I Know What You Want to Express: Sentence Element Inference by Incorporating External Knowledge Base

Xiaochi Wei, Heyan Huang, Liqiang Nie, Hanwang Zhang, Xian-Ling Mao, and Tat-Seng Chua

Abstract—Sentence auto-completion is an important feature that saves users many keystrokes in typing the entire sentence by providing suggestions as they type. Despite its value, the existing sentence auto-completion methods, such as query completion models, can hardly be applied to solving the object completion problem in sentences with the form of (subject, verb, object), due to the complex natural language description and the data deficiency problem. Towards this goal, we treat an SVO sentence as a three-element triple (subject, sentence pattern, object), and cast the sentence object completion problem as an element inference problem. These elements in all triples are encoded into a unified low-dimensional embedding space by our proposed TRANSFER model, which leverages the external knowledge base to strengthen the representation learning performance. With such representations, we can provide reliable candidates for the desired missing element by a linear model. Extensive experiments on a real-world dataset have well-validated our model. Meanwhile, we have successfully applied our proposed model to factoid question answering systems for answer candidate selection, which further demonstrates the applicability of the TRANSFER model.

Index Terms—Representation learning, external knowledge base, sentence modeling

1 Introduction

A natural language sentence is a group of grammatically linked words to express a complete thought. Beside the face to face offline communications, sentences are basic units of online human-human and human-machine interactions. On the other hand, the proliferation of Internet has propelled sentence typing to become a ubiquitous activity performed by billions of users daily, such as typing queries, comments, and emails. Cognitively formulating and physically typing sentences is, however, a time-consuming and error-prone process.[1] Spelling mistakes, forgetfulness and ambiguous intentions often make textual sentence input laborious. In order to save the intensive keystrokes of users, auto-completion is desired to subtly guide users in reformulating sentences correctly and efficiently.[2]

It is worth mentioning that several efforts have been dedicated to auto-completion. Query Auto-Completion (QAC) is a well-studied direction, which is a common visible feature that assists users in formulating queries by predicting their intended queries. Beyond queries that are sets of non-sequential keywords, sentences are much more complex and sophisticated, which follow the grammatical rules of syntax. Thereby, Existing QAC techniques cannot be directly adapted to the Sentence Auto-Completion (SAC) task. Traditional SAC methods utilize language models to infer the coming word, given the prior ones [1], [2], [3]. Despite their success in the traditional SAC task, they fail to complete the object of a sentence with the form of SVO (i.e., Subject, Verb, Object), which describes a relation between the subject and the object. A few illustrative examples for this kind of SAC are listed in Table 1. As can be seen from these examples, apart from the prior words, relations between subjects and objects in fact play a pivotal role in these cases.

As a complement to existing SAC methods, we focus on the problem of how to provide reliable element candidates for SVO sentences. In a sense, SVO sentences can be treated as an ordered triple with three elements, namely, a subject, a sentence pattern, and an object. This is somehow similar to the knowledge base representation introduced in TransE [4]. In particular, a knowledge base (KB) is a set of triples (head, relation, tail) and TransE models the relationship between the head and the tail entities by interpreting them as the translation operation on low-dimensional embeddings, (i.e, $\mathbf{head} + \mathbf{relation} \approx \mathbf{tail}$). In the light of this, object inference for SVO sentence can be naturally converted to the problem of one element inference given the other two, i.e., $\mathbf{object} \approx \mathbf{subject} + \mathbf{sentence\ pattern}$.

This inspires us to explore the knowledge representation to infer the missing element in a given sentence. However,

1. Typographical errors are very common, and the average accuracy for typists is only around 92 percent, according to eLearning Industry. http://tinyurl.com/hspmxmb

2. Take query auto-completion as an example. In 2014, global users of Yahoo search saved more than 50 percent of keystrokes when submitting English queries by selecting suggestions.

- *X. Wei, H. Huang, and X-L. Mao are with the Beijing Engineering Research Center of High Volume Language Information Processing and Cloud Computing Applications, School of Computer, Beijing Institute of Technology, Beijing 100081, China. E-mail: {wxchi, hhy63, maoxl}@bit.edu.cn.*
- *L. Nie is with the School of Computer Science, Shandong University, 1500 Sunhua Lu, High Tech Industrial Development Zone, Ji'nan 250101, China. E-mail: nieliqiang@gmail.com.*
- *H. Zhang and T.-S. Chua are with the School of Computing, National University of Singapore, Singapore 117417. E-mail: hanwangzhang@gmail.com, chuats@comp.nus.edu.sg.*

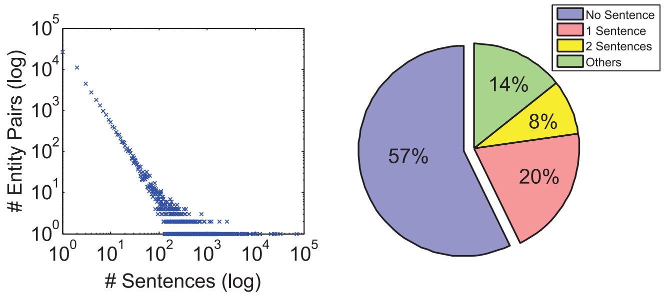
Manuscript received 10 May 2016; revised 24 Oct. 2016; accepted 25 Oct. 2016. Date of publication 27 Oct. 2016; date of current version 9 Jan. 2017. Recommended for acceptance by M. Sanderson.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TKDE.2016.2622705

TABLE 1
Examples of Object Completion in SVO Sentences

Incomplete Sentences	Objects to Complete
The first lady of America is ...	Michelle
Karachi is the largest commercial city of ...	Pakistan
Aron is important in ...	Islam
Microsoft was founded by ...	Bill Gates
Microsoft Corporation is headquartered in ...	Redmond
Michelle Obama was born in ...	Calumet Park

it is usually far from satisfactory to learn representations of entities and relations from a given sentence corpus only, because specific subject-object pairs are extremely sparse in real-world sentence data. To uncover insights into this problem, we first randomly selected 100,000 triples from Freebase.[3] Because specific relations between the head and the tail entities exist, they can be easily formulated as natural language descriptions. Thereafter, we tried to identify sentences containing any head-tail pair within the same triple from Wikipedia.[4] However, as the matching result illustrated in Fig. 1a, the number of sentences describing these given entity pairs follows a power-law distribution, i.e., only a few entity pairs have sufficient sentences to describe their relations, while most ones are mentioned in only a few sentences. This suggests that the most existing SVO triples do not have sufficient existing sentences to describe. Therefore, representations of most existing natural language relations can hardly be exactly learned via SVO sentences only. Moreover, up to 57 percent of these selected entity pairs are not mentioned in any sentences, as Fig. 1b shows. It means that natural language relations can hardly be fully covered by existing SVO sentences, which makes it impossible to infer correct subjects by traditional SAC methods. This strongly motivates us to leverage existing KB to reinforce the representation ability of sparse relations, as there exists a huge number of well-structured relational triples in the KB. Additionally, by combining existing relations in the KB, almost every natural language relation can be involved.

However, incorporating external KB into SAC is non-trivial and it faces three challenges: 1) Multiple and heterogeneous source fusion. Sentences from the given corpus are usually unstructured. In contrast, triples from external KB are well-structured. How to combine these two sources into a single homogenous structure is a largely untapped research problem. 2) Unified representation. Entities, relations, and sentences have different semantic spaces. For example, different dimensional vectors are used to describe entities, relations, and sentences in their own semantic spaces. On the other hand, different elements in these vectors may refer to different meanings. We thus have to represent them into a unified space to facilitate their translations. This is another challenge we are facing. 3) Complex relations. Some sentence patterns are complex and sophisticated, which are not necessarily archived in the KB. These are referred to hidden relations in the KB. For example, in the sentence "Michelle is the first lady of the US.", there exists the sentence pattern "... is the first lady of ...". The sentence



(a) (b)

Fig. 1. Data deficiency problem in the sentence corpus. Fig. 1a shows that the number of matched sentences given specific entity pairs follows a power-law distribution, and Fig. 1b present the ratio of entity pairs divided by the number of sentences.

pattern describes the hidden relation "First Lady", but it does not actually exist in the KB. How to learn such complex hidden relations from explicitly available relations in the KB is a hard problem.

To solve the aforementioned challenges, we devise a TRANSlation-based inFERence model, TRANSFER for short. It jointly regularizes and learns underlying sentence structures in the internal corpus, basic relational triples in the external KB, and hidden relations represented in sentences. The three involved elements mutually reinforce the learning performance. To accomplish this goal, we assume that complex hidden relations in sentences can be interpreted as the combination of existing basic relations in the KB (i.e., multi-hop relations). Entities and relations of sentences are transferred into the same semantic space by the hybrid of TransE and a deep learning model. We verified our model on Wikipedia and Freebase datasets. Experimental results demonstrated that our proposed TRANSFER model significantly outperforms several state-of-the-art methods.

Additionally, we apply our model to select answer candidates for factoid questions answering (QA). Evaluation on a manually constructed question answering set shows the applicability of the proposed model.

The main contributions of this paper are threefold:

- As far as we know, this is the first work on providing element candidates of natural language sentences via representation learning. Apart from SAC, the proposed methods can be applied to other applications, such as finding answer candidates to factoid questions.
- We propose a TRANSFER model that is capable of learning the subject, verb, and object representation by simultaneously harvesting the internal corpus and external KB.
- We have released our representative data collection to facilitate other researchers to repeat our experiments and validate their own ideas.[5]

2 Related Work

2.1 Sentence Auto-Completion

As an important research branch of SAC, QAC is one of the most prominent and visible features of modern search

3. http://www.freebase.com/
4. https://en.wikipedia.org/

5. http://datapublication.wix.com/transfer

engines. It saves laborious key strokes while user typing query into the search engine. Roughly speaking, the prior efforts on QAC fall into two categories: query-centric [5], [6], [7] and user-centric [8], [9], [10] methods. They either rely on statistical features of query contents, or leverage personal information of searchers. Though some of these efforts have been effectively applied in search engine query auto completion, they only provide specific co-occurrence query key words or phrase related to the user input. Consequently, they can hardly be applied to SAC directly.

Most SAC efforts origin from the lexical substitution track of SemEval-2007 [11]. It aims to find a replacement of a word or phrase removed from a sentence. Thereinto, two systems, KU [12] and UNT [13] stably and remarkably achieve the best performance. For KU system, only the N-gram language model is utilized, but the performance is outstanding. In contrast, the UNT employs a large varity of sources, e.g., WordNet synonym set, Microsoft Encarta encyclopedia, and bilingual dictionaries, combining with different retrieval model. But it is interesting to find that language model obtains the best performance compared with other methods in their experiments. This demonstrates the promising performance of language model in SAC. More recently, Zweig et al. [1], Mnih et al. [2] and Gubbins et al. [3] attempted to answer sentence completion questions in Scholastic Assessment Test (SAT). Different from ours, this task provides five answer candidates for each blank, and the most important is that, the missing words are neither subject nor object. Instead, they are prepositions, conjunctions or modifiers. Syntax information, rather than logical information, is more important in this task, consequently, language model based methods obtain better performance in these works. However, our task focuses on completing the object in a given SVO sentence, which requires considering the relation between the subject and the object. Existing efforts on SAC only consider internal sentences, which may suffer from data deficiency problem, as mentioned before. As a complementary work, we unify internal sentences and external KB into the same semantic representation via our proposed TRANSFER model.

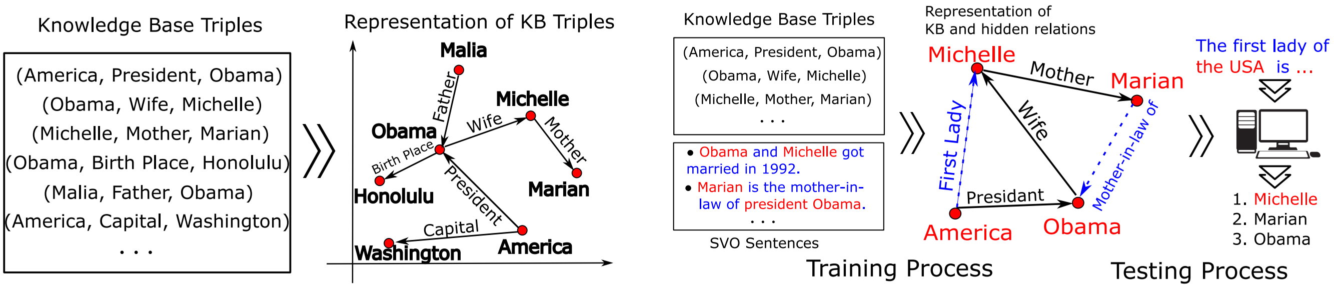
2.2 Factoid QA

Factoid QA has been popular for a long time. The TREC QA track [14], initially from 1999, is a milestone in this field. Most QA systems [15], [16], [17], [18], [19], accomplish this task by translating questions into keyword queries, in order to retrieve relevant passages from specific corpus and the Web. Then linguistic or statistical methods are utilized to extract answer candidates from the passages, and these candidates are ranked based on aggregated evidences and features to get the best answers. Different from these systems that search for answers from open text, knowledge-based QA systems [20], [21], [22], [23], [24] retrieve answers from well-structured knowledge bases. This kind of systems is springing out in recent years, following the rapid growth of KB, such as Freebase and Yago.[6] Bordes et al. [20] utilized the embedding method, representing questions and entities in KB into low dimensional embeddings, and ranked answers based on the scoring function. Yang et al. [21] joint embedded question patterns and relations in the KB, in order to convert natural language question into the KB query. Bao et al. [22] proposed a translation-based method, which regards the retrieval process as a translation from the natural language question to the KB search query. Yao et al. [23] utilized features of the KB, e.g., the entity type, to enhance the searching accuracy. Yih et al. [24] proposed a query generation method, which generates the KB search query based on the semantic parsing of the question. Due to the huge scale of KB, it is time-consuming to search these answers on the whole KB. To alleviate such problem, most existing knowledge-based QA systems only retrieve the sub-graph within at most 2-hops around the topic entity, but it may leave out many answer entities, especially in some complex questions or rarely mentioned entities. It hence may mislead the later process along the pipeline in the framework. In this paper, we try to apply the proposed method into the answer candidate selection process, to generate more reliable answer candidates.

2.3 Representation Learning

Representation learning aims to learn a transformation from complex, redundant and highly variable raw data to a representation that is mathematically and computationally convenient to process in machine learning tasks [25]. Representation learning has been widely applied in NLP, ranging from words to paragraphs. Distributed representations for symbolic data were first developed by Bengio et al. [26] in the so-called neural net language models. Mikolov et al. [27] improved the neural net language model by a recurrent layer and demonstrated the effectiveness of the recurrent neural networks among other language models. Collobert et al. [28] developed the SENNA system. By adding a convolutional architecture to the neural network, it learns the distributed representation of words, and achieves better performance on traditional NLP tasks. Additionally, some joint learning models were proposed to incorporate more information in learning word representation. Wang et al. [29] incorporated triples in the knowledge base into the neural language model, to jointly represent words and knowledge. Chen et al. [30] made use of the character and its position information in the word to improve the neural language model. As to the representation learning of sentences, several methods have been proposed, including recurrent neural networks [31], [32], recursive neural networks [33], [34], convolutional neural networks [35], [36] and sentence to vector [37]. Knowledge representation, another beneficiary of representation learning, works by vectorizing entities and direct edges between pairwise entities. Structured Embeddings (SE) [38] is a typical example, which embeds entities into a vector and relations into two matrices. Following SE, Semantic Matching Energy (SME) [39] justifies the correctness of a triple with an energy function. To reduce the complexity of SE and SME, TransE [4] learns a low-dimensional vector for each relation, and treats a relation as a translation from the head entity to the tail entity. Inspired by TransE, many translation-based models are springing up, such as TransH [40], TransR [41], and PTransE [42], to enhance the KB representation performance from different aspects. Our work is an instance of translation-based representation learning model, which is detailed in Section 3.2. Researchers also attempted to involve

6. http://tinyurl.com/m9uo24v



(a) Illustration of TransE (b) Illustration of TRANSFER

Fig. 2. Schematic illustration of TransE model and our proposed TRANSFER. In Fig. 2b, the red parts represent subjects and objects in sentences and the corresponding entities in the KB, while blue parts represent sentence patterns and the corresponding hidden relations in the KB.

text information into the knowledge embedding. For example, Toutanovan et al. [43] proposed a text and KB joint embedding method, which explores different models for text representation and KB representation. Dependency paths between two entities in sentences are used to describe relation between them. This can be regarded as an explicit representation of natural language relation. However, many complex relations can hardly be exactly expressed with dependency paths, such as the sentence "There have been 44 presidents of the United States, and Obama is leading the country now." In the light of this, we involved the path model to implicitly represent relations in sentences with relation pathes in the KB. Beyond prior efforts, we proposed a TRANSFER model, which is able to learn representations by jointly considering internal corpus and external KB in order to infer the object element in the sentence.

3 Preliminaries

3.1 Problem Definition

The major task of this paper is to infer the object of an SVO sentence, given a sentence corpus and a KB. In the sentence corpus, all subject and object entities can be matched into entities in the KB. To facilitate the element inference, all sentences in this corpus are in the form of SVO, which contains at least a subject, a verb and an object, and describes a specific relationship between the subject and the object. We denote the KB $\mathcal{T}$ as a set of triples in the form of (h, r, t), where $h \in \mathcal{E}$, $r \in \mathcal{R}$, and $t \in \mathcal{E}$. Thereinto, $\mathcal{E}$ and $\mathcal{R}$ respectively denote the set of entities, and the set of relations. Entities and relations are homogeneously encoded into vectors with the same dimension, and the dimension is denoted as k. Thereafter, the corresponding embedding representations of head, relation, and tail are denoted as $\mathbf{h} \in \mathbb{R}^k$, $\mathbf{r} \in \mathbb{R}^k$, and $\mathbf{t} \in \mathbb{R}^k$. We also represent each sentence with a triple (h_s, s, t_s), where h_s and t_s denote two entities (subject and object) mentioned in a sentence, and s denotes the *sentence pattern*, respectively. A sentence pattern is generated from a sentence by replacing the subject and the object entities with padding words, and these padding words are independent of entities. These three elements in a triple are also respectively encoded into k-dimensional vectors as $\mathbf{h_s} \in \mathbb{R}^k$, $\mathbf{s} \in \mathbb{R}^k$, and $\mathbf{t_s} \in \mathbb{R}^k$.

3.2 Knowledge Representation and TransE

Knowledge representation is to learn a low-dimensional continuous vector space. Generally, each entity in the KB is represented as a point in the space, while each relation is interpreted as an operation over entity embeddings. TransE is a simple yet effective knowledge representation model. In this model, entities are represented as data points in the embedding space, while relations are described as translations between the head entity and the tail entity in the embedding space, i.e., if the triple (h, r, t) exists in the KB, the embedding of tail entity t should be close to the embedding of the head entity h plus the embedding of the relation r. As examples illustrated in Fig. 2a, the triple $(America, President, Obama)$ exists in the KB, so it can be represented as **America** $+$ **President** $\approx$ **Obama**, where **America**, **President**, and **Obama** are corresponding embeddings of these three components in the triple. These Embeddings are obtained by minimizing the margin-based loss function,

$$L = \sum_{(h,r,t)\in\mathcal{S}} \sum_{(h',r',t')\in\mathcal{S}'} [d(\mathbf{h}+\mathbf{r},\mathbf{t}) + \epsilon - d(\mathbf{h}'+\mathbf{r}',\mathbf{t}')]_+, \quad (1)$$

where $\mathcal{S}$ is the set of triples, $\mathcal{S}'$ is the set of corrupted samples constructed by replacing a component in the triple into a random one, d indicates the distance between $\mathbf{h}+\mathbf{r}$ and $\mathbf{t}$ in the vector space, $[x]_+$ denotes the positive part of x, and $\epsilon > 0$ is a margin hyperparameter.

The simple form and its expansibility makes the model widely utilized in many joint learning models to incorporate knowledge information into existing models [29], [44]. In this work, we also utilize the TransE model to represent the KB, and hence joint represent sentences and knowledge triples into a united model.

4 Our Proposed TRANSFER Model

4.1 Assumptions

To jointly model sentences and triples, we make the following three assumptions, i.e., entity matching assumption, hidden relation assumption and relation path assumption, based on observations in our collected dataset:

- *Hidden Relation Assumption*. As mentioned in Section 3.1, SVO sentences can be regarded as triples with a subject, a sentence pattern and an object. Every sentence pattern describes a relation between the subject and the object in the sentence, which is assumed as a *hidden relation* in the KB, because it is frequently missing in the KB. In a sense, we can treat each sentence as a KB triple. This assumption homogenizes the structures between the internal corpus and the external KB.

- *Relation Path Assumption.* Hidden relations are frequently missing in the KB. In order to better describe it with existing edges, we assume that a hidden relation can be described with one or multiple relations in the KB. For example, the hidden relation "Grand Parent" is not directly available in Freebase, but it can be represented by a relation path "Parent–Parent". This assumption is able to address the missing hidden relations in the KB.
- *Entity Matching Assumption.* KB is a huge datasset containing world knowledge, and it can be regarded as a graph with each entity corresponding to a vertex and each relation corresponding to an edge. Though some studies declare that the KB is far from complete [45], [46], the missing edge problem is much more serious than missing vertex problem. Therefore it is reasonable to assume that the entities in the KB covers the most entities in the world, and the most subjects or objects commonly mentioned in SOV sentences can be matched to entities in the KB. This assumption makes it possible to jointly represent sentences and KB in a unified model.

We encode the aforementioned assumptions into our proposed TRANSFER model for sentence object inference, as illustrated in Fig. 2b. Subjects and objects in SVO sentences are matched with entities in the KB, and sentences patterns are described with hidden relations, which can be further represented with relation paths in the KB. Consider the first sentence "Obama and Michelle got married in 1992." as an example. The sentence pattern "... and ... got married in 1992." is represented as the relation "Wife" in the KB, while the sentence pattern "... is the mother-in-law of ...." of the second sentence is represented as the hidden relation "Mother-in-law of", which do not exist in the KB, but can be further represented as the relation path consisting of the existing relation "Mother" and "Wife". After training process, unstructured sentences in the internal corpus and structured relational triples in the external KB are simultaneously represented in a homogeneous graph. Consequently, in the testing process, given an incomplete SVO sentence, the TRANSFER model can return an appropriate subject entity, by inferring in the semantic space via the hidden relation of the testing sentence. As the example shows, when completing the sentence "The first lady of the USA is ...", the sentence pattern is represented as the hidden relation "First Lady", and the object entity "Michelle" is obtained via the translation operation $\mathbf{Michelle} \approx \mathbf{America} + \mathbf{First\ Lady}$. The TRANSFER model consists of three components, i.e., knowledge model, sentence model, and path model. The knowledge model is used to encode external knowledge triples. The sentence model in the second component is employed to represent SVO sentences and further make the inference. The path model represents hidden relations into the combination of direct relations existing in the KB, and bridges the gap between knowledge model and sentence model. We will detail these three parts in following sections.

4.2 Knowledge Model

We employ TransE [4] to model triples in the KB, because it has been proven to be effective yet simple in previous effort of joint representation learning. It is capable of embedding entities and relations into k-dimensional vectors in the same semantic space. The philosophy behind TransE is that the relation between two entities is represented as a translation in the embedding space. In other words, as long as a triple (h, r, t) holds, the embedding $\mathbf{t}$ of the tail entity t should be close to the embedding $\mathbf{h}$ of the head entity h plus the embedding $\mathbf{r}$ of the relation r. It can be formalized as $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$, and the energy function with the square loss of a given triple is,

$$E_k(h, r, t) = \frac{1}{2}\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_2^2. \tag{2}$$

A lower value indicates a higher probability that the given triple exists in the KB.

The drawback of TransE is that it does not perform well when dealing with the 1-to-N, N-to-1, and N-to-N relations, so the performance of the knowledge model can be enhanced by other improved translation-based knowledge representation models, such as TransH [40], TransR [41]. In this paper, considering the efficiency and the homogeneous structure with sentence triple, TransE is utilized instead of others.

4.3 Sentence Model

The sentence model is devised to capture the sentence structure. Based on the first assumption we described above, the sentence pattern s can be described as a hidden relation from the subject entity to the object entity. For a specific SVO sentence, its subject entity h_s and object entity t_s can be matched to entities in the KB. Based on the third assumption we mentioned, therefore, the subject entity and the object entity should share the same embeddings with their corresponding entities in the KB. Thus, the sentence (h_s, s, t_s) can be regarded as a *hidden triple* in the KB. It is called hidden triple mainly because the relation s may not actually exist in the KB. Considering the main idea of translation-based knowledge representation, it is reasonable to define the formula for a specific SVO sentence (h_s, s, t_s),

$$\mathbf{h_s} + \mathbf{s} \approx \mathbf{t_s}. \tag{3}$$

For a given KB, there exist finite varieties of relations in its relational triples. It is thus capable of learning a specific k-dimensional embedding for each type of relation in TransE. However, natural language is much more complex and contains richer meanings than predefined relations in the KB. Accordingly, there are infinite natural language descriptions of hidden relations. So it is impossible to train a lookup table for all sentence patterns, and utilize vectors in the lookup table to represent hidden relation embeddings as previous efforts of knowledge representation. On the other hand, each word in the sentence pattern acts as an indispensable part of the hidden relation. We expect the semantic and rich information of the whole sentence pattern can be equally encoded into the embedding $\mathbf{s}$. Considering the sentence pattern "The wife of the prime minister of ... is ...", the semantic meaning is quite different from the the sentence pattern "the prime minister of ... is ...", which only consider surrounding information of the missing entity and discard information of prior words in the sentence pattern. Towards this goal, we employ sentence vector [37]

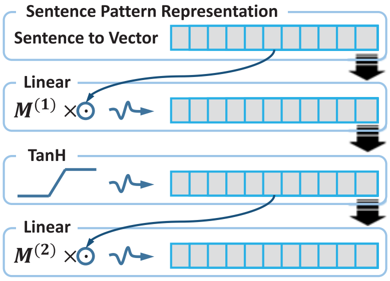


Fig. 3. The network utilized in our model consists of four layers. It utilizes the sentence vector as the input. The hidden layer contains a linear layer and a TanH layer. The output is a k-dimensional vector.

model to represent sentence patterns. In this model, every sentence is mapped to a unique vector, and concatenated with word vectors to predict the next word in a context. Softmax is employed as the classifier to make the prediction. This model maximizes the average log probability of the next word by optimizing representations of sentences and words. Compared with traditional bag-of-words methods, the extracted sentence vectors can capture the semantics and orderings of words. In our experiments, these sentence pattern embeddings are trained with all sentence patterns in the training set.

However, it is meaningless to directly perform translation on entity embeddings with sentence vectors, because entities in the KB and natural language sentence patterns are in different semantic spaces. To address such a problem, we develop a neural network model[7] $f(\mathbf{s})$ to bridge the semantic gap between the sentence space and KB space. Our network consists of four layers, i.e., an input layer, an output layer, and two hidden layers, as illustrated in Fig. 3. The hidden layers are composed of a linear fully-connected linear layer followed by a non-linear tanh layer, in which both linear features and non-linear features can be captured. The output layer is a k-dimensional vector, which plays the role of sentence pattern embedding. It is widely-accepted that the number of stacked layers in deep neural network models is a trade-off between the performance and the computational cost. Deeper networks have been explored in our experiments, but we did not observe significant performance improvement. In such a context, Eqn. (3) can be restated as

$$\mathbf{h_s} + f(\mathbf{s}) \approx \mathbf{t_s}. \tag{4}$$

Accordingly, the energy function for each sentence triple (h_s, s, t_s) is correspondingly defined as

$$E_s(h_s, s, t_s) = \frac{1}{2}\|\mathbf{h_s} + f(\mathbf{s}) - \mathbf{t_s}\|_2^2, \tag{5}$$

which encourages sentences with correct relations description between the subject entity and the object entity. Meanwhile, it penalizes the false ones.

4.4 Path Model

In the knowledge model, entities and relations are embedded, while in the sentence model, sentences are embedded, but these two parts are not seamlessly sewed up to strengthen the learning performance. Although they share parameters of entity embeddings, sentence patterns and relations are separately modeled in different components. To address such a problem, as described in the second assumption, we assume a hidden relation can be represented with a relation path $p = (r_1, r_2, \ldots, r_n)$ in the KB, linking two entities mentioned in the sentence. We thus have

$$\mathbf{h_s} + \mathbf{p} \approx \mathbf{t_s}, \tag{6}$$

where $\mathbf{p}$ is the embedding of path p. Since TransE is utilized to model the KB, i.e., $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$, we represent the embedding $\mathbf{p}$ of the path p as the sum of relation embeddings $\mathbf{r_i}$, where i indicates the index of relations in the path p. This is formalized as

$$\mathbf{p} = \mathbf{r_1} + \mathbf{r_2} + \cdots + \mathbf{r_n} \approx \sum_{i=1}^{n} \mathbf{r_i}, \tag{7}$$

where n stands for the number of hops, relying on the specific hidden relation. When the hidden relation exists in the KB, n is 1; otherwise it is equal to the number of hops from entity h_s to entity t_s. For instance, **First Lady** $\approx$ **President** $+$ **Wife**, while **Grandparents′ Birth Place** $\approx$ **Parent** $+$ **Parent** $+$ **Birth Place**. Thus the following formula is obtained, based on Eqns. (4), (6), and (7)

$$f(\mathbf{s}) \approx \sum_{i=1}^{n} \mathbf{r_i}, \tag{8}$$

where $\mathbf{s}$ is the embedding representation of sentence pattern s, $f(\mathbf{s})$ is the neural network defined in the sentence model, and $\mathbf{r_i}$ denotes the embedding of relation r_i in the relation path. For each given sentence, we define the following energy function to regularize its hidden relation representation

$$E_p(s, p) = \frac{1}{2}\|f(\mathbf{s}) - \mathbf{p}\|_2^2 = \frac{1}{2}\left\| f(\mathbf{s}) - \sum_{i=1}^{n} \mathbf{r_i} \right\|_2^2, \tag{9}$$

which is expected to get a low value when the path p is consistent with the hidden relation. In the implementation, for each training iteration, we randomly selected one predefined relation path for each sentence. The relation path selection method is detailed in Section 5.1.

4.5 Unified Model and Optimization

By unifying the aforementioned three models, we ultimately formalize the objective function of TRANSFER as

$$O = L_s(\mathcal{S}) + \lambda_k L_k(\mathcal{T}) + \lambda_p L_p(\mathcal{P}), \tag{10}$$

where $\mathcal{S}$ is the set of training sentences, $\mathcal{T}$ is the set of triples in the KB, and $\mathcal{P}$ is the set of pairs of sentence patterns extracted from $\mathcal{S}$ and the relation path identified from $\mathcal{T}$. Noticeably, each sentence pattern may correspond to more than one path, and we randomly select one for training in each iteration. $L_s(\mathcal{S})$, $L_k(\mathcal{T})$, and $L_p(\mathcal{P})$ refer to loss function of sentence model, knowledge model, and path model, respectively. Effects of external knowledge are controlled by parameters λ_k and λ_p. Inspired by TransE,

7. The network is implemented with the help of pylearn2, which is available at http://deeplearning.net/software/pylearn2/

margin-based loss functions are utilized for these three models,

$$\begin{cases} L_s(\mathcal{S}) = & \sum_{(h_s,s,t_s)\in\mathcal{S}} \sum_{(h'_s,s',t'_s)\in\mathcal{S}'} \\ & [E_s(h_s,s,t_s) + \epsilon - E_s(h'_s,s',t'_s)]_+, \\ L_k(\mathcal{T}) = & \sum_{(h,r,t)\in\mathcal{T}} \sum_{(h',r',t')\in\mathcal{T}'} \\ & [E_k(h,r,t) + \epsilon - E_k(h',r',t')]_+, \\ L_p(\mathcal{P}) = & \sum_{(s,p)\in\mathcal{P}} \sum_{(s',p')\in\mathcal{P}'} \\ & [E_p(s,p) + \epsilon - E_p(s',p')]_+, \end{cases}$$

where $[x]_+$ denotes the positive part of x and $\epsilon > 0$ is a margin hyperparameter. $\mathcal{S}'$, $\mathcal{T}'$, and $\mathcal{P}'$ are sets of corrupted samples, and they are constructed by replacing the original elements with randomly selected ones. They are formalized as

$$\begin{cases} \mathcal{S}' = \{(h',r,t)|h' \in \mathcal{E}\} \cup \\ \qquad \{(h,'r,t)|r' \in \mathcal{R}\} \cup \{(h,r,t')|t' \in \mathcal{E}\}, \\ \mathcal{T}' = \{(h'_s,s,t_s)|h'_s \in \mathcal{E}\} \cup \{(h_s,s,t'_s)|t'_s \in \mathcal{E}\}, \\ \mathcal{P}' = \{(s,p')|r \in p', r \in \mathcal{R}\}. \end{cases}$$

When it comes to optimization, stochastic gradient descent (SGD) is employed to minimize the objective function. We have to learn parameters including embeddings for entities, embeddings for relations and parameters in the neural network. We also normalize embeddings of entities and relations, after each parameter updating process. The normalization effectively prevent the training process to trivially minimize the objective function O by increasing norms of entity embeddings and relation embeddings. After training, the optimal entity embeddings, relation embeddings, and the neural network are obtained. Then TRANSFER could infer object entity given an SVO sentence with the sentence model, i.e, $\mathbf{t_s} = \mathbf{h_s} + f(\mathbf{s})$. Entities around the $\mathbf{t_s}$ in the embedding space will be returned as candidates. It is notable that, in the testing process, the sentence vector $\mathbf{s}$ is obtained by the sentence to vector model [37]. The model has been trained with sentence patterns in the training set. The subject entity in the SVO sentence can be extracted via NP chunking tools, e.g, OpenNLP[8] or OIE tools, e.g, Reverb.[9] Since we do not focus on the entity extraction, we will not detail these processes.

5 Experiments

5.1 Data Collection

To verify our proposed TRANSFER model in the object inference task, we constructed a dataset, including an SVO sentence set extracted from Wikipedia and a corresponding KB triple set extracted from Freebase. Wikipedia is the largest online encyclopedia in the world. It is maintained by crowdsourcing, and thus contains the largest amount of knowledge described with natural language in the world. Most existing KBs, such as Freebase, DBPedia, WikiData and Yago, are constructed based on the content of Wikipedia. Therefore, most entities mentioned in Wikipedia articles can be easily retrieved in the KB. As a result, it provides us a great opportunity to build an SVO sentence dataset from Wikipedia, which enables us to match subjects and objects in sentences to entities in the KB. On the other hand, Freebase is the largest KB in the world, and it contains the most amount of knowledge triples, hence it is an optimal choice for constructing the KB dataset. Moreover, we expect the hidden relations described with natural language can be represented with the combination of existing relations in Freebase. Since only SVO sentences are needed in our experiments, several heuristics were developed to guarantee that the most sentences in our dataset are SVO sentences, and their head and tail entities can be correctly linked to KB entities. Now we describe the construction of the dataset in detail.

The sentence set was constructed based on all sentences from English Wikipedia articles. We first extracted all complete sentences in English Wikipedia by removing all paragraphs without any punctuation mark at the end. Since interrogative sentences are not desired in this task, we also removed the sentences ending with a question mark. In this step, 45,018,770 sentences were obtained. In Wikipedia, anchor links in articles point to Wikipedia entities, which can be matched to the corresponding entities in the KB. So we utilized these anchor links to match entities in the later procedure. Because only SVO sentences are required in this task, where each sentence contains at least one subject and one object, all sentences in the sentence corpus should contain at least two anchor links. We thus eliminated sentences containing less than two anchor links. This ensures that most sentences in our corpus describe a relation between two entities. After this step, half of the sentences were filtered out, and 22,298,728 sentences were maintained in the dataset. Next, we used ReVerb [47], a famous Open Information Extraction (OIE) tool, to recognize SVO sentences that describe a relation between two entities in the sentence. We ran ReVerb on each sentence we obtained in the last process, and all sentences without returning any entity pairs were eliminated. Additionally, we discarded sentences with the ReVerb confidence score of less than 0.8, in order to guarantee the quality of the dataset. We have mentioned that we need to link subjects and objects to the KB, and anchor links in the Wikipedia sentences can be correctly matched with entities in the KB, in the light of this, we filtered the dataset again, remaining these sentences that both two Reverb-extracted entities are anchor text. These two extracted entities are regarded as the head entity h_s and the tail entity t_s in the sentence triple (h_s, s, t_s). Since we tried to ensure every sentence in our corpus is a SVO sentence, the rule we employed are extremely strict, so in this step, the most sentences were filtered out, and only 94,741 sentences were remained. Though a great many SVO sentences were filtered out with the strict rule, the remaining ones are of high quality. After that, we removed sentences containing sparse entities, appearing less than three times in the corpus, together with the corresponding sentences, to yield a dense dataset with 21,956 sentences.

Regarding the KB dataset, we generated a subset from Freebase. We first linked extracted entities to Freebase. For each entity, Freebase provides a link, pointing to the web pages of Wikipedia if available. These url links are stored in the "*/key/wikipedia/en_title*" value of Freebase entities. Since a Freebase entity can be linked to one Wikipedia page at most, we do not have ambiguous entity problem in our

8. http://opennlp.apache.org/
9. http://reverb.cs.washington.edu/

TABLE 2
Statistics of the Internal Corpus and the External KB

Data	# Triples	# Ent.	# Rel.	Avg. Deg.
Sentence	18,751	5,793	–	6.47
KB	140,785,671	5,170,340	7,152	27.23

experiment. However, word sense disambiguation module is required in real applications and we do not mentioned in this work. We then built a subset of Freebase, containing those entities with more than 50 edges or that can be matched to subject or object entities in the sentence corpus. A larger KB dataset theoretically can bring more information to the model, but it may also cost more time to learn the embeddings of entities. Therefore, 50 is a trade-off between the efficiency and the performance, which keeps about 10 percent (143,030,914) of Freebase triples in the dataset. This makes the KB dataset smaller but denser. Following that, we identified relation paths in the KB that link two entities in each sentence by bidirectional search, and we only counted up to four-hops in our data. Specifically, given a subject entity and an object entity in a sentence, we enumerated all entities which can link to the object entity within two hops, noted as $\mathcal{E}_{obj}$, and recorded their corresponding relation paths. We then enumerated all entities which can be linked to from the subject entity within two hops, noted as $\mathcal{E}_{sub}$, and recorded the corresponding relation paths. If there exists an entity e belongs to $\mathcal{E}_{obj}$ and $\mathcal{E}_{sub}$ simultaneously, we linked the subject and the object entity via entity e. The method is efficient, and the time complexity is $O(N^2)$, where N is the average degree of entities. If we could not find such an entity e, we deleted the sentence from the corpus. Finally we got the sentence corpus with 18,751 sentences. It is worth noting that, in order to enrich the relations, we also built an inverse relation r^{-1} for each relation. For example, if there exists a triple $(America, President, Obama)$, we would build a new triple $(Obama, President^{-1}, America)$ as the inverse, where the relation $President^{-1}$ is the inverse relation of $President$. Lastly, we removed all entities that are out of paths, and obtained a subset of the KB comprising more than 140 million triples. Table 2 summarizes our collected sentence corpus and KB.

In our experiments, sentence samples were divided into two disjoint sets, 80 percent as the training set $\mathcal{S}_{train}$, and 20 percent as the testing set $\mathcal{S}_{test}$. All extracted KB triples were used for training. For evaluation, we inferred the object entity for a given sentence, i.e., $(h_s, s, ?)$, by removing the tail entities. This was accomplished by ranking the entity candidates according to Eqn. (5) in the sentence model. Technically, TRANSFER model is able to handle the problems of missing object and subject by **object** $\approx$ **subject**$+$ **sentence pattern** and **subject** $\approx$ **object** $-$ **sentence pattern**, respectively. However, in reality, users usually type sentences from subjects to objects, and hence we only consider object inference in the SAC task. Additionally, subject inference is justified by providing answer candidates for QA systems in Section 6.

5.2 Evaluation Metrics and Baselines

Two metrics were adopted: Hits@k (H@k) and Mean Rank (MR) [4]. H@k is defined as the percentage of the test sentences where the real entity appears in the top k candidates. Specifically, k was set to 1, 3, 5, and 10 to evaluate the performance in different conditions, and it measures the performance of the top ranking candidates. The MR is the average ranking position of the real entity in the result candidate list. Therefore, a good entity prediction should achieve a higher H@k value or a lower MR value.

To verify the effectiveness of our proposed TRANSFER model, we compared it with the following baselines, and they can roughly divided into three categories according to their applications, i.e., sentence auto completion, representation learning (RL), and question answering (QA):

- *RNNLM (SAC)* [27]. RNNLM is a state-of-the-art language model, widely used in SAC tasks. it is implemented with recurrent neural network with three layers, i.e., an input layer, a hidden layer, and an output layer. Words in sentences are sequentially feed into the network for training and it utilizes the next word as the label for supervision. In our experiments, all sentences in the training set were employed to train the model. As to the testing process, words in test sentences are sequentially input into the trained model and it returns the probability of the next word. This model was trained with sentences only, while KB information can hardly be incorporated.
- *LSALM (SAC)* [1]. This method is a language model based method, initially used for answering SAT sentence completion questions. It involves semantic similarity into the language model to enhance the performance. It first obtains topic distributions of all words with a topic model, e.g., LDA in our experiments. Then it calculates the semantic similarity between the candidate words and the sentence by averaging similarities of the candidate word with other words in the sentence. At last, the final score of each candidate word is a linear combination of the semantic similarity and the probability calculated with language model. In our experiments we utilized the RNNLM as the language model. In this method, only co-occurrence information in the internal sentence corpus is considered.
- *UNT (SAC)* [13]. UNT is an outstanding sentence completion method in SemEval-2007. It leverages knowledge resources to improve the accuracy of traditional language model. Specifically, this method first extracts candidates according to knowledge resources. The original work utilized WordNet synonym sets, Microsoft Encarta encyclopedia synonyms, Roget and synonym sets generated from bilingual dictionaries as the knowledge resources. In order to make the method much more appropriate to the element inference task, we used the Freebase triples as the knowledge resource. All KB entities related to the subject entity were extracted as candidates. Then language model was used to rank these candidates. Similar as LSALM, we used the RNNLM as the language model.
- *Sentence Model (RL)*. This is the model described in Section 4.3, which only considers the internal

TABLE 3
The Overall Performance Comparison among Different Methods

Methods	MR	H@1(%)	H@3(%)	H@5(%)	H@10(%)	p-value
RNNLM (SAC)	$1{,}770 \pm 37$	1.151 ± 0.14	2.230 ± 0.35	4.211 ± 0.72	5.693 ± 1.09	$4e$-9
LSALM (SAC)	$1{,}341 \pm 34$	1.442 ± 0.14	3.582 ± 0.46	5.023 ± 0.87	8.696 ± 1.18	$3e$-9
UNT (SAC)	- -	1.044 ± 0.08	1.928 ± 0.24	2.464 ± 0.55	3.294 ± 0.98	$8e$-9
Sentence Model (RL)	$1{,}104 \pm 38$	1.403 ± 0.14	3.247 ± 0.39	4.813 ± 0.79	8.235 ± 1.51	$1e$-8
DistMult (RL)	998 ± 32	1.338 ± 0.15	3.023 ± 0.44	4.211 ± 0.79	9.854 ± 1.35	$6e$-8
Embedding (QA)	$2{,}609 \pm 35$	0.235 ± 0.10	0.706 ± 0.33	0.941 ± 0.74	1.215 ± 1.33	$9e$-14
JointEmbed (QA)	- -	0.704 ± 0.13	1.507 ± 0.38	2.245 ± 0.61	3.768 ± 1.19	$2e$-9
TRANSFER	$\mathbf{776 \pm 44}$	$\mathbf{2.187 \pm 0.21}$	$\mathbf{5.556 \pm 0.52}$	$\mathbf{7.693 \pm 0.98}$	$\mathbf{11.925 \pm 1.84}$	- -

sentence structure. Different from language models, it represents each sentence pattern as a vector, so it can describe global information of the sentence, and hence capture the relation between the subject and the object in the sentence.

- *DistMult (RL)* [43]. This is a joint embedding method, which jointly embeds sentences and KB triples. Different from our method, it uses pre-extracted dependency path from the head entity to the tail entity to represent the relation between them, and a convolutional neural network is used to embed the dependency path into the vector space. The energy function for knowledge triples and sentence triples are represented as $f(e_s, r, e_o) = \mathbf{v}(r)^T(\mathbf{v}(e_s) \circ \mathbf{v}(e_o))$, where e_s and e_o are head entities and tail entities, respectively, and r is the relation or the dependency path. Although it explores both the internal sentence information and the external KB, relation paths between entities are not considered.
- *Embedding (QA)* [20]. This is a sentence embedding method, initially developed for question answering. It represents questions and knowledge triples into a vector space in the training process by minimizing the similarity between the representation of the question, and its corresponding knowledge triple. In this model, the similarity between them is calculated with $S(q,t) = \mathbf{f}(q)^T\mathbf{g}(t)$, where q indicates the question, t indicates its corresponding KB triple, $\mathbf{f}(q)$ is the representation of the question and $\mathbf{g}(t)$ is the representation of the KB triple. It is notable that not all sentences can be matched to a KB triple, thus we only utilized the matched sentences in the training set for training. In the testing process, given a question, it ranks triples according to the similarity, and then returns the tail entity as the answer.
- *JointEmbed (QA)* [21]. This is another embedding method for question answering. This method jointly represents relations in the KB and sentence patterns into vector space by minimizing their similarities. Specifically, the similarity is calculated with $S(s,r) = \mathbf{f}(s)^T\mathbf{g}(r)$, where s is the sentence pattern, r is the corresponding KB relation, $\mathbf{f}(s)$ is the representation of the sentence pattern, and $\mathbf{g}(r)$ is the representation of the KB relation. The margin-based loss function is utilized to train these representations. In JointEmbed, not all sentence patterns can be matched to existing KB relations, so we only utilized the matched ones for training. Compared with Embedding method, JointEmbed only learns representations of KB relations rather than triples. As a result, less variables are learned in this model, which makes the model more efficient and concise than Embedding method. In the testing process, it ranks KB relations of the head entity according to the similarity to the sentence pattern, then it searches the answer entity with the KB query $(h, r, ?)$.

5.3 Overall Performance

In our experiments, we heuristically set $\epsilon = 1$, and $k = 50$ following the experiment settings of TransE [4]. We also explored different dimension k with 100, 200, 500, and 1,000, but higher dimension was not helpful to enhance the performance. We thus set $k = 50$ in further experiments to guarantee the efficiency. Table 3 shows the performance comparison between our model and baselines. Since the answer entity are not always in the final result lists of UNT method and JointEmbed method, the MR values are not available for them. From the table, we can observe that: 1) Methods integrating internal and external data (i.e., DistMult and TRANSFER) achieve better performance as compared to those that consider only internal ones. This demonstrates the effectiveness of incorporating external knowledge in the task of object inference. 2) Our TRANSFER model consistently outperforms other baselines in terms of MR and H@k. 3) Compared with traditional language model based SAC methods (i.e., RNNLM, LSALM and UNT), TRANSFER model achieves a remarkable improvement. This is because language model based methods utilize a sequential method, which can hardly represent the relation between the subject and object. In addition, RNNLM and LSALM only consider the internal sentence corpus. Even though UNT utilizes the external knowledge to extract candidates, complex natural language descriptions cause that those missing object entities are not always directly linked to head entities, thus external knowledge do not improve the performance in UNT model. In fact, the external KB is able to alleviate the data deficiency problem in TRANSFER model. 4) In contrast to DistMult, our method obtains a significant improvement. This further justifies our second assumption: complex and hidden relations can be effectively represented by relation paths in the KB. This makes the representation of hidden relation much more reliable and exploit more information in the KB. 5) The better performance of our model compared to Sentence Models signals that sentences can be better represented in the semantic space by jointly learning the internal corpus

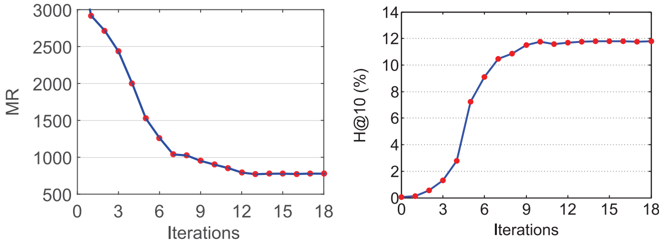


Fig. 4. The rate of convergence of TRANSFER model.

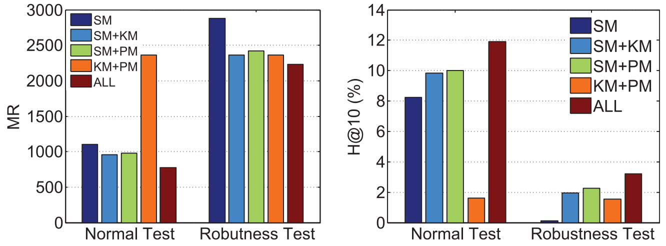


Fig. 5. Component influence in both normal test and robustness test.

and the external KB. 6) Most existing embedding-based QA methods (e.g., Embedding and JointEmbed) tries to translate natural language questions into KB query, i.e., (h, r, ?), however, most entity pairs described in sentences can hardly be directly linked in the KB, and this is the main reason that causes the poor performance of these two QA methods. This also demonstrates the importance of describing hidden relations with relation paths in the KB. 7) Results of pairwise significance tests on H@1 are all much smaller than 0.025, which indicates that the performance improvement of our model is statistically significant.

In general, the performance of all methods is very low, this is because of the following two reasons: 1) The task. Our task in this paper is to infer the object entity, rather than to complete other syntax words in SVO sentences. Thus the task is more difficult than traditional grammatical component completion, as reasoning is required. 2) The dataset. Sentences utilized in our experiments are much more complex than previous datasets (e.g., WebQ). Particularly, all sentences in previous datasets can find a direct relation from the subject entity to the object entity because complex ones are manually filtered out. However, in order to demonstrate the contribution of relation path, a challenging dataset, with less than 20 percent of sentences can be directly matched, is constructed in this work. This setting makes experiments more close to real applications.

Fig. 4 demonstrates the iteration process of our model over MR and H@10. We can see that it converges very fast within 10 iterations. Additionally, our TRANSFER model can infer an object within 0.074 seconds in the inference procedure. This demonstrates the efficiency of our method.

5.4 Robustness Analysis

In practice, some entities (subjects or objects) in SVO sentences may never appeared before. To verify the robustness of our model to cope with such scenario, we re-constructed new training and testing sets. The idea behind this is that each sentence in the testing set has at least one entity that does not appear in the training set. To accomplish this, we gathered all sentences, which have at least one entity with the frequency of less than three times, as the testing set. The remaining sentences are treated as training set. Thereafter, we obtained a training set of 14,888 sentences and a testing set of 3,863 sentences, where entities with less than three occurrence frequencies are all allocated into the test set.

It is noticeable that the language model based method, i.e., RNNLM, LSALM, and UNT, are not suitable in such experimental settings, as they cannot return the probability to an unknown word. We hence did not compare with them. The experimental results are presented in Table 4. From the table, we observed that: 1) The performance of Embedding and Sentence Models is very poor, close to random results. This reveals that these two models are not robust at all to new sentences with cold-start entities, without the assistance of external knowledge. 2) With the incorporation of external KB, DistMult and TRANSFER model achieve better performance. Again, this demonstrates the necessity of external KB. 3) The JointEmbed method does not utilize the embedding of entities, hence the performance of this method has less influence in this condition. 4) Our proposed TRANSFER model substantially outperforms DistMult. This is because our model is capable of simulating complex hidden relations with the basic relations existing in the KB. This well links internal and external data and enhances the representation performance of hidden relations.

5.5 Component-Wise Validation

We also analyzed the effectiveness of different components, Sentence Model (SM), Knowledge Model (KM), and Path Model (PM), in TRANSFER when dealing with the normal test and the robustness test, respectively. Experimental results are presented in Fig. 5. We observed the followings: 1) When there are sufficient training sentences (Normal Test) and the representation of entities in the test set have been well trained with training samples, SM plays a leading role in the object inference task, but KM+PM is unable to obtain a satisfactory performance without considering the internal corpus. 2) KM and PM can strengthen the performance of pure SM, no matter in the normal test or in the robustness test. When the KM and the PM are both regularized with SM, the best performance enhancement is obtained. 3) SM is unable to work well with

TABLE 4
Results of Robustness Analysis

Methods	MR	H@1(%)	H@3(%)	H@5(%)	H@10(%)	p-value
Sentence Model (RL)	$2{,}878 \pm 43$	0.0 ± 0.0	0.044 ± 0.01	0.078 ± 0.02	0.129 ± 0.05	$7e\text{-}7$
DistMult (RL)	$2{,}360 \pm 52$	0.387 ± 0.08	0.802 ± 0.14	1.398 ± 0.19	1.942 ± 0.27	$4e\text{-}6$
Embedding (QA)	$2{,}865 \pm 58$	0.0 ± 0.0	0.052 ± 0.01	0.072 ± 0.02	0.133 ± 0.04	$7e\text{-}7$
JointEmbed (QA)	- -	0.513 ± 0.03	1.313 ± 0.09	2.014 ± 0.14	3.116 ± 0.33	$5e\text{-}2$
TRANSFER	$\mathbf{2{,}125 \pm 74}$	$\mathbf{0.537 \pm 0.06}$	$\mathbf{1.398 \pm 0.15}$	$\mathbf{2.198 \pm 0.24}$	$\mathbf{3.222 \pm 0.37}$	- -

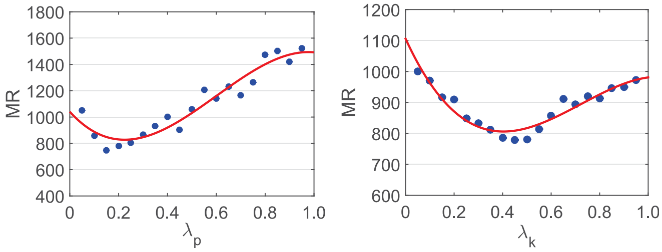


Fig. 6. Performance of TRANSFER w.r.t. varying parameters.

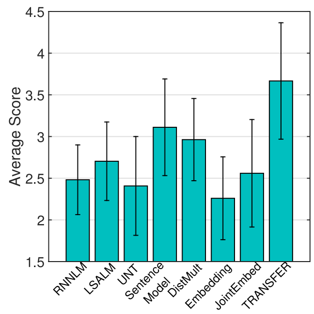
insufficient training sentences (Robustness Test) as illustrated in Section 5.4. Without the association of external knowledge, internal corpus only can hardly tackle the cold-start entities. Hence, in this situation, the performance is dominated by the external triples in the KB. 4) KM+PM gets the almost same performance in normal test and robustness test. This is because this method learns the emebddings of entities, relations and hidden relations based purely on the KB triples; it does not rely on the relations between subjects and objects in the sentence corpus; in other words, it is independent of training sentences.

5.6 Parameter Tuning and Sensitivity Analysis

In TRANSFER model, we have two key parameters as shown in Eqn. (10), i.e., λ_k and λ_p. They respectively regulate the effects of knowledge and path models. The optimal values of these two parameters were carefully tuned by grid search between 0.05 to 1.0 with the step size of 0.05. The parameters corresponding to the best MR were used to report the final results. We observed that our model reached the optimal performance when $\lambda_p = 0.15$, and $\lambda_k = 0.45$. Fig. 6 illustrates the performance of our model with respect to λ_p and λ_k, respectively. This is accomplished by fixing one and varying the other. We can see that: 1) when fixing λ_k and tuning λ_p, the MR value changes in a range of 745 between 776 and 1,521. 2) When fixing λ_p and tuning λ_k, the MR value changes in a range of 223 between 776 and 999. We concluded that the performance of TRANSFER model is non-sensitive to parameters around their optimal parameters It is notable that the performance of TRANSFER model do not decrease significantly when set $\lambda_p = 0$ or $\lambda_k = 0$. This is because that both knowledge model and path model can involve external knowledge. More specifically, the knowledge model enhances the performance of sentence model by reinforcing the representation of entities, and the path model boosts the performance by reinforcing the representation of sentence patterns. As the result shown in Fig. 5, adding either knowledge model or path model into the sentence model will obtain a performance enhancement, and a remarkable improvement will be obtained if all these three models are considered.

5.7 User Study

Finally, we comparatively validated the usability of our model on SAC. To be more specific, we randomly selected 50 sentences from the testing set as described in Section 5.1, and invited three volunteers to experience the SAC. For each sentence, the volunteers were asked to freely compare results of various SAC methods and provide their ratings on each method. We adopted the following quantization approach: 1 (poor), 2 (fair), 3 (good), 4 (very good), and 5 (excellent). The comparison results are shown in Fig. 7a. We can see that TRANSFER is able to satisfy users' experiences and achieves the best performance. Three representative sentence examples are displayed in Fig. 7b to explain



(a) Rating result of the user study.

Example Sentences	**Karachi is the largest commercial city of … and the Muhajirs are the main stakeholder in this city. (Pakistan)**	**The current website and suite of apps offered by MSN was first introduced by … in 2014. (Microsoft)**	**Aaron is important in … for his role in the events of the Exodus. (Islam)**
LSALM	India, China, Canada, NATO, Germany	Germany, France, Ravana,Google, Spain	India, England, dredd, Australia, Assyria
Sentence Model	Canada, Africa, **Pakistan**, Rome, trout	Brussels, Mumbai, Norway, Opava, Nubia	Pakistan, India, Christianity, Europe, Iran
DistMult	India, Australia, Canada, Cyprus, **Pakistan**	Copper, Hydroelectricity, gold, manhattan, middle-earth	Heracles, Mordor, buddhism, mythology, christianity
JointEmbed	Sindh, **Pakistan**, Abdullah Govt College for Women	Outlook, Web portal, **Microsoft**, Windows Live, Windows 95	Elisheba, Amram, Jochebed, Ithamar, Eleazar
TRANSFER	**Pakistan**, China, India, Japan, India	**Microsoft**, MSN, Google, Dell, Youtube	India, **Islam**, Jesus, god, Egypt

(b) Three selected examples and completion results returned by different methods, where the red words in the parentheses are the original words in blanks.

Fig. 7. Results of user study and demonstration of selected examples.

TABLE 5
Three Main Factors May Cause Errors and Corresponding Selected Examples

Issue	Examples
Sentence Constraints	**Sentence:** India defeated . . . in the first round at Bristol but lost to spain in Beckenham. (**Romania**) **TRANSFER Results:** Europe, Egypt, Germany, England, France
Multiple Relations	**Sentence:** Tezpur has direct flights from . . . (**Kolkata**) **TRANSFER Results:** Japan, Belgrade, Caracas, Mexico, Venezuela
Wrongly Extracted Sentence Triples	**Sentence:** Fabian Almazan is a jazz pianist, composer, and film score composer born in Havana, Cuba, and raised in . . ., Florida. (**Miami**) **Wrongly Recognized Head Entity:** Havana **TRANSFER Results:** Spain, Paris, Australia, Florida, California

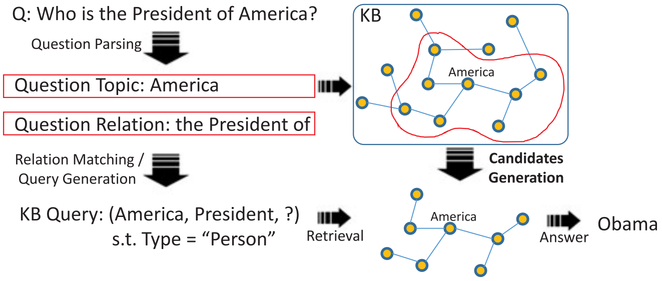


Fig. 8. Flow scheme illustration of knowledge based factoid question answering systems.

why TRANSFER gets better performance. The language model based method, LSALM, explores more information surrounding the missing words in the sentence, and hence the returned words may be irrelevant with subject entities, especially in complex sentences. Although the sentence model is able to return some words relevant to subject entities, the data deficiency problem significantly degrades the performance. By harvesting the triples in external KB, the candidates returned by DistMult are almost in the same category and are related to subject entities; however, the description of hidden relations is not accurate enough. Traditional embedding based QA methods, i.e., JointEmbed, performs well on simple questions, which can directly get answers via KB query, but complex natural language sentences cannot always be translated into simple KB query, e.g, the third sentnece, so this method can hardly get good performance stably. Finally, TRANSFER not only incorporates internal corpus and external KB, but is able to generate complex hidden relation representation, hence it returns the most relevant and accurate candidates.

Table 5 displays three examples to demonstrate the main factors that may cause errors in TRANSFER model. The first one is constraints in sentences. TRANSFER model represents the sentence pattern into the vector space. However these constraints can hardly be utilized sufficiently to filter out irrelevant entities. The second one is multiple relations. Since TransE is utilized in TRANSFER model, and it cannot address the 1-N, N-1, and N-N multiple relations well. In order to solve the problem, more complicated models, e.g., TransR and TransH, were explored in the future. And the last one is wrongly extracted sentence triples. Even though we have set strict rules to guarantee the quality of sentence triples in the data generation process, some limitations of the existing OIE method can hardly ensure all extracted triples and recognized entities are correct.

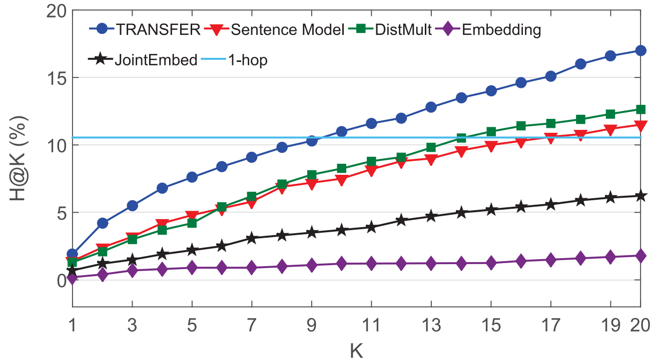


Fig. 9. Performance comparison among different methods on answer candidate selection.

6 Application: Finding Answer Candidates for Factoid QA

In this section, we applied the proposed TRANSFER model to another real-world application, finding answer candidates for factoid questions, in order to show the applicability of subject inference of the proposed model. Factoid QA system aims to answer fact-based natural language questions by providing named entities such as locations, people, organizations as answers; and it has been deeply researched for decades [48], [49], [50]. In recent years, with the popularity of the large-scale KB, researchers started to answer such questions by retrieving from the KB, and returning a KB entity as the answer. The flow scheme is illustrated in Fig. 8. Questions are first parsed to a question topic word and a question relation. Because of the huge size of current KB, it is quite inefficient to search on the complete KB for possible answers. Thus, most knowledge based factoid QA systems first select a set of answer candidates around the question topic entity; it then performs detailed analysis on these candidates based on the KB query generated from the question relation, in order to return the correct answer. Consequently, the best performance of these systems are determined by the quality of these answer candidates. Most of the prior efforts simply utilized these entities around the topic entity, i.e., one-hop [21], [22], [23], [51] or two-hop [52], [53], as candidates. These work performs well when dealing with simple factoid questions, however, they may get a lower recall when face with the complex questions; because relations described in these questions are usually beyond two hops. Consequently, correct answers cannot be included in these

TABLE 6
Illustration of Original Sentences and Their Corresponding Questions

Original Sentences	Questions	Answers
Spinach, along with other green leafy vegetables, is rich in iron.	Which vegetable, along with other green leafy vegetables, is rich in iron?	spinach
Bactria was the birthplace of Zoroastrianism.	Where was the birthplace of Zoroastrianism?	Bactria
Belgium also has a strong reputation in motocross.	Which country also has a strong reputation in motocross?	Belgium
In 1558, Akbar took possession of Ajmer, the aperture to Rajputana.	In 1558, who took possession of Ajmer, the aperture to Rajputana?	Akbar
Peru has the fourth largest area of rainforest in the world.	Which country has the fourth largest area of rainforest in the world?	Peru
Tektronix was the largest private employer in Oregon until the late 1980s.	Which company was the largest private employer in Oregon until the late 1980s?	Tektronix

TABLE 7
Selected Question Examples Which Can Be Answered by TRANSFER but Failed in One-Hop or Two-Hop Methods

Question	Topic Entity	Degree of Topic Entity	Answer	Degree of Answer Entity
Which company purchased Mannesmann?	Mannesmann	9	Vodafone	462
After her first six children had been killed, who gave birth to Krishna?	Krishna	304	Devaki	6
Who supported Eteocles, the incumbent king?	Eteocles	204	creon	6
At a very basic level, what is the liquid version of pneumatics?	pneumatics	24	hydraulics	30

candidates. In order to solve the candidate selection problem in factoid QA systems, we applied the TRANSFER model to the candidate selection module.

We randomly selected 250 sentences from the test set, and manually constructed questions based on these sentences by asking what is the subject in the sentence. Several examples of these questions and their original sentences are listed in Table 6. Since we focus on answer candidate selection, we provide these question topic entities for these questions, which are the object entities in original sentences. Since factoid questions can be convert to sentences without subject entities (e.g., the question "Where was the birthplace of Zoroastrianism?" can be convert to "... was the birthplace of Zoroastrianism."), the answer candidate selection task is considered as the subject inference task. Therefore, we utilized **subject** $\approx$ **object** $-$ **sentence pattern** to infer answer candidates. The H@k metric was utilized to evaluate the performance, where k was selected from 1 to 20. We compared our TRANSFER model with traditional one-hop methods, which were widely used for candidate selection in previous works. Those baselines used in object inference task were also compared in this experiment. The results are shown in Fig. 9. From the results, we have the following observations: 1) Only 10.6 percent of these questions can find the correct answers within one-hop relation around the topic entity. The value is similar as the H@10 of TRANSFER model. Considering that the average degree of our KB dataset is about 27, the one-hop method will return about 27 answer candidates for each question on average. However, to achieve the same accuracy, TRANSFER model only needs to return about 10 answer candidates. It thus provides more efficient method for later selection process. 2) Due to the limitation of one-hop candidate selection method, JointEmbed method can hardly achieve good performance. In addition, we explored a larger k value to inspect the upper bound of this method. It is interesting to find that this method obtained its best performance of 10.6 percent when more than 80 candidates were provided. This demonstrates that the candidate selection method determines the best performance of QA systems, and naively use one-hop methods cannot always get better performance. 3) Compared with other baselines, TRANSFER model stably outperforms them from H@1 to H@20. This demonstrates the stability of the TRANSFER model in different conditions, and the applicability of the TRANSFER model in various applications.

A case study was conducted to analyze why TRANSFER surpass traditional candidate selection methods. Several examples are listed in Table 7. Two facts mainly accounted for the failure of traditional candidate selection methods, according to the observation in our dataset. The first is the cold start entity problem. The first three questions demonstrate this problem. Relations described in these questions are simple and can be matched to existing relations in the KB. However, topic entities or answer entities are rarely mentioned in the KB triples. Thus the relation between the topic entity and the answer entity can hardly be linked via directed edges. However, TRANSFER model solves this problem by hidden relation assumption. It is capable of linking entities in the KB via hidden relations, and the missing relations are successfully fulfilled in the semantic space. The second fact is the complexity of relations described with natural languages. There are finite types of relations in the KB, while infinite types of relations exist and can be described with natural languages. The last question in Table 7 illustrates this condition. The relation "the liquid version of" is complex and cannot be directly matched with existing relations in the KB, but it can be simulated with the relation path "$Type - Strict\ Included - Included^{-1} -\ Type^{-1}$" from "hydraulics" to "pneumatics". This demonstrates the effectiveness of modeling relation paths in enhancing the embedding performance.

7 Conclusion and Future Work

This paper presents a novel representation learning model for sentence completion. It jointly unifies sentence structure in the internal corpus, relation triples in the external KB, and hidden relations in sentences into the same semantic space. In the light of this, external knowledge can be incorporated into sentence completion model to alleviate the data deficiency problem. Extensive experiments on a real-world dataset demonstrated the effectiveness and efficiency of the proposed method, and user study has shown that our approach greatly meets user expectations. Furthermore, we successfully applied the TRANSFER model to select answer candidates for factoid questions, which demonstrates the applicability of the model.

In the future, we plan to strengthen our model from the following aspects: 1) We will distinguish relation paths for given hidden relations to reduce the impact of noisy paths. 2) We will explore more effective neural networks to better represent hidden relations. 3) Beyond TransE, we will investigate other knowledge representation models, such as TransH and TransR, for TRANSFER model.

Acknowledgments

This work was supported by the State Key Program of National Natural Science of China (Grant No. 61132009), the National High Technology Research and Development Program of China (863 Program, No. 2015AA015404), the National Natural Science Foundation of China (61402036), and International Graduate Exchange Program of Beijing Institute of Technology. NExT research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its IRC@SG Funding Initiative. Xiaochi Wei is a visiting student at the National University of Singapore, supervised by Dr. Liqiang Nie and Prof. Tat-Seng Chua. Heyan Huang is the corresponding author.

References

[1] G. Zweig, J. C. Platt, C. Meek, C. J. Burges, A. Yessenalina, and Q. Liu, "Computational approaches to sentence completion," in *Proc. 50th Annu. Meeting Assoc. Comput. Linguistics*, 2012, pp. 601–610.

[2] A. Mnih and Y. W. Teh, "A fast and simple algorithm for training neural probabilistic language models," in *Proc. 29th Int. Conf. Mach. Learn.*, 2012, pp. 1751–1758.

[3] J. Gubbins and A. Vlachos, "Dependency language models for sentence completion," in *Proc. Conf. Empirical Methods Natural Language Process.*, 2013, pp. 1405–1410.

[4] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Proc. Advances Neural Inf. Process. Syst.*, 2013, pp. 2787–2795.

[5] U. Ozertem, O. Chapelle, P. Donmez, and E. Velipasaoglu, "Learning to suggest: A machine learning framework for ranking query suggestions," in *Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2012, pp. 25–34.

[6] D. Beeferman and A. Berger, "Agglomerative clustering of a search engine query log," in *Proc. 6th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2000, pp. 407–416.

[7] B. Mitra and N. Craswell, "Query auto-completion for rare prefixes," in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage.*, 2015, pp. 1755–1758.

[8] H. Cao, D. Jiang, J. Pei, E. Chen, and H. Li, "Towards context-aware search by learning a very large variable length hidden markov model from search logs," in *Proc. 18th Int. Conf. World Wide Web*, 2009, pp. 191–200.

[9] Q. Mei, D. Zhou, and K. Church, "Query suggestion using hitting time," in *Proc. 17th ACM Conf. Inf. Knowl. Manage.*, 2008, pp. 469–478.

[10] Z. Zhao, R. Song, X. Xie, X. He, and Y. Zhuang, "Mobile query recommendation via tensor function learning," in *Proc. 24th Int. Conf. Artif. Intell.*, 2015, pp. 4084–4090.

[11] D. McCarthy and R. Navigli, "SemEval-2007 task 10: English lexical substitution task," in *Proc. 4th Int. Workshop Semantic Eval.*, 2007, pp. 48–53.

[12] D. Yuret, "KU: Word sense disambiguation by substitution," in *Proc. 4th Int. Workshop Semantic Eval.*, 2007, pp. 207–213.

[13] S. Hassan, A. Csomai, C. Banea, R. Sinha, and R. Mihalcea, "UNT: SubFinder: Combining knowledge sources for automatic lexical substitution," in *Proc. 4th Int. Workshop Semantic Eval.*, 2007, pp. 410–413.

[14] E. M. Voorhees, "The TREC-8 question answering track report," in *Proc. 8th Text Retrieval Conf.*, 1999, pp. 77–82.

[15] C. Kwok, O. Etzioni, and D. S. Weld, "Scaling question answering to the Web," *ACM Trans. Inf. Syst.*, vol. 19, no. 3, pp. 242–262, 2001.

[16] J. Lin, "An exploration of the principles underlying redundancy-based factoid question answering," *ACM Trans. Inf. Syst.*, vol. 25, no. 2, 2007, Art. no. 6.

[17] D. Ravichandran and E. Hovy, "Learning surface text patterns for a question answering system," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguistics*, 2002, pp. 41–47.

[18] L. Nie, M. Wang, Z. Zha, G. Li, and T. Chua, "Multimedia answering: Enriching text QA with media information," in *Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2011, pp. 695–704.

[19] L. Nie, M. Wang, Y. Gao, Z. Zha, and T. Chua, "Beyond text QA: Multimedia answer generation by harvesting Web information," *IEEE Trans. Multimedia*, vol. 15, no. 2, pp. 426–441, Feb. 2013.

[20] A. Bordes, J. Weston, and N. Usunier, "Open question answering with weakly supervised embedding models," in *Proc. Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2014, pp. 165–180.

[21] M.-C. Yang, N. Duan, M. Zhou, and H.-C. Rim, "Joint relational embeddings for knowledge-based question answering," in *Proc. Conf. Empirical Methods Natural Language Process.*, 2014, pp. 645–650.

[22] J. Bao, N. Duan, M. Zhou, and T. Zhao, "Knowledge-based question answering as machine translation," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, 2014, pp. 967–976.

[23] X. Yao and B. van Durme, "Information extraction over structured data: Question answering with freebase," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, 2014, pp. 956–966.

[24] W.-T. Yih, M.-W. Chang, X. He, and J. Gao, "Semantic parsing via staged query graph generation: Question answering with knowledge base," in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Language Process.*, 2015, pp. 1321–1331.

[25] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[26] Y. Bengio, R. Ducharme, P. Vincent, and C. Janvin, "A neural probabilistic language model," *J. Mach. Learn. Res.*, vol. 3, pp. 1137–1155, 2003.

[27] T. Mikolov, M. Karafiát, L. Burget, J. Cernockỳ, and S. Khudanpur, "Recurrent neural network based language model," in *Proc. 11th Annu. Conf. Int. Speech Commun. Assoc.*, 2010, pp. 1045–1048.

[28] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12, pp. 2493–2537, 2011.

[29] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph and text jointly embedding," in *Proc. Conf. Empirical Methods Natural Language Process.*, 2014, pp. 1591–1601.

[30] X. Chen, L. Xu, Z. Liu, M. Sun, and H. Luan, "Joint learning of character and word embeddings," in *Proc. 24th Int. Conf. Artif. Intell.*, 2015, pp. 1236–1242.

[31] N. Kalchbrenner and P. Blunsom, "Recurrent continuous translation models," in *Proc. Conf. Empirical Methods Natural Language Process.*, 2013, pp. 1700–1709.

[32] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Advances Neural Inf. Process. Syst.*, 2014, pp. 3104–3112.

[33] R. Socher, C. C. Lin, C. Manning, and A. Y. Ng, "Parsing natural scenes and natural language with recursive neural networks," in *Proc. 26th Int. Conf. Mach. Learn.*, 2011, pp. 129–136.

[34] R. Socher, et al., "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proc. Conf. Empirical Methods Natural Language Process.*, 2013, pp. 1631–1642.

[35] B. Hu, Z. Lu, H. Li, and Q. Chen, "Convolutional neural network architectures for matching natural language sentences," in *Proc. Advances Neural Inf. Process. Syst.*, 2014, pp. 2042–2050.

[36] J. Zhang, D. Zhang, and J. Hao, "Local translation prediction with global sentence representation," in *Proc. 24th Int. Conf. Artif. Intell.*, 2015, pp. 1398–1404.

[37] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proc. 31st Int. Conf. Mach. Learn.*, 2014, pp. 1188–1196.

[38] A. Bordes, J. Weston, R. Collobert, and Y. Bengio, "Learning structured embeddings of knowledge bases," in *Proc. 25th AAAI Conf. Artif. Intell.*, 2011, pp. 301–306.

[39] A. Bordes, X. Glorot, J. Weston, and Y. Bengio, "A semantic matching energy function for learning with multi-relational data," *Mach. Learn.*, vol. 94, no. 2, pp. 233–259, 2014.

[40] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph embedding by translating on hyperplanes," in *Proc. 28th AAAI Conf. Artif. Intell.*, 2014, pp. 1112–1119.

[41] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 2181–2187.

[42] L. Yankai, L. Zhiyuan, L. Huan-Bo, S. Maosong, R. Siwei, and L. Song, "Modeling relation paths for representation learning of knowledge bases," in *Proc. Conf. Empirical Methods Natural Language Process.*, 2015, pp. 705–714.

[43] K. Toutanova, D. Chen, P. Pantel, H. Poon, P. Choudhury, and M. Gamon, "Representing text for joint embedding of text and knowledge bases," in *Proc. Conf. Empirical Methods Natural Language Process.*, 2015, pp. 1499–1509.

[44] C. Xu, et al., "RC-NET: A general framework for incorporating knowledge into word representations," in *Proc. 23rd ACM Int. Conf. Conf. Inf. Knowl. Manage.*, 2014, pp. 1219–1228.

[45] R. Socher, D. Chen, C. D. Manning, and A. Ng, "Reasoning with neural tensor networks for knowledge base completion," in *Proc. Advances Neural Inf. Process. Syst.*, 2013, pp. 926–934.

[46] Z. Li, M. A. Sharaf, L. Sitbon, X. Du, and X. Zhou, "CoRE: A context-aware relation extraction method for relation completion," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 4, pp. 836–849, Apr. 2014.
[47] A. Fader, S. Soderland, and O. Etzioni, "Identifying relations for open information extraction," in *Proc. Conf. Empirical Methods Natural Language Process.*, 2011, pp. 1535–1545.
[48] W. A. Woods and R. Kaplan, "Lunar rocks in natural english: Explorations in natural language question answering," *Linguistic Structures Process.*, vol. 5, pp. 521–569, 1977.
[49] L. Nie, M. Wang, L. Zhang, and S. Yan, "Disease inference from health-related questions via sparse deep learning," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 8, pp. 2107–2119, Aug. 2015.
[50] L. Nie, Y. L. Zhao, X. Wang, J. Shen, and T. S. Chua, "Learning to recommend descriptive tags for questions in social forums," *ACM Trans. Inf. Syst.*, vol. 32, no. 1, pp. 479–496, 2014.
[51] J. Berant, A. Chou, R. Frostig, and P. Liang, "Semantic parsing on freebase from question-answer pairs," in *Proc. Conf. Empirical Methods Natural Language Process.*, 2013, pp. 1533–1544.
[52] A. Bordes, S. Chopra, and J. Weston, "Question answering with subgraph embeddings," in *Proc. Conf. Empirical Methods Natural Language Process.*, 2014, pp. 615–620.
[53] J. Berant and P. Liang, "Semantic parsing via paraphrasing," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, 2014, pp. 1415–1425.

Xiaochi Wei received the BE degree in computer science and technology from Liaoning University, Shenyang, in 2012. Currently, he is working toward the PhD degree in the School of Computer Science, Beijing Institute of Technology. His research interests include question answering, information retrieval, and natural language processing. He is a student member of the CCF and the ACM.

Heyan Huang received the BE degree in computer science from Wuhan University, in 1983, the ME degree in computer science and technology from the National University of Defense Technology, in 1986, and the PhD degree from the China Academy of Sciences, Institute of Computer Technology, in 1989. Now, she is a professor, doctoral tutor, and president in School of Computer, Beijing Institute of Technology, and the director of the Research Center of High Volume Language Information Processing and Cloud Computing. Her current research interest mainly focus on machine translation.

Liqiang Nie received the BE degree from Xi'an Jiaotong University, Xi'an, China, in 2009, and the PhD degree from the National University of Singapore, in 2013. He is currently a professor in the School of Computer Science, Shandong University. His research interests include information retrieval and healthcare analytics. Various parts of his work have been published in top forums including ACM SIGIR, ACM SIGMM, IJCAI, the *ACM Transactions on Information Systems*, the *ACM Transactions on Intelligent Systems and Technology*, and the *IEEE Transactions on Multimedia*. He has been served as reviewers for various journals and conferences.

Hanwang Zhang is a research fellow in the School of Computing, National University of Singapore (2014-present). His research interests include multimedia and computer vision-developing techniques for efficient search and recognition in visual contents. He received the Best Demo runner-up award in ACM MM 2012 and the Best Student Paper award in ACM MM 2013. He is the winner of Best PhD Thesis Award of School of Computing, NUS, 2014

Xian-Ling Mao received the PhD degree from Peking University, in 2013. Now, he is an assistant professor of computer science with the Beijing Institute of Technology. He works in the fields of machine learning and information retrieval. His current research interests include topic modeling, learning to hashing, and question answering.

Tat-Seng Chua received the PhD degree from the University of Leeds, United Kingdom. He is the KITHCT chair professor in the School of Computing, National University of Singapore. He was the acting and founding dean of the School from 1998-2000. His main research interests include multimedia information retrieval and social media analytics. In particular, his research focuses on the extraction, retrieval, and question-answering (QA) of text and rich media arising from the Web and multiple social networks. He is the co-director of NExT, a joint Center between NUS and Tsinghua University to develop technologies for live social media search. He is the 2015 winner of the prestigious ACM SIGMM award for Outstanding Technical Contributions to Multimedia Computing, Communications and Applications. He is the chair of the steering committee of the ACM International Conference on Multimedia Retrieval (ICMR) and Multimedia Modeling (MMM) conference series. He is also the general co-chair of the *ACM Multimedia* 2005, ACM CIVR (now ACM ICMR) 2005, ACM SIGIR 2008, and ACM Web Science 2015. He serves on the editorial boards of four international journals. He is the co-founder of two technology startup companies in Singapore.