Appendix for Paper 3549

1 Initial knowledge representation Learning

1.1 Top-N Entity Linking Layer

We perform entity mention detection (entity linking) by performing n-gram matching between a input document and entity mentions in KB. In particular, given a document X and all entity names from Freebase, we use each word x_k of X to retrieve entities in KB containing this word and obtain a set of top-N entity candidates from KB for x_k. We refer to the set of all these retrieved candidate entities as $E_k = \{e_{k1}, e_{k2}, ..., e_{kN}\} \in \mathbb{R}^{N \times d_{kb}}$, where d_{kb} is the dimension of the entity embedding in KB and e_{ki} represents the i-th candidate entity mention for the k-th word of X. The embedding of each entity (i.e., e) in KB is learned by a KB embedding layer via DeepWalk algorithm. Next, we will elaborate the KB embedding layer in details.

1.2 KB Embedding Layer

Let us formally define the problem of KB representation learning. Suppose there is a knowledge base $G_{kb} = (V^{kb}, E^{kb})$, where V^{kb} is the set of all vertices and E^{kb} is the set of connections between these vertices. KB representation learning aims to build a low-dimensional representation $e \in \mathbb{R}^{d_{kb}}$ for each vertex $v^{kb} \in V^{kb}$, where d_{kb} is the dimension of representation space and expected much smaller than $|V^{kb}|$. The learned representations encode semantic roles of vertices in the KB.

In this paper, we use DeepWalk (Perozzi et al., 2014) algorithm to learn the graph embeddings of KB. DeepWalk is a typical representation learning method that performs random walks over a graph to build vertex sequences. By regarding vertex sequences as word sequences, it adopts Skip-Gram (Mikolov et al., 2013), a widely-used word representation algorithm, to learn graph representations. Motivated by Skip-Gram, DeepWalk aims to maximize the co-occurrence probability between a target vertex and its context vertices within a random-walk window. Formally, we set a context window size K, and for each vertex v_i^{kb} we define its context vertices $\mathbf{c}_i = \{v_{i-K}^{kb}, \dots, v_{i-1}^{kb}, v_{i+1}^{kb}, \dots, v_{i+K}^{kb}\}$. Thus the objective of DeepWalk can be formalised as:

$$L_{DW} = \frac{1}{|V^{kb}|} \sum_{v_i^{kb} \in V^{kb}} \sum_{v_j^{kb} \in \mathbf{c}_i} \log Pr(v_j^{kb} | v_i^{kb}) \tag{1}$$

Here, the probability $Pr(v_j^{kb} | v_i^{kb})$ is computed using softmax function:

$$Pr(v_j^{kb} | v_i^{kb}) = \frac{\exp(e_j \cdot e_i)}{\sum_{t \in |V_{kb}|} \exp(e_t \cdot e_i)} \tag{2}$$

where e_i and e_j are the representation vectors of the vertices v_j^{kb} and v_i^{kb}. $(\cdot)$ is the the inner product between vectors.

Finally, we use stochastic gradient descent (SGD) to learn the parameters of the KB embedding model by maximizing the objective function of DeepWalk (i.e., L_{DW}), and get the vertex embedding e_i for each vertex v_i^{kb}.

1.3 CNN-based Knowledge Representation Layer

The candidate entities are averaged to form the knowledge representation for the k-th word in the document: $\bar{E}_k = \frac{1}{N} \sum_{i=1}^{N} e_{ki} \in \mathbb{R}^{d_{kb}}$. After obtaining the knowledge representation for each entity mention in the document, a CNN layer is then employed to capture the local n-gram information and learn a higher level knowledge representation $E^{init} \in \mathbb{R}^{n \times d_k}$ (d_k is the number of hidden states of CNN):

$$E^{init} = \text{CNN}(\bar{E}) \tag{3}$$

where $\bar{E} = \{\bar{E}_1, \bar{E}_2, \dots, \bar{E}_n\}$, and n is the length of the input document.

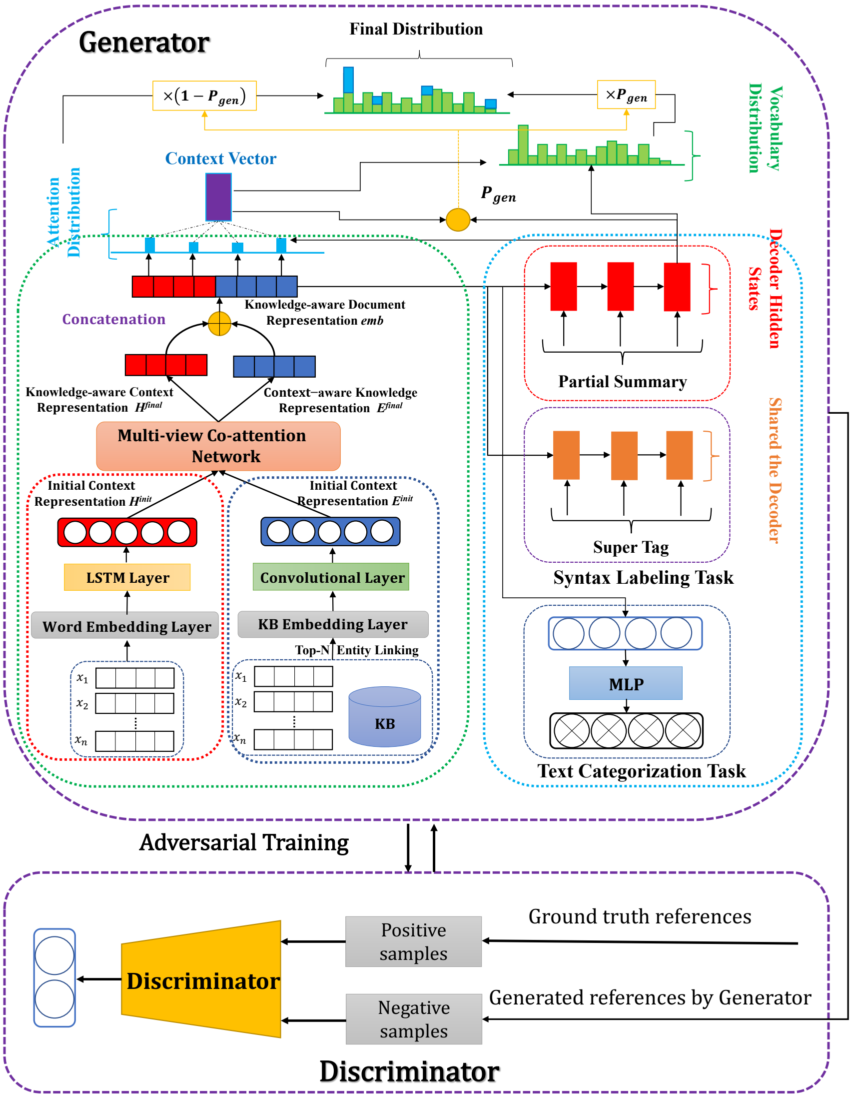

Figure 1: The overall architecture of our model.

2 Architecture of HHS-ATS

We slightly modified the architecture figure of HHS-ATS, as shown in Figure 1, to better describe the knowledge-based attention module. Specifically, the initial knowledge representation learning consists of three components: a top-N entity linking layer, a KB embedding layer, and a convolutional layer.

3 Introduction of CCG supertag annotation

Combinatory Category Grammar (CCG) (Steedman and Baldridge, 2011) uses a set of lexical categories to represent constituents, which provides a connection between syntax and semantics of natural language. In particular, a fixed finite set of CCG categories is described in Table A.1. The basic categories could be used to generate an infinite set $\mathbb{C}$ of functional categories by applying the following recursive definition: (i) $N, NP, PP, S \in \mathbb{C}$; (2) $X/Y, X\backslash Y \in \mathbb{C}$ if $X, Y \in \mathbb{C}$. Each functional category specifies some arguments. Combining the arguments can form a new category according to the orders (Steedman and Baldridge, 2011). The argument could be either basic or functional, and the orders are determined by the forward slash/ and the backward slash \. A category X/Y is a forward functor which could accept an argument Y to the right and get X, while the backward functor $X\backslash Y$ should accept its argument Y to the left.

CCG supertag annotation (Clark, 2002) is a task to assign lexical categories to each word in a piece of text. Formally, CCG supertag annotation can be formulated by $P(Z|X)$, where $X = \{x_1, \ldots, x_n\}$ indicates the n words in a document, and $Z = \{z_1, \ldots, z_n\}$ indicates the corresponding lexical categories. Notice that the length of the words and lexical categories are the same. We provide two examples of documents and the corresponding CCG supertags in Table A.2.

CCG Category	Description
N	noun
NP	noun phrase
PP	prepositional phrase
S	sentence

Table A.1: The description of basic categories used in CCG.

Captions:	a	kitchen	with	two	pots	sitting	on	a	stove	
CCG:	NP/N	N	PP/NP	NP/N	N	(S[ng]\NP)/(S\NP)\(S\NP)	(S\NP)\(S\NP)/NP	NP/N	N	
Captions:	a	suitcase	filled	with	lots	of	items	on	a	bed
CCG:	NP/N	N	(S[pss]\NP)/PP	PP/NP	NP/PP	PP/NP	NP	(S\NP)\(S\NP)/NP	NP/N	N

Table A.2: Examples of captions and corresponding CCG supertags generated by our model.

4 Case Study

First, the knowledge-attention network, which mimics the pre-reading process, learns the commonsense knowledge form the KB as prior knowledge to distinguish the important information from the input text and determine the focus of the summary. For example, HHS-ATS can successfully focus on the words “Olivier Rousteing” that are contained in KB (see Table A.3). After discarding the pre-reading (KB), the generated summary looses salient information, as shown in Table A.3. **Second**, according to what we observe, summary styles in different categories can significantly vary. Two common categories (i.e., *Sports* and *Politics*) in CNN/Daily Mail Corpus are taken as an example, demonstrated in Table A.4. To summarize a politic event, people have a tendency to emphasize the subject of the event, and the result or influence of the event. In contrast, a sport summary is expected to include the teams and scores of the sport event. Obviously, the generated summaries should pay particular attention to different aspects of the topics which belong to the corresponding categories. However, existing methods apply a uniform model to generate summaries for the source texts in different categories, which tend to generate trivial and generic summaries that easily miss or under-represent salient aspects of the original document. Furthermore, in an ablation study of our model (see Table A.3), a model which does not recognize the text categorization could generate a descriptive summary missing a salient entity in the text. **Third**, syntactic information plays a crucial role in sentence generation. Enforcing syntactic conformance addresses issues like incomplete sentences and duplicated words. As shown in Table A.3, the model which is unaware of the text syntax could generate a broken summary for the given document. Yet, an improved system whose component has been co-trained with syntax annotation task could generate a more satisfied sentence for accurately and correctly condensing the raw document. **Forth**, as shown in Table A.3, GAN can further refine the summarization performance and generate more plausible, high-quality and human-like abstractive summaries by adding/deleting information and rephrasing the generated words.

References

Stephen Clark. 2002. Supertagging for combinatory categorial grammar. In *Proceedings of the Sixth International Workshop on Tree Adjoining Grammar and Related Frameworks (TAG+ 6)*. pages 19–24.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. pages 3111–3119.

Category: Show
Article (truncated): "they are one of the world's most famous couples-and have quickly gained respect among the fashion elite. and now, one esteemed designer has revealed why kim kardashian and kanye west have the midas touch. olivier rousteing has revealed that he chose kim and kanye to star in balmain's latest campaign because they 'represent a family for the new world'. scroll down for video. fashion's most well-connected designer, olivier rousteing, has revealed why he snapped kim kardashian and kanye west up to front his balmain campaign. (......) the 29-year-old creative director says he snapped up the duo, seen wearing his designs, because they are among the most talked-about people and embody the idea of a modern family. olivier-who regularly dresses kim, 34, and her siblings for the red carpet-explained that when kendall jenner and kim wear his clothes, they look like a 'fashion army'. the whole family seem enamoured with rousteing's designs and kim and kanye often sport matching outfits by the french fashion house. kim and kanye this week made trips to france and armenia with their daughter, north west. the trip to the religious mecca reportedly included north being baptised in the country where her late father's side of the family originated from. kim kardashian , kanye west and north visit the geghard monastery in armenia and take in the sights. kim, kanye and north have become a fashionable family. pictured here with alia wang, aimie wang and nicki minaj at the alexander wang show in february 2014."
Reference summary: "olivier rousteing has revealed why he chose kim and kanye for balmain. designer says the couple are among the most talked-about people. fashionable couple love wearing matching designs by balmain designer."
Summary by our model: "olivier rousteing has revealed why he chose kim and kanye to star in balmain's latest campaign because they represent a family for the new world. french designer says the couple are among the most talked-about people."
Summary by our model without KB: "kanye west kim kardashian rocking balmain. the 29-year-old creative director has revealed that he was inspired to feature the couple - who have a 22-month-old daughter north - in the label's spring/summer 2015 men's campaign"
Summary by our model without text categorization: "olivier rousteing has seen kim kardashian and kanye west. kim kardashian and kanye west have worn his clothes."
Summary by our model without syntax annotation: "olivier rousteing has revealed why (he chose) kim and kanye west (up to front) his balmain campaign. french designer (says) the couple (~~couple~~) are among the most talked-about (people)."
Summary by our model without GAN: "olivier rousteing has revealed why he chose kim and kanye to star in balmain's latest campaign because they represent a family for the new world. the 29-year-old creative director has revealed that he was inspired to feature the couple-who have a 22-month-old daughter north - in the label's spring/summer 2015 men's campaign."

Table A.3: An example of article from *Show* category and its summaries by different models. The words in red indicate the incomplete or redundant phrases.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pages 701–710.

Mark Steedman and Jason Baldridge. 2011. Combinatory categorial grammar. *Non-Transformational Syntax: Formal and Explicit Models of Grammar* pages 181–224.

Category: Politics
Article (truncated): “isis claimed it controlled part of iraq’s largest oil refinery sunday, posting images online that purported to show the storming of the facility, fierce clashes and plumes of smoke rising above the contested site. the group said it launched an assault on the baiji oil refinery late saturday. by sunday, isis said its fighters were inside the refinery and controlled several buildings, but iraqi government security officials denied that claim and insisted iraqi forces remain in full control. cnn couldn’t independently verify isis’ claim. it wouldn’t be the first time that militants and iraqi forces have battled over the refinery, a key strategic resource that has long been a lucrative target because the facility refines much of the fuel used by iraqis domestically. if an attack damaged oil fields or machinery, it could have a significant impact. the refinery is just 40 kilometers (25 miles) from the northern iraqi city of tikrit, which iraqi forces and shiite militias wrested from isis less than two weeks ago. cnn’s jennifer deaton and catherine. shoichet contributed to this report.”
Reference summary: “isis says it controls several buildings at the baiji oil refinery. iraqi government security officials say iraqi forces remain in full control the refinery, iraq ’s largest, has long been a lucrative target for militants.”

Category: Sports
Article (truncated): “Article (truncated): serena williams claimed her eighth miami open title in 14 years after ruthlessly brushing aside the challenge of 12th seed carla suarez navarro in saturday’s final. the world no 1 won the final 10 games in a 6-2 6-0 demolition of her spanish opponent to claim her third straight title at an event she has dominated since winning her first crown back in 2002. serena williams poses on the beach with the championship trophy after defeating carla suarez navarro. williams poses with the road to singapore sign post on crandon park beach after her straight-sets victory. (......) williams saved a break point in her opening service game and then broke to love in the next to leave suarez navarro with a mountain to climb. there would be no way back, with suarez navarro winning just two points on serve in the second set. williams, meanwhile, won 21 of 22 points on her first serve. spain’s suarez navarro started strongly but was no match for serena williams. williams once again broke to love to move 5-0 ahead before clinically wrapping up the match inside 57 minutes. suarez navarro was quick to hail williams, adding in quotes broadcast by bt sport 1: all that you have, you deserve and for me you are the number one right now.”
Reference summary: “serena williams won her eighth miami open title on saturday. she won the final 10 games in a 6-2 6-0 demolition in miami. the unbeaten world no 1 has now won 12 consecutive finals.”

Table A.4: Two example articles and their summaries from *politics* and *sports* categories, respectively.