Master Thesis Plan **Master Artificial Intelligence**

Date	:	**12th December, 2016**
Name and student number	:	**Stefanie Deckers, i6099223**
Name of the thesis examiner	:	**Kurt Driessens**
Name of the daily supervisor (if different from the thesis examiner)	:	
Title of the master thesis	:	**Development of a Recommender System for multiple online shops**
Expected final date	:	**25th January, 2016**

Signature thesis examiner **Signature master student**

.... Stefanie Deckers ....

Assessment
The assessment will be based on the contents and form of the thesis, the presentation of this thesis and any product and/or software that accompanies the thesis.

Research plan

Background
The master thesis will be performed at the e-business company Braintags GmbH in Willich, Germany. They are hosting various online shops for different customers and want a universal recommender system for all shops. This recommender system will be developed in the scope of the master thesis.

The recommender system is supposed to work as an independent software component, such that every online shop Braintags is hosting can request recommendations from it. It will be implemented in Java using the tool-kit vert.x *(http://vertx.io/)*.

Description of the assignment
The goal of the research is to develop a recommendation system that is able to provide recommendations for different online shops.
The two most common kinds of algorithms described in [2] will be investigated:
Content-based techniques make recommendations based on similarities between items, i.e. by finding items that are similar to the items that the user liked in the past.
Collaborative approaches rely on preferences of similar users, where user similarities are calculated by similar purchases/clicks/ratings in the past. There are two different types of collaborative filtering approaches. Memory-based, also called neighbour-based, approaches find neighbours of a user or an item at runtime and use them to make recommendations. Model-based techniques learn an alternative model in advance by transforming users and items into the latent factor space, e.g. by singular value decomposition.

Implicit feedback
Most research activity is based on datasets with explicit feedback data, where users intentionally give some kind of rating to express their preferences as in the Netflix price [1]. For this thesis only implicit feedback data is available that consists of purchase histories, page views and shop cart entries. This kind of data is more noisy, because those implicit user interaction might not correlate with the user's real preferences. Moreover it lacks negative feedback. A user who did not buy or view an item may either not know or not be interested in that item. Due to those difficulties most common recommendation techniques are not applicable. So the task of this thesis is to find a recommendation technique that can manage the e-commerce data in real-time.

Generally a model-based approach is preferable, because after a model is trained recommendations can be made faster and less memory intensive than using a memory-based approach. This is especially important for large datasets.
The most common model-based approach aims to learn user and item features similar to singular value decomposition (SVD), such that a rating of a user for an item can be calculated by

$$\hat{r}_{u,i} = p_u^T q_i \, .$$

This approach is problematic for e-commerce data, because there are no ratings, but only one-class positive feedback. Sampling or weighting schemes as in [3] will be examined to adapt the model to e-commerce data.

The performance of different recommendation techniques depends on the data it is applied on. Since the developed recommender system should be able to serve different shops, hybrid approaches like in [4] are investigated to allow to tune the system for different datasets.

Evaluation
Evaluation will be performed in terms of the mean average precision (MAP)

$$MAP_n = \frac{1}{U} \sum_u^U \frac{1}{n} \sum_i^n Precision(i)$$

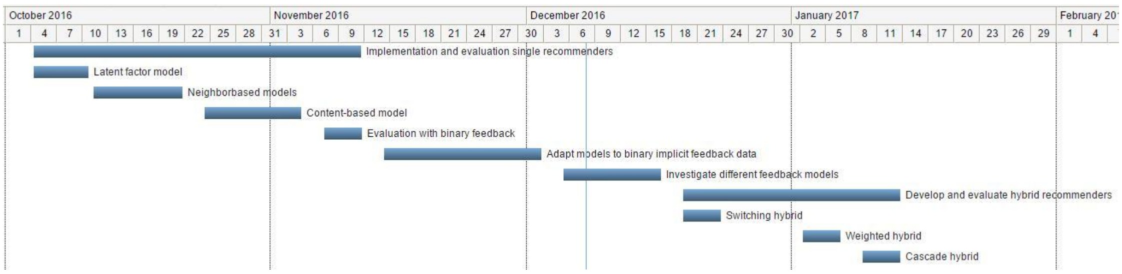
where U indicates the number of users, n the number of recommended items. Precision(i) measures the precision at the list of recommendations cut off at position i.

Research questions:
What kind of algorithms perform well on implicit feedback e-commerce datasets?
How can model-based recommendation approaches be applied to implicit feedback data?
Are hybrid approaches able to improve performance?

References:
[1] Bennett, James, and Stan Lanning. "The netflix prize." *Proceedings of KDD cup and workshop*. Vol. 2007. 2007.
[2] Ricci, Francesco, Lior Rokach, and Bracha Shapira. *Introduction to recommender systems handbook*. Springer US, 2011.
[3] Pan, Rong, et al. "One-class collaborative filtering." *2008 Eighth IEEE International Conference on Data Mining*. IEEE, 2008.
[4] Burke, Robin. "Hybrid web recommender systems." *The adaptive web*. Springer Berlin Heidelberg, 2007. 377-408.

Planning:

	Name	Start	Finish
1	Implementation and evaluation single recommenders	04/10/2016	11/11/2016
2	Latent factor model	04/10/2016	10/10/2016
3	Neighborbased models	11/10/2016	21/10/2016
4	Content-based model	24/10/2016	04/11/2016
5	Evaluation with binary feedback	07/11/2016	11/11/2016
6	Adapt models to binary implicit feedback data	14/11/2016	02/12/2016
7	Investigate different feedback models	05/12/2016	16/12/2016
8	Develop and evaluate hybrid recommenders	19/12/2016	13/01/2017
9	Switching hybrid	19/12/2016	23/12/2016
10	Weighted hybrid	02/01/2017	06/01/2017
11	Cascade hybrid	09/01/2017	13/01/2017



In case of an external training period

Name of the company	:	Braintags GmbH
External supervisor	:	Martin Plücker
Size of the project	:	
Working hours	:	40h / week
Frequency of contact	:	Daily
Agreement about payment (if applicable)	:	400€ gross / month
Confidentiality	:	Confidentiality agreement regarding personal data only
Special agreements	:	

Signature external supervisor

Martin Plücker

....