Towards Intelligent Recommendations Using the Neural Knowledge DNA

Guangjian Ning [a], Chunwang Wu [a], Yuan Li [a], Haoxi Zhang [a], Edward Szczerbicki [b]

[a] School of Cybersecurity, Chengdu University of Information Technology, Chengdu, China; [b] Faculty of Management and Economics, Gdansk University of Technology, Gdansk, Poland

Abstract. In this paper we propose a novel recommendation approach using past news click data and the Neural Knowledge DNA (NK-DNA). The Neural Knowledge DNA is a novel knowledge representation method designed to support discovering, storing, reusing, improving, and sharing knowledge among machines and computing systems. We examine our approach for news recommendation tasks on the MIND benchmark dataset. By taking advantages of NK-DNA, deep learning, and the SOEKS technologies, our approach can learn from users' past behaviours to form reading preference of the user, and reuse learned knowledge for improving the recommendation performance.

Address correspondence to Chunwang Wu, School of Cybersecurity, Chengdu University of Information Technology, No. 24 Block 1, Xuefu Road, Chengdu, China, 610225. E-mail: ajexwu@cuit.edu.cn

Keywords: Recommendation systems, Recommender systems, Neural Knowledge DNA, Set of Experience Knowledge Structure, Knowledge Representation, Deep Learning.

INTRODUCTION

Nowadays, both the amount of available information and the number of users on the Internet are growing explosively. Consequently, problems of timely access to contents of interest on the Internet and information overload arise. This has driven the recommendation systems to become a hot and important research topic.

Recommendation systems tackle the problem of information overload (Konstan & Riedl 2012) by filtering decisive information fragments out of a huge amount of dynamically generated information based on user's interests, preferences, and past decisions on different items (Pan & Li 2010). By utilizing different filtering and recommending approaches, recommendation systems are able to predict whether a certain item would be interesting to a particular user or not based on the user's profile. Recommendation systems not only allow users to find items of interest more efficiently but also enable the service providers to actively send items to users to improve their service quality and users' experience (Isinkaye, Folajimi & Ojokoh 2015; Pu, Chen, & Hu 2011).

In this paper, we propose a novel approach utilizing the Neural Knowledge DNA (Zhang et al. 2017) to exploit users' past preferences and decisions for better

recommending performance. The rest of this paper is organized as follows: section two describes the idea of recommendation systems; section three introduces the concept of the Neural Knowledge DNA, section four introduces our proposed approach; section five presents the experimental results on an open benchmark dataset. Finally, in section six, concluding remarks are drawn.

RECOMMENDATION SYSTEMS

The recommendation system is defined as a decision-making method to deal with the problem of information overload by providing users with personalized, exclusive content and service recommendations related to users' interests and preferences (Rashid et al. 2002; Isinkaye, Folajimi & Ojokoh 2015; Schafer, Konstan, & Riedl 1999). The goal of the recommendation systems is to produce suggestions that are meaningful and interesting to a group of users for contents, items, or products (Resnick & Varian 1997). In general, the recommendation problem is distilled to the problem of rank estimating for the items that have not been seen by a user. Intuitively, this estimation is usually based on the ranks given by this user or similar users to other items. As a result, the system is able to recommend meaningful items to users according to the estimated unseen-item ranks (Adomavicius & Tuzhilin 2005). Examples of such applications include recommending shoes, clothes, and other products at e-commercial websites, short videos and movies at online video websites, and news at ByteDance Technology.

Recently, researchers have developed various approaches for creating recommendation systems. Based on the key technologies they are built with, recommendation systems can be broadly classified into three categories, namely Collaborative Filtering (Acilar & Arslan 2009), content-based recommending (Chen et al. 2008), and Hybrid approaches (Jalali et al. 2010):

• *Collaborative Filtering*: In collaborative filtering systems, the user is recommended items that people with similar tastes and preferences liked in the past;

• *Content-based recommending*: In these approaches, the user is recommended items similar to the ones the user preferred in the past;

• *Hybrid approaches*: These methods combine collaborative and content-based methods.

However, existing recommendation systems still require further improvements to make recommendation approaches more effective and applicable to an even broader range of real-life applications. For instance, better methods for representing user behaviours and the information about the items to be recommended are demanded for improving current recommendation systems (Adomavicius & Tuzhilin 2005). Therefore, in this paper, we introduce a novel representing method to enhance the performance of news recommendations. Motivated by the knowledge representation technology, we propose to apply a newly developed knowledge representation method, the neural knowledge DNA (NK-DNA), to support the task of news recommendation. By utilizing the NK-DNA, the user behaviours and the news information can be well

4

represented, which helps the recommender suggest more interesting news to users.

THE NEURAL KNOWLEDGE DNA

The Neural Knowledge DNA (NK-DNA) was introduced to represent and carry knowledge by using artificial neural networks (Zhang et al. 2017). It utilises the ideas underlying the success of deep learning (LeCun et al. 2015) to the scope of knowledge representation. The NK-DNA is designed in a similar way of how natural DNA formed that is built up by four essential elements. Like the DNA produces phenotypes, the Neural Knowledge DNA stores information and knowledge via its four components: Experiences, Actions, States, and Networks.

The NK-DNA's four-element combination is constructed to carry detailed information of decisions: *Experiences* are domain's historical operation segments with feedbacks from outcomes. *Actions* are used to store the decisions or motions the domain can select. While *States* are situations in which a decision or a motion can be made or performed. Finally, the learned knowledge is retained in the *Networks*.

Generally, the knowledge is obtained as a model through training in deep learning systems. The model usually saves values of weights and biases of the connections between neurons in a neural network, and the hierarchical information about the neural network. Once the neural network has been trained, the network will give results straightforward through the computation of its network layers after feeding it with inputs. Based on deep learning technology, the NK-DNA keeps and reuses knowledge

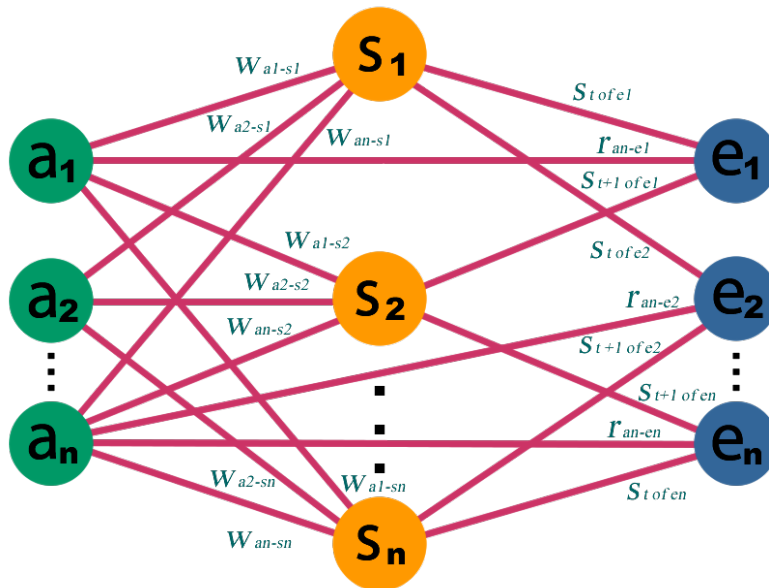in the same way. Figure 1 shows the concept of knowledge carried by the NK-DNA.

Fig. 1. Concept of the NK-DNA-carried knowledge.

Another important feature of this approach is that the NK-DNA uses past experience of preferences as the key to collect and develop intelligence for future decision making. Experience in the NK-DNA is stored as the Set of Experience Knowledge Structure (SOEKS) ((Sanin et al. 2012; Sanin & Szczerbicki 2006). Usually, the agent transitions from one state to another during its operation, and it makes decisions (picks actions) in each state and receives feedbacks from its operation; these states, actions, feedbacks, and transitions makes up the so-called 'experience'. In recommendation domain, the agent is the user, and the decisions are users' choices. For example, the user chose to read one article over another.

THE NK-DNA BASED RECOMMENDATION

In recent years, there has been a significant increase in the number of people accessing the news via digital means. However, due to the large number of news generated and published every day, it is difficult for users to find their own contents of interest from a large amount of news in a limited time. The proposed NK-DNA-based news recommender is not only able to capture and store the user's personal interests from the user's past news reading behaviours but also extract the representation of the news content for better candidate news rating.

A) The NK-DNA based Information Representation

In order to establish the recommendation model between users and news, we first need to generate representations of users, news, and reading experience. Because whether the news is highly interesting to a user, is not just depending on the news itself, either on the user's basic profile, but also based on the user's past news reading behaviours. For example, two users can have very different preferences on reading news, but in the meantime, they are quite similar in terms of basic profiles, like age, gender, education, etc. This suggests that if we want to give more accurate recommendations, we have to consider conditions of features of news, user profiles, plus individual reading experience, and link them together. For this reason, in the NK-DNA based recommendation method we store the news features, user profiles, and reading experience by using SOEKS and NK-DNA, Figure 2 shows how the information of these three parts are organized and kept in our proposed approach.
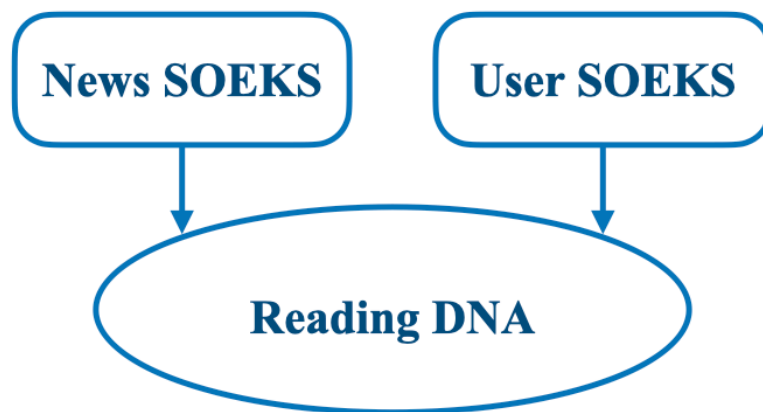
Fig. 2. Information representation in the NK-DNA method.

There are News SOEKS, User SOEKS, and Reading DNA in our system. The News SOEKS keep the data of the news, such as title, summary, body, author, press, type, and link. The User SOEKS store information about users, including the user's age, gender, nationality, education, etc. And the Reading DNA (to be acquired through training) carries the experiential knowledge on users' past reading decisions.

B) Reading Preference Extraction

Based on SOEKS representation, a reading preference model for news recommendation can be learnt through the NK-DNA training. In the NK-DNA, a neural network is used to carry the relation between *states* and *actions*: in our case, each *state* composes of the user (i.e. User SOEKS) who is going to click/pick the news and what are the news titles available for the user to pick (i.e. News SOEKS). And the user's click is saved as an *action* (i.e. Reading SOEKS). Finally, the neural network is trained to contain the experiential knowledge linking *states* and *actions*. Here, the recurrent neural network (RNN) is chosen to be the neural network for our recommender. RNN is a type of

8

recursive neural network that is very effective for learning the temporal and semantic information from sequential data (LeCun et al. 2015).

Figure 3 shows the overview of the recommender of our method. It consists of two modules: On the right (the red dashed rectangle), there is the NK-DNA carrying the reading preferences by utilizing an RNN. On the left (the blue dashed rectangle), there is a probability matrix decomposition module that generates the final recommendations to users based on features and information provided by the NK-DNA module.
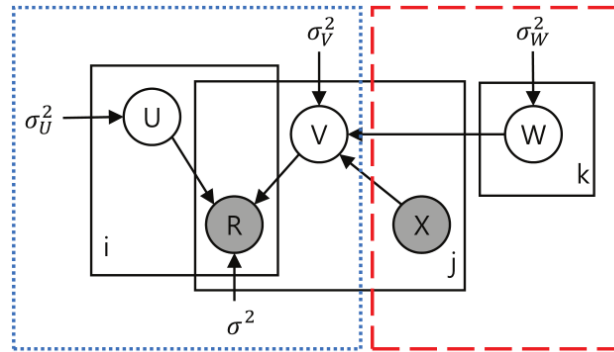

Fig. 3. Overview of the proposed recommender.

Suppose we have N users and M items, so that the observed ratings between users and items can be represented by a $R \in \mathbb{R}^{N \times M}$ matrix. Our goal is to find user-item latent models $U \in \mathbb{R}^{k \times N}$ and $V \in \mathbb{R}^{k \times M}$ whose product $U^T V$ reconstructs the rating matrix $\mathbb{R}$. In a probabilistic point of view, the conditional distribution over observed ratings is given by

$$\mathcal{P}(R|U,V,\sigma^2) = \prod_i^N \prod_j^M N\left(r_{ij} \mid u_i^T v_j, \ \sigma^2 \right)^{I_{ij}} \tag{1}$$

where $N(x \mid \mu, \sigma^2)$ is the probability density function of the Gaussian normal distribution with mean μ and variance σ^2, and I_{ij} is an indicator function.

For user latent models, as a generative method, we place the conventional priori, a zero-mean spherical Gaussian prior on it with variance σ_U^2.

$$\mathcal{P}(U \mid \sigma_U^2) = \prod_i^N N(U_i \mid 0, \sigma_U^2 I) \tag{2}$$

However, unlike user latent models, for item latent models in our approach, we assume that an item latent model is generated from three variables: 1) internal weights W in our RNN, 2) X_j representing the jth document, and 3) epsilon variable as Gaussian noise, which enables us to further optimize the item latent models for gaining better ratings. Thus, the final item latent models are obtained by the following equations:

$$\mathcal{V}_j = RNN(W, X_j) + \epsilon_j \tag{3}$$

$$\epsilon_j \sim N(0, \sigma_v^2 I) \tag{4}$$

For each weight w_k in W, we place zero-mean spherical Gaussian prior, the most commonly used prior:

$$\mathcal{P}(W \mid \sigma_w^2) = \prod_k N(w_k \mid 0, \sigma_w^2) \tag{5}$$

EXPERIMENTAL TESTING OF THE CONCEPT

In order to obtain more convincing experimental results, a high-quality baseline dataset, the MIND dataset (Wu et al. 2020), is used to test our proposed approach. The MIND dataset is built from anonymous news click history for Microsoft News users over a six-week week, contains more than 160, 000 news items, more than 15 million impressions, and more than 24 million clicks from 1 million anonymous users. In the

MIND data set, each news item has rich text information, such as titles, summaries, bodies, categories, and entities, as shown in table 1.

Table 1. Features and Items in MIND dataset.

#News	161,013	#Users	1,000,000
#News category	20	#Impression	15,777,377
#Entity	3,299,687	#Click behavior	24,155,470
Avg. title length	11.52	Avg. abstract length	43.00
Avg. body length.	585.05		

In the experiment, two existing methods are added to the comparison, which are Latent Dirichlet Allocation(LDA) and Convolutional Matrix Factorization (ConvMF). LDA (Blei et al. 2003) is a generative probabilistic model for collections of discrete data such as text corpora. LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modelled as a finite mixture over an underlying set of topics. Each topic is, in turn, modelled as an infinite mixture over an underlying set of topic probabilities. In the context of text modelling, the topic probabilities provide an explicit representation of a document. ConvMF (Kim et al. 2016) integrates convolutional neural network (CNN) into probabilistic matrix factorization (PMF). Consequently, ConvMF captures contextual information of documents and further enhances the rating prediction accuracy.

Moreover, to evaluate these recommenders' performance, three widely used indicators are used to measure the experimental results from different perspectives. These three indicators are namely AUC (Area Under the Curve), MRR (mean reciprocal rank), and NDCG@k (Normalized Discounted Cumulative Gain for k): AUC (Fawcett 2006) provides an aggregate measure of performance across all possible

classification thresholds. In other words, the AUC represents the probability that a random positive example is rated higher than a random negative example. MRR (Radev et al. 2002) is a statistic measure for evaluating any process that produces a list of possible responses to a sample of queries, ordered by the probability of correctness. The reciprocal rank of a query response is the multiplicative inverse of the rank of the first correct answer: 1 for first place, $\frac{1}{2}$ for second place, $\frac{1}{3}$ for third place and so on. The mean reciprocal rank is the average of the reciprocal ranks of results for a sample of queries. The NDCG@k (Wang et al. 2013) measures the usefulness, or gain, of a document based on its position in the result list taking into account the graded relevance values. The gain is accumulated from the top of the result list to the bottom, with the gain of each result discounted at lower ranks. NDCG@k is the top-k version of NDCG, which refers to the fixed number k of shown recommendations. By using NDCG@k the tested recommender system is penalized for not ranking relevant items.

The recommenders are tested with three different types of users, which are all users, overlapping users, and unseen users. The experimental results (see Table 2) show that our proposed method achieves better performance than another two methods in news recommendation tasks.

Table 2. Experiment results.

	All Users			Overlapping Users			Unseen Users		
	AUC	MRR	NDCG@5	AUC	MRR	NDCG@5	AUC	MRR	NDCG@5
LDA	59. 93	28. 23	30. 05	60. 23	28. 08	29. 94	59. 72	28. 35	30. 14
ConvMF	60. 16	30. 01	31. 38	62. 35	29. 22	31. 33	61. 89	29. 38	31. 41

| OURS | 64.31 | 30.47 | 33.86 | 64.70 | 30.39 | 32.84 | 64.01 | 30.53 | 33.88 |

Finally, we evaluate the performance of these recommenders taking into account of user interactions. Figure 4 shows how the recommenders' performances change with the increase in the number of user clicks.
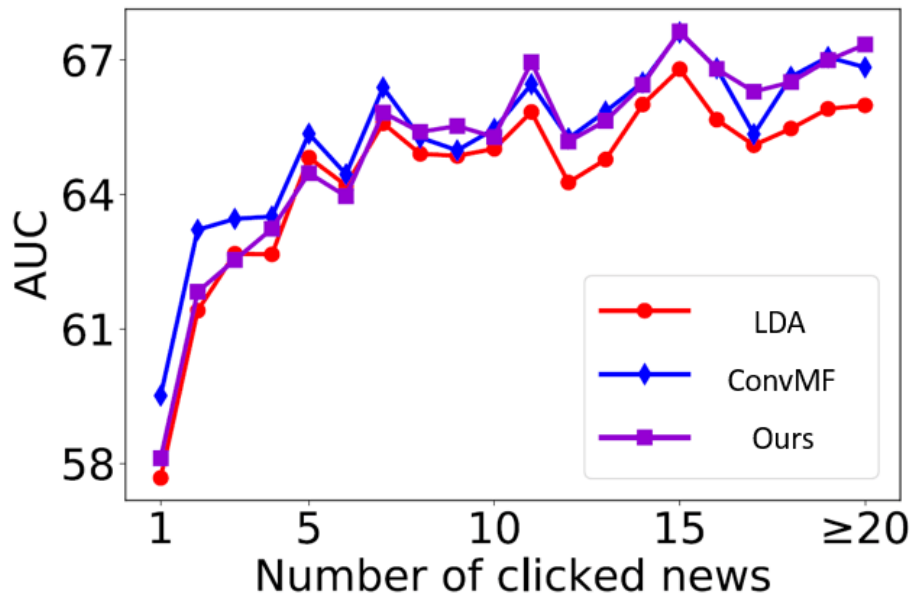

Fig. 4. Recommendation accuracy change under user click changes.

As we can see from the figure, the recommendation accuracies increase with the increasing in number of clicks. In general, all three recommenders perform poorly when there are few clicks. With user clicks raise, all recommendation accuracies improve. However, our proposed method outperforms the other two when the number of user clicks becomes bigger and bigger. The reason behind this is that the past user decision behaviour can largely enhance the NK-DNA for experimental knowledge extraction, which also helps improve the final performance of recommendation accuracy of our proposed method.

13

CONCLUSIONS AND FUTURE WORK

In this paper, we introduce a novel recommendation approach using past news click data and the Neural Knowledge DNA. By taking advantages of NK-DNA, deep learning, and the SOEKS technologies, our approach can learn from users' past behaviours to form reading preference of the user, and reuse learned knowledge for improving the recommendation performance.

As this research is at its early stage, there are many further work and refinements remaining to be done, some of them are:

- Further design and development of the NK-DNA based recommendation framework.

- Further design and development of the user-item representation.

- Refinement and further development of the reading preference learning algorithm.

ACKNOWLEDGEMENT

REFERENCES

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE transactions on knowledge and data engineering, 17(6), 734-749.

Acilar, A. M., & Arslan, A. (2009). A collaborative filtering method based on artificial immune network. Expert Systems with Applications, 36(4), 8324-8332.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

Chen, L. S., Hsu, F. H., Chen, M. C., & Hsu, Y. C. (2008). Developing recommender systems with the consideration of product profitability for sellers. Information Sciences, 178(4), 1032-1048.

Fawcett, T. (2006). An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874.

Haoxi Zhang, Cesar Sanin, Edward Szczerbicki, and Ming Zhu, Towards Neural Knowledge DNA. 2017. Journal of Intelligent & Fuzzy Systems, vol. 32, no. 2, pp. 1575-1584

Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. Egyptian informatics journal, 16(3), 261-273.

Jalali, M., Mustapha, N., Sulaiman, M. N., & Mamat, A. (2010). WebPUM: A Web-based recommendation system to predict user future movements. Expert Systems with Applications, 37(9), 6201-6212.

Kim, D., Park, C., Oh, J., Lee, S., & Yu, H. (2016, September). Convolutional matrix factorization for document context-aware recommendation. In Proceedings of the 10th ACM conference on recommender systems (pp. 233-240).

Konstan, J. A., & Riedl, J. (2012). Recommender systems: from algorithms to user experience. User modeling and user-adapted interaction, 22(1-2), 101-123.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep learning." Nature 521.7553: 436-444.

Pan, C., & Li, W. (2010, June). Research paper recommendation with topic analysis. In Computer Design and Applications (Vol. 4, pp. V4-264). IEEE.

Pu, P., Chen, L., & Hu, R. (2011, October). A user-centric evaluation framework for recommender systems. In Proceedings of the fifth ACM conference on Recommender systems (pp. 157-164), ACM, New York, NY, USA.

Radev, D. R., Qi, H., Wu, H., & Fan, W. (2002, May). Evaluating Web-based Question Answering Systems. In LREC.

Rashid, A. M., Albert, I., Cosley, D., Lam, S. K., McNee, S. M., Konstan, J. A., & Riedl, J. (2002, January). Getting to know you: learning new user preferences in recommender systems. In Proceedings of the 7th international conference on Intelligent user interfaces (pp. 127-134).

Resnick, P., & Varian, H. R. (1997). Recommender systems. Communications of the ACM, 40(3), 56-58.

Sanin, C., C. Toro, Z. Haoxi, E. Sanchez, E. Szczerbicki, E. Carrasco, W. Peng and L. Mancilla-Amaya. 2012. Decisional DNA: A Multi-Technology Shareable Knowledge Structure for Decisional Experience. Neurocomputing 88(0): 42–53.

Sanin C., Szczerbicki E., (2006) Using Set of Experience in the Process of Transforming Information into knowledge, International Journal of Enterprise Information Systems, 2(2), pp. 40-55.

Schafer, J. B., Konstan, J., & Riedl, J. (1999, November). Recommender systems in e-commerce. In Proceedings of the 1st ACM conference on Electronic commerce (pp. 158-166).

Wang, Y., Wang, L., Li, Y., He, D., Chen, W., & Liu, T. Y. (2013, June). A theoretical analysis of NDCG ranking measures. In Proceedings of the 26th annual conference on learning theory , PMLR 30:25-54.

Wu, F., Qiao, Y., Chen, J. H., Wu, C., Qi, T., Lian, J., ... & Zhou, M. (2020, July). Mind: A large-scale dataset for news recommendation. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 3597-3606).