Title	正例･弱い負例･強い負例を考慮した情報推薦
Author(s)	持田, 春嘉
Citation	
Issue Date	2024-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/18919
Rights	
Description	Supervisor: 白井 清昭, 先端科学技術研究科, 修士(情報科学)

Recommendation System Considering Positive, Weak Negative, and Strong Negative Samples

2030415 Haruka Mochida

With the development of information technology, online product transactions have increased, making it challenging to find desired products among a vast array of options. To address this issue, recommender systems have been widely studied. It is a technique that automatically extracts items that users may be interested in, based on users' purchase history and similar users' behaviors. Methodology of recommender systems can be divided into two categories: methods based on explicit feedback (e.g. rating explicitly given by users) and implicit feedback (e.g. purchase and browsing history). Although explicit feedback more accurately represents users' preference, implicit feedback are commonly used in commercial recommender systems due to availability of a large amount of data.

Traditional recommender models using implicit feedback often treated items as two classes: positive samples (items clicked by a user) and negative samples (items not clicked by a user). However, all negative samples are not homogeneous. On the one hand, some negative samples may not have been clicked because the user was unaware of them, despite potentially finding them favorable. On the other hand, some negative samples are genuinely of no interest. In traditional methods where items are classified into only two classes as positive and negative samples, these items are treated identically. However, by distinguishing different types of negative samples, it is expected that a recommender system could capture users' preference more precisely.

This study proposes a new recommender system that treats items as three classes: positive samples, weak negative samples, and strong negative samples. First, two kinds of negative samples are defined using data called "exposure", which indicates whether an item was displayed on the user's screen. Strong negative samples are items that were displayed on the user's screen but not clicked by the user, while weak negative samples are items that were not displayed on the user's screen and not clicked as well. Strong negative samples, which a user definitely see but did not click, are more likely to be unfavorable items compared to weak negative samples. By distinguishing weak and strong negative samples, which were treated as the same in previous traditional models, more accurate recommendation can be provided to users.

Two existing models, Factorization Machines (FM) and Neural news recommendation with attentive multi-view learning (NAML), are extended as a "three-classes model" where positive, weak negative and strong negative samples are handled separately. FM is a method that efficiently learns regression parameters using matrix factorization. Its has many advantages including the ability to incorporate attributes of users and items arbitrarily, robustness against sparse data, and a low computational cost compared to deep learning models. In this study, a trinary extended FM is proposed, where the target value for positive samples are set to 1, weak negative samples to 0, and strong negative samples to -1, in order to give higher recommendation scores in the order of positive, weak negative, and strong negative samples. Besides, NAML is a deep learning method that focuses on news article recommendation. Its characteristic is ability to obtain precise latent representation or embedding of users and items by using Attention and Convolutional Neural Network (CNN). In this study, a trinary extended NAML is proposed, where its final layer is replaced with a fully connected layer that outputs three classes: positive, weak negative, and strong negative samples. Furthermore, Negative User Encoder, which obtains the latent representation of items disliked by a user from strong negative samples, is newly introduced and incorporated into the original NAML.

Microsoft News Dataset (MIND) that includes exposure data is used for the implementation and evaluation of the proposed method. This dataset consists of behavioral logs on a news site operated by Microsoft, including information of news articles (such as titles and abstracts) and IDs of news articles clicked by users. Normalized Discounted Cumulative Gain (NDCG) and Area Under the Curve (AUC) are used as evaluation criteria for the proposed recommender systems. NDCG is measured on two tasks: ranking a large number of items in the order of user's preference, and ranking a small number of items displayed on the user's screen in the order of user's preference. AUC evaluates the binary classification task to guess whether a target item will be clicked by a user or not. The traditional binary FM and NAML that handle postive and negative samples are used as the baseline. They are compared with the trinary extended FM and NAML proposed in this study.

The experimental results showed that in the task of ranking a large number of items according to users' preference, both the trinary extended FM and NAML models achieved higher NDCG than the binary baseline models. Specifically, the trinary extended FM model showed an improvement of 0.004 to 0.039 points compared to the baseline, and the NDCG of the trinary extended NAML model was improved by approximately 0.004 points. In addition, in the task of ranking a small number of items displayed on the screen, the trinary extened FM could achieve better performance than the binary FM with 0.007 to 0.044 points, but the trinary extened NAML did not show the improvement consistently. Comparing FM and NAML, NDCG of FM was better than NAML in both tasks. Specifically, FM outperformed

NAML by approximately 0.015 points in the task of ranking a large number of items, while the NDCG of FM was better than NAML by 0.107 points in the task of ranking a small number of items. However, in the evaluation of the binary classification of NAML with respect to AUC, no improvement by the proposed trinary NAML was found. These results suggest that distinguishing between weak and strong negative samples in the training of the recommender systems can improve the performance of recommendation in the ranking task. They also indicate that deep learning methods (NAML) do not always outperform traditional methods (FM).

One of the future directions in this study is to investigate a method to incorporate the information of negative samples more precisely into the model by considering the strength of negative preference of them. In addition, a way how to measure the strength of these negative preference as evaluation scores to be incorporated into a model should also be explored.