Connecting Emerging Relationships from News via Tensor Factorization

Jingyuan Zhang[‡], Chun-Ta Lu[*], Bokai Cao[*], Yi Chang[†] and Philip S. Yu[*§]
[‡]*Big Data Lab (BDL), Baidu Research, Sunnyvale, CA, USA*
[*]*Department of Computer Science, University of Illinois at Chicago, IL, USA*
[†]*Huawei Research America, Santa Clara, CA, USA*
[§]*Institute for Data Science, Tsinghua University, Beijing, China*
{*zhangjingyuan03*}*@baidu.com*, {*clu29, caobokai, psyu*}*@uic.edu*, {*yichang*}*@huawei.com*

***Abstract*—Knowledge graphs (KGs) have been widely used to represent relationships among entities, while KGs cannot capture new relationships between entities emerging along time. Since news often provides the latest information regarding the new entities and relationships, there is an opportunity to connect emerging relationships from news timely. However, it is a challenging task due to the source heterogeneity of structured KGs and unstructured news texts. In order to address the issue, we propose a tensor-based framework to capture the complex interactions among multiple types of relations, entities and text descriptions. We further develop an efficient Text-Aware MUlti-RElational learning method (TAMURE) that can learn the embedding representations of entities and relation types from both KGs and news, by jointly factorizing the interaction parameters. Furthermore, the complexity of TAMURE is linear in the number of parameters, which makes it suitable to large-scale KGs and news texts. Extensive experiments via TensorFlow demonstrate the effectiveness of the proposed TAMURE model compared with nine state-of-the-art methods on real-world datasets.**

***Keywords*-Information Extraction; Emerging Relationships; Tensor Factorization; Embedding**

I. Introduction

Knowledge graphs (KGs), such as Freebase[1] and DBpedia[2], have been widely used to represent relationships between entities in the form of triplets (h, r, t). Here h and t are two entities (head and tail) and r is a relation. Each triplet is a relation instance[3]. Entities can be persons, organizations, locations, etc., and examples of relations can be person-affiliation and organization-location. The KGs with relations and entities are useful sources for many real-word applications in information extraction, natural language understanding and information retrieval. However, current KGs have limited coverage of real-world relationships [1], especially for new entities that arise with new relationships emerging over time [2, 3] .

Fortunately, with the latest information in news, there is an opportunity to connect emerging relationships from news timely. Consider Figure 1 as an example, where an emerging relationship appears when the publisher "Walt Disney Studios" produces a new movie "Zootopia". Although there is no information regarding Zootopia in KGs, many pieces of news are talking about this new movie, the publisher and the directors. Detecting such emerging relationships has many benefits in practice. For example, the current KGs can be expanded and updated with emerging relationships. In addition, emerging relationships can help news related tasks, such as news retrieval and ranking, event detection, etc..

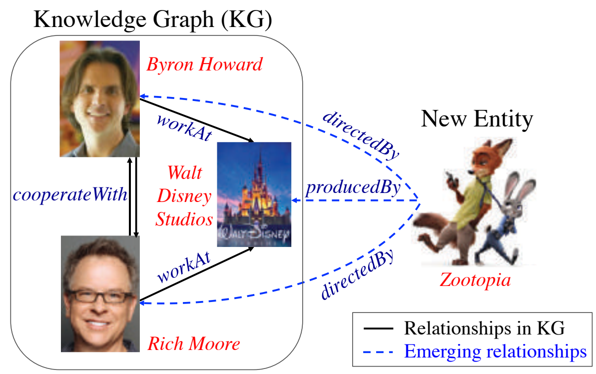


Figure 1. An example of emerging relationships.

However, learning emerging relationships from text descriptions is a challenging task due to the source heterogeneity of structured KGs and unstructured news texts. Although many research works try to mitigate the problem of knowledge sparsity in KGs [4–9], the main focus of these works is to utilize the structural information to fill in the missing relationships in KGs. For instance, Path Ranking Algorithm (PRA) [10, 11] completes KGs by performing random walk techniques; some other studies embed entities and relations into a low-dimensional space and infer missing relationships by translating relations from head entities to tail entities in KGs [9, 12, 13]. However, it is nontrivial to incorporate news texts into these structure-based methods to connect emerging relationships.

From the other aspect, there are several studies attempting to embed large-scale texts [14–17]. For example, LINE is proposed in [16] to embed texts into a low-dimensional space by constructing a homogeneous word co-occurrence network from texts. Later PTE is proposed in [17] to improve

[1]https://www.freebase.com/
[2]http://wiki.dbpedia.org/
[3]For simplicity, a relation means a relation type in KGs, and a relationship means a triple instance (h, r, t).

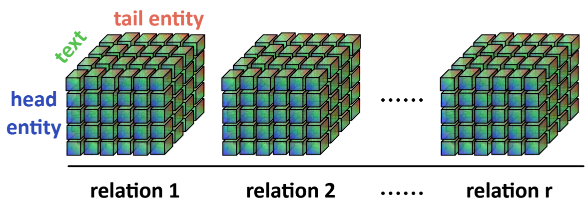


Figure 2. Tensors built for relations, the associated head entities, tail entities and news texts.

the LINE method by building a heterogeneous text network. These methods focus on embedding every single word in texts but ignoring the semantic relations among words. Therefore, they would fail to capture emerging relationships from news texts. Recently, some work tries to embed the KG and the texts jointly [3, 18]. However, the method in [18] embeds the KG and texts separately. So their indirect inference according to texts cannot help connect emerging relationships effectively. Though the method in [3] aims to detect emerging relationships, it fails to distinguish among different relation types.

To address the issue, in this paper, we propose a tensor-based framework to combine KGs and news texts effectively for detecting not only emerging relationship connection, but also the relation type. A fourth-order tensor structure is used to capture the hidden connection between multiple relations in KGs and multiple text descriptions of relations in news. Specifically, we model the multimodal interactions among head entities, relations, tail entities, and text descriptions as a tensor structure, by taking the tensor product of feature spaces of entities, relations and texts. Figure 2 shows the tensor structure constructed from KGs and news. As the interactions (i.e., tensor product) of entities, relations and text descriptions can reflect the connection between KGs and news, we use them to learn the embedding representations of entities and relations. In this manner, it can deal with multiple relation types without difficulty.

Noteworthily, directly learning from the tensor structure would be problematic. First, the space complexity of building the fourth-order tensor is polynomial to the numbers of entities, relations and text descriptions, making it challenging to fit the tensor into memory. Second, due to the large number of interaction parameters, it is time-consuming to decompose the built tensor directly and it is prone to overfitting. Last but not least, how to learn a meaningful representation of a given news sentence, which consists of the head and tail entities with the other text descriptions, and connect the representation to the relation types in KGs is challenging.

In order to solve the above challenges, we further develop a **T**ext-**A**ware **MU**lti-**RE**lational learning method (TAMURE) to learn the embeddings of entities and relation types from both news and KGs. By factorizing the interaction parameters, the proposed TAMURE method can efficiently learn the latent representation of entities and text descriptions, without physically building the tensor. Since the parameters are learned jointly through the factorization, it also makes the parameter estimation more accurate under sparsity and renders the model with the capacity to avoid overfitting. In summary, our contributions are as follows:

- We formulate a new task, connecting emerging relationships, which is to discover relation types with new entities by fusing information from heterogeneous sources, i.e., the structured KGs and the unstructured news texts.
- We introduce a novel tensor-based framework to connect emerging relationships from news. The news texts and KGs are incorporated into an elegant fourth-order tensor formulation, where the complex multiple interactions among relation types, entities and text descriptions are embedded within the tensor structure.
- The proposed TAMURE method can effectively recognize emerging relationships from news, by capturing not only the fourth- but also lower-order interactions in the built tensor. The lower-order interactions can explore hidden compatibility among entities, relations and text descriptions (see section III-B2 for details). Furthermore, the complexity of TAMURE is linear in the number of parameters, which makes it suitable to large-scale applications.
- We demonstrate the effectiveness of TAMURE by comparing it with nine state-of-the-art methods via TensorFlow on real-world KG and news data.

The rest of the paper is organized as follows. Section II formulates the problem; Section III introduces the details of the proposed TAMURE method; Section IV presents the experimental setup and the results; Section V briefly reviews related work; and Section VI concludes this study.

II. Preliminary

In this paper, we study the problem of connecting emerging relationships from news. Before proceeding, we first introduce the related concepts, and then state the problem of emerging relationship detection from news. Table I lists basic symbols that will be used throughout the paper.

A. Basic Concepts

Definition 1: **Entity, Relation and Relationship:** An entity e can represent a person, an organization, or a location, etc.. A relation can be a person-affiliation type or an organization-location type. A relationship is defined in the form of triplets (h, r, t), where h is a head entity, t is a tail entity and r is a relation. For each possible triple (h, r, t), we use $y \in \{0, 1\}$ to indicate whether the triple exists.

Definition 2: **Knowledge Graph (KG):** A *knowledge graph* is denoted as a directed graph $\mathcal{G}_{kg} = (E_{kg}, \mathcal{E}_{kg})$,

Table I
LIST OF BASIC SYMBOLS.

Symbol	Definition and description
(h, r, t)	a relationship
(h, r, t, d)	a relationship with a text description d
x	each lowercase letter represents a scale
$\mathbf{x}$	each boldface lowercase letter represents a vector
$\mathbf{X}$	each boldface uppercase letter represents a matrix
$\mathcal{X}$	each calligraphic letter represents a tensor, set or space
$[1 : M]$	a set of integers in the range of 1 to M inclusively
$\langle \cdot, \cdot \rangle$	denotes inner product
$\circ$	denotes tensor product (outer product)
$*$	denotes Hadamard (element-wise) product

where E_{kg} is the set of entities and $\mathcal{E}_{kg}$ is the set of known relationships. Each directed edge in $\mathcal{E}_{kg}$ can be represented by a triplet (h, r, t), where the entities h and $t \in E_{kg}$, and the relation r is one type of the relations in the relation set R.

Nowadays, due to the rapid growth of real-world knowledge, large volumes of *emerging relationships* are arising with time. An emerging relationship is defined as follows:

Definition 3: **Emerging Relationship:** An emerging relationship (h, r, t) exists, if its label $y = 1$ in the real world and at least one entity is not included in the given KG (i.e., $h \notin E_{kg}$ or $t \notin E_{kg}$).

For example, in Figure 1, (*Zootopia, producedBy, Walt Disney Studios*) is an emerging relationship with $y = 1$ since *Zootopia* is a new movie entity and it is produced by the publisher entity *Walt Disney Studios*. Similarly, (*Zootopia, directedBy, Byron Howard*) and (*Zootopia, directedBy, Rich Moore*) are also examples of emerging relationships.

The key of this work is to apply the tensor structure to fuse the KG and the news for connecting emerging relationships. In the following, we introduce some related concepts and notations about the tensor.

Definition 4: **Tensor:** Tensors are higher order arrays that generalize the notion of vectors (first order) and matrices (second order). Following [19], an M-th order tensor is denoted by $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_M}$ and its elements by $x_{i_1, \cdots, i_M}$. An index is denoted by a lowercase letter, spanning the range from 1 to the uppercase letter of the index, *e.g.*, $i = 1, 2, \cdots, I$. All vectors are column vectors unless otherwise specified.

For an arbitrary matrix $\mathbf{X} \in \mathbb{R}^{I \times J}$, its i-th row and j-th column vector are denoted by $\mathbf{x}^i$ and $\mathbf{x}_j$, respectively. The inner product of two same-sized tensors $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{I_1 \times \cdots \times I_M}$ is defined by $\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i_1=1}^{I_1} \cdots \sum_{i_1=1}^{I_M} x_{i_1, \cdots, i_M} y_{i_1, \cdots, i_M}$. The outer product of M vectors $\mathbf{x}^{(m)} \in \mathbb{R}^{I_m}$ for $m \in [1 : M]$ is an M-th order tensor and defined elementwise by $\left(\mathbf{x}^{(1)} \circ \cdots \circ \mathbf{x}^{(M)}\right)_{i_1, \cdots, i_M} = x_{i_1}^{(1)} \cdots x_{i_M}^{(M)}$ for all values of the indices. In particular, for $\mathcal{X} = \mathbf{x}^{(1)} \circ \cdots \circ \mathbf{x}^{(M)}$ and $\mathcal{Y} = \mathbf{y}^{(1)} \circ \cdots \circ \mathbf{y}^{(M)}$, it holds that

$$\left\langle \mathcal{X}, \mathcal{Y} \right\rangle = \prod_{m=1}^{M} \left\langle \mathbf{x}^{(m)}, \mathbf{y}^{(m)} \right\rangle = \prod_{m=1}^{M} \mathbf{x}^{(m)\mathrm{T}} \mathbf{y}^{(m)}. \quad (1)$$

For a general tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_M}$, its CANDECOM / PARAFAC (CP) factorization [19–21] is

$$\mathcal{X} = \sum_{k=1}^{K} \mathbf{x}_k^{(1)} \circ \cdots \circ \mathbf{x}_k^{(M)} = [\![\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(M)}]\!], \quad (2)$$

where for $m \in [1 : M]$, $\mathbf{X}^{(m)} = [\mathbf{x}_1^{(m)}, \cdots, \mathbf{x}_K^{(m)}]$ are factor matrices of size $I_m \times K$, K is the number of factors, and $[\![\cdot]\!]$ is used for shorthand.

B. Problem Statement

Given a large collection of news and the existing KG $\mathcal{G}_{kg}$, the task of connecting emerging relationships aims to determine the existence of multiple types of relations among entities. Assume we can extract a set of relationship candidates, each of which is represented by a triplet (h, r, t) and is associated with a vector of text descriptions $\mathbf{d}$ from news. Since the entities in the candidates may not exist in the entity set E_{kg} of the KG, we denote E_{news} as the set of entities that appear in the news, and denote $E = E_{kg} \cup E_{news}$ as the set of all the entities. The task of connecting emerging relationships is to learn a score function $f : (h, r, t, \mathbf{d}) \rightarrow \{0, 1\}$ that correctly predicts the label of the test instance, where the entities h and $t \in E$, $r \in R$ and $(h, r, t) \notin \mathcal{E}_{kg}$.

III. PROPOSED METHOD

In this section, we first introduce how to design the tensor-based score function for fusing the KG and the news texts. Then we derive an efficient **T**ext-**A**ware **MU**lti-**RE**lational learning method (TAMURE) that learns the embeddings of entities and relation types in linear complexity.

A. Tensor-based Score Function

We begin by introducing how to extract entities and relations from news. We then describe how to design the score function if only one of the sources (the KG or the news texts) is available. After that, we show that both sources can be integrated into an elegant tensor-based model.

Given a large collection of news, entities can be extracted via popular Named Entity Recognition (NER) techniques in [22, 23]. The entities that cannot be exactly matched to the KG are new entities. For existing relationships in the KG, we can associate their relation types with text descriptions from news. Given emerging relationships, we aim to determine their relation types with the help of text descriptions from news.

Without considering the existence of the KG, the most common approach is using a linear score function for each relation type r based on the text description vector $\mathbf{d} \in \mathbb{R}^{I_D}$ extracted from news:

$$f_{\text{news}}(r, \mathbf{d}) = \sum_{i=1}^{I_D} u_{r,i} d_i = \mathbf{d}^{\mathrm{T}} \mathbf{u}_r, \tag{3}$$

where $\mathbf{u}_r \in \mathbb{R}^{I_D}$ is the weight vector for relation type r. For learning multiple relation types at the same time, we let $I_R = |R|$ denote the number of relation types and $\mathbf{U} \in \mathbb{R}^{I_D \times I_R}$ denote the weight matrix to be learned, whose columns are the vector $\mathbf{u}_r$. Let $\mathbf{e}_r \in \mathbb{R}^{I_R}$ denote the relation type indicator vector

$$\mathbf{e}_r = [\underbrace{0, \cdots, 0}_{\text{r-1}}, 1, 0, \cdots, 0]^{\mathrm{T}}.$$

We can observe that $\mathbf{U}$ is actually the weight matrix of a bilinear feature map for modeling the second-order interactions between the text description vector and the relation type:

$$f_{\text{news}}(r, \mathbf{d}) = \mathbf{d}^{\mathrm{T}} \mathbf{U} \mathbf{e}_r = \langle \mathbf{U}, \mathbf{d} \circ \mathbf{e}_r \rangle. \tag{4}$$

Given the KG only, we can use a similar score function for modeling the interactions between the head entity, tail entity and relation type. Let $I_E = |E|$ denote the number of entities, and let $\mathbf{e}_h \in \mathbb{R}^{I_E}$ and $\mathbf{e}_t \in \mathbb{R}^{I_E}$ denote the head and tail entity indicators, respectively. We can form the multilinear score function by

$$f_{\text{KG}}(h, r, t) = \mathbf{e}_h^{\mathrm{T}} \mathbf{V}_r \mathbf{e}_t = \langle \mathcal{V}, \mathbf{e}_r \circ \mathbf{e}_h \circ \mathbf{e}_t \rangle. \tag{5}$$

where $\mathbf{V}_r \in \mathbb{R}^{I_E \times I_E}$ is the weight matrix between head entities and tail entities for relation type r, and $\mathcal{V} \in \mathbb{R}^{I_R \times I_E \times I_E}$ is a stacked tensor with each slice $\mathcal{V}_{r,:,:} = \mathbf{V}_r$.

Obviously, we can fuse the news and the KG together by formulating the score function as follows:

$$f(h, r, t, \mathbf{d}) = \langle \mathcal{W}, \mathbf{e}_r \circ \mathbf{e}_h \circ \mathbf{e}_t \circ \mathbf{d} \rangle, \tag{6}$$

where $\mathcal{W} \in \mathbb{R}^{I_R \times I_E \times I_E \times I_D}$ is the weight tensor to be learned.

It is worthy to be noted that some existing relationships in the KG might not have the text description $\mathbf{d}$, such that the text description $\mathbf{d} = \mathbf{0}$. Besides, in order to include relationships that are extracted from the news but not available in the KG, we add a "co-occurrence" relation type to the existing relations during the learning process and do not test this relation type in the testing phase.

B. Text-aware Multi-relational Learning

Now the news texts and the KG have been incorporated into an elegant tensor formulation, such that the complex multiple interactions among relation types, entities and text descriptions are embedded within the tensor structure. However, the space complexity of building the fourth-order tensor is $O(I_R \times I_E \times I_E \times I_D)$. With large volumes of emerging relationships in news, it is impractical to physically build the tensor. Moreover, directly learning the weight tensor $\mathcal{W}$ would be problematic. First, due to the large number of parameters $I_R \times I_E \times I_E \times I_D$, the learning procedure is prone to overfitting and less effective coupled with its sparse counterpart $\mathbf{e}_r \circ \mathbf{e}_h \circ \mathbf{e}_t \circ \mathbf{d}$. Second, since the weight parameters are learned independently, it cannot model the interactions that never appear. To address such issues, we propose an efficient method based on tensor factorization.

1) Efficient Tensor Decomposition Framework: Assume that the effect of interactions has a low rank, the weight tensor $\mathcal{W}$ can be factorized as

$$\mathcal{W} = [\![\mathbf{M}^{(r)}, \mathbf{M}^{(h)}, \mathbf{M}^{(t)}, \mathbf{M}^{(d)}]\!],$$

where $\mathbf{M}^{(r)} \in \mathbb{R}^{I_R \times K}$ and $\mathbf{M}^{(h)}$, $\mathbf{M}^{(t)} \in \mathbb{R}^{I_E \times K}$ represent the embedding matrices for relation types, head entities and tail entities, respectively; $\mathbf{M}^{(d)} \in \mathbb{R}^{I_D \times K}$ represents the weight matrix for text descriptions. From Eq. 1 and Eq. 6, we can easily derive that

$$\begin{aligned} f(h, r, t, \mathbf{d}) &= \sum_{k=1}^{K} \left\langle \mathbf{M}_{:,k}^{(r)} \circ \mathbf{M}_{:,k}^{(h)} \circ \mathbf{M}_{:,k}^{(t)} \circ \mathbf{M}_{:,k}^{(d)} \,,\ \mathbf{e}_r \circ \mathbf{e}_h \circ \mathbf{e}_t \circ \mathbf{d} \right\rangle \\ &= \sum_{k=1}^{K} \left(\mathbf{e}_r^{\mathrm{T}} \mathbf{M}_{:,k}^{(r)}\right) \left(\mathbf{e}_h^{\mathrm{T}} \mathbf{M}_{:,k}^{(h)}\right) \left(\mathbf{e}_t^{\mathrm{T}} \mathbf{M}_{:,k}^{(t)}\right) \left(\mathbf{d}^{\mathrm{T}} \mathbf{M}_{:,k}^{(d)}\right) \\ &= \left(\mathbf{e}_r^{\mathrm{T}} \mathbf{M}^{(r)}\right)^{\mathrm{T}} \left(\left(\mathbf{e}_h^{\mathrm{T}} \mathbf{M}^{(h)}\right) * \left(\mathbf{e}_t^{\mathrm{T}} \mathbf{M}^{(t)}\right) * \left(\mathbf{d}^{\mathrm{T}} \mathbf{M}^{(d)}\right) \right) \\ &= \mathbf{r}^{\mathrm{T}} \left(\mathbf{h} * \mathbf{t} * (\mathbf{d}^{\mathrm{T}} \mathbf{M}^{(d)}) \right), \end{aligned} \tag{7}$$

where $*$ is the Hadamard (elementwise) product, $\mathbf{r} = \mathbf{e}_r^{\mathrm{T}} \mathbf{M}^{(r)}$, $\mathbf{h} = \mathbf{e}_h^{\mathrm{T}} \mathbf{M}^{(h)}$ and $\mathbf{t} = \mathbf{e}_t^{\mathrm{T}} \mathbf{M}^{(t)}$ are the embedding vectors learned for the relation type r, head entity h and tail entity t, respectively.

2) Lower-order Constraint on Relation Types: One can notice that Eq. 7 only models the fourth-order interactions between the specific relation type r, entities h and t, and the associated text description $\mathbf{d}$. However, the lower-order interactions can also be discriminative for determining the existence of the relationship. For example, if the head entity is a person, e.g., "Rich Moore", the relation type r is unlikely to be "producedBy" no matter which tail entity t is chosen in the sample instance $(h, r, t, \mathbf{d})$. In this case, even the pairwise interaction between h and r can be discriminative. Thus, we consider to incorporate the lower-order interactions in the predictive model. This can be done by adding bias vectors in Eq. 7 as follows:

$$f(h, r, t, \mathbf{d}) = \mathbf{r}^{\mathrm{T}} \left((\mathbf{h} + \mathbf{b}_h) * (\mathbf{t} + \mathbf{b}_t) * \left(\mathbf{d}^{\mathrm{T}} \mathbf{M}^{(d)} + \mathbf{b}_v \right) \right), \tag{8}$$

where $\mathbf{b}_v \in \mathbb{R}^{I_D}$, $\mathbf{b}_h$ and $\mathbf{b}_t \in \mathbb{R}^{I_E}$ are the bias vectors that are independent to the given instance. To illustrate why the bias vectors can help model the lower-order interactions, we can decompose Eq. 8 into two parts. The first part $\mathbf{r}^{\mathrm{T}} \left(\mathbf{h} * (\mathbf{t} + \mathbf{b}_t) * \left(\mathbf{d}^{\mathrm{T}} \mathbf{M}^{(d)} + \mathbf{b}_v \right) \right)$ models the fourth-order interactions between $(h, r, t, \mathbf{d})$, while the second part

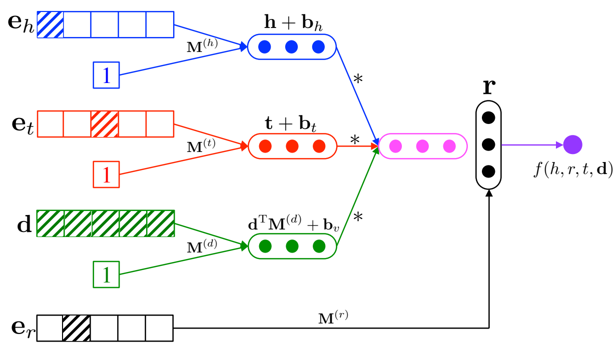


Figure 3. The work flow of TAMURE.

$\mathbf{r}^{\mathrm{T}}\left(\mathbf{b}_h * (\mathbf{t}+\mathbf{b}_t) * (\mathbf{d}^{\mathrm{T}}\mathbf{M}^{(d)}+\mathbf{b}_v)\right)$ with the bias vector $\mathbf{b}_h$ models the third-order as well as the lower-order interactions between $(r, t, \mathbf{d})$ without the head entity h involved. Other types of the third-order interactions and the lower-order interactions can be derived in a similar way.

We name the model in Eq. 8 as **T**ext-Aware **MU**lti-**RE**lational learning method (TAMURE). The work flow of TAMURE is illustrated in Figure 3. After mapping the head entity, tail entity and text description extracted from news into a common embedding space via $(\mathbf{h}+\mathbf{b}_t) * (\mathbf{t}+\mathbf{b}_t) * \left(\mathbf{d}^{\mathrm{T}}\mathbf{M}^{(d)}+\mathbf{b}_v\right)$, TAMURE learns a meaningful representation to connect relation types in KGs with news texts via $\mathbf{r}^{\mathrm{T}}\left((\mathbf{h}+\mathbf{b}_h) * (\mathbf{t}+\mathbf{b}_t) * \left(\mathbf{d}^{\mathrm{T}}\mathbf{M}^{(d)}+\mathbf{b}_v\right)\right)$.

Clearly, the parameters of the interactions among multiple relation types, entities, text descriptions are jointly factorized. The joint factorization benefits parameter estimation under sparsity, since dependencies exist when the interactions share the same entities or text descriptions. Therefore, the model parameters can be effectively learned without direct observations of such interactions especially in highly sparse data. Further, since the interactions are modeled with the bias vectors, this joint factorization model can easily deal with missing information and even incomplete information for relationships.

Another appealing property of TAMURE comes from the main characteristics of multilinear analysis. After factorizing the weight tensor $\mathcal{W}$, there is no need to construct the input tensor physically. Moreover, the model complexity is $O(K(I_R+I_E+I_D))$, which is linear in the number of parameters. This multilinear property can help save memory and also speed up the learning procedure.

C. Learning Procedure of TAMURE

Given the training set $\mathcal{D}=\{(h, r, t, \mathbf{d})\}$, TAMURE learns vector embeddings of the entities, relation types and text descriptions via the score function f. Following the same strategy as in [8, 12], we minimize a margin-based ranking criterion over the training set:

Algorithm 1 The TAMURE algorithm

Input: The training set $\mathcal{D}$, entities E and relation types R, margin γ, embedding dimension K.
Output: Embeddings of entities, relation types and texts.

```
Initialize embeddings h, r, t ← uniform(−1/√K, 1/√K) for each h ∈ E, t ∈ E and r ∈ R
Initialize weight parameters randomly
Normalize entity embeddings
while not convergence do
    //sample a mini-batch of size m
    D_batch ← sample(D, m)
    S_batch ← ∅
    for (h, r, t, d) ∈ D_batch do
        // Sample a corrupted instance
        (h', r, t', d') ← sample(D')
        S_batch ← ∪{((h, r, t, d), (h', r, t', d'))}
    end for
    Update embeddings and weight parameters w.r.t. the gradient of Eq. 9 on S_batch
    Constrain entity embeddings with the max-norm regularization
end while
```

$$\mathcal{L}=\sum_{(h,r,t,\mathbf{d})\in\mathcal{D}}\ \sum_{(h',r,t',\mathbf{d}')\in\mathcal{D}'}\max(0, f(h,r,t,\mathbf{d})+\gamma-f(h',r,t',\mathbf{d}')), \tag{9}$$

where $\gamma > 0$ is a margin hyper-parameter, and

$$\mathcal{D}'=\{(h', r, t, \mathbf{d}')|h' \in E\} \cup \{(h, r, t', \mathbf{d}')|t' \in E\}. \tag{10}$$

The set of corrupted relationships $\mathcal{D}'$ is composed of training data with either the head or tail replaced by a random entity (but not both at the same time). For a relationship from the KG, the corrupted data is checked to make sure that it is not in the KG. Notice that since the text description $\mathbf{d}$ is always associated with the entity pairs, when one of the entities is replaced, the text description vector will also be replaced accordingly. The loss function in Eq. 9 favors higher scores for training data than for corrupted data.

The learning process of TAMURE is carried out by Adam optimizer [24] in mini-batch mode, with the additional max-norm regularization constraint [12, 25], which constrains the L_2-norm of the embeddings of the entities to be no larger than 1. No regularization or norm constraints are given to the relation type and text description embeddings. The detailed optimization procedure is described in Algorithm 1. At each main iteration of the algorithm, a small set of data is sampled from the training set and servers as the training data of the mini-batch. For each existing relationship, a single corrupted relationship is sampled accordingly. The parameters are then updated by taking a gradient step with a constant learning rate. Before the next iteration, the embedding vectors of the entities are normalized via the max-norm regularization constraint.

IV. Experiments

In this section, we conduct extensive experiments to evaluate the proposed TAMURE method. After introducing the datasets and the experimental settings, we compare different baseline methods.

A. Data Processing

We connect emerging relationships from the ClueWeb12 corpus in the FB15k-237[4] dataset [26]. FB15k-237 contains 200 million sentences in the ClueWeb12 corpus coupled with Freebase mention annotations. There are around 3.9 million text descriptions corresponding to the relation types in Freebase. These texts are represented as full lexicalized dependency paths. For example, given the news text "Barack Obama is the 44th and current President of United States.", the dependency path for the relation between entity "Barack Obama" and "United States" is represented as "SUBJECT $\xleftarrow{\text{nsubj}}$ *president* $\xrightarrow{\text{prep}}$ *of* $\xrightarrow{\text{obj}}$ OBJECT". In the experiment, we extracted n-gram (n = 1, 2, 3) features from the lexicalized dependency paths and concatenated them together as the features of the text descriptions. In the process of n-gram feature extraction, we considered the lexical dependencies such as "nsubj" and "prep" as words. Those n-gram features with frequencies less than 10 are removed in the experiment. In FB15k-237, Freebase is used as the KG. A subset of entities and relations are extracted from Freebase. Almost all entities occur in the ClueWeb12 corpus.

In order to evaluate the effectiveness of the proposed TAMURE method, we randomly select half of the entities in FB15k-237 as new entities to avoid the high-cost of human labeling. Those relationships associated with new entities are emerging relationships. In the experiment, relationships with existing entities are considered as the training set from the KG. The emerging relationships are equally split into the validation set and the testing set. We generate the negative emerging relationships for the validation and testing sets by replacing each relation type with a random one. The validation set is used to find the threshold for each relation by thresholding the real-valued output scores of the proposed TAMURE model as in [8]. The statistics of the dataset are summarized in Table II.

B. Compared Methods

In order to show that the proposed TAMURE model can effectively connect emerging relationships, we compare the following nine methods:

- TransE: It is the classic KG embedding model by treating relation types as translations from head entities to tail entities [12]. Each entity is embedded into a low-dimensional vector space and each relation type is represented as a translation vector. The score function of TransE is $f(h, r, t) = ||\mathbf{h} + \mathbf{r} - \mathbf{t}||_{\ell_2}$.
- Skip-Gram: It is the state-of-the-art word embedding model [14]. It considers the KG and news texts together as a large corpus and learns embedding vectors for each word where each entity or relation type is regarded as a word.
- DeepWalk: It is an embedding model for homogeneous graphs with binary edges [15]. It learns embeddings of nodes by applying truncated random walks on the graph. By viewing entities, relation types and words as the same type of nodes, we can build a homogeneous graph from news texts and KGs. Then we apply the DeepWalk model on this graph.
- LINE: It is the **L**arge-scale **I**nformation **N**etwork **E**mbedding method (LINE) [16]. Similar to DeepWalk, LINE treats entities, relation types and words as one type of nodes but considers the weights of edges when learning the embeddings. The weights are the frequencies of two nodes co-occurring in news or KGs.
- PTE: It is the **P**redictive **T**ext **E**mbedding method (PTE) [17]. It learns embeddings of nodes from a heterogeneous graph built from news texts and KGs. The heterogeneous graph includes four types of graphs: a word-word co-occurrence graph, an entity-entity graph, a word-entity graph and an entity-relation graph.
- TransE+SG: It is based on the pTransE method for relationship inference from news texts and knowledge graphs [18]. It first applies TransE for entity embeddings from KGs and Skip-Gram for word embeddings from news texts. Then the two models are combined via aligning the embeddings into the same space. Since the dataset FB15k-237 has already annotated entities in news, there is no need to apply the alignment model.
- RESCAL: It is a relational learning approach based on tensor factorization [27]. It focuses on the KG and embeds relation types into a matrix space that operates as a bilinear operator on entity embeddings. RESCAL applies a more flexible tensor decomposition than CP, as the relation embedding matrix introduces interaction terms for entity embeddings.
- TAMURE-KG: It is the proposed tensor-based framework on the KG only. We first build a third-order tensor about head entities, relation types and tail entities from the KG. Then the proposed multi-relational factorization model is applied to learn embeddings. The score function of TAMURE-KG is $f(h, r, t) = \mathbf{r}^{\mathrm{T}}((\mathbf{h} + \mathbf{b}_h) * (\mathbf{t} + \mathbf{b}_t))$.
- TAMURE-HI: It is the proposed tensor-based framework with the high-order (fourth-order) interactions only. The lower-order constraint on relation types is not considered in the predictive model of TAMURE-HI. The score function is shown in Eq. 7.
- TAMURE: It is the **T**ext-**A**ware **MU**lti-**RE**lational learning model (TAMURE) proposed in this paper. We build a fourth-order tensor structure to combine the KG and news texts together for connecting emerging relationships.

We implement TAMURE via TensorFlow. For fair comparisons, we set the dimensionality of embeddings as 20 for

[4]https://www.microsoft.com/en-us/download/details.aspx?id=52312

Table II
STATISTICS OF THE DATASET.

News			Entities			Relationships			Train / Validation / Test		
# Texts	# Text Descriptions	# Entities	# Entities	# Existing Entities	# Emerging Entities	# Relationships	# Existing Relationships	# Emerging Relationships	# Training Relationships	# Validation Relationships	# Testing Relationships
200,000,000	3,978,014	13,937	14,541	7,270	7,271	310,116	82,062	228,054	82,062	114,027	114,027

all the above methods. For a given entity, its embedding vector is the same when the entity appears as the head or as the tail of a relation type. The Adam algorithm [24] within TensorFlow is applied as the optimizer for TransE, RESCAL, TAMURE-KG, TAMURE-HI and TAMURE. The learning rate for Adam optimizer is set as 0.01, the mini-batch size is set as 1000, the maximum number of epochs is set as 20, and the margin is set as 1 in the experiments. For the text embedding baselines, we follow the settings in [14–17].

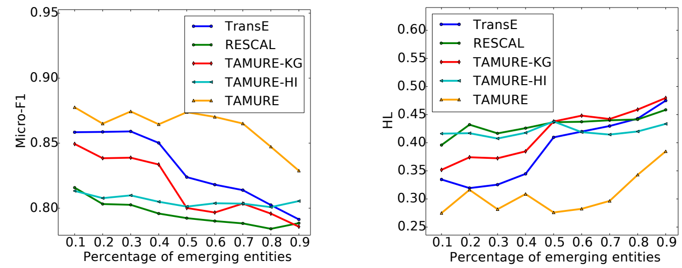
To evaluate the performance of the compared approaches, we turn the proposed TAMURE model into a binary classifier as in [8] by thresholding the real-valued output scores where the thresholds for each relation type are found on the validation set. For the text embedding baselines (i.e., Skip-Gram, DeepWalk, LINE and PTE), we follow [17] to apply the logistic regression model in the LibLinear package[5] after learning the embeddings. The embedding concatenation of head entities, relation types and tail entities is considered as the feature during the classification phase. Since there might be multiple relation types between two entities, we measure the performance by adopting five popular multi-label evaluation metrics in the literature: *Micro F1* and *Macro F1* that evaluate a classifier's label set prediction performance and consider the micro/macro-average of $precision$ and $recall$ on all binary labels with equal importance [28]; *Average Accuracy* ($AvgAcc$), *Average AUC* ($AvgAuc$) and *Hamming Loss* (HL) that evaluate the average accuracy, AUC and error rate over all the binary labels (relation types) [29].

C. Performance Evaluation

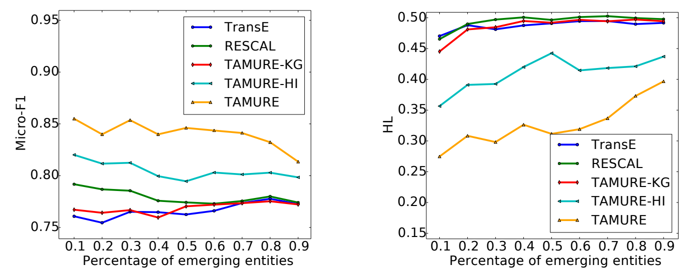
In this section, we report the performance of the compared methods. The testing emerging relationships can be categorized into two groups: one is about those with only one emerging entity and the other is about those relationships with both entities not in the KG. In order to show the effectiveness of the proposed TAMURE method, we not only show the performance over all emerging relationships, but also show the results for each group of emerging relationships. The performance with the rank of each method is reported in Table III. It can be observed that the proposed TAMURE method consistently outperforms all the other nine baselines regardless of the groups of emerging relationships. Specifically, compared with the second best baseline, TAMURE achieves an improvement of 47% on the HL metric and 24% on the $AvgAcc$ metric. Furthermore, in the case where both entities are emerging entities, TAMURE achieves a significant performance improvement (42% and 27% higher than the second best baseline on HL and $AvgAuc$, respectively), showing the capability of TAMURE capturing the interactions between the relation types in KGs and the text descriptions in news.

Compared with the KG embedding baselines, TAMURE performs the best since it incorporates the news texts and the KG into an elegant fourth-order tensor formulation to capture complex interactions among relation types, entities and text descriptions. For those emerging relationships with only one new entity, TransE achieves the second best results.

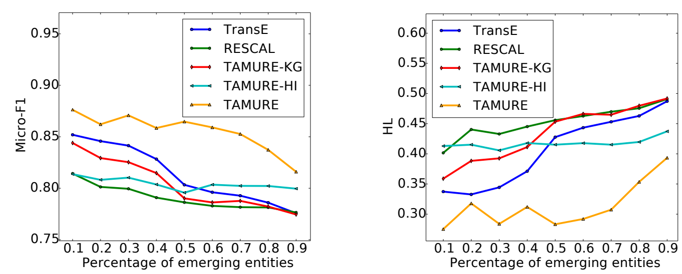
[5]http://www.csie.ntu.edu.tw/~cjlin/liblinear/



(a) Relationships with one emerging entity



(b) Relationships with two emerging entities



(c) All testing relationships

Figure 4. The performance with different percentages of emerging entities. For the $Micro\text{-}F1$ and $Macro\text{-}F1$ metrics, the larger the value, the better the performance. For the HL metric, the smaller the better.

Table III

THE CLASSIFICATION PERFORMANCE "VALUE (RANK)" ON CONNECTING EMERGING RELATIONSHIPS. "↑" INDICATES THE LARGER THE VALUE, THE BETTER THE PERFORMANCE. "↓" INDICATES THE SMALLER THE BETTER.

(a) Results on emerging relationships with only one entity not in the KG.

	Methods									
	KG Embedding			Text Embedding				KG+Text Embedding		
Criteria	TransE	RESCAL	TAMURE-KG	Skip-Gram	DeepWalk	LINE	PTE	TransE+SG	TAMURE-HI	TAMURE
Micro-F1↑	0.8239 (2)	0.7923 (5)	0.8001 (4)	0.0723 (10)	0.1019 (9)	0.2792 (8)	0.4263 (7)	0.7893 (6)	0.8012 (3)	**0.8738** (1)
Macro-F1 ↑	0.7022 (2)	0.6746 (5)	0.6810 (4)	0.1189 (10)	0.1603 (9)	0.4296 (7)	0.3215 (8)	0.6654 (6)	0.6881 (3)	**0.7827** (1)
AvgAcc ↑	0.5905 (2)	0.5635 (3)	0.5623 (4)	0.3777 (9)	0.3907 (8)	0.5184 (7)	0.3072 (10)	0.5188 (6)	0.5622 (5)	**0.7240** (1)
AvgAuc ↑	0.6425 (2)	0.4860 (6)	0.5759 (3)	0.1420 (10)	0.1860 (9)	0.4752 (7)	0.1947 (8)	0.5187 (5)	0.5634 (4)	**0.7370** (1)
HL ↓	0.4096 (2)	0.4365 (3)	0.4377 (4)	0.6223 (9)	0.6093 (8)	0.4816 (7)	0.6928 (10)	0.4812 (6)	0.4378 (5)	**0.2760** (1)

(b) Results on emerging relationships with both entities not in the KG.

	Methods									
	KG Embedding			Text Embedding				KG+Text Embedding		
Criteria	TransE	RESCAL	TAMURE-KG	Skip-Gram	DeepWalk	LINE	PTE	TransE+SG	TAMURE-HI	TAMURE
Micro-F1↑	0.7625 (6)	0.7742 (3)	0.7704 (4)	0.1703 (10)	0.2540 (9)	0.3751 (8)	0.6125 (7)	0.7656 (5)	0.7946 (2)	**0.8461** (1)
Macro-F1 ↑	0.6450 (6)	0.6564 (3)	0.6502 (5)	0.1936 (10)	0.2637 (9)	0.5103 (8)	0.5748 (7)	0.6513 (4)	0.6827 (2)	**0.7555** (1)
AvgAcc ↑	0.5090 (4)	0.5035 (6)	0.5080 (5)	0.3243 (10)	0.3438 (9)	0.5025 (7)	0.4266 (8)	0.5150 (3)	0.5575 (2)	**0.6885** (1)
AvgAuc ↑	0.5028 (6)	0.5060 (4)	0.5049 (5)	0.1577 (10)	0.2025 (9)	0.4758 (7)	0.2045 (8)	0.5199 (3)	0.5583 (2)	**0.7081** (1)
HL ↓	0.4910 (4)	0.4965 (6)	0.4920 (5)	0.6757 (10)	0.6562 (9)	0.4975 (7)	0.5734 (8)	0.4850 (3)	0.4425 (2)	**0.3115** (1)

(c) Results on all emerging relationships.

	Methods									
	KG Embedding			Text Embedding				KG+Text Embedding		
Criteria	TransE	RESCAL	TAMURE-KG	Skip-Gram	DeepWalk	LINE	PTE	TransE+SG	TAMURE-HI	TAMURE
Micro-F1↑	0.8033 (2)	0.7863 (5)	0.7902 (4)	0.1083 (10)	0.1599 (9)	0.3139 (8)	0.4948 (7)	0.7819 (6)	0.7957 (3)	**0.8647** (1)
Macro-F1 ↑	0.6877 (2)	0.6675 (5)	0.6700 (4)	0.1414 (10)	0.1918 (9)	0.4459 (7)	0.4129 (8)	0.6632 (6)	0.6830 (3)	**0.7754** (1)
AvgAcc ↑	0.5723 (2)	0.5442 (5)	0.5467 (4)	0.3658 (9)	0.3796 (8)	0.5138 (7)	0.3439 (10)	0.5210 (6)	0.5580 (3)	**0.7168** (1)
AvgAuc ↑	0.6050 (2)	0.4932 (6)	0.5517 (4)	0.1531 (10)	0.1966 (9)	0.4768 (7)	0.2514 (8)	0.5073 (5)	0.5606 (3)	**0.7326** (1)
HL ↓	0.4277 (2)	0.4558 (5)	0.4533 (4)	0.6342 (9)	0.6204 (8)	0.4862 (7)	0.6562 (10)	0.4791 (6)	0.4420 (3)	**0.2832** (1)

However, without the help of news texts, TransE cannot perform well if the relationship has both entities not in the KG. Since emerging relationships usually first appear in news, it is important and necessary to consider news during the detection process. With specific designs for the KG embedding, it is difficult to incorporate news texts into TransE and RESCAL. In contrast, the proposed tensor-based framework (TAMURE-KG) can be easily adapted to add news information (TAMURE), thereby connecting emerging relationships effectively.

Another observation is that text embedding baselines do not perform well for connecting emerging relationships. The reason lies in the two-step learning of emerging relationships for these models. First the embedding vectors are learned and then the emerging relationships are classified based on the embedding vectors. With two separate steps, it is difficult to connect emerging relationships with the learning of embeddings. Though PTE considers the heterogeneity among entities, relation types and words in news texts, it still performs worse than the KG embedding models.

Furthermore, compared with TransE+SG, TAMURE-HI and TAMURE still perform better. TransE+SG combines the KG and news texts together, thereby achieving reasonable performance on emerging relationships with both entities not in the KG. However, TransE+SG embeds the KG and the news separately. The interactions between the relation types in KGs and the text descriptions in news are not captured. Hence, it cannot help detect emerging relationships effectively. The proposed TAMURE-HI and TAMUE methods in this paper builds a fourth-order tensor to capture the hidden connections between the KG and the news texts, and thereby achieving better performance. Since TAMURE considers the lower-order interactions in the built tensor, it outperforms TAMURE-HI to a large extent.

In summary, with a tensor structure, the proposed TAMURE helps connect emerging relationships from news effectively and outperforms all the nine baseline methods.

D. *Effects of Emerging Entities*

As mentioned previously, we randomly select half of the entities in FB15k-237 as new entities to evaluate the effectiveness of TAMURE. In the following, we assess the performance of TAMURE with different percentages of new entities, in the range of 10% to 90%. We select the best five models according to Table III (TransE, RESCAL, TAMURE-KG, TAMURE-HI and TAMURE) and represent their performance in Figure 4. Due to space limit, we

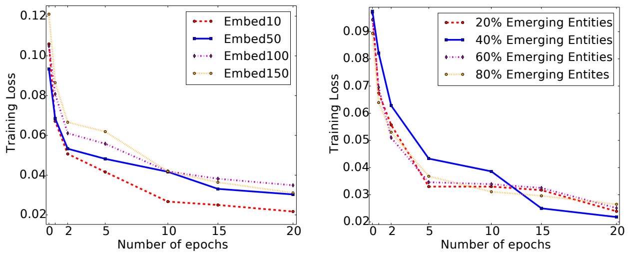


Figure 5. The performance with different numbers of epochs.

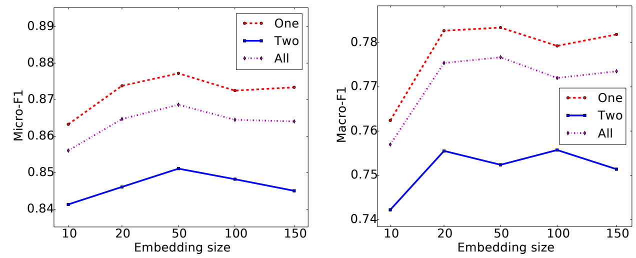


Figure 6. The performance with different embedding sizes.

only show the performance on $Micro$-$F1$ and HL. Similar performance is achieved on the other metrics of $Macro$-$F1$, $AvgAcc$ and $AvgAuc$.

From Figure 4, we can observe that TAMURE significantly outperforms other baselines regardless of how many percentages of emerging entities we select. In addition, TAMURE can achieve a stable performance when there are less than 70% of emerging entities. With less than 30% of existing entities in the KG, little information can be provided for connecting emerging relationships from news, thereby resulting in a drop of performance for TAMURE. However, TAMURE still achieves the best result compared with other baseline methods since it captures the hidden connections between relation types in the KG and text descriptions in the news.

E. Parameter Analysis

In the following, we analyze the performance of TAMURE with different embedding sizes and epochs.

1) Influence of Embedding Size: We demonstrate the performance of TAMURE with different embedding sizes by fixing the other parameters. In Figure 6, we show the results on $Micro$-$F1$ and $Macro$-$F1$ metrics, where "One" means the results for emerging relationships with only one entity not in the KG, "Two" is about the relationships with neither entities in the KG, and "All" is about all the emerging relationships. It can be observed that, with a larger embedding size, TAMURE achieves better performance on both $Micro$-$F1$ and $Macro$-$F1$. However, when the embedding size is larger than 20, the performance of TAMURE becomes more stable with small improvement. Therefore, in our experiment, we set the embedding size as 20. Due to space limit, we do not show the results on $AvgAcc$, $AvgAuc$ and HL metrics, on which similar patterns can be observed.

2) Influence of Epochs: We now investigate the performance of TAMURE with different numbers of epochs. Figure 5 shows the convergence process of the training loss of TAMURE, with different embedding sizes and different percentages of emerging entities, respectively. "Embed$k$" on the left of Figure 5 indicates the results with embedding size k. We can observe that the training loss drops quickly at first few epochs and becomes stable after 10 epochs, regardless of the embedding size or the percentage of emerging entities. It demonstrates a fast convergence of the proposed method TAMURE. In our experiment, we set the number of epochs as 20 to achieve a stable performance.

V. Related Work

Knowledge graphs (KGs) of real-world relations about entities are useful sources for a lot of important applications in information extraction, natural language understanding and information retrieval [30, 31]. However, knowledge graph (KG) is incomplete with large amounts of missing relation types [1].

Due to the limited coverage of KGs, the task of KG completion has received a lot of attention [6, 7, 9, 13, 32]. Some work learns embedding representations of entities and relation types in the KG and use these embeddings to infer missing relationships [4, 5, 8, 12, 33]. In addition, some studies predict missing relationships from a graph view [10, 11, 34, 35]. For instance, the Path Ranking Algorithm (PRA) [10, 11] performs link prediction in the KG via a random walk inference technique. Recently, the research work [35] uses a recursive neural network to create embedded representations of paths learned from [10, 11]. In our work, the emerging relationships have new entities that are not included in the KG. Hence, it is impossible to apply these techniques directly.

Furthermore, some work tries to embed the KG and the texts jointly [3, 18]. However, the method in [18] embeds the KG and texts separately. So their indirect inference according to texts cannot help detect emerging relationships effectively. Though the method in [3] aims to connect emerging relationships, it fails to distinguish among different relation types. In contrast, our proposed TAMURE model handles multi-label relation types by building a fourth-order tensor structure.

Our work is also related to the problem of information network modeling and mining [36–39]. Recently, there are some work about embedding large-scale texts [14–17]. For example, DeepWalk and LINE are proposed in [15] and [16] to embed texts into a low-dimensional space by constructing a homogeneous word co-occurrence network from texts. Later PTE is proposed in [17] to improve the LINE method by building a heterogeneous text network. These methods focus on embedding every single word in texts but ignoring the semantic relation types among words. Therefore, they cannot help connect emerging relationships from news texts effectively.

VI. Conclusion

In this paper, we formulate a new task of connecting emerging relationships from news and propose a novel tensor-based framework to combine KGs and news texts effectively for connecting emerging relationships. With an efficient **T**ext-**A**ware **MU**lti-**RE**lational learning method (TAMURE), the complex interactions among relation types, entities and text descriptions are jointly factorized without physically building the tensor. Extensive experiments via TensorFlow demonstrate the effectiveness of the proposed TAMURE model compared with nine state-of-the-art methods on real-world datasets.

Acknowledgement

This work is supported in part by NSF through grants IIS-1526499, CNS-1626432, and NSFC 61672313. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X GPU used for this research.

References

[1] X. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmann, S. Sun, and W. Zhang, "Knowledge vault: A web-scale approach to probabilistic knowledge fusion," in *KDD*, 2014.

[2] J. Hoffart, Y. Altun, and G. Weikum, "Discovering emerging entities with ambiguous names," in *WWW*, 2014.

[3] J. Zhang, C. Lu, M. Zhou, S. Xie, Y. Chang, and P. Yu, "Heer: Heterogeneous graph embedding for emerging relation detection from news," in *IEEE BigData*, 2016.

[4] K. Chang, W. Yih, B. Yang, and C. Meek, "Typed tensor decomposition of knowledge bases for relation extraction," in *EMNLP*, 2014.

[5] A. Bordes, J. Weston, R. Collobert, and Y. Bengio, "Learning structured embeddings of knowledge bases," in *AAAI*, 2011.

[6] A. García-Durán, A. Bordes, and N. Usunier, "Effective blending of two and three-way interactions for modeling multi-relational data," in *ECML/PKDD*, 2014.

[7] M. Nickel, X. Jiang, and V. Tresp, "Reducing the rank in relational factorization models by including observable patterns," in *NIPS*, 2014.

[8] R. Socher, D. Chen, C. Manning, and A. Ng, "Reasoning with neural tensor networks for knowledge base completion," in *NIPS*, 2013.

[9] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph embedding by translating on hyperplanes," in *AAAI*, 2014.

[10] M. Gardner and T. Mitchell, "Efficient and expressive knowledge base completion using subgraph feature extraction," in *EMNLP*, 2015.

[11] N. Lao, T. Mitchell, and W. Cohen, "Random walk inference and learning in a large scale knowledge base," in *EMNLP*, 2011.

[12] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *NIPS*, 2013.

[13] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion." in *AAAI*, 2015.

[14] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *NIPS*, 2013.

[15] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *KDD*, 2014.

[16] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "Line: Large-scale information network embedding," in *WWW*, 2015.

[17] J. Tang, M. Qu, and Q. Mei, "Pte: Predictive text embedding through large-scale heterogeneous text networks," in *KDD*, 2015.

[18] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph and text jointly embedding," in *EMNLP*, 2014.

[19] T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM review*, 2009.

[20] B. Cao, H. Zhou, G. Li, and P. S. Yu, "Multi-view machines," in *WSDM*, 2016.

[21] C.-T. Lu, L. He, W. Shao, B. Cao, and P. S. Yu, "Multilinear factorization machines for multi-task multi-view learning," in *WSDM*, 2017.

[22] D. Nadeau and S. Sekine, "A survey of named entity recognition and classification," *Lingvisticae Investigationes*, vol. 30, no. 1, 2007.

[23] *Stanford Named Entity Recognizer*, http://nlp.stanford.edu/software/CRF-NER.shtml.

[24] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *ICLR*, 2015.

[25] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting." *JMLR*, 2014.

[26] K. Toutanova, D. Chen, P. Pantel, P. Choudhury, and M. Gamon, "Representing text for joint embedding of text and knowledge bases," in *ACL*, 2015.

[27] M. Nickel, V. Tresp, and H. Kriegel, "A three-way model for collective learning on multi-relational data," in *ICML*, 2011.

[28] N. Ghamrawi and A. McCallum, "Collective multi-label classification," in *CIKM*, 2005.

[29] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining multi-label data," in *Data Mining and Knowledge Discovery Handbook*, 2009.

[30] G. Angeli, J. Tibshirani, J. Wu, and C. Manning, "Combining distant and partial supervision for relation extraction," in *EMNLP*, 2014.

[31] D. Zeng, K. Liu, Y. Chen, and J. Zhao, "Distant supervision for relation extraction via piecewise convolutional neural networks," in *EMNLP*, 2015.

[32] M. Gardner, P. Talukdar, J. Krishnamurthy, and T. Mitchell, "Incorporating vector space similarity in random walk inference over knowledge bases," in *EMNLP*, 2014.

[33] G. Ji, S. He, L. Xu, K. Liu, and J. Zhao, "Knowledge graph embedding via dynamic mapping matrix," in *ACL*, 2015.

[34] M. Gardner, P. Talukdar, B. Kisiel, and T. Mitchell, "Improving learning and inference in a large knowledge-base using latent syntactic cues," in *EMNLP*, 2013.

[35] A. Neelakantan, B. Roth, and A. McCallum, "Compositional vector space models for knowledge base completion," in *ACL*, 2015.

[36] B. Cao, X. Kong, J. Zhang, P. S. Yu, and A. B. Ragin, "Mining brain networks using multiple side views for neurological disorder identification," in *ICDM*, 2015.

[37] J. Zhang, X. Kong, L. Jie, Y. Chang, and P. Yu, "Ncr: A scalable network-based approach to co-ranking in question-and-answer sites," in *CIKM*, 2014.

[38] J. Zhang, L. Jie, A. Rahman, S. Xie, Y. Chang, and P. Yu, "Learning entity types from query logs via graph-based modeling," in *CIKM*, 2015.

[39] J. Zhang, B. Cao, S. Xie, C.-T. Lu, P. S. Yu, and A. B. Ragin, "Identifying connectivity patterns for brain diseases via multi-side-view guided deep architectures," in *SDM*, 2016.