ESA: Entity Summarization With Attention

中国科学院信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING,CAS

中国科学院大学
University of Chinese Academy of Sciences

Dongjun Wei [1,2] , Yaxin Liu[1,2] , Fuqing Zhu[1], Liangjun Zang[1] , Jizhong Han[1], Wei Zhou[1], and Songlin Hu[1]

[1] Institute of Information Engineering, Chinese Academy of Sciences

[2] School of Cyber Security, University of Chinese Academy of Sciences

Introduction

- Query a subject named ***Hagar Wilde*** in DBpedia with SPARQL：

RDF triples:
<Subject, Predicate, Object>

Search Hager Wilde in DBpedia:
<**Hagar Wilde**, Predicate, Object>
...
Total **65** RDF triples

Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)

http://dbpedia.org

Query Text

```
Select distinct ?Predicate ?Object where
{
<http://dbpedia.org/resource/Hagar_Wilde>
?Predicate
?Object
}
```

Predicate	Object
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2002/07/owl#Thing
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://xmlns.com/foaf/0.1/Person
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://dbpedia.org/ontology/Person
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#Agent
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#NaturalPerson
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.wikidata.org/entity/Q215627
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.wikidata.org/entity/Q24229398
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.wikidata.org/entity/Q28389
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.wikidata.org/entity/Q36180
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.wikidata.org/entity/Q5
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://dbpedia.org/ontology/Agent
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://dbpedia.org/ontology/ScreenWriter
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://dbpedia.org/ontology/Writer
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://schema.org/Person

Introduction

- Select *Top-5* triples from total 65 triples mentioned above：

```
<http://dbpedia.org/ontology/basedOn> <http://dbpedia.org/resource/Hagar_Wilde> .
<http://dbpedia.org/ontology/birthDate> "1905-07-07"^^<http://www.w3.org/2001/XMLSchema#date> .
<http://dbpedia.org/ontology/deathDate> "1971-09-25"^^<http://www.w3.org/2001/XMLSchema#date> .
<http://xmlns.com/foaf/0.1/name> "Hagar Wilde"@en .
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/class/yago/WomenTelevisionWriters> .
```

Table 1. Top-5 predicate-object pairs of subject ***Hagar Wilde***

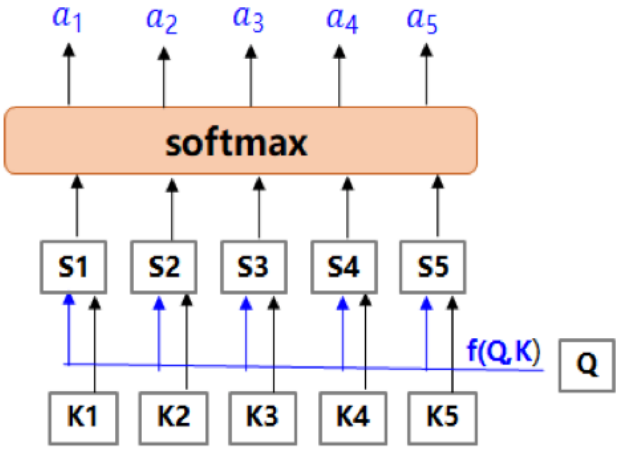
Predicate	Object
Name	Hagar Wilde
Type	Women Televison Writers
Birthdate	1905-07-07
Deathdate	1971-09-25

ESA: Machine Attention

- Attention Mechanism [1]

- Construct machine attention vector [2]

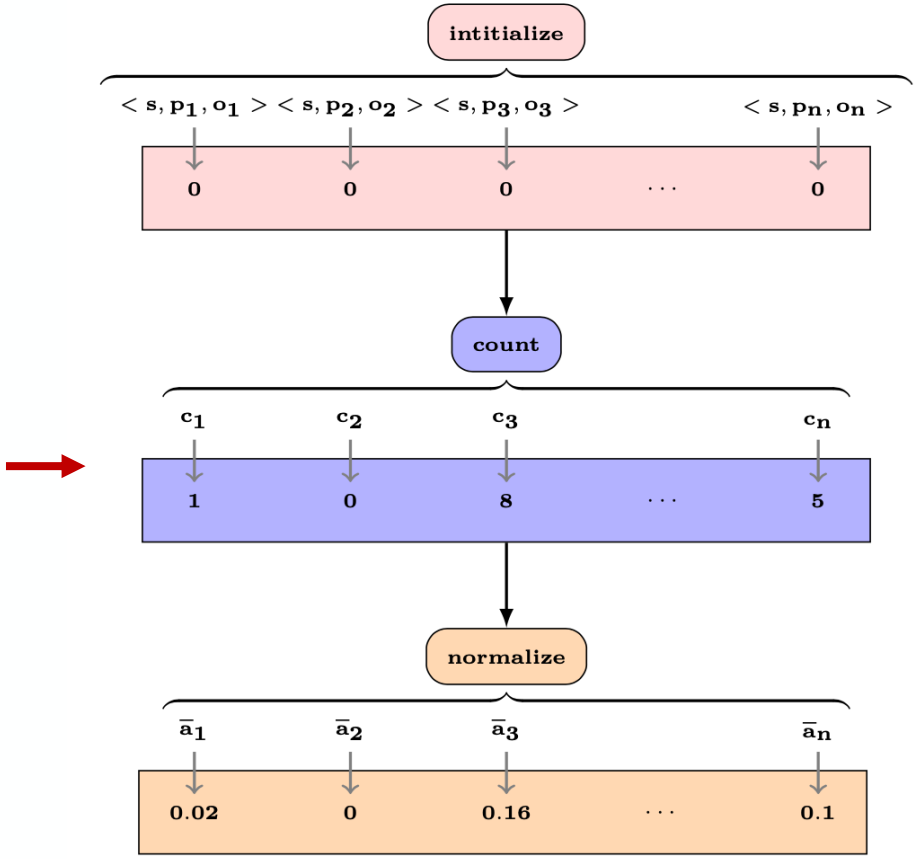
$$f(Q,K)=\begin{cases} Q^T K_i & \textit{dot} \\ Q^T W_a K_i & \textit{general} \\ W_a[Q, K_i] & \textit{concat} \\ V^T_a \tanh(W_a Q + U_a K_i) & \textit{perceptron} \end{cases}$$



ESA: Gold Attention

- Construct gold attention vector

```
|--ESBM_benchmark
        |-- dbpedia
                |-- 1
                        |-- 1_desc.nt
                        |-- 1_gold_top5_0.nt
                        |-- 1_gold_top5_1.nt
                        |-- ...
                        |-- 1_gold_top5_5.nt
                        |-- 1_gold_top10_0.nt
                        |-- 1_gold_top10_1.nt
                        |-- ...
                        |-- 1_gold_top10_0.nt
                |-- 2
                |-- ...
                |-- 100
                |-- 141
                |-- ...
                |-- 165
        |-- lmdb
                |-- 101
                        |-- 101_desc.nt
                        |-- 101_gold_top5_0.nt
                        |-- ...
                |-- 102
                |-- ...
                |-- 140
                |-- 166
                |-- ...
                |-- 175
```



ESA: Model Architecture

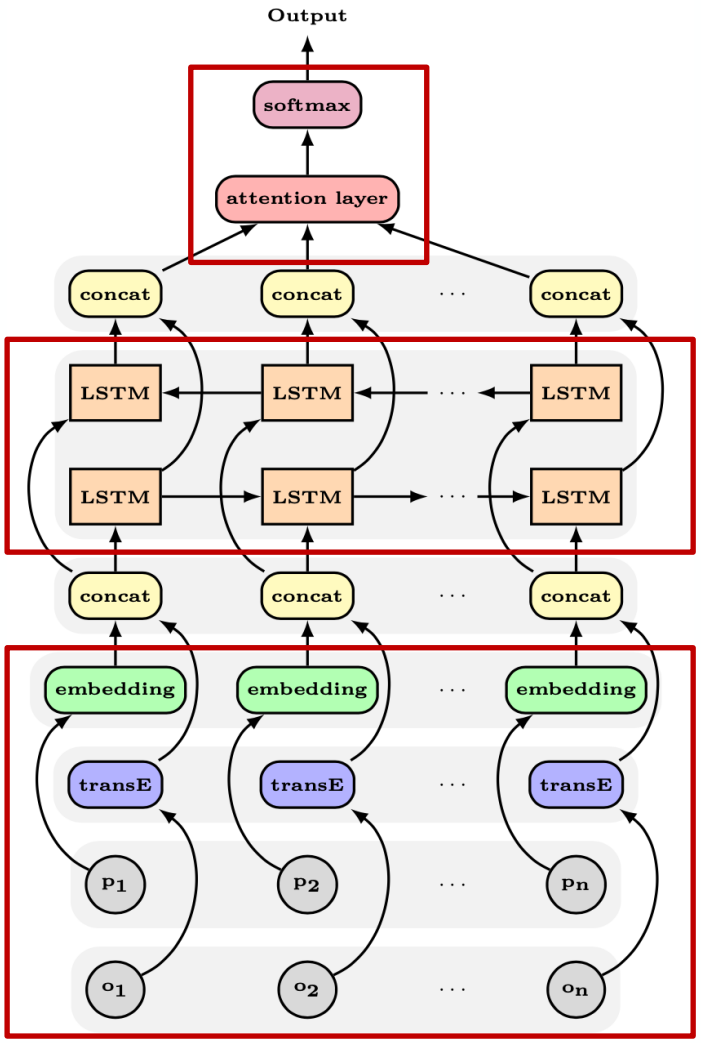
- Architecture
  - Attention Mechanism

    $\alpha = softmax(\boldsymbol{h}_s^T \boldsymbol{h})$

    $\boldsymbol{h}_{s_i} = [\boldsymbol{h}_{L_i}, \boldsymbol{h}_{R_i}]$
  - Bidirectional Network

    $\boldsymbol{h}_{L_i} = LSTM_L(x_i, \boldsymbol{h}_{L_{i-1}})$

    $\boldsymbol{h}_{R_i} = LSTM_R(x_i, \boldsymbol{h}_{R_{i-1}})$
  - Knowledge Representation
    - TransE (Predicates)
    - Word Embedding (Objects)



Experiment

- ESBM Datasets: DBpedia & LMDB

	DBpedia		LinkedMDB		ALL	
	k=5	k=10	k=5	k=10	k=5	k=10
RELIN [3]	0.242	0.455	0.203	0.258	0.231	0.399
DIVERSUM [4]	0.249	0.507	0.207	0.358	0.237	0.464
CD [5]	0.287	0.517	0.211	0.328	0.252	0.455
FACES−E [6]	0.280	0.485	0.313	0.393	0.289	0.461
FACES [7]	0.270	0.428	0.169	0.263	0.241	0.381
LinkSUM [8]	0.274	0.479	0.140	0.279	0.236	0.421
ESA	**0.310**	**0.525**	**0.320**	**0.403**	**0.312**	**0.491**

Table 1. Experimental Results on ESBM benchmark v1.1 of F-measure

	DBpedia		LinkedMDB		ALL	
	k=5	k=10	k=5	k=10	k=5	k=10
RELIN [3]	0.342	0.519	0.241	0.355	0.313	0.466
DIVERSUM [4]	0.310	0.499	0.266	0.390	0.298	0.468
CD [5]	-	-	-	-	-	-
FACES−E [6]	0.388	0.564	0.341	0.435	0.375	0.527
FACES [7]	0.255	0.382	0.155	0.273	0.227	0.351
LinkSUM [8]	0.242	0.271	0.141	0.279	0.213	0.345
ESA	**0.392**	**0.582**	**0.367**	**0.465**	**0.386**	**0.549**

Table 2. Experimental Results on ESBM benchmark v1.1 of MAP

Conclusion

- Neural network is applied into entity summarization task.
- A novel pattern is designed to construct machine attention vectors for modelling supervised attention mechanism.
- Both F-measure and MAP achieves a competitive level in ESBM benchmark v1.1.

References

[1] Sneha Chaudhari *et al.* An Attentive Survey of Attention Models. In arXiv:1904.02874, 2019.

[2] Thang Luong *et al.* Effective Approaches to Attention-based Neural Machine Translation. In EMNLP, 2015.

[3] Gong Cheng *et al*. RELIN: Relatedness and Informativeness-Based Centrality for Entity Summarization. In ESWC, 2011.

[4] Marcin Sydow *et al.* DIVERSUM: Towards diversified summarisation of entities in knowledge graphs. In ICDEW, 2010.

[5] Danyun Xu *et al.* Generating Characteristic and Diverse Entity Summaries. In ENSEC, 2016.

[6] Kalpa Gunaratna *et al.* Gleaning Types for Literals in RDF Triples with Application to Entity Summarization. In ESWC, 2016.

[7] Kalpa Gunaratna *et al.* FACES: Diversity-Aware Entity Summarization using Incremental Hierarchical Conceptual Clustering. In AAAI, 2015.

[8] Andreas Thalhammer *et al.* LinkSUM: Using Link Analysis to Summarize Entity Data. In ICWE, 2016.

中国科学院信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING,CAS

Thanks!
Q&A.