, 1 conference

USER MODEL
Robust and generic abstraction method for click-stream data in domains with textual representation of items.
Model for demography prediction and its comparison with two state-of-the-art models.
User model for recommendation and research in the influence of individual features to recommendation results.

Dataset

Discount portal ZľavaDňa
- 3 mil. of transactions
- 20,8 mil. of events
- questionnaires
- data from project HIBER

Event abstraction

The main source of information for the recommendation in e-commerce is user activity - often in form of click-stream.

Problem 1: Click-stream consists of a huge amount of events that are linked to different items. Raw click-stream is not usable for machine learning tasks, e.g. clustering as a part of the recommendation.

Solution: Pattern-based event abstraction method
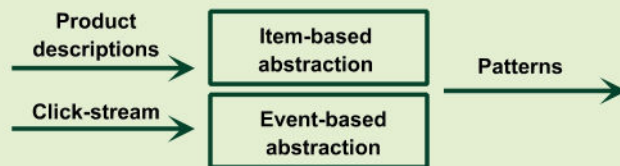
Results
Item-based abstraction in comparison with domain expert abstraction achieved comparable or better results in machine learning task (depends on ML task).

Combination of different pattern methods can improve the result of prediction - combination of N-grams and sequence patterns increased the recall of gender prediction by 4%.

Demography prediction

As research has shown, the demography is a factor that influences people shopping behaviour. At the same time, one of the main types of recommendation is the demographic recommendation.

Problem 2: Demographic characteristics are the type of information which users do not like to share with. Even when they share them, they are often misleading.

Solution: Model for demography prediction

Model consists of 7 562 features
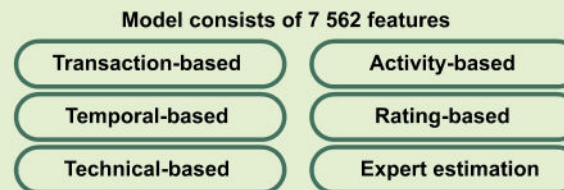
Results
Our model, in general, achieved better results than other 2 state-of-the-art models (that we have implemented) on our dataset - increase in recall by 5.2% for net income, 7.5% for address, 10.1% for relationship and 10.5% for child in family.

A complex model combining multiple data sources can achieve better results in demography prediction tasks.

User modelling

Most of the recommendation types are based on the user model. User modelling is currently one of the most discussed themes in the context of recommendation.

Problem 3: User model in general consists of user preferences for individual products. However, there is potential to contain other different types of information (e.g. demography, personality) that can significantly improve the recommendation.

Solution: Robust user model for recommendation

Results
The results of the recommendation were mainly influenced by user product preferences. However, user personality traits (extroversion and openness) in combination with category preferences improved the recommendation by 1.14% in MAP and 1.3% in nDCG.

Creation of robust user models also open the question of an intelligent method for feature selection - as can be seen in our result, most of the features made the recommendation worse.