

The Use of Impressions in Recommender Systems: Improving Complete and Semi Cold-Start

Matteo Garavaglia^{1,3,*,\dagger}, Alessandro Solinas^{2,4,\dagger}, Ricardo Anibal Matamoros Aragon^{2,3,\dagger}, Stefania Bandini³ and Francesco Epifania²

¹*AITech4T S.r.l.*

²*Social Things S.r.l., Milan, Italy*

³*Department of Informatics, Systems and Communication, University of Milano-Bicocca (ITALY)*

⁴*Politecnico di Milano, Milan, Italy*

Abstract

Recommender Systems (RS) are tools that are often utilized and need constant development in both the structure and the data to use. Impressions Data are a new type of information that is underused and can be helpful in various scenarios. Therefore, we propose a hybrid RS that uses Impressions to mitigate the significant issues in our original system, Knowledge Graph Attention Network (KGAT). The first problem is the situation of complete cold-start, for which we propose the use of questions on selected meaningful attributes and a BERT-based Content-Based RS to perform recommendations following the user's choices. After that, when in a framework of semi cold-start, the recommendations will be enhanced by using Impressions to rerank the following ones and, from these interactions, to build a profile to use with KGAT. The last issue we will address is the need for more interpretation of negative interactions through the Knowledge Graph, that is, recommendations presented but not chosen. To solve this issue, we use the Impression Discounting model on the set of recommendations produced by KGAT.

Keywords

Recommender Systems, Impressions, Cold Start, Real-time Recommendations

1. Introduction

Research in Recommender Systems (RS) mainly builds recommendation models using historical feedback (e.g., clicks, purchases, watching actions) of products collected from users, but the community is always looking to improve the recommendation quality by leveraging other data sources [1]. Impressions represent an emerging concept in RS, but their full potential is limited by a lack of reliable and open-source data. This research aims to develop a comprehensive RS framework that combines Knowledge Graph Attention Network (KGAT)[2], a content-based (CB) recommender system utilizing BERT, and impressions. KGAT, empowered by the

*Workshop on Artificial Intelligence and Applications for Business and Industries (ALABI 2023) co-located with 22th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2023), Rome, Italy*

*Corresponding author.

\dagger These authors contributed equally.

✉ m.garavaglia20@campus.unimib.it (M. Garavaglia); alessandro.solinas@socialthingum.com (A. Solinas); ricardo.matamoros@socialthingum.com (R. A. M. Aragon); stefania.bandini@unimib.it (S. Bandini); francesco.epifania@socialthingum.com (F. Epifania)

🆔 0009-0006-7451-3986 (M. Garavaglia); 0000-0002-5428-3187 (A. Solinas); 0000-0002-1957-2530 (R. A. M. Aragon); 0000-0002-7056-0543 (S. Bandini)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

information captured in the knowledge graph (KG), aspires to provide personalized and precise recommendations, leveraging both user and item characteristics. However, this model does not include the interpretation of negative interactions, implying that the model does not learn anything new if an item is recommended but not chosen by the user, moreover, the training phase of KGAT is time-and-resource-consuming, making quick recurrent training unfeasible. For situations where no prior user data is available, or a specific user has not been included in KGAT, we introduce CB-BERT, which leverages resource titles and abstracts to match user preferences effectively. To further refine these recommendation systems, we incorporate real-time impressions. In the current literature, exist two main classes of impression models: (i) re-ranking and (ii) impressions as user profiles. The first class, which is the one we will employ, allows the system to dynamically re-rank recommendations, while the second may treat impressions as user interactions are traditionally used [1]. We propose a slightly modified Impression Discounting (ID) model, as detailed in [3], which easily adapts to both KGAT and CB scenarios thanks to its peculiarity of being a plug-in approach, completely independent of the system that generates the recommendations to re-rank. ID utilizes impressions to facilitate real-time updates, enabling the system to continually refine its recommendations. Summarizing, we aspire to propose a framework in which we use a CB RS and the ID to mitigate the drawbacks of KGAT, namely poor performance in full and semi cold-start, a long training time and the disregard of the user's refusal of recommendations.

2. Overview of our system components

In this section we will briefly explain the components of our system, highlighting a few key aspects of each of them.

2.1. Data

The dataset used for this work was the **Microsoft News Recommendation Dataset (MIND)**. MIND is a dataset created by Microsoft Researcher to advance the research on news recommendation constructed from user click logs of Microsoft News. The dataset is in English, has one million of users, 161013 news, 24155470 clicks and, for each news has title, abstract, body and category. Fundamental for this work, the Dataset also contains **Impressions** of the users in the form of click history. Each impression contains the ID of the user who generated it, the timestamp, the click history of the user before the impression and the actual impressions, that is the list of id of the recommended items with two possible suffixes: "-1" for items clicked by the user and "-0" for ones not clicked by the user [5].

2.2. KGAT

The core element of this recommender system is KGAT, a model designed to handle complex relationships in a knowledge graph (KG). KGAT does this by recursively updating node embeddings, like objects, users, or attributes, and using attention to determine the significance of these connections. The system starts by creating a KG that represents various object relationships. For example, in movie recommendations, a film may be linked to actors, directors, genres, and

more. To make this graph usable for neural networks, an adjacency matrix is formed, where 1s represent existing edges, and 0s indicate no connection. Then, a multi-layer neural network processes KG data to improve recommendations. The first layer takes the adjacency matrix and converts it into numerical embeddings that capture object characteristics. These embeddings are generated using proximity information to map objects into a lower-dimensional space. The model calculates attention coefficients, assigning importance to relationships based on proximity and relevance. Finally, recommendations are produced using object embeddings and graph attention. These recommendations are presented to users as suggestions for products, movies, music, etc., that might match their interests.

The key components of KGAT are:

- **Embedding construction:** entities and relations present in the KG are parametrized using TransR [4], optimizing the translation principle $e_h^r + e_r \approx e_t^r$, if the triple (h, r, t) exists in the graph. Where $e_h, e_t \in \mathbb{R}^d, e_r \in \mathbb{R}^k$ are h,t and r embeddings, respectively, while e_h^r, e_t^r are the projections of e_h, e_t in the relations space.
- **Attentive Embedding Propagation Layers:** the set of attention coefficients is generated to obtain the importance of high-order connectivities. Each attention layer consists of 3 components: information propagation, knowledge-aware attention, and information aggregation. Considering entity h and the triples for which h is the starting node (also known as the ego network), $\mathcal{N}_h = \{(h, r, t) | (h, r, t) \in \mathcal{G}\}$, it is possible to characterize the 1st-order connectivity of entity h with a linear combination of its ego network:

$$e_{\mathcal{N}_h} = \sum_{(h,r,t) \in \mathcal{N}_h} \pi(h, r, t) e_t$$

The term $\pi(h, r, t)$ controls the flow of information from h to t through r. This coefficient is implemented using a relational attention mechanism and is calculated using the *tanh* activation function, followed by normalization through a softmax function.

As a final step, for each entity, the information from the embedding e_h is aggregated with its representation in the ego network $e_{\mathcal{N}_h}$. The aggregation function used is:

$$f(e_h, e_{\mathcal{N}_h}) = \text{LeakyReLU}(W_1(e_h + e_{\mathcal{N}_h})) + \text{LeakyReLU}(W_2(e_h \odot e_{\mathcal{N}_h}))$$

Where W_1 and W_2 are trainable weight matrices in the model, and \odot denotes element-wise multiplication. This principle is used to model 1st-order connections, but it can be generalized to higher-order connections by constructing the ego network of order l.

2.3. CB-BERT

Since KGAT could not handle effectively a user in a situation of cold start, we implemented a Content Based Recommender System (CBRS) to use when a new user asks for recommendations. A typical CBRS works by creating a profile based on the attributes of items with which the user has interacted. Then, to perform recommendations, the system has to compute a similarity score between each item and the user profile, to, finally, recommend the ones that are the most similar to the profile [6]. In our case the user profile is built using the embedding of abstracts (or titles, if abstracts are not present in the data) of the articles with which the user has interacted. The

embeddings are given by a **Bidirectional Encoder Representations from Transformers** (BERT) model. The Transformer model, which is defined by a stack of self-attention layers and feedforward neural networks, is the foundation of the architecture of BERT. With the help of this architecture, BERT gathers context from both preceding and following words, allowing it to understand the context [7].

BERT must go through a two step procedure to create context-aware embeddings:

1. Pre-Training:

- *Masked Language Modeling (MLM)*: BERT is pre-trained on a corpus of text of large dimensions in our case Wikipedia and Books Corpus. During this phase BERT learns to predict missing words in a sentence by masking random words in the input text and setting BERT's objective to predict what those words should be.
- *Next Sentence Prediction (NSP)*: BERT is then trained to predict if there is a sequential relationship between two sentences. This phase enable BERT to comprehend sentence-level contextual dependencies, which is vital for tasks such as document classification and question answering.

2. **Fine-Tuning**: BERT is fine-tuned on a task-specific dataset relevant to a downstream application after pre-training. With this modification, BERT is now able to use its previously learnt contextual embeddings for tasks like sentiment analysis, named entity recognition, and text categorization. This process allows BERT to be in line with task-specific goals.

In our case we build the profile as the combination of the embedding of the abstracts (if present, if not the titles) of the article with which the user has previously interacted. The embedding of users are used to compute the *cosine similarity* with the ones of the items. The cosine similarity is defined as

$$\text{sim}(A, B) = \cos(\theta) = \frac{A \cdot B}{|A| \cdot |B|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Where A and B are the embeddings.

2.4. Impression Discounting

To model impression we opted for a slight modification of the model presented in [3]. ID is a model used as a plug-in for an existing Recommender System (RS), which, through impressions, re-ranks the resources provided by the RS. An **Impression** is a set containing a user-object pair and the behaviors that occurred between them in a session (logs), defined as $\tau = \{user, item, conversion, behaviors, t, R\}$. When grouping by (user, item), a series of impressions τ_1, \dots, τ_n are obtained and used as data. Within the impression set τ , there are three additional elements: **t** represents the timestamp associated with the impression, **R** is the score vector provided by the existing RS for the user, and finally, **conversion** is a boolean indicating whether the recommendation was accepted or not. It is worth noting that *conversion* = 1 is only possible in the last element of an impression sequence, as we assume that an accepted item will no longer be shown to the user. The possible behaviors are:

- LastSeen: the day difference (in days) between the current and last impressions, associated with the same (user, item).

- ImpCount: the number of previous impressions associated with the same (user, item).
- Position: the position of the item in the RS recommendation list
- UserFreq: the frequency of the user’s interactions with the RS

The objective of the ID model is to identify a discounting factor d , which updates the ranking associated with every (user, item) couple to maximize the number of converted items. The relation between each behavior and the discounting factor is modeled through a discounting function that is chosen independently for a single behavior through a Bayesian optimization algorithm. The set of possible discounting functions is the following:

- Linear: $f_L(x) = \alpha_1 + x \cdot \alpha_2$
- Inverse: $f_I(x) = \frac{\alpha_1}{x} + \alpha_2$
- Exponential: $f_E(x) = e^{\alpha_1 \cdot x + \alpha_2}$
- Quadratic: $f_Q(x) = \alpha_1(x - \alpha_2)^2 + \alpha_3$

So, given an impression, the discounting factor d is computed as a linear combination of the function applied to each measured behavior x , as follows:

$$d = \sum_{x \in X} f_x(x)$$

Where X is considered the set of the behaviors {LastSeen, ImpCount, Position, UserFreq} and f_x is the discounting function associated with the behavior $x \in X$, among the above-indicated ones.

3. Results

In this section, we present the preliminary results obtained by adding the impression framework both on KGAT recommendations and on colt-start, using CB-Bert recommendations.

3.1. KGAT experiment

Concerning the KGAT standard utilization, we wanted to see how impressions add value to the existing recommendation. To put this theory to the test, we performed the following experiment. We initialized the KG with the items present in each user history, which we recall is univocal and contains past positive interactions with the existing items. With this setup, we proceeded to train KGAT and generate a recommendation list for each user. We then used this list to optimize and test the ID model. Looking at the recommendations’ metrics (such as Precision, Recall, NDCG @20) before and after the use of impressions we could see a significant improvement, suggesting a better user representation, but a deeper and thorough analysis is needed.

3.2. Cold Start experiment

Similarly to the KGAT experiment, we want to test how valuable are impressions in improving the performances in the case of complete cold-start, recommending using CB-BERT. To test it we present the following pipeline:

From the user's history, we collected the categories and sub-categories of the news with positive interactions and used them to perform a CB-BERT recommendation. This was done to simulate a sort of questionnaire about the user's preferences. The NDCG@20 registered in this case (without the Impressions Discounting) is 0.001, which is a good starting point considering the scenario. With the use of impressions we were able to improve it to 0.013.

4. Conclusions and future developments

In conclusion, this research addresses the objective of enhancing the recommendation quality in RS. In general, a single RS model suffers from different problems, ranging from cold-start to the partial use of the information available. Proposing a coherent and personalized experience to the user on the recommendation remains a challenge. We propose an ensemble of three different techniques to manage at best the possible problematics, focusing on cold start and the management of negative interactions. We were able to increase the model's performance in both the standard and cold-start experiments. From our results, it is evident that impressions add a significant value with ease of use and flexibility in many of these frameworks while not penalizing the computational burden. We want to underline that these are only preliminary results, a more complex modeling and a deeper analysis could lead to more promising results.

For future work, we plan to introduce more suited data gathered from our systems to perform more extensive evaluations. We also have the ambition to extend the impression scheme by improving the set of behaviors and exploring other techniques to exploit them.

References

- [1] Fernando Benjamin Perez Maurera, Maurizio Ferrari Dacrema, and Paolo Cremonesi. *Towards the Evaluation of Recommender Systems with Impressions*. In *Proceedings of the 16th ACM Conference on Recommender Systems (RecSys '22)*, Association for Computing Machinery, New York, NY, USA, 2022. ISBN: 978-1450392785. Pages 610-615. DOI: 10.1145/3523227.3551483.
- [2] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. *KGAT: Knowledge Graph Attention Network for Recommendation*. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*, Association for Computing Machinery, New York, NY, USA, 2019. ISBN: 978-1450362016. Pages 950-958. DOI: 10.1145/3292500.3330989.
- [3] Pei Lee, Laks V.S. Lakshmanan, Mitul Tiwari, and Sam Shah. *Modeling Impression Discounting in Large-Scale Recommender Systems*. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*, Association for Computing Machinery, New York, NY, USA, 2014. ISBN: 978-1450329569. Pages 1837-1846. DOI: 10.1145/2623330.2623356.
- [4] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. *Learning Entity and Relation Embeddings for Knowledge Graph Completion*. *Proceedings of the AAAI Conference on Artificial Intelligence*, February 2015. DOI: 10.1609/aaai.v29i1.9491.

- [5] Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, and Ming Zhou. *MIND: A Large-scale Dataset for News Recommendation*. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Online, July 2020. Pages 3597-3606. DOI: 10.18653/v1/2020.acl-main.331.
- [6] Pasquale Lops, Dietmar Jannach, Cataldo Musto, Toine Bogers, and Marijn Koolen. *Trends in content-based recommendation*. In *User-Adap Inter 29*. 2019. Pages 239–249. DOI: 10.1007/s11257-019-09231-w
- [7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. *Attention Is All You Need*. In *31st Conference on Neural Information Processing System*. 2017. DOI: 10.5555/3295222.3295349.