# News Recommendation with Topic-Enriched Knowledge Graphs

Dongho Lee
Yonsei University
Seoul, South Korea
dongholee@yonsei.ac.kr

Byungkook Oh
Yonsei University
Seoul, South Korea
bkoh@icl.yonsei.ac.kr

Seungmin Seo
Yonsei University
Seoul, South Korea
smseo91@yonsei.ac.kr

Kyong-Ho Lee
Yonsei University
Seoul, South Korea
khlee89@yonsei.ac.kr

## ABSTRACT

News recommendation systems' purpose is to tackle the immense amount of news and offer personalized recommendations to users. A major issue in news recommendation is to capture the precise news representations for the efficacy of recommended items. Commonly, news contents are filled with well-known entities of different types. However, existing recommendation systems overlook exploiting external knowledge about entities and topical relatedness among the news. To cope with the above problem, in this paper, we propose *Topic-Enriched Knowledge Graph Recommendation System*(TEKGR). Three encoders in TEKGR handle news titles in two perspectives to obtain news representation embedding: (1) to extract meaning of news words without considering latent knowledge features in the news and (2) to extract semantic knowledge of news through topic information and contextual information from a knowledge graph. After obtaining news representation vectors, an attention network compares clicked news to the candidate news in order to get the user's final embedding. Our TEKGR model is superior to existing news recommendation methods by manipulating topical relations among entities and contextual features of entities. Experimental results on two public datasets show that our approach outperforms state-of-the-art deep recommendation approaches.

## CCS CONCEPTS

• **Information systems** → **Information systems applications**.

## KEYWORDS

Recommendation System, Knowledge Graphs, Neural Networks, News Recommendation

**ACM Reference Format:**
Dongho Lee, Byungkook Oh, Seungmin Seo, and Kyong-Ho Lee. 2020. News Recommendation with Topic-Enriched Knowledge Graphs. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20), October 19–23, 2020, Virtual Event, Ireland.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3340531.3411932


*CIKM '20, October 19–23, 2020, Virtual Event, Ireland*

ACM ISBN 978-1-4503-6859-9/20/10…$15.00
https://doi.org/10.1145/3340531.3411932

## 1 INTRODUCTION

Online news websites, such as MSN News and Google News, collect news contents from various sources to provide them to users and have attracted a massive number of users [2, 40]. However, it is nearly impossible for users to read all the articles since an enormous amount of news is generated every day. Although news articles cater to diverse personal interests, users often experience a hard time to choose interesting articles corresponding to their interests manually. Thus, it is essential for users to automatically receive personalized recommendations based on their interests [2, 21, 25, 29, 31, 35].

Recent studies focus on a comprehensive understanding of items (music, news, movie, books, etc.) to increase the accuracy of recommendation systems. To do so, state-of-the-art methods utilize knowledge graphs (KGs) to obtain explainability and informative connections among the items recommended. Typically, KGs are directed heterogeneous graphs where nodes correlate with entities and edges correlate with the relations of entities.

There exist several nonprofit knowledge graphs such as NELL[1] and DBpedia[2] as well as commercial ones such as Google Knowledge Graph[3] and Microsoft Satori[4]. Utilizing the relational facts of KGs in a recommendation scenario is necessary since it could offer rich and meaningful information about entities. Through exploring KGs, recommendation models can capture an item's relationship with its neighbors.

Since KGs are composed of semantic relatedness among items, recommendation models can benefit from learning their latent relationships and enhance the accuracy of recommended items. Also, with the relations between entities, a recommendation system can explain why a user may like a recommended item. For example, if a user watched and liked the movie "Whiplash", which was directed by Damien Chazelle, he or she may also enjoy "La La Land" since it shares the same director. Recommendation models can take practical advantages from KGs' relations between entities in exploring a user's potential interests with explanations via utilizing semantic relations.

[1] http://rtw.ml.cmu.edu/rtw/
[2] http://wiki.dbpedia.org/
[3] https://www.google.com/intl/bn/insidesearch/features/search/knowledge.html
[4] https://searchengineland.com/library/bing/bing-satori

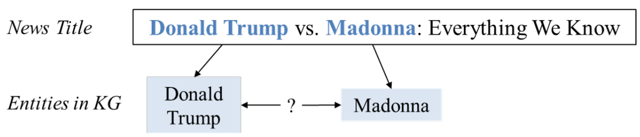


**Figure 1: Illustration of news title with multiple entities in KG.**

However, unlike other recommendation domains such as music, books, and movies, news recommendation has its own unique characteristics. First, news recommendation is highly time-dependent. As shown in [40], about 90% of news is clicked within just two days and then replaced by newer news. Thus, a thorough understanding of news content is necessary to overcome the limitations of conventional ID-based collaborative filtering methods.

Second, as shown in Figure 1, news titles may contain multiple entities, e.g., politicians, celebrities, companies, or organizations, which usually play crucial roles in the title. Furthermore, these entities are not independent to each other but can be linked to other entities through various relationships and are organized as a graph. Such characteristic makes the recommendation task more challenging. From the example news, we are able to suppose that Donald Trump and Madonna have conflicting political perspectives. However, such semantic relation between the two entities does not exist in common KGs.

Thus, approaches with exploiting KGs in news recommendations must be delicate in handling entities and their hidden relationships. It is hard for unsophisticated methods to model the rich information of news and the user's reading pattern. Furthermore, missing to capture such features would cause failure in discovering latent knowledge-level connections among the news. For the better understanding of news contents, subtle yet effective news modeling, which considers multiple entities with relation paths, is required.

Deep Knowledge-aware Network (DKN) [40] exploits entities and word embeddings for each individual news in a multi-channel approach, then designs a CNN model to aggregate features together. However, this method does not consider the semantic topics of the news itself. Topic information is also a vital factor in attracting users' interests and can increase the efficiency of the recommendation model if well employed.

Table 1 shows a brief example of a user's clicked news, which is highly related to marijuana stocks. This shows the importance of capturing the user's interested topics, which can affect the performance of a news recommendation model. However, with DKN, news titles with *Politics* and *Crime* topics received high probability scores even the user did not read them (labeled as 0 with the interaction label). This is because DKN ignored semantic topic information of news but only utilized recognized entities in news titles such as "Cannabis" or "Marijuana", while the actual crucial words to capture in the cases are "Prices" or "Stocks". Not only ignoring semantic topics, but DKN also does not consider the importance of words and semantic topics in news titles to learn more precise news representations.

LSTUR[1] learns the representations of news from their titles and explicitly given topic information. Although explicit topic labels can accurately represent the information of the news, a deficiency is that the simple data-driven topic information may not be meticulous enough to characterize the news topics, especially when the news titles can be listed as two or more different topics. LSTUR only uses explicit topic information, but for detailed news modeling, latent topic information is required.

For example, the following news title, "Donald Trump vs. Madonna: Everything We Know", appears as a Music topic. However, the actual article content is more related to Politics. Such misinterpretation in news modeling can cause a severe error in modeling users' interested topics. Thus, depending on explicit topic information and overlooking latent topic information of news can degrade the accuracy of news recommendation systems.

To address the limitations of existing approaches, we propose a Topic-Enriched Knowledge Graph Recommendation System (TEKGR) for efficient KG-aware recommendation. TEKGR is designed for click-through rate (CTR) prediction, which takes a piece of candidate news and a user's click history as input, and outputs the probability of the user clicking the news titles. In this paper, as input, we solely take news titles since they can be a pivotal factor for attracting users to read. Nevertheless, in addition to the title, our model is also feasible to use full article contents or abstracts of news. Note that our method can be simply utilized in any kind of recommendation scenario with short texts.

Specifically, we use three encoders to obtain news representation vectors: (1) word-level news encoder, (2) knowledge encoder, and (3) KG-level news encoder. The word-level news encoder learns the vector representations of news from their titles with three layers: word embedding layer, Bi-GRU layer, and attention layer. Through the layers, we focus on the words that play significant roles to represent the news.

The knowledge encoder uses the relational facts of KGs to extract the topics of a news title. Concepts of entities contain the "isA" relationship of entities in a KG. For example, an entity "Donald Trump" has the concepts of a politician, president, CEO, etc. From the concepts of entities in a news title, we extract topic information vector for detailed news modeling.

The KG-level news encoder constructs a topic-enriched subgraph from the entities of news titles by adding their 2-hop neighbors with topical relations, which are learned from knowledge encoder. Figure 2 concisely illustrates the example subgraph of Figure 1. Green nodes (*Trump* and *Madonna*) are the recognized entities connected with their 2-hop neighbor nodes. The topical relation vector from the knowledge encoder is added between the entities to imply *Politics* topic exists between the entities, which generic KGs do not have. Additionally, while DKN was not able to identify whether the entities are appearing in the same news, the topical relation can indicate that the two entities are appearing in the same news title.

After constructing the topic-enriched subgraph, we apply graph neural networks (GNNs) to get a news embedding vector. We then concatenate news representation vectors from each encoder to get the final representation of a news title. With extracted news representation, we use an attention layer to compare clicked news to the candidate news in order to get a user's final embedding.

Extensive experiments on a real-world dataset collected from various news platforms validate the effectiveness of our approach to news recommendation. The outcomes show that TEKGR attains

**Table 1: Sample user's news click log example.**

| | News Title | Interaction Label | Probability Score | Topic Label |
|---|---|---|---|---|
| Training | Will Falling Cannabis Prices Hurt Marijuana Stocks? | 1 | 1 | Investing |
| | This New Survey Highlights America's Changing View on Marijuana and Marijuana Stocks | 1 | 1 | Investing |
| | Marijuana Stocks: Cannabis is Outpacing the Chocolate Industry | 1 | 1 | Investing |
| | 10 Reasons Marijuana Stocks Still Go Higher | 1 | 1 | Investing |
| | 3 Top Cannabis Stocks to Watch in August | 1 | 1 | Investing |
| | Marijuana Stocks Drive Corporate Value by Delivering Automation | 1 | 1 | Stocks |
| | Marijuana Stock Predictions Expect Weed Stocks | 1 | 1 | Stocks |
| Test | Nevada Makes $30 Million In Marijuana Taxes DuringFirst Six Months Of Sales | 1 | 0.875 | Business |
| | Marijuana Stock Powers Higher On African Joint Venture | 1 | 0.736 | Stocks |
| | California Votes Legalize Marijuana | 0 | 0.724 | Politics |
| | Largest North American Marijuana IPO Soars in First Week | 1 | 0.756 | Stocks |
| | 3 Things Investors Really Need to Know About Marijuana Stock Aurora Cannabis | 0 | 0.882 | Investing |
| | Irmo Man Arrested Receiving Pound Marijuana Mail | 0 | 0.766 | Crime |
| | The 1 Word That Should Keep Marijuana Stock Investos Up At Night | 1 | 0.802 | Investing |

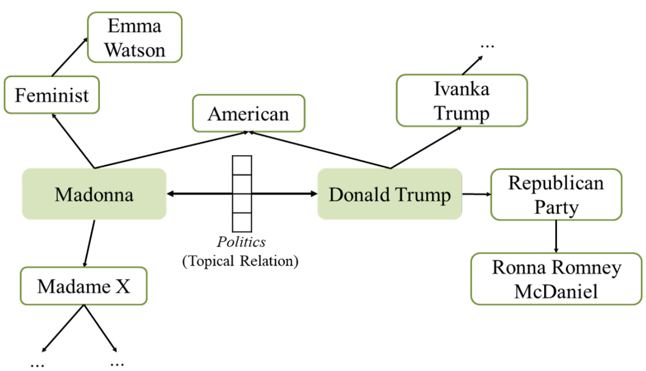


**Figure 2: Illustration of subgraph with added topical relation.**

considerable improvements over state-of-the-art deep-learning-based methods for news recommendation. Specifically, TEKGR notably outperforms baselines by 11.87% on F1 score and 12.36% on AUC. The results also demonstrate that the usage of contextual knowledge and implicit topic classification can bring additional improvement in the TEKGR framework.

## 2 RELATED WORK

Lately, recommendation systems have been taking more advantage of deep learning, which resulted in better performance in various recommendation cases. Different from other recommendation domains, news recommendation has to be much more complicated in text mining. In general, news recommendation systems can be broadly classified into content-based methods [2, 23, 25, 31, 48] and collaborative filtering (CF) methods [20, 22]. For example, Liu et al. [25] suggested a CF-based news recommendation method based on users' click behavior. The model used the Bayesian model to calculate the user interest features based on their click distributions of news articles on a different topic of news. On the other hand, Okura et al. [29]suggested learning the distributed representations of the news based on their similarity. After learning the distribution, the model learned user representations and clicked probabilities through recurrent neural networks (RNN) with users' click histories.

Content-based methods recommend news that is similar in content to the user-clicked news log [40]. To achieve this, recommendation systems model user profiles with the user's clicked news log while the news profile is also generated with the contents of the news. The similarity score between the user feature and the news feature is calculated to recommend the most suitable news for users. These approaches can improve their accuracy based on how the profile of the user and the news is generated. For example,[40] uses external information of knowledge graph in news recommendation to model news titles for differentiating from ID-based methods such as collaborative filtering. Nevertheless, such methods cannot model the preferences of users effectively - most news recommendation methods are based on CF techniques.

However, conventional CF approaches commonly suffer from several fundamental issues such as data sparsity and cold-start problems, since news items are constantly replaced. To address these limitations, studies have used content-based techniques to offer complementary information to CF [2, 23, 25, 29, 31, 35, 40]. For example, Okura et al. [29]proposed to learn news embeddings based on the news similarities considering topic information while using recurrent neural networks to learn user representations from their click histories. Thus, in our model, we model both user and news representations with external knowledge graph and neural networks to consider both issues from CF methods and content-based methods.

In general, there are two forms of usage of deep neural networks in deep recommendation systems: processing the raw features of users or items and modeling the interaction amongst users and items. Such approaches include DSSM[16], DeepFM[11], DMF[47], and Neural Collaborative Filtering[37] models. In addition, Multi-view Deep Learning[8] and SHINE[38] are also great examples of deep recommendation systems. In addition to the deep neural-based recommendation system, KGs have been added as side information in order to attain user's preferences of targeted items. KGs such as YAGO[14], Freebase[3], and Google Knowledge Graph[34] includes entities from real-world objects and concepts such as places, people, and organizations.

An entity may have various relationships and properties to other entities. Such knowledge bases have been proved to be critical for many real-world tasks, such as question answering[6], text classification[42], and even in regional similarity search[18]. Since knowledge bases can provide fruitful information on items, various recommendation researchers have used additional information to support their models. For example, Wang et al. [41] proposed an end-to-end recommendation model, KGNN-LS, that employs a KG as side information to expand the efficiency of the recommendation system. For a given user, KGNN-LS utilizes a trainable function to highlight valuable relations in KG and transforms the KG into a personalized weighted graph. Meanwhile, Fan et al. [9] proposed a graph neural network recommendation model, GraphRec, for a social recommendation. In general, the social recommendation KG contains two graphs: a graph of users with their relationships and a graph of user-item with their interactions.

Along with the general recommendation domain, news recommendation has also been gaining increasing attention in the text mining field and been widely researched for past years. In recommendation scenarios such as movies, books, or music, targeted items exist in KGs as entities. Thus, such circumstances mainly focus on relation path modeling between item and user or utilize item's neighbor information. On the other hand, in the news recommendation domain, models have to process semantic connections of entities with different types in KGs. To do so, text mining of news is crucial to make use of news entities. For example, DKN[40] proposed incorporating knowledge representation in news recommendations. RippleNet[39] also included the usage of a knowledge graph into recommendation systems to automatically propagate users' potential preferences while using a memory neural network to capture users' high-order preferences based on knowledge entities. KGCN[41] proposed knowledge graph convolutional networks for recommendation systems, which extends non-spectral Graph Convolutional Network[7, 12, 28] approaches to the knowledge graph through selectively aggregating neighborhood node features.

In addition to the knowledge graph, knowledge graph representation has also been extensively researched in recent years. Knowledge graph representation aims to learn a low-dimensional vector for each entity and relation in the knowledge graph while preserving the original graph structure. For example, Graph Convolutional Network utilizes localized graph convolutions for a classification task while KGAT[43] uses a self-attention network for information propagation, which utilizes a multi-head attention mechanism to increase model capacity. Recently, translation-based knowledge graph embedding methods have gained huge attention due to their concise models and superior performance. TransE[4], TransH[46], TransD[17], and TransR[24] utilize distance-based scoring functions when learning representations of entities and relations.

The major difference between these methods and ours is that we consider topical information during the process of recommendation through deep neural networks. Not only considering users and items, but we also took an additional step to consider the item's additional features with KGs. Specially, we employed multitask learning by utilizing the loss function of short-text modules in both the classification task and recommendation task. In conclusion, through contemplating various levels of news, we aim to model news and users as precisely as possible.

# 3 PROBLEM FORMULATION

The news recommendation with topic-enriched knowledge graphs problem is formulated as follows. The inputs of the system are candidate news title and clicked news titls while the output is a probability of the user clicking the candidate news. There are total M number of users and N number of news while the set of users would be denoted as $\mathcal{U}$={$u_1$,$u_2$,...,$u_M$} and set of news would be denoted as $\mathcal{V}$={$v_1$,$v_2$,...,$v_N$}. In the online news platform, user i's click history can be described as {$v_{i1}$, $v_{i2}$,...$v_{in}$}, where $v_{ij}$ (j=1,..., n) represents the *j*-th, clicked news' title by user *i*. We also implement knowledge graph $\mathcal{G}$, which contains of entity-relation-entity triples (*h*, *r*, *t*). Note that *h*, *r*, and *t* are the head, relation, and tail of a triple in a KG. For example, the triple(La La Land, film.film.director, Damien Chazelle) illustrates that Damien Chazelle is the director of the film "La La Land". Also, the head or tail can be associated with one or more entities in $\mathcal{G}$. For instance, Damien Chazelle is also linked with the movie "Whiplash" as a director. In a news recommendation scenario, the news title "Elon Musk unveils SpaceX's timeline for sending people to Mars", is connected with entities "Elon Musk", "SpaceX", and "Mars". The user-news interaction matrix Y = {$y_{uv}$| $u \in \mathcal{U}$, $v \in \mathcal{V}$} is defined according to user's clicked history, where:

$$y_{uv} \begin{cases} 0, & \text{if user } u \text{ did not click news } v \\ 1, & \text{if user } u \text{ clicked news } v \end{cases} \tag{1}$$

Denote the sequence of words in a news title $t$ as $t = [w_1, w_2, \ldots]$, where each word $w$ may be correlated with an entity $e$ in the KG. For example, in the news title "Lady Gaga Protests Donald Trump outside Trump Tower", $w_1$ and $w_2$ ("Lady" and "Gaga") are linked with the entity "Lady Gaga", while $w_4$ and $w_5$ ("Donald" and "Trump") are linked with the entity "Donald Trump". The topical relation $tr$, represents the newly added relation between entities which are extracted from the topic classification of news titles. Using users' click history along with the link between words in the titles and entities in the knowledge graph, we aim to predict whether user $i$ has a potential interest in news title $t_j$, which has not been clicked before.

# 4 RECOMMENDATION SYSTEM WITH TOPIC ENRICHED KNOWLEDGE GRAPH

In this section, we present the proposed TEKGR model in detail. We first introduce the overall framework of TEKGR, as illustrated in Figure 3, then discuss the process of each layer with encoders. TEKGR is composed of three layers: (1) KG-Based News Modeling Layer, (2) Attention Layer, and (3) Scoring Layer. For each piece of news, we extract a news representation vector through the KG-Based News Modeling Layer, which uses three encoders to extract features of the news, allowing us to obtain embedding vectors set for a user's clicked news. In the Attention Layer, we use an attention-based approach to dynamically compare the single user's clicked news to the candidate news through aggregating the user's interests with weights to get the user's final embedding regarding the candidate news. In Scoring Layer, we use the scoring function to calculate the probability of a user clicking the candidate news.

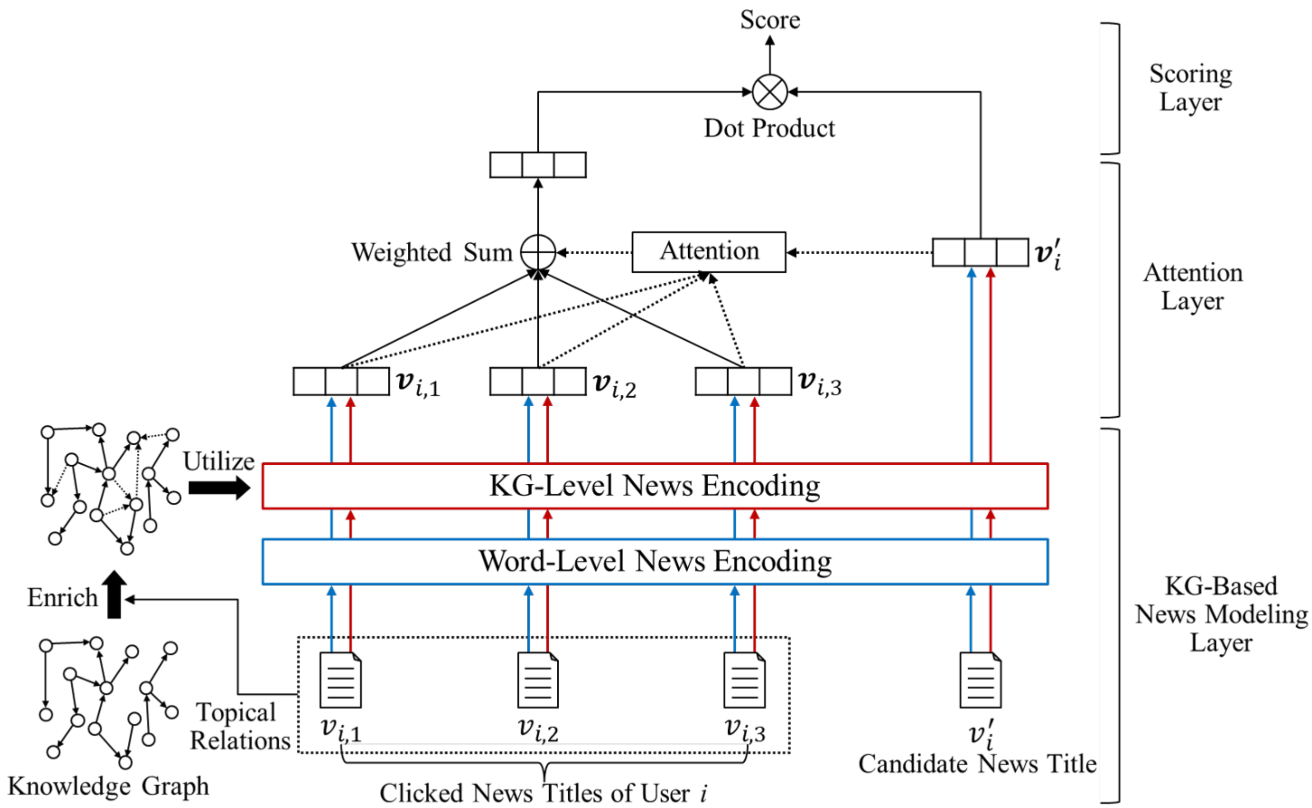


**Figure 3: Illustration of TEKGR framework.**

## 4.1 KG-based News Modeling Layer

As shown in Figure 4, the KG-based news modeling layer has three encoders to extract the news representation vector.

*4.1.1 Word-Level News Encoder.* The word-level news encoder has three layers to learn representations of news titles, which focuses on the meaning of words rather than considering latent knowledge representations in the titles. Word embedding, the first layer, converts a news title from a sequence of words to a sequence of dense semantic vectors (i.e., Through a word embedding matrix, the sequence $t$=[$w_1$,$w_2$,…,$w_n$] is transformed into [$\mathbf{w}_1$,$\mathbf{w}_2$,…,$\mathbf{w}_n$]). Bidirectional GRU (Bi-GRU)[13], the second layer in the word-level news encoder, captures the contextual information of the sequence. Local contexts of words play a significant role in order to understand the semantic meaning of the news title. For example, a user with a huge interest in marijuana investment would click marijuana news titles that have financial terms in them, such as "10 Reasons Marijuana Stocks Continue to Soar" or "Marijuana Stocks: Cannabis is Outpacing the Chocolate Industry." On the other hand, the user would have relatively low interests in marijuana news titles with non-financial terms in them, such as "California Votes to Legalize Marijuana" or "Irmo Man Arrested for Receiving One Pound Marijuana Mail". Since recurrent neural networks(RNN) is proved to be effective for sentence modeling[36], we apply Bi-GRU as Hao et al.,[13] does, which uses both forward and backward GRU to learn local context information:

$$\overrightarrow{h_t} = \overrightarrow{\mathrm{GRU}}(\mathbf{w}_t, h_{t-1}) \tag{2}$$

$$\overleftarrow{h_t} = \overleftarrow{\mathrm{GRU}}(\mathbf{w}_t, h_{t-1}) \tag{3}$$

To extract a hidden state $h_t$, we concatenate each $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ . Let the hidden unit number for each unidirectional GRU be u. For short, we denote all the $h_t$s as $H \in \mathbb{R}^{\mathrm{n} \times 2\mathrm{u}}$:

$$H = (h_1, h_2, ..., h_n) \tag{4}$$

Attention Layer, the third layer, focuses on important words in news titles to extract more informative features in news representations since each word in the news title may have varying levels of importance. For instance, if a user read articles "Amber Heard Speaks About Split From SpaceX CEO Elon Musk On Instagram" and "Elon Musk Inspired an Industry of Hyperloop Startups. Now He's Building His Own", the word "Elon Musk" plays a significant role regardless of the title's topic. Hence, we apply an attention network to choose keywords in titles. Through summing contextual representations of words weighted by the attention weights, we can obtain the final representation of a news title in word-level news encoder:

$$e_t = \sum_{i=1}^{N} \alpha_i h_t$$

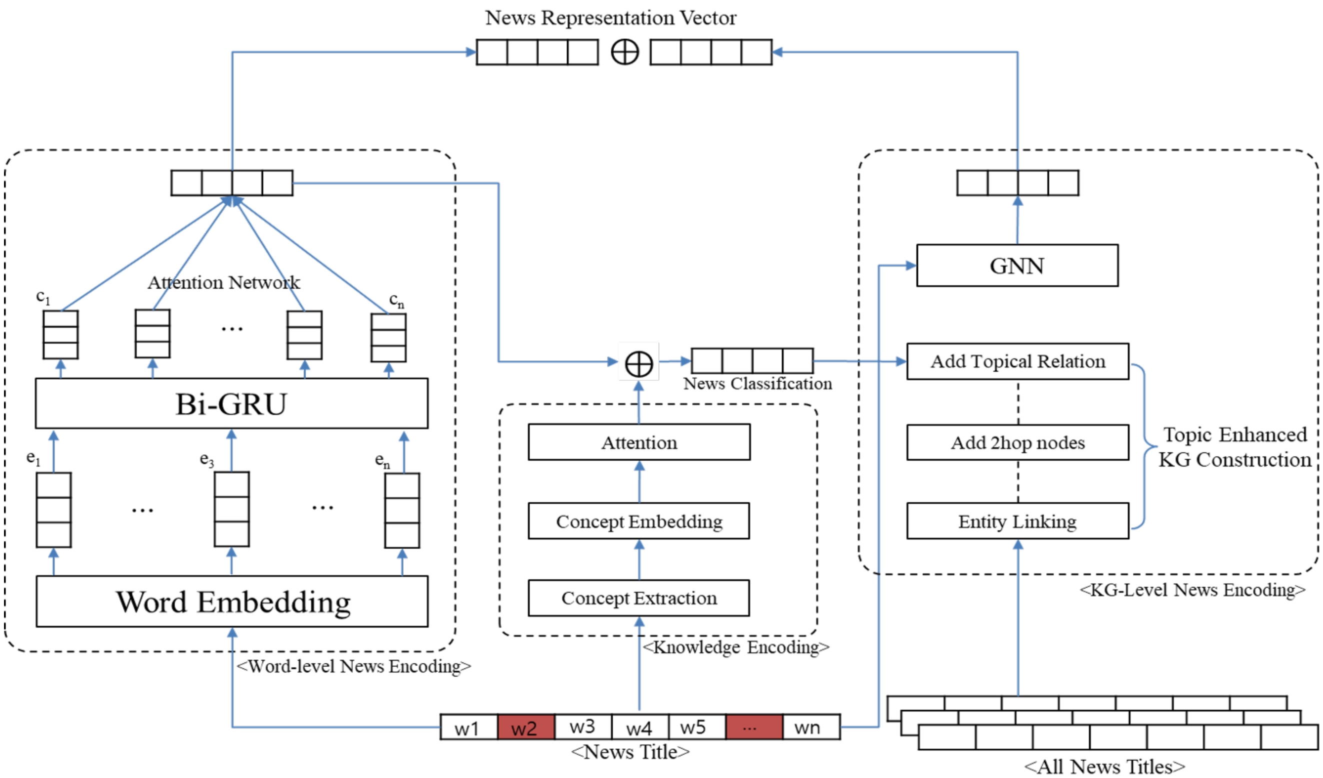


**Figure 4: Illustration of KG-based news modeling layer in the TEKGR.**

The representation of a news title will be concatenated later with the representation vector from the KG-level news encoder.

*4.1.2 Knowledge Encoder .* The knowledge encoder also has three layers to learn topical information on the news title. Since knowledge graphs provide more fruitful information to decide the news title topic, we used conceptual information from the knowledge graphs in the knowledge encoder. In general, short texts are mapped in implicit spaces and are represented as dense vectors[26]. While the implicit method with deep neural networks captures semantic information in a short text, it overlooks essential semantic relations in knowledge graphs such as "*isA*". To interact with unseen news titles, such contextual information is vital to understand the semantic meanings of the news.

For example, in the news title "Donald Trump vs. Madonna, Everything We Know," the implicit model is unable to recognize the following facts: Donald Trump is a politician and a CEO while Madonna is a singer and a feminist. Knowing such information can be helpful to classify the news as political news. Furthermore, utilizing such topical relations can represent hidden relations among the entities in the news title. Although previous approaches [39–41] showed that using knowledge graphs can alleviate data sparsity problems, they are unable to capture novel relations among entities (e.g., Donald Trump and Madonna), which are only illustrated in the news title but not in knowledge graphs. Thus, we employ "*isA*" relation and associate each news title with its relevant concepts in knowledge graphs through Concept Extraction.

With extracted concepts from a news title, the concept embedding layer maps each concept and word to a high-dimensional vector space. After the Concept Embedding, we use the self-attention network to assign a larger weight in proper concept; For example, in political news, Madonna should have a higher weight with a feminist concept than a singer concept. We denote each concept's self-attention network as follows: We concatenate the concept vector with a news representation vector (from Word-level News Encoder) to extract the topic information vector. In the end, we use the topic information vector of the news title as a relation among entities in KG-level news encoder for better use of knowledge graphs.

*4.1.3 KG-Level News Encoder.* The KG-level news encoder is composed of two parts: Topic Enhanced KG Construction and Graph Neural Network (GNN). The goal of Topic Enhanced KG Construction is to consider contextual knowledge information along with topical information for the recommendation system. The first step of Topic Enhanced KG Construction is to differentiate knowledge entities in news titles through entity linking method with predefined entities in knowledge graphs[27, 33].

With recognized entities in news titles, we extort all relations among the entities from a KG to build a sub-graph. However, since the relations among recognized entities may not be dense enough, we enlarge the sub-graph through adding entities within two hops of recognized entities. To balance between data sparsity and data overflowing, we add neighbor entities based on entity selection algorithm: we search at most 2-hop neighbors in the above-mentioned

KG for entities that appear in news titles. During the search, if another recognized entity appears, we assign more weight to the relation that connects the entities.

After adding neighbor nodes, to solve the unseen relation problem where there are no predefined relations among entities in KG, we add a topical information vector as a relation between entities that were extracted from the knowledge encoder. For example, there will be a new implicit topical relation between Donald Trump and Madonna. After constructing Topic Enhanced KG, to extract a news representation with topic and knowledge, we use GNN[41]. Through GNN, we can extract the news representation vector. The final representation of a news title is the concatenation of the representations from the word-level news encoder and KG-level news encoder.

## 4.2 Attention Layer & Scoring Layer

Now that we obtain news representation vector from multiple encoders, we use an attention layer to compare clicked news to the candidate news in order to get the user's final embedding. We denote user i's clicked history as $\{v_{i1}, v_{i2}, \ldots v_{iN}\}$ and its embeddings can be denoted as $\{e(v_{i1}), e(v_{i2}), \ldots e(v_{iN})\}$. Since a user can be attracted by various news, we employed the Attention Network[45, 49]. Attention Networks can capture user's diverse interests in news titles through assigning different weights on clicked news. Motivated from DKN[40], we first concatenate user's clicked news embedding and candidate news embedding before employing a Deep Neural Network as the attention network. Then, we use softmax function to evaluate the normalized impact weight.

In Scoring Layer, to train the model, we use straightforward dot production in accordance with[29] to calculate the probability score of a user clicking candidate news. Although the scoring layer can get complicated in various situations, news recommendation circumstance allows us to extract user and news representations in advance. Thus, we kept the scoring function as simple as possible for latency reduction as follows. The probability score $s(u, c_x)$ of the user clicking news is computed as $s(u, c_x) = u\top v_x$

# 5 EXPERIMENTS

In this section, we discuss our experiments and results with dataset descriptions and baseline models.

## 5.1 Dataset and Experimental Settings

For experiments, we use the following datasets for news and academic paper recommendation, respectively: Bing News[40] is the briefly released dataset from DKN since there is no off-the-shelf dataset for news recommendations. Each line in the dataset includes user id, news title, and click (0 for no click and 1 for click). The feedback of Bing News is collected from October 16, 2016, to August 11, 2017. Adressa News Dataset[10] is a news dataset that includes news articles in Norwegian in connection with multiple users. The dataset is collected from Adresseavisen's news portal with the help of the Norwegian University of Science and Technology (NTNU). As shown in Table 2, there is a total of 11,207 articles with 561,733 total number of users.

For experiments, we applied pre-trained GloVe embedding[30] for the initialization of the word embeddings, which has 200 dimensions. Besides, Microsoft Satori was used to construct the subgraph for a given dataset. A subset of triples that have a higher confidence level than 0.8 was selected during the experiments. For optimizing our model, we used Adam Optimizer[19], while the learning rate was 0.01. The initial batch size was set to 64. The ratio of training, evaluation, and test set were 6: 2: 2 for each dataset. For the click-through rate (CTR) prediction, we employed the trained model to each line of the dataset in the test dataset and obtained the predicted click probability. We chose the Area Under Curve (AUC)[5] to evaluate the recommended results. For top-K recommendation, we applied the trained model to choose K items with the highest predicted click probability for each user in the test set. To evaluate the K recommended list, we chose the F1 score metric. We ran each experiment 10 times independently and reported the average and maximum deviation as results.

## 5.2 Baselines

We use the following state-of-the-art methods as baselines in our experiments:

- DKN[40] is a deep news recommendation model which utilizes KG for obtaining knowledge features from news title and uses CNN and Attention Network for news modeling.
- LSTUR[1]is a neural news recommendation method that can learn both long and short-term user representations through a news encoder and user encoder. Furthermore, through the GRU network, LSTUR learns short-term representations of users from their recent news clicks.
- LibFM[32] is a matrix factorization model in various CTR scenarios. We set the concatenation of news features and user features as the input of the experiment. The news features involve TF-IDF features from its title and one-hot vectors of its topic and subtopic. The user features are a concatenation of normalized count features from clicked news' topics and subtopics and TF-IDF features from the clicked news.
- DeepFM[11] is a deep recommendation model that merges parts from factorization machines and deep neural networks. We used identical input for DeepFM in the same way that LibFM uses.

Note that except for LibFM, other baselines are all based on deep neural networks since we aim to compare our approach with state-of-the-art deep learning models.

## 5.3 Results

In this section, we illustrate the experiment results of an evaluation of various models and the comparison of TEKGR variants.

The results of a comparison of different models are shown in Table 4, and several observations stand out from the results: All the deep-learning-based baselines outperform LibFM by 3.0% to 10.0% on F1 and by 2.0% to 10.0% on AUC. This result illustrates that the deep models are efficient in handling the non-linear relations and dependencies in news titles. Since manually created features are usually not ideal, but neural networks can capture both global and local semantic context in the news, neural networks based

**Table 2: Dataset Statistics**

| Bing News Dataset | | | | Adressa News Dataset | |
|---|---|---|---|---|---|
| #users | 141,487 | Avg. #words per title | 7.9 | #articles | 11,207 |
| #news | 535,145 | News with no entity % | 2.04% | # entires | 2,286,835 |
| #logs | 1,025,192 | Avg. # entities per title | 3.7 | #users | 561,733 |
| #entities | 336,350 | News with more than 6 entity & | 17% | #Subscriber Users | 126,723 |
| #relations | 4,668 | | | Ratio of Subscriber users to total users | 0.2256 |

**Table 3: Comparison of different models.**

| Models | Bing News | | Adressa | |
|---|---|---|---|---|
| | F1 | AUC | F1 | AUC |
| DKN | 68.91 | 65.92 | 63.45 | 59.11 |
| LSTUR | 67.56 | 64.10 | 65.56 | 61.19 |
| LibFM | 58.33 | 56.52 | 58.67 | 55.93 |
| DeepFM | 61.15 | 58.83 | 57.11 | 59.64 |
| **TEKGR** | **70.20** | **68.88** | **67.82** | **64.06** |

models can learn more precise news and user representations for the recommendation model.

LSTUR and DKN show similar performances yet performs better than other baselines. This is probably because LSTUR captures the importance of words in each news title through attention networks. Also, LSTUR uses the GRU network to gain user's both long-term preferences and short-term interests in news reading while other baselines only learn a single representation for each user. Meanwhile, DKN's usage of the entity and contextual embedding through KGs helped the model to have sufficient performance.

TEKGR performs best among all methods in the datasets. Specifically, TEKGR outperforms other baselines by 3.0% to 9% on F1 and by 3.0% to 9% on AUC. Our TEKGR model can capture news title's topical feature and contextual knowledge to capture the sophisticated and varied user interests in news reading, while the baseline approaches overlook topic features and lack sufficient contextual knowledge to model news. In other words, our TEKGR uses various encoders to choose important words, important topics, and important contextual knowledge, which can aid in learning more informative news representations.

To evaluate the efficiency of our encoders, we conducted an additional experiment on our encoders. As shown in table 4, experiment results show that our encoders are working properly with respect to their goals. For example, in the table, there was a major drop of 5.08% in performance when we extract news vector without KG-level encoder. Since KG-level encoder captures contextual knowledge information such as hop neighbors and the hidden relation between news entities, news modeling misses important information related to the news title. For example, for the news title "Ford Mustang vs. Chevy Camaro vs. Dodge Challenger", the news modeling layer would have a hard time to realize the title is related to car performances.

Table 4 also proves the effectiveness of knowledge encoder. Without a knowledge encoder, our TEKGR model performs about 66.50%. As mentioned above, the knowledge encoder captures concepts of news entities and classifies news titles with the concept information. Furthermore, the encoder gives more weights on the relatively significant concept since one news entity can have multiple concepts (Madonna can have 'Singer', 'American', and 'Feminist' concepts, but in political news, Madonna should be considered as a 'Feminist' rather than a 'Singer'). Without knowing the knowledge information, the news modeling layer cannot recognize the news' topics. Since the topic information plays an important role in attracting users' reading patterns, the absence of knowledge encoder causes performance drop in our model.

The absence of a word-level encoder also caused a performance drop in our model. Word-level encoder focuses on the meaning of words rather than considering latent knowledge representations in the titles. This encoder was built to focus on words that are not recognized as entities in KGs. For example, the news title "10 Reasons Marijuana Stocks Continue to Soar" is attractive to users who have interests in marijuana investment. However, the word "Stocks" is not recognized as an entity while "Marijuana" does. With the sole entity information of Marijuana, our news modeling layer cannot be confident enough that users are interested in marijuana investment but rather recommend unrelated news titles such as "California Votes to Legalize Marijuana".

From the performance drop results, we can assume that each encoder plays a vital role for their purposes. Through the three encoders, TEKGR's news modeling layer can treat news titles from various perspectives, which ultimately improves the performance of our model.

We also conducted an experiment on the maximal hop number $h$ to see how outcomes change in TEKGR. As shown in Table 5, the model reaches its best performance when h is getting larger. Evidently, TEKGR with higher $h$ achieved the highest AUC compare to others on the Bing News dataset, which implies that modeling high-order semantic relations among entities can improve the accuracy of the model. Stacking additional hops in TEKGR, however, only achieves slight improvements, which means that TEKGR with 2 or $3h$ is sufficient to learn the semantic relations between entities, which is constant with the findings in [15, 44]. Conjointly evaluating Table 3 and Table 5, TEKGR with $2h$ steadily performs better than other baselines in most cases. It again confirms the efficacy of the knowledge encoder and KG-level news encoder, empirically showing that they model the contextual knowledge of news better. Based on the results, we can infer that large $h$ can drag noises that distract positive entities and relations while the $1h$ can hardly explore semantic relations among entities.

*5.3.1 Error Analysis.* From the experiment results, we will discuss error analysis in our model in this section. Even our model outperforms other baselines; there is still space to further investigate - False positive and false negative.

For false positive, where the recommendation system recommended news even the user did not read, our model was not able to

**Table 4: Effect of encoders in KG-Based News Modeling Layer.**

| Models | Bing News | | Adressa | |
|---|---|---|---|---|
| | F1 | AUC | F1 | AUC |
| w/o KG-Level Encoder | 65.12 | 62.15 | 59.11 | 56.94 |
| w/o Knowledge Encoder | 66.50 | 63.78 | 61.19 | 58.02 |
| w/o Word-Level Encoder | 67.43 | 65.81 | 62.93 | 60.19 |

**Table 5: The results of AUC with different hop numbers.**

| Hop Number $h$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Bing News | 67.66 | 68.88 | 69.02 | 69.96 |
| Adressa | 63.17 | 64.06 | 64.98 | 65.31 |

handle users' interests that contain the time or temporal demands. Most of the news that is recommended was not chosen by users due to the lowered interests in specific contexts. After reviewing the experiment result log, we found that such a problem occurs in all topics of news.

For example, in a weather news scenario, a user who reads "Tornado Warning Issued Multiple Counties Storms Blow Houston" already obtained the knowledge that there is a storm coming to the state Texas. Thus, related articles "National Weather Service Confirms Tornadoes in Texas" were not read by the user; even the news titles share the same place and weather forecast. Furthermore, since most users read weather news only once a day, it is rare for them to click weather news twice.

Sports result news also shows similar results. Once the users know what the results of sports events are, they rarely click related news again. A similar phenomenon happens with sports entities' transfer news – whether the news is true or not, most of the users did not revisit such news after reading them once. However, the news recommendation system misinterprets that a user is interested in specific entities and recommends related news. For false negative, where the recommendation system did not recommend news even the user read the news, there was an example of breaking and trending news. Specifically, social news took a portion of 55.64% in this error, while entertainment news took a portion of 28.21%. This result implies that our model is not sufficient enough to capture the user's sophisticated and improvised reading pattern.

For example, the news "US Navy calls off search for 7 missing sailors; several bodies found inside the destroyer", is largely read by the users from the dataset. The news was getting huge attention from the crowd and thus got high views from the users. Nevertheless, in the recommendation system, most users do not have entity interaction with "US Navy". Our TEKGR recommended the example news to users who read news related to military or politics mostly. Such breaking news gets close attention from users even though the recommendation system understands the news completely unrelated to the user profiles.

On the other hand, for the entertainment news, users clicked news titles based on celebrity entities. For example, browsing of the news on "Oscars Award" causes users to read several related news such as "Five decades and 200 films later, Jackie Chan finally wins Oscar" since "Jackie Chan" won the Oscar award, even users may never read news about "Jackie Chan" before. Such a reading pattern can be derived from users' short-term interest in the Oscars award. Although our model was able to capture the topic "Entertainment" in such scenarios, we were unable to capture users' inconsistent interests in various celebrity entities.

In short, TEKGR's error comes from not pleasingly capturing users' diverse reading patterns. Although our model focused enough on modeling news representation, we fall behind in modeling user representations. To solve such errors, we are considering implementing Collaborative Filtering (CF) methods with Recurrent Neural Networks (RNN) to capture users' diverse interests, which are mostly related to time and temporal demands.

## 6 CONCLUSIONS & FUTURE WORK

In this paper, we proposed the *Topic-Enriched Knowledge Graph Recommendation System* (TEKGR), which takes advantage of contextual knowledge and topic knowledge from KGs through encoders in news recommendation. TEKGR tackles a major challenge of overseeing topic feature of news from news recommendation scenarios. Through concept embedding, it extracts topic features from news titles then utilizes them during modeling news representation. Furthermore, we adopted attention networks in the encoders to select important words and the concepts of entities from a news title.

The experimental results show the substantial superiority of TEKGR compared with recent baselines, as well as the effectiveness of the topic feature extraction. For future work, we aim to consider a more diverse reading pattern and wider coverage of news content.

## ACKNOWLEDGMENTS

This work was supported by the Korea Electric Power Corporation (Grant number:R18XA05). Kyong-Ho Lee is the corresponding author.

## REFERENCES

[1] Mingxiao An, Fangzhao Wu, Chuhan Wu, Kun Zhang, Zheng Liu, and Xing Xie. 2019. Neural news recommendation with long-and short-term user representations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 336–345.
[2] Trapit Bansal, Mrinal Das, and Chiranjib Bhattacharyya. 2015. Content driven user profiling for comment-worthy recommendations of news and blog articles. In *Proceedings of the 9th ACM Conference on Recommender Systems*. 195–202.
[3] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. 1247–1250.
[4] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*. 2787–2795.
[5] Andrew P Bradley. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition* 30, 7 (1997), 1145–1159.
[6] Li Dong, Furu Wei, Ming Zhou, and Ke Xu. 2015. Question answering over freebase with multi-column convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 260–269.
[7] David K Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P Adams. 2015. Convolutional

networks on graphs for learning molecular fingerprints. In *Proceedings of the 29th Conference on Neural Information Processing Systems*. 2224–2232.

[8] Ali Mamdouh Elkahky, Yang Song, and Xiaodong He. 2015. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In *Proceedings of the 24th International Conference on World Wide Web*. 278–288.

[9] Wenqi Fan, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang, and Dawei Yin. 2019. Graph neural networks for social recommendation. In *Proceedings of the 2019 world wide web conference*. 417–426.

[10] Jon Atle Gulla, Lemei Zhang, Peng Liu, Özlem Özgöbek, and Xiaomeng Su. 2017. The Adressa dataset for news recommendation. In *Proceedings of the international conference on web intelligence*. 1042–1048.

[11] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. DeepFM: a factorization-machine based neural network for CTR prediction. *arXiv preprint arXiv:1703.04247* (2017).

[12] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *Proceedings of the 31st Conference on Neural Information Processing Systems*. 1024–1034.

[13] Yanchao Hao, Yuanzhe Zhang, Kang Liu, Shizhu He, Zhanyi Liu, Hua Wu, and Jun Zhao. 2017. An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 221–231.

[14] Johannes Hoffart, Fabian M Suchanek, Klaus Berberich, and Gerhard Weikum. 2013. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artificial Intelligence* 194 (2013), 28–61.

[15] Binbin Hu, Chuan Shi, Wayne Xin Zhao, and Philip S Yu. 2018. Leveraging meta-path based context for top-n recommendation with a neural co-attention model. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1531–1540.

[16] Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, and Larry Heck. 2013. Learning deep structured semantic models for web search using clickthrough data. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. 2333–2338.

[17] Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 687–696.

[18] Xiongnan Jin, Byungkook Oh, Sanghak Lee, Dongho Lee, Kyong-Ho Lee, and Liang Chen. 2019. Learning Region Similarity over Spatial Knowledge Graphs with Hierarchical Types and Semantic Relations. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 669–678.

[19] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[20] Joseph A Konstan, Bradley N Miller, David Maltz, Jonathan L Herlocker, Lee R Gordon, and John Riedl. 1997. GroupLens: applying collaborative filtering to Usenet news. *Commun. ACM* 40, 3 (1997), 77–87.

[21] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. 2010. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*. 661–670.

[22] Lei Li, Dingding Wang, Tao Li, Daniel Knox, and Balaji Padmanabhan. 2011. SCENE: a scalable two-stage personalized news recommendation system. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. 125–134.

[23] Jianxun Lian, Fuzheng Zhang, Xing Xie, and Guangzhong Sun. 2018. Towards better representation learning for personalized news recommendation: a multi-channel deep fusion approach. In *Proceedings of the International Joint Conferences on Artificial Intelligence*. 3805–3811.

[24] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. [n.d.]. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*. 2187–2187.

[25] Jiahui Liu, Peter Dolan, and Elin Rønby Pedersen. 2010. Personalized news recommendation based on click behavior. In *Proceedings of the 15th international conference on Intelligent user interfaces*. 31–40.

[26] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[27] David Milne and Ian H Witten. 2008. Learning to link with wikipedia. In *Proceedings of the 17th ACM conference on Information and knowledge management*. 509–518.

[28] Mathias Niepert, Mohamed Ahmed, and Konstantin Kutzkov. 2016. Learning convolutional neural networks for graphs. In *Proceedings of the International Conference on Machine Learning*. 2014–2023.

[29] Shumpei Okura, Yukihiro Tagami, Shingo Ono, and Akira Tajima. 2017. Embedding-based news recommendation for millions of users. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1933–1942.

[30] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.

[31] Owen Phelan, Kevin McCarthy, and Barry Smyth. 2009. Using twitter to recommend real-time topical news. In *Proceedings of the third ACM conference on Recommender systems*. 385–388.

[32] Steffen Rendle. 2012. Factorization machines with libfm. *ACM Transactions on Intelligent Systems and Technology (TIST)* 3, 3 (2012), 1–22.

[33] Avirup Sil and Alexander Yates. 2013. Re-ranking for joint named-entity recognition and linking. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. 2369–2374.

[34] Amit Singhal. 2012. Introducing the knowledge graph: things, not strings. *Official google blog* 16 (2012).

[35] Jeong-Woo Son, A-Yeong Kim, and Seong-Bae Park. 2013. A location-based news article recommendation with explicit localized semantic analysis. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. 293–302.

[36] Kai Sheng Tai, Richard Socher, and Christopher D Manning. 2015. Improved semantic representations from tree-structured long short-term memory networks. *arXiv preprint arXiv:1503.00075* (2015).

[37] Hao Wang, Naiyan Wang, and Dit-Yan Yeung. 2015. Collaborative deep learning for recommender systems. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 1235–1244.

[38] Hongwei Wang, Fuzheng Zhang, Min Hou, Xing Xie, Minyi Guo, and Qi Liu. 2018. Shine: Signed heterogeneous information network embedding for sentiment link prediction. In *Proceedings of the 11th ACM International Conference on Web Search and Data Mining*. 592–600.

[39] Hongwei Wang, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2018. Ripplenet: Propagating user preferences on the knowledge graph for recommender systems. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 417–426.

[40] Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. 2018. DKN: Deep knowledge-aware network for news recommendation. In *Proceedings of the 2018 world wide web conference*. 1835–1844.

[41] Hongwei Wang, Fuzheng Zhang, Mengdi Zhang, Jure Leskovec, Miao Zhao, Wenjie Li, and Zhongyuan Wang. 2019. Knowledge-aware graph neural networks with label smoothness regularization for recommender systems. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 968–977.

[42] Jin Wang, Zhongyuan Wang, Dawei Zhang, and Jun Yan. 2017. Combining Knowledge with Deep Convolutional Neural Networks for Short Text Classification. In *Proceedings of the International Joint Conferences on Artificial Intelligence*. 2915–2921.

[43] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 950–958.

[44] Xiang Wang, Dingxian Wang, Canran Xu, Xiangnan He, Yixin Cao, and Tat-Seng Chua. 2019. Explainable reasoning over knowledge graphs for recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 5329–5336.

[45] Xuejian Wang, Lantao Yu, Kan Ren, Guanyu Tao, Weinan Zhang, Yong Yu, and Jun Wang. 2017. Dynamic attention deep model for article recommendation by learning human editors' demonstration. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2051–2059.

[46] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*.

[47] Hong-Jian Xue, Xinyu Dai, Jianbing Zhang, Shujian Huang, and Jiajun Chen. 2017. Deep Matrix Factorization Models for Recommender Systems. In *Proceedings of the International Joint Conferences on Artificial Intelligence*. 3203–3209.

[48] Guanjie Zheng, Fuzheng Zhang, Zihan Zheng, Yang Xiang, Nicholas Jing Yuan, Xing Xie, and Zhenhui Li. 2018. DRN: A deep reinforcement learning framework for news recommendation. In *Proceedings of the 2018 World Wide Web Conference*. 167–176.

[49] Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. 2018. Deep interest network for click-through rate prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1059–1068.