Project name.

Santander Product Recommendation[1]

Team.

Anisimov Aleksandr, Sushchinskaya Svetlana, Mazyavkina Nina.

Background.

Under their current system, a small number of Santander's customers receive many recommendations while many others rarely see any resulting in an uneven customer experience. In their second competition, Santander is challenging Kagglers to predict which products their existing customers will use in the next month based on their past behavior and that of similar customers.

With a more effective recommendation system in place, Santander can better meet the individual needs of all customers and ensure their satisfaction no matter where they are in life.

Linear algebra works as a computation engine in ML. Most ML algorithms use a classifier or regressor and train it by minimising error between the value calculated by the nascent classifier and the actual value from the training data. This can be done either iteratively or using linear algebra techniques. If the latter, then the technique is usually SVD or some variant.

In the recommendation systems we usually handle a big amount of data and all of the techniques in current use involve some type of matrix decomposition, a fundamental class of linear algebra techniques (e.g., non-negative matrix approximation, and positive-maximum-margin-matrix approximation).

Problem formulation.

Let us consider without loss of generality a task of product recommendations. The main goal of this task is, given some prior information about users and items (products), try to predict what particular items will be the most relevant to a selected user. The relevance is measured with some relevance score (or utility) function fR that is estimated from the user feedbacks. More formally,

$$f_R : User \times Item \text{ -> } Relevance\ Score$$

where User is a domain of all users, Item is a domain of all items. The feedback can be either explicit or implicit, depending on whether it is directly provided by a user (e.g. ratings, likes/dislikes, etc.) or implicitly collected through an observation of his/her actions (e.g. page clicks, product purchases, etc.).
The type of prior information available in RS model defines what class of techniques will be used for building recommendations. If only an observation history of interactions can be accessed, then this is a task for collaborative filtering (CF) approach. If RS model uses intrinsic properties of items as well as profile attributes of users in order to find the best matching (user, item) pairs, then this is a content-based (CB) approach.[2]

Data.

In this competition, we were provided with 1.5 years of customers behavior data from Santander bank to predict what new products customers will purchase. The data starts at 2015-01-28 and has monthly records of products a customer has, such as "credit card", "savings account", etc. We will predict what additional products a customer will get in the last month, 2016-06-28, in addition to what they already have at 2016-05-28. These products are the columns named: ind_(xyz)_ult1, which are the columns #25 - #48 in the training data. We will predict what a customer will buy in addition to what they already had at 2016-05-28.

The test and train sets are split by time, and public and private leaderboard sets are split randomly.[1]

Related work.

The building of a recommender system is a very popular task, which is implemented in different ways on the various websites. Consequently, there are a lot of various examples and methods for product recommendations such as this project, or this two patented systems - Intelligent performance-based product recommendation system and Online shopping support method and system for sales promotions based on the purchase history of users . [3],[4],[5]
On top of that there are several similar challenges with the similar problems and objectives by different e-companies like, for example, on kaggle Million Song Dataset Challenge or Event Recommendation Engine Challenge. [6],[7]
However, there is no uniform solution to the recommendation system building problem as the datasets and data patterns vary from problem to problem and require an individual approach. So that’s why find our project theme very interesting and relevant.

Scope.

The result of this project will be trying to build the system that predicts which products Santender bank's existing customers will use in the next month, based on their past behavior and that of similar customers, and then submit our solution to kaggle to compare the results with the other challenge participants. We will split our work in several stages.
Firstly, we will analyze the data given to us, by bringing it to some canonical form and finding the existing data patterns. Next, we will try to solve our problem by implementing the standard models, like, for example, the ones based on SVD. We will then analyze the received results and try to use more advanced methods, for instance, using Tensor Factorization in a way explained in these two articles - Factorization Models for Context-/Time-Aware Movie Recommendations and Temporal Collaborative Filtering with Bayesian Probabilistic Tensor Factorization. [8],[9] Then we will choose which of the methods gave us the best results.

Evaluation.

Submissions are evaluated according to the Mean Average Precision @ 7 (MAP@7):

$$MAP@7 = \frac{1}{|U|}\sum_{u=1}^{|U|}\frac{1}{min(m,7)}\sum_{k=1}^{min(n,7)}P(k)$$

where |U| is the number of rows (users in two time points), P(k) is the precision at cutoff k, n is the number of predicted products, and m is the number of added products for the given user at that time point. If m = 0, the precision is defined to be 0.

References.

[1] Santander product recommendation. Available at: https://www.kaggle.com/c/santander-product-recommendation (Accessed: 18 November 2016).

[2] Frolov E., Oseledets I. Tensor Methods and Recommender Systems //arXiv preprint arXiv:1603.06038. – 2016.

[3] Hu J., Zhang B. Product Recommendation System //CS224W Project Report. – 2012.

[4] Shaya S. A. et al. Intelligent performance-based product recommendation system : пат. 7809601 США. – 2010.

[5] Ono T. et al. Online shopping support method and system for sales promotions based on the purchase history of users : пат. 5909023 США. – 1999.

[6] *Million Song Dataset challenge*. Available at: https://www.kaggle.com/c/msdchallenge (Accessed: 18 November 2016).

[7] *Event recommendation engine challenge*. Available at: https://www.kaggle.com/c/event-recommendation-engine-challenge (Accessed: 18 November 2016).

[8] Gantner Z., Rendle S., Schmidt-Thieme L. Factorization models for context-/time-aware movie recommendations //Proceedings of the Workshop on Context-Aware Movie Recommendation. – ACM, 2010. – С. 14-19.

[9] Pragarauskas H., Gross O. Temporal collaborative filtering with bayesian probabilistic tensor factorization. – 2010.