Entity Summarization via Exploiting Description Complementarity and Salience

Liyi Chen, Zhi Li, Weidong He, Gong Cheng, *Member, IEEE*, Tong Xu, *Member, IEEE*, Nicholas Jing Yuan, *Senior Member, IEEE*, and Enhong Chen, *Senior Member, IEEE*

***Abstract*— Entity summarization is a novel and efficient way to understand real-world facts and solve the increasing information overload problem in large-scale knowledge graphs (KG). Existing studies mainly rely on ranking independent entity descriptions as a list under a certain scoring standard such as importance. However, they often ignore the relatedness and even semantic overlap between individual descriptions. This may seriously interfere with the contribution judgment of descriptions for entity summarization. Actually, the entity summary is a whole to comprehensively integrate the main aspects of entity descriptions, which could be naturally treated as a set. Unfortunately, the exploration of these set characteristics for entity summarization is still an open issue with great challenges. To that end, we draw inspiration from a *set completion* perspective and propose an entity summarization method with complementarity and salience (ESCS) to deeply exploit description complementarity and salience in order to form a summary set for the target entity. Specifically, we first generate entity description representations with textual features in the description embedding module. For the purpose of learning complementary relationships within the entire summary set, we devise a bi-directional long short-term memory structure to capture global complementarity for each summary in the summary complementarity learning module. Meanwhile, in order to estimate the salience of individual descriptions, we calculate similarities between semantic embeddings of the target entity and its property-value pairs in the description salience learning module. Next, with a joint learning stage, we can optimize ESCS from a set completion perspective. Finally, a summary generation strategy is designed to infer the entire summary set step-by-step for the target entity. Extensive experiments on a public benchmark have clearly demonstrated the effectiveness of ESCS and revealed the potential of set completion in entity summarization task.**

***Index Terms*— Entity description, entity summarization, knowledge graph (KG), neural network.**

Manuscript received 5 July 2021; revised 21 November 2021 and 18 January 2022; accepted 30 January 2022. Date of publication 23 February 2022; date of current version 30 October 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1402600, in part by the National Natural Science Foundation of China under Grant 62072423, and in part by the Huawei–USTC Joint Innovation Program. *(Corresponding author: Tong Xu.)*

Liyi Chen, Zhi Li, Weidong He, Tong Xu, and Enhong Chen are with the School of Data Science, University of Science and Technology of China, Hefei 230027, China (e-mail: liyichencly@gmail.com; zhili03@mail.ustc.edu.cn; hwd@mail.ustc.edu.cn; tongxu@ustc.edu.cn; cheneh@ustc.edu.cn).

Gong Cheng is with the State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China (e-mail: gcheng@nju.edu.cn).

Nicholas Jing Yuan is with Huawei Cloud & AI, Hangzhou 310007, China (e-mail: nicholas.jing.yuan@gmail.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2022.3149047.

Digital Object Identifier 10.1109/TNNLS.2022.3149047

I. Introduction

RECENT years have witnessed the enormous development of knowledge graphs (KGs) which benefit a lot of downstream applications, such as recommender systems [1]–[3], dialog systems [4], [5], and question answering [6], [7]. Meanwhile, real industry demands and massive knowledge facts have motivated significant growth of the KG scale. Along this line, the increasing information overload problem in large-scale KGs becomes a critical topic in both the industry and academia. With this in mind, many researchers present entity summarization task to depict a target entity with a summary of numerous descriptions [8]. Entity summarization assists people in understanding real-world facts and prevents them from being overwhelmed by the glut of information.

In the literature, much of the early work is devoted to addressing entity summarization problem by unsupervised methods [8]–[10]. These studies mainly aim at mining linguistics [8] or feature relations [11] for entity properties and generating a summary with clustering techniques [10] or probabilistic topic models [12]. With the rapid development of deep learning methods in many task, supervised techniques have manifested promising improvements in generating description summary of entity [13]–[15]. However, they principally consider entity summarization as a ranking problem to form a description list under a certain scoring standard, such as importance [16]. And then top-k description triples are selected as the entity summary.

The greatest limitation of existing work is that descriptions are treated as independent individuals to separately measure their semantic contributions to the entity. However, there may be relatedness and even semantic overlap between individual descriptions. In this case, separately measuring the semantic contribution of each description would leave out some descriptions that are more representative and provide richer information. For instance, Fig. 1(a) shows an entity "Andrew Kippis" comprising many descriptions. And the list ranking result of entity summarization for "Andrew Kippis" is shown in Fig. 1(b). It is noticed that treating each description as an independent individual to rank a list cannot guarantee the comprehensiveness of the summary. Even though "⟨Name⟩ Andrew Kippis" and "⟨Given Name⟩ Andrew" are useful descriptions, they are not expected to be summarized together. This requires descriptions combined with each other as a whole instead of independent individuals.

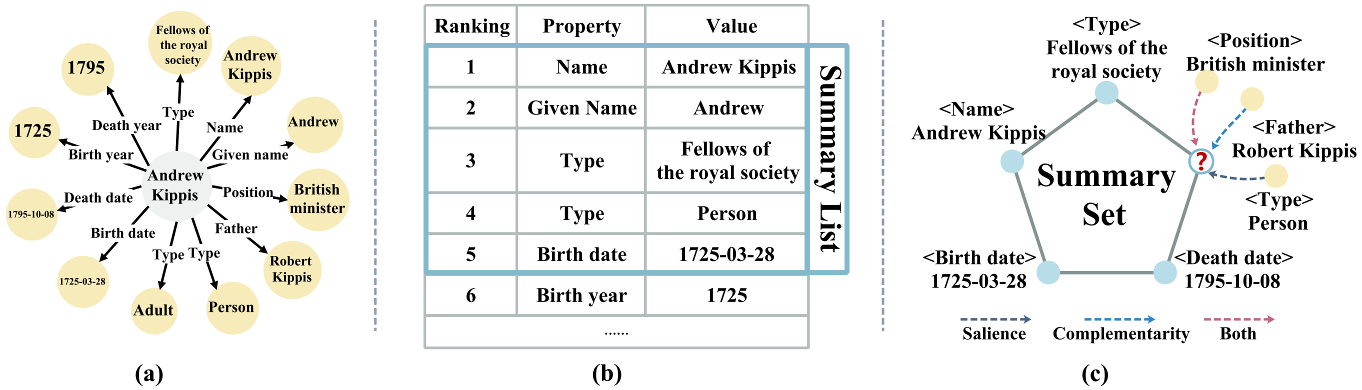


Fig. 1. A toy example of entity summarization. (a) All descriptions of the entity "Andrew Kippis" in a KG. (b) Summarized result for "Andrew Kippis" from the list ranking perspective. (c) Summarized result for "Andrew Kippis" in terms of three different choices from the set completion perspective.

In fact, the entity summary is a whole to comprehensively integrate the main aspects of entity descriptions, which could be naturally treated as a set. Each summarized description from an ideal summary set is complementary to others and salient to the target entity simultaneously. On the one hand, complementarity means that descriptions of target entity provided in the entire summary set are diverse and related. On the other hand, salience means that descriptions in the entire summary set are all important and represent the core features of the target entity. When we deeply analyze the complementarity and salience of summarized descriptions in the entire summary set, we can tackle this task from a set completion perspective and complete the summary set by inferring summarized descriptions step-by-step. As shown in Fig. 1(c), given several selected summarized descriptions of "Andrew Kippis" in the summary set, we need to continue selecting another description to fill the summary set until it is completed, and three different choices are considered in the following.

1) When only considering description **salience**, we may choose "⟨Type⟩ Person" since "⟨Type⟩" is always a critical property to depict an entity. However, "⟨Name⟩ Andrew Kippis" and "⟨Type⟩ Fellows of the royal society" in the existing summary set have already indicated that "Andrew Kippis" is a person. Therefore, "⟨Type⟩ Person" is not semantically complementary to the existing summary set.
2) When only conditioned on description **complementarity**, we may select "⟨Father⟩ Robert Kippis" because it can supply a lack of information in the existing summary set. Nevertheless, "⟨Father⟩" is not salient to the target entity, we prefer this person's own information.
3) When incorporating description **salience** with **complementarity**, "⟨Position⟩ British minister" is more complementary to the existing summary set and also has much more salient information about this person than "⟨Type⟩ Person" and "⟨Father⟩ Robert Kippis," i.e., not only position but also nationality.

By comparison, "⟨Position⟩ British minister" is a superior choice for "Andrew Kippis."

Based on these intuitions, in this article, we present a focused study on exploiting description complementarity and salience for entity summarization. To begin with, we address entity summarization from a novel set completion perspective. Set completion for entity summarization is to select descriptions to fill the summary set step-by-step until the size of the summary set equals a constrained value according to certain criteria. Afterward, we propose an entity summarization method with complementarity and salience (ESCS), to adaptively complete entity summary set in line with the criteria of complementarity and salience. Specifically, we first generate entity description representations with textual features by multilayer perceptrons (MLP) in the description embedding (DE) module. Then, we devise a summary complementarity learning (SCL) module to learn the complementary relationship within the entire summary set via a bi-directional long short-term memory (Bi-LSTM) structure. Meanwhile, we invent a description salience learning (DSL) module to measure the salience of individual descriptions within the entire summary set by calculating the similarity between semantic embeddings of entity and property. Next, we jointly learn description complementarity and salience in ESCS to model the process of summary set completion. Finally, a summary set generation approach is designed in the inference stage to automatically complete the entire entity summary set step-by-step, which solves the problem of no initialized set available and makes full use of global complementarity and salience within the entire summary set. Extensive experiments are conducted on a public benchmark. The results not only confirm the effectiveness of ESCS compared with state-of-the-art methods but also reveal the potential of set completion in entity summarization task.

The major contributions of this article are listed as follows.

1) *Problem View:* We tackle entity summarization task from a novel set completion perspective, which may bring new insights to relevant research in this area.
2) *Method View:* We propose an end-to-end framework named ESCS to exploit description complementarity and salience within the entire summary set.
3) *Application View:* We design a training strategy to jointly learn the description complementarity and salience, and

an inference strategy to adaptively simulate the set completion process of the entity summarization task.

4) *Result View:* We conduct evaluation experiments and make extensive analyses on results, which clearly demonstrate the effectiveness of ESCS.

The remainder of this article is structured as follows. In Section II, we briefly introduce the related work of our study. In Section III, we list notations and formally define the problem of entity summarization from the perspective of set completion. Technical details of our proposed model will be specified in Section IV. Then, we evaluate the model on a public benchmark in Section V, with some further discussions on experimental results. In Section VI, we conclude this article and provide some ideas for future work in this area.

II. Related Work

In this section, we briefly provide a review of relevant approaches. Specifically, we group the related work into two directions of literature: *1) entity summarization* and *2) KG embedding*.

A. Entity Summarization

Entity summarization is a technique to generate an optimal and concise entity summary to characterize the entity in KGs. Early research on entity summarization mainly concentrates on unsupervised methods. RELIN [8] leverages the relatedness and informativeness of description elements for ranking. Besides, DIVERSUM [9] attaches the importance of diversity-aware algorithm. Both FACES [10] and FACES-E [17] are clustering approaches. FACES mimics human conceptual clustering techniques to group facts and pick representative facts, while FACES-E extends FACES to glean types for datatype properties by processing their object values. Meanwhile, CD [11] treats entity summarization as a quadratic knapsack problem (QKP). LinkSUM [18] optimizes the combination of the PageRank algorithm with an adaption of the Backlink method together with new approaches for predicate selection. Afterward, balancing frequency and rarity metrics for all entity properties is considered in BAFREC [19]. KAFCA [20] extracts tokens from objects and forms a hierarchical relationship. Then, MPSUM [12] extends a probabilistic topic model to integrate the idea of predicate-uniqueness and object-importance for ranking triples.

In recent years, deep learning methods have gradually emerged in entity summarization task for KGs, and most of them deem this problem as a ranking task [13], [14], [16]. The first neural network method called ESA [16] uses a supervised attention mechanism with Bi-LSTM and calculates attention weights for facts in each entity, then ranks facts to generate reliable summaries. Some researchers deem that chosen order influences model performance with Bi-LSTM and textual semantics is more useful than graph structure; therefore, Liu *et al.* [13] propose DeepLENS which satisfies permutation invariance. Due to the difficulty of obtaining labeled data in this task, NEST [21] utilizes programmatically labeled data as a weak supervision training method. Besides, it jointly encodes graph structure and text in KGs and generates high-quality diversified summaries. For overcoming the drawbacks of previous models, Wei *et al.* [14] integrate automatic feature extraction and multiuser preference simulation in AutoSUM. A cooperative environment where a user reads the current entity summary providing feedback to correct an entity summarizer is constructed in DRESSED [15] modeled as a reinforcement learning agent.

On the whole, prior work mainly treats entity summarization as a ranking problem, whereas we tackle this task in a novel set completion view to exploit description complementarity and salience within the entire summary set.

B. KG Embedding

Although effective in representing structured data, the underlying symbolic nature of such triples usually makes KGs hard to manipulate. Therefore, KG embedding has been proposed [22]–[25], which benefits a lot of applications [26]–[28]. Generally speaking, KG embedding methods can be classified into three categories: translational distance models [23], [24], [29], [30], tensor factorization models [22], [31], [32], and neural network models [33]–[36].

For translational distance models, TransE [23] introduces translation invariance into this field. Afterward, many methods extend the TransE model to handle more complex relations. For example, TransH [24] applies relation-specific hyperplanes. TransR [37] utilizes linear transformations to heterogeneous relation spaces. TransD [29] employs dynamic matrices. TranSparse [30] enforces sparseness on the projection matrix. And LineaRE [38] regards KG embedding as a simple linear regression.

Whereas for tensor factorization methods, Nickel *et al.* [22] propose RESCAL model to perform collective learning via the latent components of the factorization. Then, there are many extensions on basic RESCAL. DistMult [39] simplifies RESCAL by restricting the diagonal matrices. HolE [40] combines the expressive power of RESCAL with the efficiency and simplicity of DistMult, and ComplEx [41] introduces complex-valued embeddings for better modeling asymmetric relations by extending DistMult. Besides, SimplE [42] proposes a simple enhancement of canonical polyadic decomposition. Enlightened by Euler's identity, RotatE [43] introduces the rotational Hadmard product, and QuatE [44] utilizes the Hamilton product in complex space. For the purpose of modeling the non-commutative composition pattern, Rotate3D [45] maps entities to the 3-D space.

From the neural network viewpoint, SME [33] defines an energy function for semantic matching, and NTN [46] designs a neural tensor network to calculate the energy score. Meanwhile, SLM [46] is a simpler version of NTN. Due to a large number of parameters in NTN, MLP [47] makes all relations share the same parameters. With the prevalence of deep learning, more complicated architectures are devised, such as NAM [48] with a deep layer, RMNN [48] generating knowledge-specific connections, ConvKB [49] introducing convolutional neural networks, R-GCN [50] utilizing graph

TABLE I
MATHEMATICAL NOTATIONS IN THE PROBLEM STATEMENT

Symbol	Description
G	The knowledge graph.
E	The set of entities.
T	The triple set of the knowledge graph.
e	The entity in the knowledge graph.
s	The subject of the subject-predicate-object triple.
p	The predicate of the subject-predicate-object triple.
p^{-1}	The inverse of p.
o	The object of the subject-predicate-object triple.
t	The property-value pair of entity e.
$prop(t)$	The property in triple t.
$val(t)$	The value in triple t.
k	The integer size constraint of the entity summary set.
$D(e)$	The description triple set of the entity e.
$\hat{S}(e)$	The summary set of the entity e.

neural networks, NSCaching [51] applying generative adversarial learning, and NAS [52] about neural architecture search.

To conclude, KG embedding methods are applied based on the requirements of specific application scenarios. Thus, we employ the textual names and average their embedding vectors in the entity summarization task.

III. Problem Statement

In this section, we introduce the notations and definitions associated with our task. For facilitating illustration, Table I lists the mathematical notations.

A. Entity Description

In a KG G, we denote T as the set of subject–predicate–object triples in the form of $\langle s, p, o\rangle$ where s, p, and o denote the subject, predicate, and object of subject–predicate–object triple, respectively. $D(e)$ is the description triple set of entity e as follows:

$$D(e) = \{\langle e, p, o\rangle \in T\} \cup \{\langle s, p, e\rangle \in T\}. \tag{1}$$

For example, in ⟨"Andrew Kippis", "Type", "Person"⟩, e is "Andrew Kippis", p is "Type", and o is "Person". In addition, in ⟨"Robert Kippis", "Son", "Andrew Kippis"⟩, s is "Robert Kippis", p is "Son", and e is "Andrew Kippis". And t describes the property-value pair of entity e

$$t = \langle \text{prop}(t), \text{val}(t)\rangle = \begin{cases} \langle p, o\rangle, & \langle e, p, o\rangle \in D(e) \\ \left\langle p^{-1}, s\right\rangle, & \langle s, p, e\rangle \in D(e) \end{cases} \tag{2}$$

where prop(t) and val(t) represent the property and value in t, and p^{-1} is the inverse of predicate p. And we denote $\langle e, t\rangle$ or $\langle e, \text{prop}(t), \text{val}(t)\rangle$ as the description triple.

B. Entity Summarization

Given an entity e, a description triple set $D(e)$, and an integer size constraint k, a summary of entity e is defined as a subset of description triples $\hat{S}(e) \subseteq D(e)$ such that $|\hat{S}(e)| = k$. Entity summarization aims to provide a summary of the entity from descriptions by selecting an optimal subset of description triples in the KG.

In this article, considering the complementarity and salience of summarized descriptions, we address the entity summarization in a *set completion* view. Specifically, we start with an empty set and need to step-by-step select the summarized description of entity e from $D(e)$ to fill the summary set until the size of the summary set equals k. And for each step $n (n < k)$, we have already selected $n - 1$ summarized descriptions. With existing $n - 1$ summarized descriptions, at the nth step, we will choose another description based on both the complementarity and salience to fill the existing summary set. Until step $n = k$, the summary set is completed successfully.

IV. Methodology

In this section, we first introduce an overview of our proposed model, i.e., ESCS, to address entity summarization task from a new set completion perspective. Then, we detail three major components in ESCS, i.e., DE, SCL, and DSL. Besides, we jointly learn the description complementarity and salience for training the summary set completion process. Finally, the summary generation approach in the inference stage is presented.

A. Framework Overview

In this article, we propose a novel entity summarization method named ESCS from a set completion perspective, to exploit description complementarity and salience within the entire summary set to automatically generate a summary set for the target entity in the KG. As shown in Fig. 2, our proposed ESCS consists of three major components.

1) *Description Embedding:* In the DE module, we consider textual semantics and generate the representations of description triples by MLPs.
2) *Summary Complementarity Learning:* In the SCL module, we learn the complementary relationship within the entire summary set by a Bi-LSTM structure.
3) *Description Salience Learning:* In the DSL module, we measure the salience of individual descriptions within the entire summary set by calculating the similarity between semantic embeddings of entity and property.

For the purpose of simulating the process of summary set completion, we develop a joint learning strategy for training and a summary set generation strategy for inference.

B. Description Embedding

To depict the target entity with a summary set of massive descriptions, we first need to generate the representations of the target entity and all descriptions. In light of prior work concluding that textual semantics is more useful than graph structure in entity summarization task [13], we only consider the textual semantics for representations. Therefore, all target entities and corresponding descriptions are converted into a low-dimensional embedding space with textual features. To begin with, we utilize pretrained word embedding vectors [53] to generate initial description triple representations. Given a description triple $\langle e, \text{prop}(t), \text{val}(t)\rangle$, the entity representation $\mathbf{e}$, property representation $\mathbf{p}$ and value

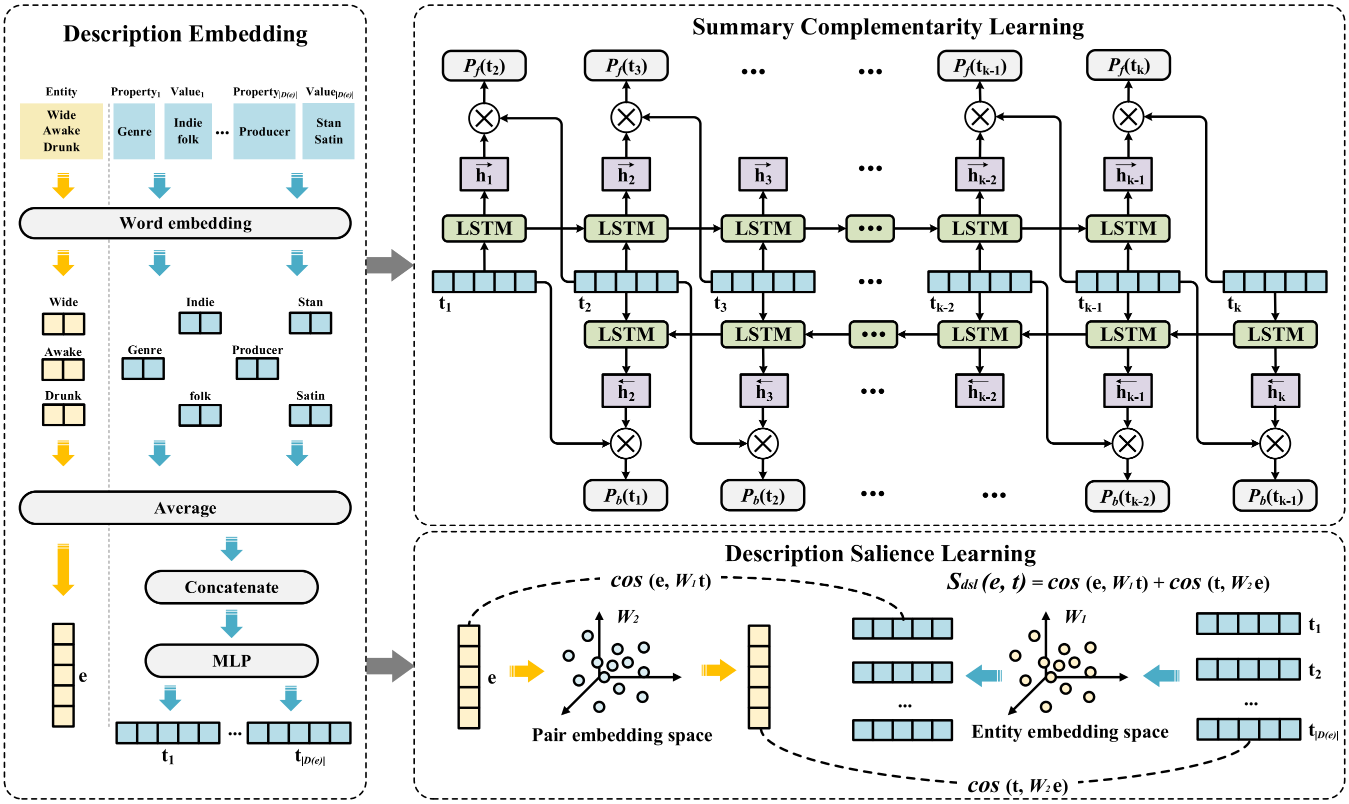


Fig. 2. Framework overview of the proposed ESCS, which consists of three vital modules, i.e., DE, SCL, and DSL.

representation $\mathbf{v}$ are derived from averaging all pretrained embedding vectors over literal words of e, prop(t) and val(t) to extract semantic features. Then, we concatenate $\mathbf{p}$ and $\mathbf{v}$ to form $\mathbf{h}$ for combining the property and value in the description triple. Here, $\mathbf{e}$, $\mathbf{p}$ and $\mathbf{v}$ are d-dimensional vectors, and $\mathbf{h}$ is a $2d$-dimensional vector. Finally, we feed $\mathbf{h}$ into a MLP to generate final representation of property-value pair t

$$\mathbf{h} = [\mathbf{p}; \mathbf{v}] \tag{3}$$
$$\mathbf{t} = \mathrm{MLP}_{\mathrm{pair}}(\mathbf{h}) \tag{4}$$

where $\mathbf{t}$ is an m-dimensional vector to represent t.

C. Summary Complementarity Learning

Normally, entity descriptions in a comprehensive summary of the target entity are mutually complementary. Different from several prior methods which present studies on the diversity or novelty for the summarization [54], we focus on the set complementary of the entity summary. For example, a book is novel for a set of cell phones, but phone chargers are complementary for cell phones. In this module, we intend to utilize Bi-LSTM to learn complementary relationships among entity summaries. On the one hand, Bi-LSTM has been proven to learn complementary relationships well on various task [55]–[58]. And its structure can capture global information about all others for each summary triple through bi-directional propagation. On the other hand, previous studies [59], [60] also demonstrate that recurrent neural networks can cope with the permutation invariance of the set. To be concrete, the structure of one unidirectional LSTM is described as follows:

$$\mathbf{f}_n = \sigma\left(\mathbf{W}_{xf}\mathbf{x}_n + \mathbf{W}_{hf}\mathbf{h}_{n-1} + \mathbf{W}_{cf}\mathbf{c}_{n-1} + \mathbf{b}_f\right) \tag{5}$$
$$\mathbf{i}_n = \sigma\left(\mathbf{W}_{xi}\mathbf{x}_n + \mathbf{W}_{hi}\mathbf{h}_{n-1} + \mathbf{W}_{ci}\mathbf{c}_{n-1} + \mathbf{b}_i\right) \tag{6}$$
$$\mathbf{c}_n = \mathbf{f}_n\mathbf{c}_{n-1} + \mathbf{i}_n \tanh(\mathbf{W}_{xc}\mathbf{x}_n + \mathbf{W}_{hc}\mathbf{h}_{n-1} + \mathbf{b}_c) \tag{7}$$
$$\mathbf{o}_n = \sigma\left(\mathbf{W}_{xo}\mathbf{x}_n + \mathbf{W}_{ho}\mathbf{h}_{n-1} + \mathbf{W}_{co}\mathbf{c}_n + \mathbf{b}_o\right) \tag{8}$$
$$\mathbf{h}_n = \mathbf{o}_n \tanh(\mathbf{c}_n) \tag{9}$$

where $\mathbf{x}_n$ and $\mathbf{h}_n$ are the input and hidden vectors of the nth time step, $\mathbf{f}_n$, $\mathbf{i}_n$, $\mathbf{c}_n$, and $\mathbf{o}_n$ are the result vectors of the forget gate, input gate, memory cell, and output gate of the nth time step, $\mathbf{W}_{\alpha\beta}$ is the weight matrix between vector α and β, $\mathbf{b}_\alpha$ is the bias term of α, $\sigma(\cdot)$ is the sigmoid activation function, and $\tanh(\cdot)$ is another kind of activation function.

Our focus is on the complementary relationships among descriptions instead of the descriptions themselves within the entire summary set. Therefore, we take the entity summary set with constraint size k as the summary sequence $\mathrm{Embed}(\hat{S}(e)) = \{\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_k\}$ in a random order and feed them into the Bi-LSTM as the input of each time step. Afterward, we extract hidden vectors derived from the Bi-LSTM as follows:

$$\overrightarrow{\mathbf{h}_1}, \overrightarrow{\mathbf{h}_2}, \ldots, \overrightarrow{\mathbf{h}_k} = \overrightarrow{\mathrm{LSTM}}(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_k) \tag{10}$$
$$\overleftarrow{\mathbf{h}_1}, \overleftarrow{\mathbf{h}_2}, \ldots, \overleftarrow{\mathbf{h}_k} = \overleftarrow{\mathrm{LSTM}}(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_k). \tag{11}$$

For each summary triple, e.g., the nth triple, property-value pair t_n is predicted depending on the complementarity to both forward sequence $\{\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{n-1}\}$ and backward sequence $\{\mathbf{t}_k, \mathbf{t}_{k-1}, \ldots, \mathbf{t}_{n+1}\}$, i.e., all other summary triples. In this case, the global complementarity can be learned through bi-directional propagation. Then, we add a softmax layer to evaluate the probability of t_n by the following equations:

$$P_f(\mathbf{t}_n \mid \mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{n-1}) = \frac{\exp\left(\overrightarrow{\mathbf{h}_{n-1}}\mathbf{t}_n\right)}{\sum_{t \in D(e)} \exp\left(\overrightarrow{\mathbf{h}_{n-1}}\mathbf{t}\right)} \tag{12}$$

$$P_b(\mathbf{t}_n \mid \mathbf{t}_k, \mathbf{t}_{k-1}, \ldots, \mathbf{t}_{n+1}) = \frac{\exp\left(\overleftarrow{\mathbf{h}_{n+1}}\mathbf{t}_n\right)}{\sum_{t \in D(e)} \exp\left(\overleftarrow{\mathbf{h}_{n+1}}\mathbf{t}\right)} \tag{13}$$

where $D(e)$ is the description triple set of entity e, and $\overrightarrow{\mathbf{h}_{n-1}}$ and $\overleftarrow{\mathbf{h}_{n+1}}$ are the hidden vectors of forward LSTM at $n-1$th time step and backward LSTM at $n+1$th time step. Here, $\overrightarrow{\mathbf{h}_{n-1}}$ and $\overleftarrow{\mathbf{h}_{n+1}}$ have already contained the complementarity of $\{\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{n-1}\}$ and $\{\mathbf{t}_k, \mathbf{t}_{k-1}, \ldots, \mathbf{t}_{n+1}\}$, respectively. Specially, $\mathbf{t}_1$ and $\mathbf{t}_k$ are borderline and only depend on the backward $\overleftarrow{\mathbf{h}_2}$ and the forward $\overrightarrow{\mathbf{h}_{k-1}}$, respectively. The objective functions are designed to maximize the probability of complementarity in the following:

$$L_f(\mathbf{t}_n) = -\log P_f(\mathbf{t}_n \mid \mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{n-1}) \tag{14}$$

$$L_b(\mathbf{t}_n) = -\log P_b(\mathbf{t}_n \mid \mathbf{t}_k, \mathbf{t}_{k-1}, \ldots, \mathbf{t}_{n+1}). \tag{15}$$

For k summary triples of entity e in $\hat{S}(e)$, we design the loss function as follows:

$$L_{\text{scl}} = \frac{1}{|\hat{S}(e)|} \sum_{t \in \hat{S}(e)} (L_f + L_b). \tag{16}$$

D. Description Salience Learning

Descriptions in a KG contain massive semantic information about the target entity. If we merely rely on complementarity to choose summarized descriptions, we may have some information unimportant or uncommon to the target entity itself, such as "⟨Father⟩" in Fig. 1 or "⟨Waistline⟩" of a person. Although these descriptions can supply additional semantics, summarized descriptions in the entire summary set may not be salient. In this module, to weigh the salience of description triples in $D(e)$, we calculate cosine similarities between semantic embeddings of the target entity and its property-value pairs. Note that the entity representation $\mathbf{e}$ and property-value pair representation $\mathbf{t}$ locate in different embedding spaces; hence, we project the entity embedding to property-value pair embedding space. Identically, the property-value pair embedding is transferred into entity embedding space. The salience score function of description triples is defined in the following:

$$\cos(\mathbf{e}, W_1\mathbf{t}) = \frac{\mathbf{e} \cdot W_1\mathbf{t}}{\|\mathbf{e}\| \|W_1\mathbf{t}\|} \tag{17}$$

$$\cos(W_2\mathbf{e}, \mathbf{t}) = \frac{W_2\mathbf{e} \cdot \mathbf{t}}{\|W_2\mathbf{e}\| \|\mathbf{t}\|} \tag{18}$$

$$S_{\text{dsl}}(e, t) = \cos(\mathbf{e}, W_1\mathbf{t}) + \cos(W_2\mathbf{e}, \mathbf{t}) \tag{19}$$

where W_1 and W_2 are the transformation matrices.

For each target entity, there are several summary sets selected by human experts on the basis of their judgment.

Algorithm 1 Model Training Strategy

Require: Entity e, set of entity descriptions $D(e)$, set of entity summaries $\hat{S}(e)$, summary set size k, learning rate η, dimension sizes d and m

Ensure: Model parameters θ

Average word embeddings to get d-dimensional entity embedding $\mathbf{e}$, property embedding $\mathbf{p}$, value embedding $\mathbf{v}$ of each description triples

$\mathbf{h} \leftarrow$ Concatenate $\mathbf{p}$ and $\mathbf{v}$

$\mathbf{t} \leftarrow \text{MLP}_{pair}(\mathbf{h})$

Compute the frequency c of each property-value pairs

repeat

  Input k and summary set sequence $Embed(\hat{S}(e))$ of $\hat{S}(e)$ in a random order and compute $L_{scl}(\hat{S}(e))$ according to Equation 16

  Input all the description triples in $D(e)$ and compute L_{dsl} according to Equation 20

  $L \leftarrow L_{dsl} + L_{scl}$

  Update θ by minimizing the loss function L

until convergence

We can utilize them to calculate the frequency of descriptions in summary sets provided by all experts. Intuitively, the more frequently a description triple occurs in summary sets that all experts provide for the target entity, the more salient it is. Therefore, the loss function is designed as follows:

$$L_{\text{dsl}} = \frac{1}{|D(e)|} \sum_{t \in D(e)} (S_{\text{dsl}}(e, t) - c_t)^2 \tag{20}$$

where c_t is the frequency of property-value pair t, and $S_{\text{dsl}}(e, t)$ and c_t are normalized.

E. Model Training Stage

So far, three modules of our proposed ESCS have been discussed. Through SCL and DSL modules, we can learn complementary relationships among summarized descriptions and estimate the salience of descriptions for summary generation from the perspective of set completion. To attach importance to both aspects together, we design a joint learning method and assign equal weights to L_{dsl} and L_{scl} for set completion training. The overall objective function is as follows:

$$L = L_{\text{dsl}} + L_{\text{scl}}. \tag{21}$$

Then, we can minimize the overall objective function to optimize our proposed ESCS via backpropagation. Algorithm 1 presents the process of set completion training.

F. Model Inference Stage

In the inference stage, we intend to simulate the summary set completion. However, it faces two critical issues. On the one hand, initial summary set is an empty set. Hence, there lacks summarized description contained in the set to judge complementary relationships to determine first one. On the other hand, the complementarity utilized during step-by-step filling is a little different from the complementarity utilized during the training stage, because the global complementarity among k summarized descriptions can be leveraged only after the set is completed. To fully exploit description complementarity and salience within the entire summary set, we devise a summary set generation approach to tackle these issues.

In response to the cold-start issue, choosing a random one from entity description triples may make entity summary set completion error-prone iteratively. Moreover, detecting all types of set composition is not practical due to a large number of entity description triples. Therefore, we solve this issue by calculating the salience score of description triples, as the description triple with the highest score usually describes vital features of the target entity. After the first one is determined, we can follow the forward LSTM to gain the complementarity to the existing set. In this case, with the complementarity obtained from the forward LSTM and salience score computing, we can generate a complete summary set. To address the complementarity issue, with the complete summary set generated by the above-mentioned operations, we combine the forward LSTM and backward LSTM to leverage the global complementarity within the entire summary set step-by-step.

More specifically, given an entity description set $D(e)$ of entity e and a constraint size k of the entity summary, we select the description in D_e with the highest salience score as the first one in the following:

$$t_1 = \arg\max_{t \in D(e)} S_{\mathrm{dsl}}(e, t). \tag{22}$$

Then, we feed t_1 into the forward LSTM to generate the following summarized descriptions. For each step, e.g., the nth triple, the hidden vector $\overrightarrow{\mathbf{h}_{n-1}}$ of the forward LSTM is applied to predict which triple is the next summarized description t_n conditioned on the complementarity to $\{t_1, t_2, \dots, t_{n-1}\}$ in existing summary set and description salience. We decide t_n according to the result as follows:

$$t_n = \arg\max_{t \in D(e)} \left(S_{\mathrm{dsl}}(e, t) + \overrightarrow{\mathbf{h}_{n-1}} \mathbf{t} \right) \tag{23}$$

where $S_{\mathrm{dsl}}(e, t)$ and $\overrightarrow{\mathbf{h}_{n-1}}\mathbf{t}$ are normalized. Along this line, a complete summary set denoted as $\{t_1, t_2, \dots, t_k\}$ is generated by the complementary relationships and description salience.

After that, we leverage Bi-LSTM to take advantage of global complementarity within the entire summary set to check each summarized description again. We follow the reverse order to start with t_k, because t_k has already been inferred from all other summary triples $\{t_1, t_2, \dots, t_{k-1}\}$ based on global complementarity which is only calculated by $\overrightarrow{\mathbf{h}_{k-1}}$ from the forward LSTM such that $\hat{t}_k = t_k$. Then, for each following step from $k-1$ to 2, e.g., the nth triple, $\hat{t}_n$ is chosen based on the complementarity to all others obtained by Bi-LSTM, i.e., $\{t_1, t_2, \dots, t_{n-1}\}$ from the forward LSTM and $\{\hat{t}_k, \hat{t}_{k-1}, \dots, \hat{t}_{n+1}\}$ from the backward LSTM as follows:

$$\hat{t}_n = \arg\max_{t \in D(e)} \left(S_{\mathrm{dsl}}(e, t) + \operatorname{avg}\left(\overrightarrow{\mathbf{h}_{n-1}}\mathbf{t}, \overleftarrow{\mathbf{h}_{n+1}}\mathbf{t} \right) \right) \tag{24}$$

where $\operatorname{avg}(\cdot)$ denotes the average value of parameters. Especially, we predict $\hat{t}_1$ conditioned on all others $\{\hat{t}_k, \hat{t}_{k-1}, \dots, \hat{t}_2\}$ by $\overleftarrow{\mathbf{h}_2}$ derived from the backward LSTM as follows:

$$\hat{t}_1 = \arg\max_{t \in D(e)} \left(S_{\mathrm{dsl}}(e, t) + \overleftarrow{\mathbf{h}_2}\mathbf{t} \right). \tag{25}$$

Algorithm 2 Model Inference Strategy

Require: Entity e, set of entity descriptions $D(e)$, summary set size k, model parameters θ
Ensure: Set of entity summary $\hat{S}(e)$
Initial an empty summary set $\hat{S}(e)$
Choose t_1 as the first input of forward LSTM according to Equation 22 and get the hidden vector $\overrightarrow{\mathbf{h}_1}$
$\hat{S}(e) \leftarrow \{t_1\}$
for $n = 2; n <= k; n++$ **do**
  Choose t_n as the next input of forward LSTM according to Equation 23 and get the hidden vector $\overrightarrow{\mathbf{h}_n}$
  $\hat{S}(e) \leftarrow \{t_1, t_2, \dots, t_n\}$
end for
$\hat{t}_k \leftarrow t_k$
$\hat{S}(e) \leftarrow \{t_1, \dots, t_{k-1}, \hat{t}_k\}$
for $n = k-1; n > 1; n--$ **do**
  Choose $\hat{t}_n$ as the next input of backward LSTM according to Equation 24 and get the hidden vector $\overleftarrow{\mathbf{h}_n}$
  $\hat{S}(e) \leftarrow \left\{t_1, \dots, t_{n-1}, \hat{t}_n, \hat{t}_{n+1}, \dots, \hat{t}_k\right\}$
end for
Choose $\hat{t}_1$ as according to Equation 25
Get the set of entity summary $\hat{S}(e) \leftarrow \left\{\hat{t}_1, \hat{t}_2, \dots, \hat{t}_k\right\}$

As a consequence, we succeed to complete entire entity summary set $\hat{S}(e) = \{\hat{t}_1, \hat{t}_2, \dots, \hat{t}_k\}$ based on the description complementarity and salience within entity summary set. Algorithm 2 describes the process of summary set generation. Moreover, we will prove the effectiveness of our summary generation approach in experiments.

V. Experiments

In this section, we evaluate ESCS with state-of-the-art methods on a public benchmark and demonstrate that ESCS takes advantage of description complementarity and salience within entity summary set in a set completion view. Meanwhile, many discussions on experimental results will be illustrated.

A. Experimental Settings

1) Datasets: We conduct the experiments on a public benchmark: Entity Summarization BenchMark (ESBM) v1.2[1] [61]. This series of benchmarks have already been applied in many previous studies [14]–[16], [21]. ESBM v1.2 contains 1500 summaries for 125 entities in the DBpedia dataset which is an encyclopedic dataset and 600 summaries for 50 entities in the LinkedMDB dataset which is a domain-specific dataset. Entities in the DBpedia dataset are classified into five kinds, i.e., agent, event, location, species, and work. And entities in the LinkedMDB dataset are classified into two kinds, i.e., film and person. Descriptions of each entity in ESBM are labeled with six types of summaries for $k = 5$ and

[1] https://w3id.org/esbm/

six types of summaries for $k = 10$ provided by different human experts on the basis of their judgment. With regard to k, we denote them as the DBpedia-Top5 dataset, LinkedMDB-Top5 dataset, DBpedia-Top10 dataset, and LinkedMDB-Top10 dataset, respectively. We use the train-valid-test split specified on the benchmark.

2) Evaluation Metric: In accordance with the ESBM benchmark, we compare model-generated summaries with ground-truth summaries by calculating the F1 score. Similar to NEST [21], we only use the F1 score as the evaluation metric, because our proposed model is not a ranking model, which makes prior metrics, such as mAP in [16] and NDCG in [15] unsuitable for ESCS. For each method to evaluate, we calculate its mean F1 over all entities and ground-truth summaries on each dataset. The calculation formula of F1 score is as follows:

$$\mathrm{P} = \frac{|S_m \cap S_h|}{|S_m|}, \quad \mathrm{R} = \frac{|S_m \cap S_h|}{|S_h|} \tag{26}$$

$$\mathrm{F1} = \frac{2\mathrm{PR}}{\mathrm{P} + \mathrm{R}} \tag{27}$$

where S_m and S_h denote the model-generated entity summary and the human-made ground-truth entity summary, respectively. In our experiments, we execute all the models to output k triples as entity summaries and set $k = |S_h|$ to be same as prior work [16], [21], so that we have $\mathrm{P} = \mathrm{R} = \mathrm{F1}$. The higher value of F1 reveals the better performance of the method.

3) Implementation Details: We implement our model using the deep learning framework Tensorflow.[2] And hyperparameters are tuned on all datasets in the ESBM v1.2 benchmark. Three hundred-dimensional fastText [62] word embedding vectors trained on Wikipedia are applied to generate initial description triple representations, and hence, d is 300. In $\mathrm{MLP}_{\mathrm{pair}}$, the numbers of hidden units are $[512, 256, 128, 64]$ so that m is 64, and the hidden layers select Leaky ReLU as the activation function whose slope is set to 0.2. The dimension of hidden vectors in Bi-LSTM is 64. All weight parameters are initialized with the Xavier method [63]. We set the learning rate to 0.001 and minimize the loss function by Adam optimizer [64]. The number of training epochs is set to 50, and we perform early stopping on the validation sets to prevent overfitting.

B. Compared Methods

To evaluate the performance of the proposed model, we compare our model with the following representative methods which could be roughly grouped into unsupervised and supervised methods.

1) Unsupervised Methods:

1) **RELIN** [8], which is a variant of the random surfer model that leverages the relatedness and informativeness of description elements for ranking.
2) **DIVERSUM** [9], which focuses on the problem of diversification of entity summary and the context of limited presentation budget.

[2]https://www.tensorflow.org/

TABLE II

COMPARISON WITH THE STATE-OF-ART METHODS FOR ENTITY SUMMARIZATION (F1)

Model	DBpedia		LinkedMDB	
	k=5	**k=10**	**k=5**	**k=10**
RELIN	0.242	0.455	0.203	0.258
DIVERSUM	0.249	0.507	0.207	0.358
FACES	0.270	0.428	0.169	0.263
FACES-E	0.280	0.488	0.313	0.393
CD	0.283	0.513	0.217	0.331
LinkSUM	0.287	0.486	0.140	0.279
BAFREC	0.335	0.503	0.360	0.402
KAFCA	0.314	0.509	0.244	0.397
MPSUM	0.314	0.512	0.272	0.423
ESA	0.331	0.532	0.350	0.416
NEST	0.351	0.528	0.342	0.477
DeepLENS	0.402	0.574	0.474	0.493
ESCS	**0.415**	**0.582**	**0.494**	**0.512**

3) **FACES** [10], which groups conceptually similar facts, in order to select the ranking is the highest feature based on uniqueness and popularity from each group to form a faceted entity summary.
4) **FACES-E** [17], which gleans types for datatype properties by processing their object values and utilizes them to group facts on the basis of their semantics in computing diversified entity summaries.
5) **CD** [11], which formulates entity summarization as a binary QKP to solve and centers on information overlap between features.
6) **LinkSUM** [18], which is a lightweight link-based approach for the relevance-oriented entity summarization.
7) **BAFREC** [19], which balances frequency and rarity metrics for all entity properties in a sophisticated manner.
8) **KAFCA** [20], which creates a graph of objects related to the predicate to determine how knowledge hierarchy reflects intrinsic relationships between triples.
9) **MPSUM** [12], which extends a probabilistic topic model by integrating the idea of predicate-uniqueness and object-importance for ranking triples.

2) Supervised Methods:

1) **ESA** [16], which is the first neural network method and calculates attention weights for facts in each entity, then ranking facts to generate reliable summaries.
2) **NEST** [21], which feeds KGs into a neural encoder and combine neural triple and summary scorers.
3) **DeepLENS** [13], which exploits textual semantics for encoding triples and scores each candidate triple based on its interdependence on other triples.

C. Results and Analyses

Table II reports the results of performance comparison with baselines for entity summarization on the ESBM benchmark.

From the overview, ESCS achieves state-of-the-art performance. Specifically, there are several observations.

1) ESCS performs better than all the other methods. Compared with these methods, the F1 scores are improved by around 3% on all datasets. The results indicate that the proposed ESCS is more suitable for entity summarization than these methods.
2) Supervised solutions generate better results than unsupervised solutions in most cases. Both ESCS and DeepLENS outperform all unsupervised solutions. Compared with unsupervised solutions, the F1 scores of ESCS are at least improved by 13.5%, and the F1 scores of DeepLENS are at least improved by 11.9%. This reveals the superiority and stability of the supervised solutions.
3) ESCS surpasses ESA, NEST, and DeepLENS absolutely, suggesting our supervised method for entity summarization in a view of set completion is more effective. It also demonstrates that learning the complementarity and salience within the entire summary set benefits the entity summarization task a lot.
4) FACES and LinkSUM on LinkedMDB show clear gaps with others. The F1 scores of FACES and LinkSUM both are lower than 0.2. This may be caused by their excessive attention to predicate diversity in descriptions, but many same predicates appear in summarized descriptions of film entities in LinkedMDB.
5) it is unachievable for any method to make the F1 score reach 1. Because there are six different human-made ground-truth summaries in the ESBM benchmark and the summary result that models output cannot match six types of ground-truth summaries simultaneously. Besides, F1 score is much lower than 1 under most conditions, owing to the evaluation metric rather than the algorithm.

In conclusion, ESCS outperforms these baselines which were shown to achieve state-of-the-art performance in entity summarization task. And experimental results suggest the effectiveness of our proposed model.

D. Ablation Study

To further assess the robustness of ESCS and validate the contributions of each component in the proposed model, we design four variants for the ablation study as follows.

1) **SCL-U**, which only uses the SCL module and replaces the Bi-LSTM with the unidirectional LSTM.
2) **SCL**, which only uses the SCL module to generate entity summaries.
3) **DSL**, which only uses the DSL module to generate entity summaries.
4) **ESCS-U**, which is the same as ESCS except that the Bi-LSTM is replaced by the unidirectional LSTM.

The results of the ablation study are shown in Fig. 3. On the whole, ESCS outperforms its four variants, which clearly clarifies the effectiveness of SCL and DSL modules in ESCS. To be specific, we have the following findings.

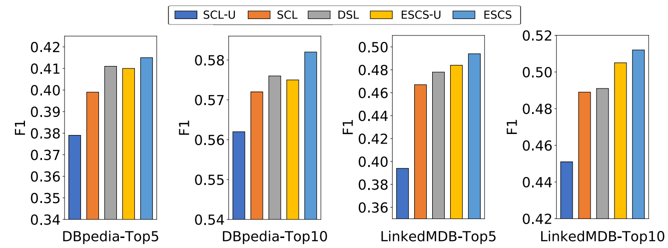


Fig. 3. Ablation study on ESBM benchmark to evaluate the contributions of each component in our proposed ESCS.

1) Learning description salience boosts the performance. Compared with ESCS, the performance of **SCL** drops significantly. In addition, **ESCS-U** performs better than **SCL-U**. These results suggest that selecting descriptions with vital features of the target entity is crucial, and verify the effectiveness of the DSL module.
2) Learning complementary relationships within the entire summary set improves the model effect. This is evidenced by the inferiority of **DSL** to ESCS, which apparently indicates the effectiveness of the SCL module.
3) The effect of salience learning surpasses complementarity learning. According to the comparison between **SCL** and **DSL**, we find that salience learning is more effective than complementarity learning. This is mainly because the first of information needs is salience and the next is complementarity. In addition, selecting descriptions simply conditioned on the complementary may have some information unimportant or uncommon.
4) Applying a Bi-LSTM structure to exploit global complementarity within the entire summary set contributes to performance improvement. From the results that ESCS surpasses **ESCS-U**, we find that the Bi-LSTM structure performs better than the unidirectional LSTM structure. Besides, **SCL** outperforms **SCL-U** obviously which also suggests the effectiveness of the Bi-LSTM structure for exploiting global complementarity.
5) The effects on the two datasets show the difference. **DSL** performs better than **ESCS-U** on DBpedia but not on LinkedMDB, which is caused by the data characteristics. Local complementarity obtained by unidirectional LSTM is not enough to generate satisfying results for the entities in DBpedia, but Bi-LSTM solves this issue by harnessing the global complementarity.

The above-mentioned findings illustrate that our proposed model with SCL module and DSL module can exploit description complementarity and salience within the entire summary set in a view of set completion to generate better results.

E. Inference Strategy Study

In the inference stage, we invent a novel strategy for a summary set generation. To validate the influence of initial triple selection on summary results and the necessity of Bi-LSTM for global complementarity, we design five solutions to cope with summary set generation in ESCS and investigate their effects.

1) **ESCS-I**, which inputs zero vectors to Bi-LSTM to get the results at the beginning.

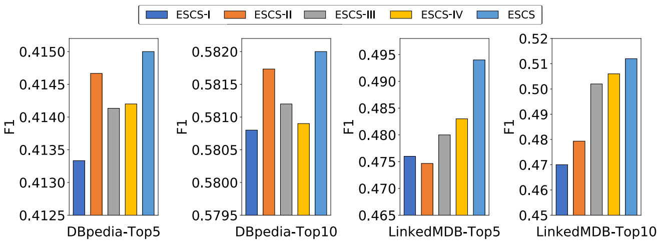


Fig. 4. Results of diverse inference strategies, which verify the effectiveness of our summary generation method.

2) **ESCS-II**, which chooses the triple among entity descriptions randomly at the beginning.
3) **ESCS-III**, which initializes hidden vectors with standard normal distribution randomly and takes the triple whose complementary probability is highest at the beginning.
4) **ESCS-IV**, which is the same as ESCS except that the summary generation only depends on the forward LSTM.
5) **ESCS**, which selects the triple with the highest salience score at the beginning.

From the results shown in Fig. 4, we can easily observe that ESCS outperforms its all variants ESCS-I–IV. This demonstrates the effectiveness of our inference strategy for summary set generation in ESCS. ESCS-I inputs zero vectors to Bi-LSTM, which may make some information ignored. ESCS-II and ESCS-III are both random methods with great uncertainty. ESCS-IV without global information discards a part of complementarity from the backward LSTM. In particular, we observe that the trends of experimental results on DBpedia and LinkedMDB are not consistent. This may be due to different amounts of entity descriptions and different categories of entities. In conclusion, the results of five different inference strategies declare the effectiveness of the proposed ESCS.

F. Parameter Sensitivity Study

We now discuss the parameter sensitivities in our proposed model. When applying ESCS to a new dataset, we need to figure out the impact of several important hyperparameters. Here, we investigate the performance change of ESCS with regard to the dimension of hidden vectors in LSTM. Specifically, Fig. 5 visualizes the performance with the increasing values of the dimension of hidden vectors $\mathrm{dim}_{\mathbf{h}}$ in LSTM and we select $[16, 32, 64, 128, \ldots, 1024, 2048]$ to conduct the experiments. From the overview, the experimental results on all datasets indicate that when $\mathrm{dim}_{\mathbf{h}} = 64$, the performance is good enough, which will be difficult to further improve and tend to decline as enlarging the value of $\mathrm{dim}_{\mathbf{h}}$. As for the results on the LinkedMDB-Top5 dataset according to Fig. 5(c), the F1 value is relatively low when $\mathrm{dim}_{\mathbf{h}} = 16$ and 32. The reason may be that the vector dimension is too small to fit well. Until $\mathrm{dim}_{\mathbf{h}} = 64$, the performance of ESCS reaches the best situation. About the results on the LinkedMDB-Top10 dataset shown in Fig. 5(d), although the F1 value fluctuates slightly, the maximum value is still obtained with $\mathrm{dim}_{\mathbf{h}} = 64$.

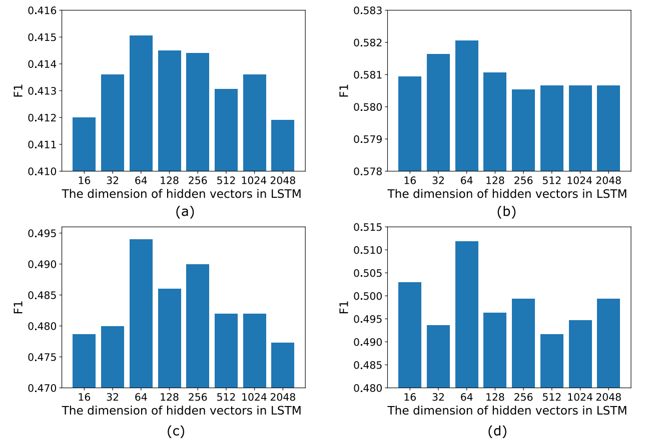


Fig. 5. Performance of ESCS with different dimensions of hidden vectors in Bi-LSTM. (a) DBpedia-Top5. (b) DBpedia-Top10. (c) LinkedMDB-Top5. (d) LinkedMDB-Top10.

Entity: 2011 Kor Royal Cup

Property	Value
Type	Event
Type	SportsEvent
Type	SocietalEvent
Type	FootballMatch
Previous event	2010 Kor Royal Cup
City	Bangkok
Subject	Kor Royal Cup seasons
...	...
Team	Chonburi F.C.
Team	Muangthong United F.C.
Location	National Stadium(Thailand)
Date	2011/1/30
Label	"2011 Kor Royal Cup"@en

DeepLENS

Property	Value
Label	"2011 Kor Royal Cup"@en
Team	Muangthong United F.C.
Team	Chonburi F.C.
Location	National Stadium(Thailand)
Date	2011/1/30

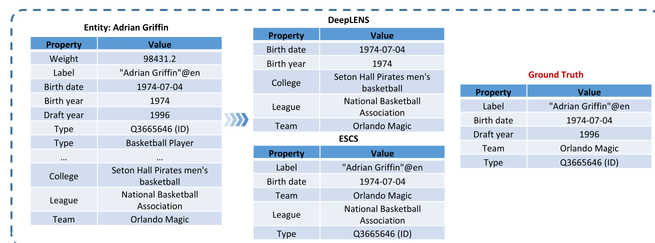
ESCS

Property	Value
Type	FootballMatch
City	Bangkok
Team	Chonburi F.C.
Location	National Stadium(Thailand)
Date	2011/1/30

Ground Truth

Property	Value
Type	FootballMatch
City	Bangkok
Team	Chonburi F.C.
Location	National Stadium(Thailand)
Date	2011/1/30

(a)



Entity: Adrian Griffin

Property	Value
Weight	98431.2
Label	"Adrian Griffin"@en
Birth date	1974-07-04
Birth year	1974
Draft year	1996
Type	Q3665646 (ID)
Type	Basketball Player
...	...
College	Seton Hall Pirates men's basketball
League	National Basketball Association
Team	Orlando Magic

DeepLENS

Property	Value
Birth date	1974-07-04
Birth year	1974
College	Seton Hall Pirates men's basketball
League	National Basketball Association
Team	Orlando Magic

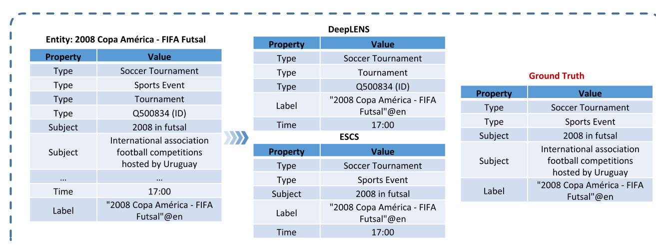
ESCS

Property	Value
Label	"Adrian Griffin"@en
Birth date	1974-07-04
Team	Orlando Magic
League	National Basketball Association
Type	Q3665646 (ID)

Ground Truth

Property	Value
Label	"Adrian Griffin"@en
Birth date	1974-07-04
Draft year	1996
Team	Orlando Magic
Type	Q3665646 (ID)

(b)



Entity: 2008 Copa América - FIFA Futsal

Property	Value
Type	Soccer Tournament
Type	Sports Event
Type	Tournament
Type	Q500834 (ID)
Subject	2008 in futsal
Subject	International association football competitions hosted by Uruguay
...	...
Time	17:00
Label	"2008 Copa América - FIFA Futsal"@en

DeepLENS

Property	Value
Type	Soccer Tournament
Type	Tournament
Type	Q500834 (ID)
Label	"2008 Copa América - FIFA Futsal"@en
Time	17:00

ESCS

Property	Value
Type	Soccer Tournament
Type	Sports Event
Subject	2008 in futsal
Label	"2008 Copa América - FIFA Futsal"@en
Time	17:00

Ground Truth

Property	Value
Type	Soccer Tournament
Type	Sports Event
Subject	2008 in futsal
Subject	International association football competitions hosted by Uruguay
Label	"2008 Copa América - FIFA Futsal"@en

(c)

Fig. 6. Several typical cases, which present the entity summaries generated by ESCS, DeepLENS, and a kind of ground truth. (a) First case about the target entity "2011 Kor Royal Cup." (b) Second case about the target entity "Adrian Griffin." (c) Third case about the target entity "2008 Copa América - FIFA Futsal."

G. Case Study

Fig. 6 shows several typical cases from testing results. Due to the limited space, we select the results of the strongest baseline DeepLENS and our proposed ESCS, and a kind of ground truth. From these cases, we have the following observations.

First, our method performs better than DeepLENS. In the first case, our method gives exactly the same summary as the ground truth, while the summary generated by DeepLENS lacks salient and complementary description "⟨City⟩ Bangkok" explaining the location of this event. Furthermore, for a long-term event, we only need to know a representative team but DeepLENS outputs two teams and neglects other vital information. Second, our method attaches higher importance to the complementarity within the entire summary set. As mentioned earlier, existing research mainly focuses on the ranking of individual triples but leaves out complementary relationships within the entire summary set. For example, in the second case, the summary provided by DeepLENS contains duplicate information, i.e., "⟨Birth date⟩ 1974-07-04" and "⟨Birth year⟩ 1974," since the latter can be inferred from the former. Besides, DeepLENS pays too much attention to some similar aspects in the third case, and includes three descriptions about "⟨Type⟩." Among them, "⟨Type⟩ Soccer Tournament" has already expressed the semantics of "⟨Type⟩ Tournament." Third, although sometimes summaries generated by our method are not completely consistent with the ground truth, it provides more comprehensive information. Taking the third case as an example, compared with the ground truth, ESCS provides not only "⟨Subject⟩ 2008 in futsal" but also "⟨Time⟩ 17:00" of the futsal, which makes summary result complementary and salient at the same time. The above-mentioned typical cases indicate the effectiveness of ESCS once again.

VI. Conclusion

In this article, we first tackled entity summarization from a novel set completion perspective. Then, we proposed an ESCS to adaptively complete the entire summary set. Specifically, the description representations were generated by a DE module. Afterward, we learned the complementary relationship within the entire summary set by a Bi-LSTM structure in the SCL module. Simultaneously, we devised a DSL module to estimate the salience of individual descriptions by calculating the similarity between semantic embeddings of entity and property. In order to simulate the process of summary set completion, two novel training and inference strategies for our proposed ESCS were designed. Extensive experiments on a public benchmark clearly demonstrated the effectiveness of our proposed ESCS and explained the potential of set completion in entity summarization task.

Compared with list ranking methods, ESCS also indicated the limitation in inference efficiency, since it needed to determine each summary step-by-step and could not output all elements to form a complete set at once. In future work, we will further explore more set characteristics to improve the model efficiency. For example, we can solve entity summarization in a binary set element selection view. By this means, the entire set is able to be efficiently generated at once. In addition, we will consider the combination of multimodal information to provide entity summaries in more abundant forms. And the interpretability of the results will also be focused on.

References

[1] X. Wang, X. He, Y. Cao, M. Liu, and T.-S. Chua, "KGAT: Knowledge graph attention network for recommendation," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 950–958.

[2] H. Wang, F. Zhang, X. Xie, and M. Guo, "DKN: Deep knowledge-aware network for news recommendation," in *Proc. World Wide Web Conf.*, 2018, pp. 1835–1844.

[3] F. Zhang, N. J. Yuan, D. Lian, X. Xie, and W.-Y. Ma, "Collaborative knowledge base embedding for recommender systems," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 353–362.

[4] L. Liao, Y. Ma, X. He, R. Hong, and T.-S. Chua, "Knowledge-aware multimodal dialogue systems," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018, pp. 801–809.

[5] X. Li, L. Mou, R. Yan, and M. Zhang, "StalemateBreaker: A proactive content-introducing approach to automatic human-computer conversation," 2016, *arXiv:1604.04358*.

[6] S. Hu, L. Zou, J. X. Yu, H. Wang, and D. Zhao, "Answering natural language questions by subgraph matching over knowledge graphs," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 5, pp. 824–837, May 2018.

[7] Y. Lan, S. Wang, and J. Jiang, "Multi-hop knowledge base question answering with an iterative sequence matching model," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2019, pp. 359–368.

[8] G. Cheng, T. Tran, and Y. Qu, "RELIN: Relatedness and informativeness-based centrality for entity summarization," in *Proc. 10th Int. Semantic Web Conf.* Bonn, Germany: Springer-Verlag, 2011, pp. 114–129.

[9] M. Sydow, M. Pikuła, and R. Schenkel, "The notion of diversity in graphical entity summarisation on semantic knowledge graphs," *J. Intell. Inf. Syst.*, vol. 41, no. 2, pp. 109–149, Oct. 2013.

[10] K. Gunaratna, K. Thirunarayan, and A. Sheth, "FACES: Diversity-aware entity summarization using incremental hierarchical conceptual clustering," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 116–122.

[11] D. Xu, L. Zheng, and Y. Qu, "CD at ENSEC 2016: Generating characteristic and diverse entity summaries," in *Proc. SumPre@ ESWC*, 2016, pp. 1–3.

[12] D. Wei, S. Gao, Y. Liu, Z. Liu, and L. Hang, "MPSUM: Entity summarization with predicate-based matching," 2020, *arXiv:2005.11992*.

[13] Q. Liu, G. Cheng, and Y. Qu, "DeepLENS: Deep learning for entity summarization," 2020, *arXiv:2003.03736*.

[14] D. Wei *et al.*, "AutoSUM: Automating feature extraction and multi-user preference simulation for entity summarization," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*. Singapore: Springer, 2020, pp. 580–592.

[15] Q. Liu, Y. Chen, G. Cheng, E. Kharlamov, J. Li, and Y. Qu, "Entity summarization with user feedback," in *Proc. Eur. Semantic Web Conf.* Heraklion, Greece: Springer, 2020, pp. 376–392.

[16] D. Wei *et al.*, "ESA: Entity summarization with attention," 2019, *arXiv:1905.10625*.

[17] K. Gunaratna, K. Thirunarayan, A. Sheth, and G. Cheng, "Gleaning types for literals in RDF triples with application to entity summarization," in *Proc. Eur. Semantic Web Conf.* Heraklion, Greece: Springer, 2016, pp. 85–100.

[18] A. Thalhammer, N. Lasierra, and A. Rettinger, *LinkSUM: Using Link Analysis to Summarize Entity Data*. Lugano, Switzerland: Springer, 2016.

[19] H. Kroll, D. Nagel, and W.-T. Balke, "BAFREC: Balancing frequency and rarity for entity characterization in open linked data," in *Proc. 1st Int. Workshop Entity Retr. (EYRE)*, 2018, pp. 1–4.

[20] E. Kim and K.-S. Choi, "Entity summarization based on formal concept analysis," in *Proc. 1st Int. Workshop Entity Retr. (EYRE)*, 2018, p. 4.

[21] J. Li *et al.*, "Neural entity summarization with joint encoding and weak supervision," 2020, *arXiv:2005.00152*.

[22] M. Nickel, V. Tresp, and H.-P. Kriegel, "A three-way model for collective learning on multi-relational data," in *Proc. 28th Int. Conf. Int. Conf. Mach. Learn.*, 2011, pp. 809–816.

[23] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Proc. Neural Inf. Process. Syst. (NIPS)*, 2013, pp. 1–9.

[24] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph embedding by translating on hyperplanes," in *Proc. AAAI Conf. Artif. Intell.*, 2014, vol. 28, no. 1, pp. 1112–1119.

[25] J. A. Starzyk and A. Horzyk, "Episodic memory in minicolumn associative knowledge graphs," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3505–3516, Nov. 2019.

[26] H. Wu *et al.*, "Deep technology tracing for high-tech companies," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2019, pp. 1396–1401.

[27] Y. Du *et al.*, "Inheritance-guided hierarchical assignment for clinical automatic diagnosis," in *Proc. 26th Int. Conf. Database Syst. Adv. Appl. (DASFAA)* (Lecture Notes in Computer Science), vol. 12683. Taipei, Taiwan: Springer, Apr. 2021, pp. 461–477.

[28] Z. Li *et al.*, "Cross-oilfield reservoir classification via multi-scale sensor knowledge transfer," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 5, pp. 4215–4223.

[29] G. Ji, S. He, L. Xu, K. Liu, and J. Zhao, "Knowledge graph embedding via dynamic mapping matrix," in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Conf. Natural Lang. Process.*, vol. 1, 2015, pp. 687–696.

[30] G. Ji, K. Liu, S. He, and J. Zhao, "Knowledge graph completion with adaptive sparse transfer matrix," in *Proc. AAAI Conf. Artif. Intell.*, 2016, vol. 30, no. 1, pp. 985–991.

[31] A. P. Singh and G. J. Gordon, "Relational learning via collective matrix factorization," in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2008, pp. 650–658.

[32] M. Nickel, V. Tresp, and H.-P. Kriegel, "Factorizing YAGO: Scalable machine learning for linked data," in *Proc. 21st Int. Conf. World Wide Web*, 2012, pp. 271–280.

[33] 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.

[34] W. Zeng, X. Zhao, J. Tang, and X. Lin, "Collective entity alignment via adaptive features," in *Proc. IEEE 36th Int. Conf. Data Eng. (ICDE)*, Apr. 2020, pp. 1870–1873.

[35] L. Chen, Z. Li, Y. Wang, T. Xu, Z. Wang, and E. Chen, "MMEA: Entity alignment for multi-modal knowledge graph," in *Proc. Int. Conf. Knowl. Sci., Eng. Manage.* Hangzhou, China: Springer, 2020, pp. 134–147.

[36] C. Ren *et al.*, "Ontological concept structure aware knowledge transfer for inductive knowledge graph embedding," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2021, pp. 1–8.

[37] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," in *Proc. AAAI Conf. Artif. Intell.*, 2015, vol. 29, no. 1, pp. 2181–2187.

[38] Y. Peng and J. Zhang, "LineaRE: Simple but powerful knowledge graph embedding for link prediction," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2020, pp. 422–431.

[39] B. Yang, W.-T. Yih, X. He, J. Gao, and L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," 2014, *arXiv:1412.6575*.

[40] M. Nickel, L. Rosasco, and T. Poggio, "Holographic embeddings of knowledge graphs," in *Proc. AAAI Conf. Artif. Intell.*, 2016, vol. 30, no. 1, pp. 1955–1961.

[41] T. Trouillon, J. Welbl, S. Riedel, E. Gaussier, and G. Bouchard, "Complex embeddings for simple link prediction," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 2071–2080.

[42] S. Mehran Kazemi and D. Poole, "SimplE embedding for link prediction in knowledge graphs," 2018, *arXiv:1802.04868*.

[43] Z. Sun, Z.-H. Deng, J.-Y. Nie, and J. Tang, "RotatE: Knowledge graph embedding by relational rotation in complex space," 2019, *arXiv:1902.10197*.

[44] S. Zhang, Y. Tay, L. Yao, and Q. Liu, "Quaternion knowledge graph embeddings," 2019, *arXiv:1904.10281*.

[45] C. Gao, C. Sun, L. Shan, L. Lin, and M. Wang, "Rotate3D: Representing relations as rotations in three-dimensional space for knowledge graph embedding," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2020, pp. 385–394.

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

[47] X. Dong *et al.*, "Knowledge vault: A web-scale approach to probabilistic knowledge fusion," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2014, pp. 601–610.

[48] Q. Liu *et al.*, "Probabilistic reasoning via deep learning: Neural association models," 2016, *arXiv:1603.07704*.

[49] D. Quoc Nguyen, T. Dinh Nguyen, D. Quoc Nguyen, and D. Phung, "A novel embedding model for knowledge base completion based on convolutional neural network," 2017, *arXiv:1712.02121*.

[50] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in *Proc. Eur. Semantic Web Conf.* Heraklion, Greece: Springer, 2018, pp. 593–607.

[51] Y. Zhang, Q. Yao, Y. Shao, and L. Chen, "NSCaching: Simple and efficient negative sampling for knowledge graph embedding," in *Proc. IEEE 35th Int. Conf. Data Eng. (ICDE)*, Apr. 2019, pp. 614–625.

[52] X. Kou, B. Luo, H. Hu, and Y. Zhang, "NASE: Learning knowledge graph embedding for link prediction via neural architecture search," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, 2020, pp. 2089–2092.

[53] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, *arXiv:1301.3781*.

[54] J. Carbonell and J. Goldstein, "The use of MMR, diversity-based reranking for reordering documents and producing summaries," in *Proc. 21st Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*, 1998, pp. 335–336.

[55] W. Li, W. Nie, and Y. Su, "Human action recognition based on selected spatio-temporal features via bidirectional LSTM," *IEEE Access*, vol. 6, pp. 44211–44220, 2018.

[56] W. Ge, X. Lin, and Y. Yu, "Weakly supervised complementary parts models for fine-grained image classification from the bottom up," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3034–3043.

[57] R. Alzaidy, C. Caragea, and C. L. Giles, "Bi-LSTM-CRF sequence labeling for keyphrase extraction from scholarly documents," in *Proc. World Wide Web Conf. (WWW)*, 2019, pp. 2551–2557.

[58] Y. Ma, H. Peng, and E. Cambria, "Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM," in *Proc. AAAI Conf. Artif. Intell.*, New Orleans, LA, USA, 2018, vol. 32, no. 1.

[59] X. Han, Z. Wu, Y.-G. Jiang, and L. S. Davis, "Learning fashion compatibility with bidirectional LSTMs," in *Proc. 25th ACM Int. Conf. Multimedia*, Oct. 2017, pp. 1078–1086.

[60] Y. Li, L. Cao, J. Zhu, and J. Luo, "Mining fashion outfit composition using an end-to-end deep learning approach on set data," *IEEE Trans. Multimedia*, vol. 19, no. 8, pp. 1946–1955, Aug. 2017.

[61] Q. Liu, G. Cheng, K. Gunaratna, and Y. Qu, "ESBM: An entity summarization benchmark," in *Proc. Eur. Semantic Web Conf.* Heraklion, Greece: Springer, 2020, pp. 548–564.

[62] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Trans. Assoc. Comput. Linguistics*, vol. 5, pp. 135–146, Dec. 2017.

[63] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

[64] H. Zhong *et al.*, "Adam revisited: A weighted past gradients perspective," *Frontiers Comput. Sci.*, vol. 14, no. 5, pp. 1–16, Oct. 2020.

Liyi Chen received the B.E. degree in computer science and technology from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2020. She is currently pursuing the M.E. degree with the School of Data Science, University of Science and Technology of China, Hefei, China.

Her research interests include data mining, knowledge graph, and multimodal learning.

Ms. Chen was a recipient of the KSEM 2020 Best Paper Award.

Zhi Li received the B.E. degree in software engineering from Xi'an Jiaotong University, Xi'an, China, in 2015. He is currently pursuing the Ph.D. degree with the School of Data Science, University of Science and Technology of China, Hefei, China.

He has authored or coauthored papers in refereed conference proceedings, such as KDD, IJCAI, AAAI, MM, and SIGIR. His main research interests include data mining, industrial intelligence, and recommender systems.

Weidong He received the B.E. degree in computer science and technology from the University of Science and Technology of China, Hefei, China, in 2016, where he is currently pursuing the Ph.D. degree with the School of Computer Science and Technology.

He has authored or coauthored several papers in refereed conference proceedings, such as ACM MM, IJCAI, and ICME. His research interests include multimodal learning and natural language processing.

Gong Cheng (Member, IEEE) received the Ph.D. degree in computer software and theory from Southeast University, Nanjing, in 2010.

He is currently an Associate Professor with the Department of Computer Science and Technology, Nanjing University, Nanjing, China. His research interests include semantic search, data summarization, and question answering. He has authored or coauthored more than 40 papers in major venues in these areas, such as TKDE, TWEB, WWW, AAAI, IJCAI, and ISWC.

Tong Xu (Member, IEEE) received the Ph.D. degree from the University of Science and Technology of China (USTC), Hefei, China, in 2016.

He is currently an Associate Professor with the School of Data Science, USTC. He has authored more than 70 top-tier journal and conference papers in the fields of social networks and management computing analysis, including TKDE, TMC, TMM, TOMM, KDD, SIGIR, WWW, AAAI, and IJCAI.

Dr. Xu has served as a Program Committee Member for several international conferences, such as KDD, AAAI, WWW, and IJCAI. He was a recipient of the Best Paper Award of KSEM 2020. He has served on numerous conferences, such as the Track Chair for the IEEE ICKG 2020, the Session Chair for the CCKS 2018 and SMP 2020, the Co-Organizers for workshops held in KDD 2018–2021 and SDM 2020–2021.

Nicholas Jing Yuan (Senior Member, IEEE) was the Deputy GM of the AI Creation Business Team, a Senior Applied Scientist Lead/Manager with Microsoft, Beijing, China. He is currently the Vice President of AI, the Chief Scientist, and the Director of the Language and Speech Innovation Laboratory, Huawei Cloud & AI, Hangzhou, China. He has authored or coauthored more than 60 papers in top-tier conferences and journals.

Dr. Yuan is a Senior Member of ACM and CCF. He has received several best paper awards, such as SIGSPATIAL in 2010, ICDM in 2013, and SIGKDD in 2016 and 2018. He was a recipient of several distinguished awards, such as the Natural Science Price of the State Ministry of Education (First Class), the Distinguished Doctoral Dissertation Award of the Chinese Academy of Sciences, and the Microsoft Fellowship. He regularly served as a Program Committee Member in top-tier conferences, such as SIGKDD, WWW, ACL, and AAAI.

Enhong Chen (Senior Member, IEEE) is currently a Professor and the Vice Dean with the School of Computer Science, University of Science and Technology of China, Hefei, China. His research was supported by the National Science Foundation for Distinguished Young Scholars of China. He has authored or coauthored more than 100 papers in refereed conferences and journals, including *Nature Communications*, the IEEE/ACM TRANSACTIONS, KDD, *Neural Information Processing Systems*, IJCAI, and AAAI. His research interests include data mining and machine learning, social network analysis, and recommender systems.

Dr. Chen was on program committees of numerous conferences, including KDD, ICDM, and SDM. He received the Best Application Paper Award on KDD in 2008, the Best Research Paper Award on ICDM in 2011, and the Best of SDM in 2015.