KHAN: Knowledge-Aware Hierarchical Attention Networks for Accurate Political Stance Prediction

Yunyong Ko
yyko@illinois.edu
University of Illinois at Urbana-Champaign, IL, USA

Seongeun Ryu
ryuseong@hanyang.ac.kr
Hanyang University
Seoul, Republic of Korea

Soeun Han
sosilver@hanyang.ac.kr
Hanyang University
Seoul, Republic of Korea

Yeongseung Jeon
ysj@g.ucla.edu
University of California
Los Angeles, CA, USA

Jaehoon Kim
jaehoonkimm@hanyang.ac.kr
Hanyang University
Seoul, Republic of Korea

Sohyun Park
sally5004@ajou.ac.kr
Ajou University
Suwon, Republic of Korea

Kyungsik Han
kyungsikhan@hanyang.ac.kr
Hanyang University
Seoul, Republic of Korea

Hanghang Tong
htong@illinois.edu
University of Illinois at Urbana-Champaign, IL, USA

Sang-Wook Kim*
wook@hanyang.ac.kr
Hanyang University
Seoul, Republic of Korea

（WWW-2023)

code： https://github.com/yy-ko/khan-www23

Leibniz-Institut für Sozialwissenschaften

Reported by Zhaoze Gao

1. Introduction
2. Approach
3. Experiments

gesis
Leibniz-Institut für Sozialwissenschaften

Introduction

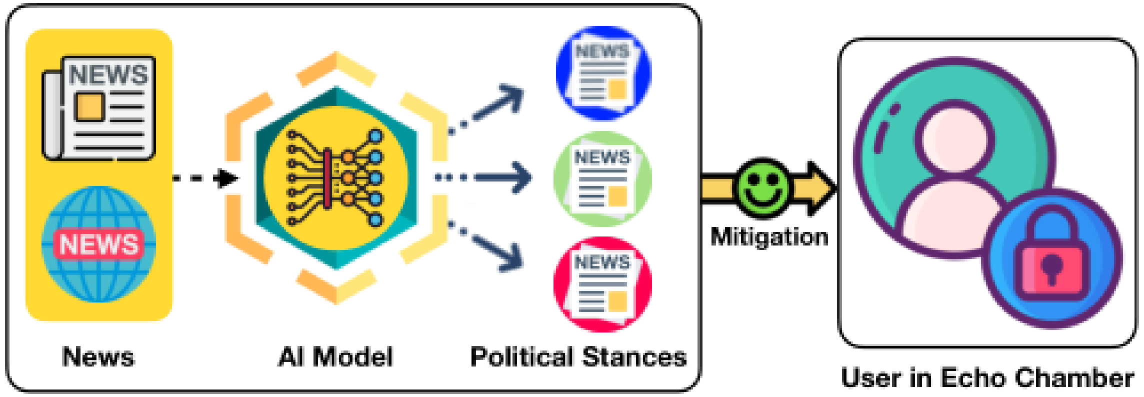


Figure 1: Accurate provision of diverse political stances to mitigate the echo chamber effect.

The political stance prediction for news articles has been widely studied to mitigate the echo chamber effect – people fall into their thoughts and reinforce their pre-existing beliefs.

Approach

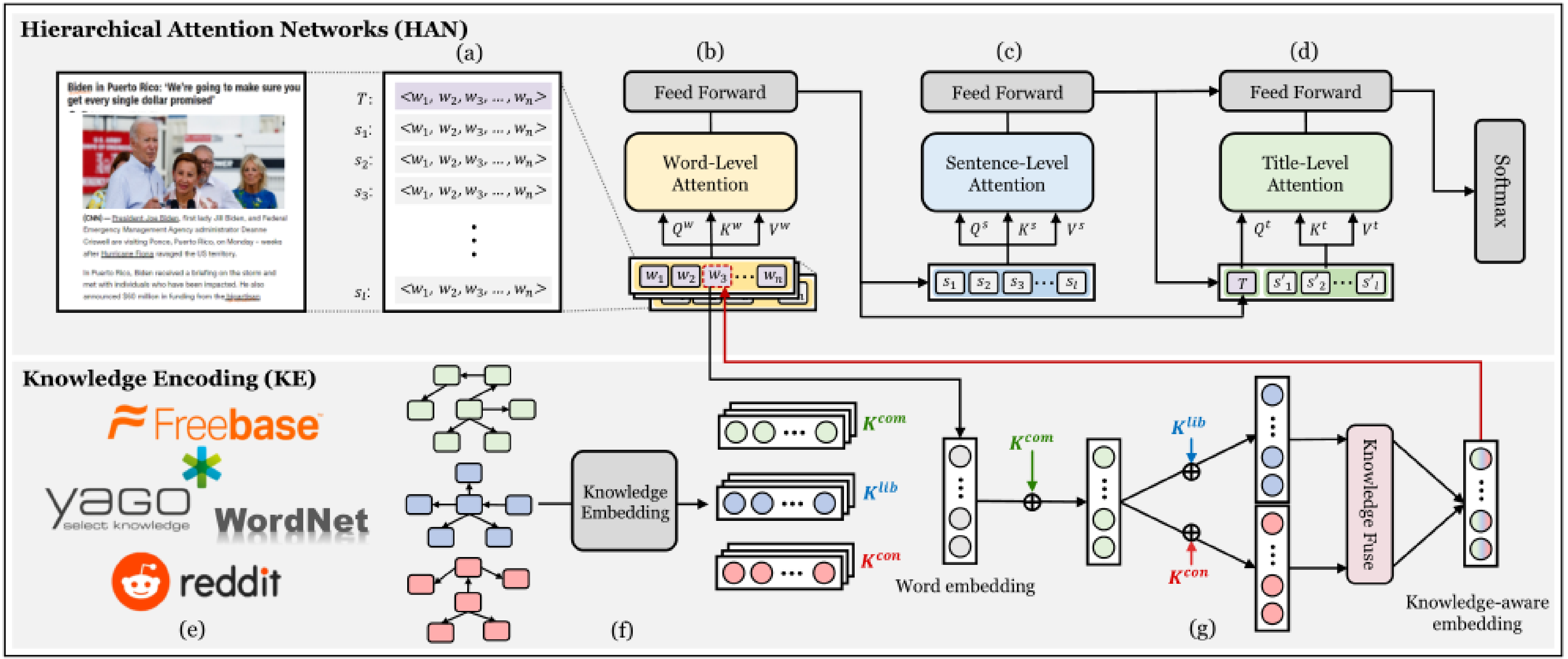


Figure 2: The overview of KHAN: hierarchical attention networks (upper) and knowledge encoding (lower).

Approach

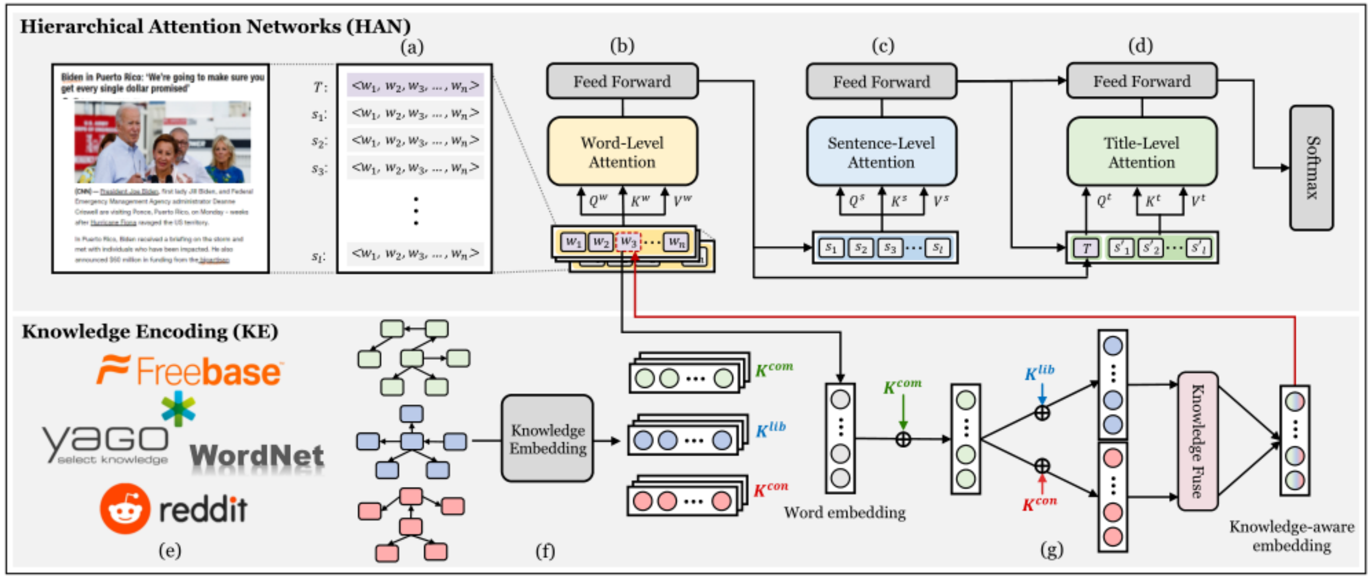


Figure 2: The overview of KHAN: hierarchical attention networks (upper) and knowledge encoding (lower).

Notation	Description
W	a set of word embeddings
S	a set of sentence embeddings
T	the title embedding
w_i, s_j	i^{th} word embedding, j^{th} sentence embedding
d	the embedding dimensionality
N	the total number of words in a dataset
n	the maximum number of words in a sentence
l	the maximum number of sentences in an article
K^{com}	a set of common knowledge embeddings
K^{lib}	a set of liberal knowledge embeddings
K^{con}	a set of conservative knowledge embeddings
α, β	common and political knowledge factors
A, a	news article dataset and a news article
X	a set of learnable parameters
$F(\cdot)$	loss function (i.e., cross-entropy loss)
η	user-defined learning rate

$$S = \{s_1, s_2, \dots, s_l\}$$

Approach

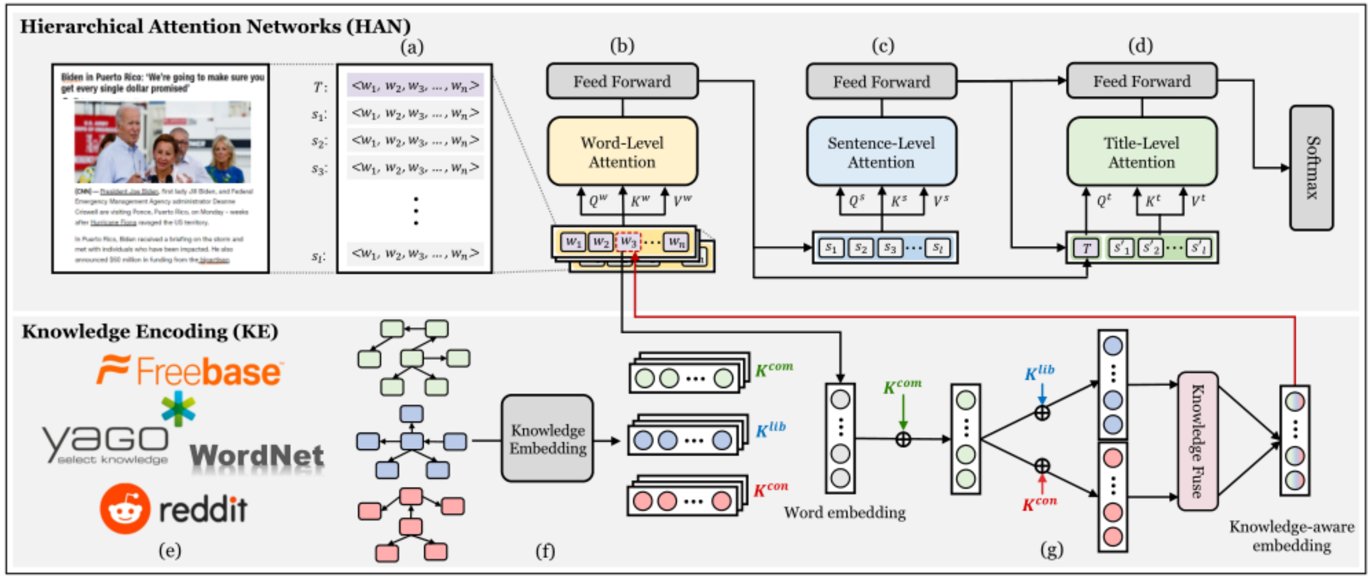


Figure 2: The overview of KHAN: hierarchical attention networks (upper) and knowledge encoding (lower).

$$\min_{w \in \mathbb{R}} \frac{1}{|A|} \sum_{a \in A} F(X, a) + ||X||^2, \quad (1)$$

$$\tilde{W}_j = MultiHead(Q^w, K^w, V^w), \quad (2)$$

$$Q^w = K^w = V^w = W_j$$

$$s_j = Avg(\tilde{W}_j)$$

$$\tilde{S}_k = MultiHead(Q^s, K^s, V^s), \quad (3)$$

$$Q^s = K^s = V^s = S_k$$

Approach

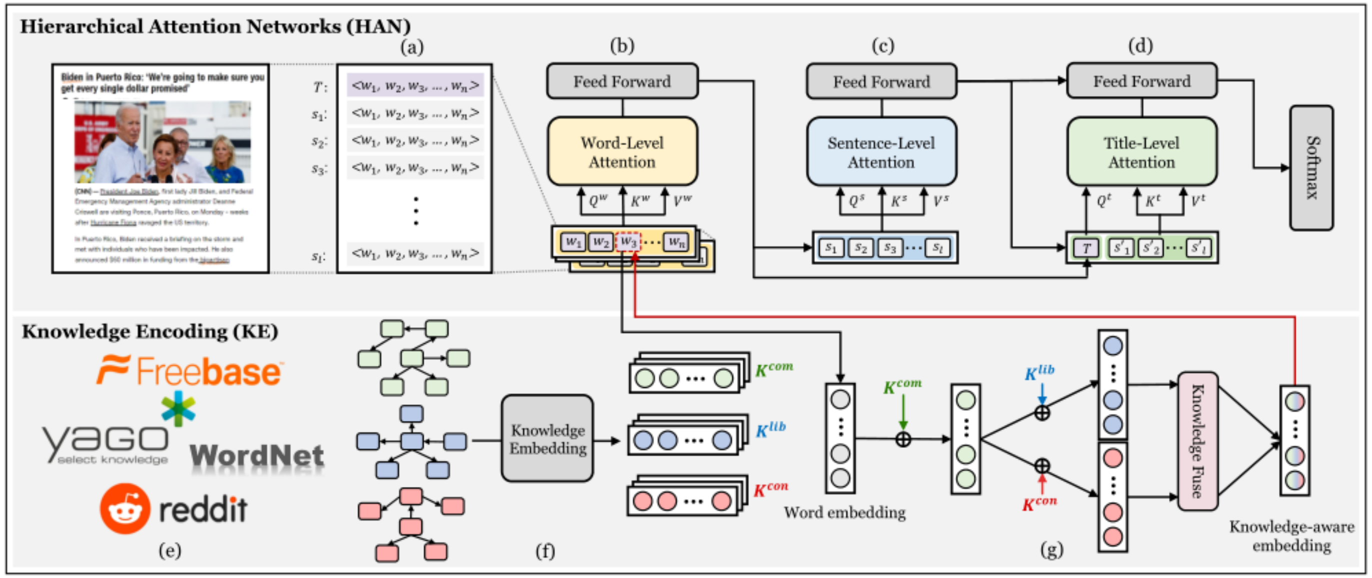


Figure 2: The overview of KHAN: hierarchical attention networks (upper) and knowledge encoding (lower).

$$\tilde{S}_k^T = MultiHead(Q^t, K^t, V^t), \tag{4}$$

$$\tilde{S}_k^* = \tilde{S}_k^T + \tilde{S}_k, \tag{5}$$

$$\hat{y} = Predict(a_k), \quad a_k = Avg(\tilde{S}_k^*), \tag{6}$$

Approach

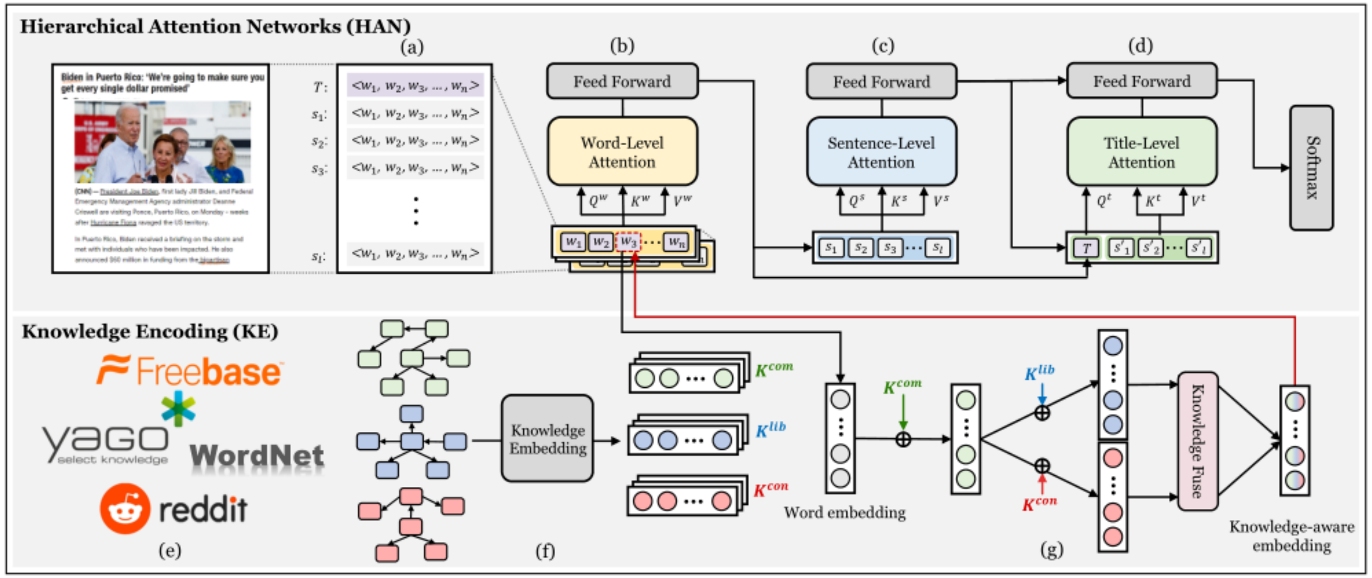


Figure 2: The overview of KHAN: hierarchical attention networks (upper) and knowledge encoding (lower).

$$e \leftarrow W[i]$$
$$e^{com} \leftarrow (1-\alpha)\cdot e \oplus \alpha \cdot K^{com}[i]$$
$$e^{lib} \leftarrow (1-\beta)\cdot e^{com} \oplus \beta \cdot K^{lib}[i]$$
$$e^{con} \leftarrow (1-\beta)\cdot e^{com} \oplus \beta \cdot K^{con}[i]$$
$$W^{*}[i] \leftarrow \mathrm{Fuse}([e^{lib} \| e^{con}]) \oplus e$$

'$\oplus$' means the element-wise addition

Experiments

Table 2: Statistics of political KGs.

	KGAP [13]	KG-lib	KG-con
# of source posts	-	219,915	276,156
# of entities	1,071	5,581	6,316
# of relations	10,703	29,967	33,207
Political stances	Both	Liberal	Conservative

Table 3: Statistics of political news article datasets.

Dataset	# of articles	Class distribution
SemEval	645	407 / 238
AllSides-S	14.7k	6.6k / 4.6k / 3.5k
AllSides-L	719.2k	112.4k / 202.9k / 99.6k / 62.6k / 241.5k

Experiments

Table 4: Comparison of the model accuracy on three real-world datasets (The bold font indicates the best results).

Method	Dataset		
	SemEval	AllSides-S	AllSides-L
Word2Vec [50]	0.7027	0.4858	0.4851
GloVe [55]	0.8071	0.7101	0.6354
ELMo [56]	0.8678	0.8197	0.7483
BERT [29]	0.8692	0.8246	0.7812
RoBERTa [48]	0.8708	0.8535	0.8222
KGAP [13]	0.8956	0.8602	N/A
KCD [69]	0.9087	0.8738	N/A
KHAN-RotatE	0.9426	0.9151	0.8584
KHAN-HAKE	0.9395	0.9216	0.8563
KHAN-ModE	**0.9521**	**0.9256**	**0.8617**

Experiments

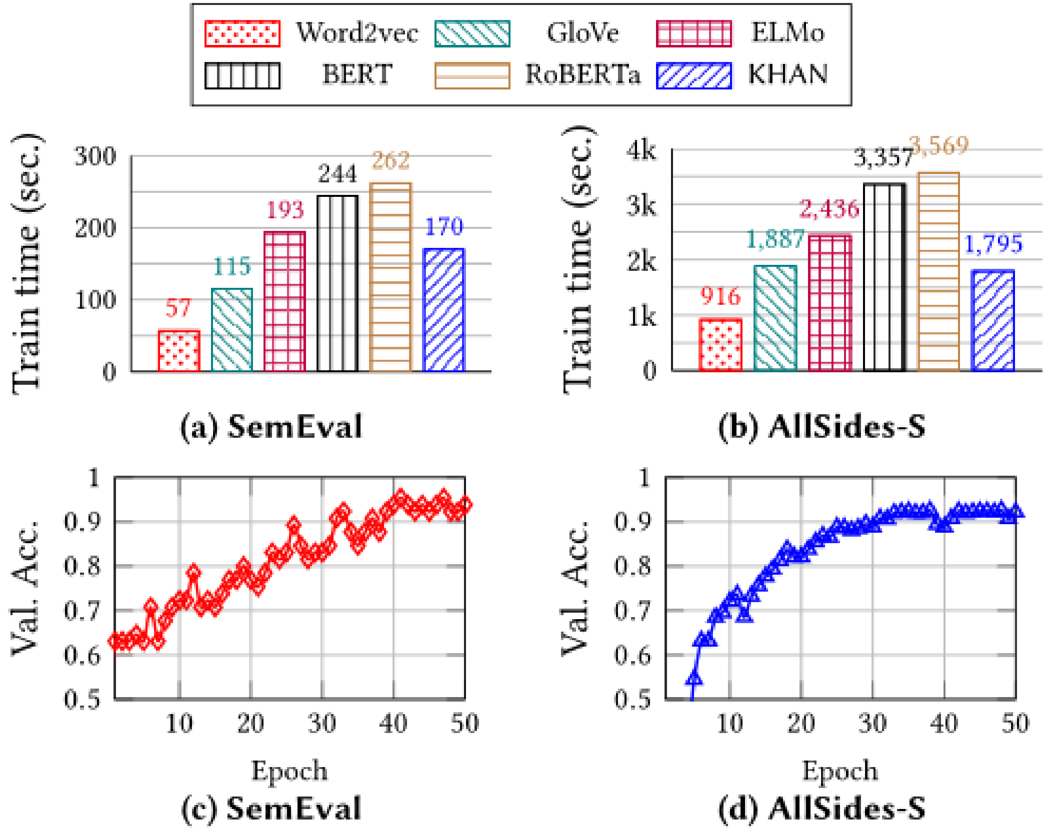


Figure 3: The training time and convergence rate with respect to training epochs on SemEval and AllSides-S.

Experiments

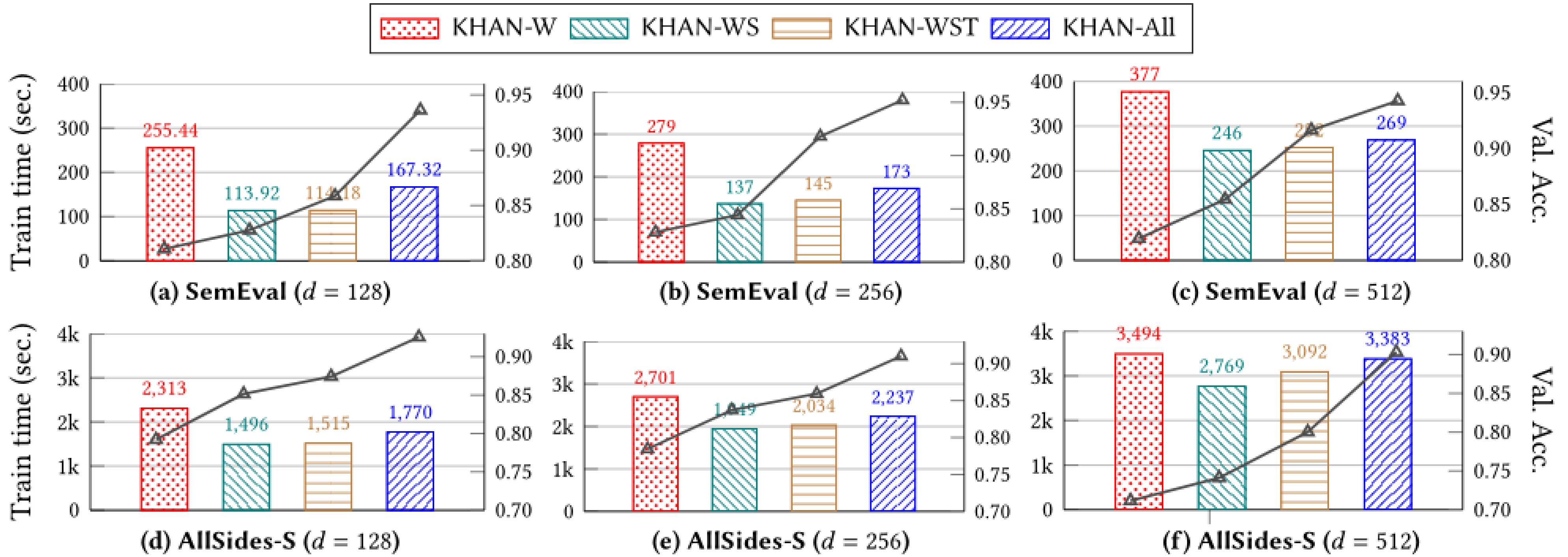
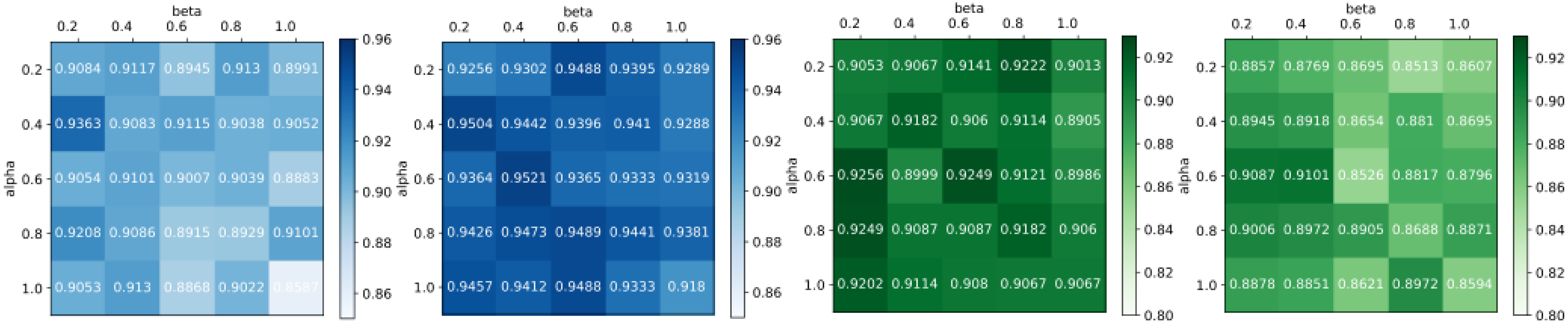


Figure 4: Effectiveness of the main components of KHAN in terms of the training time (bar) and model accuracy (line).

Experiments



(a) SemEval ($d = 128$) (b) SemEval ($d = 256$) (c) AllSides-S ($d = 128$) (d) AllSides-S ($d = 256$)

Figure 5: The impact of hyperparameters α and β on the model accuracy of KHAN in political stance prediction.

Experiments

Table 5: Result of the user study: the top-5 factors in political stance predictions and their importance scores.

Rank	Factor name	Importance score (1-5)
1	Context	4.19 ± 0.94
2	Keyword	4.01 ± 0.88
3	Person	3.94 ± 0.96
4	Tone	3.93 ± 1.13
5	Freq. used word	3.35 ± 1.07

Table 6: The knowledge graph (KG) completion accuracy of RotatE [62] on KG-lib and KG-con.

RotatE						
Metric	KG-lib			KG-con		
	$d = 128$	$d = 256$	$d = 512$	$d = 128$	$d = 256$	$d = 512$
MR	632.69	573.84	567.85	728.78	654.26	640.45
MRR	0.1312	0.1700	0.1859	0.1079	0.1494	0.1633
HITS@1	0.0842	0.1089	0.1209	0.0692	0.0974	0.1093
HITS@3	0.1316	0.1801	0.1985	0.1059	0.1549	0.1693
HITS@10	0.2133	0.2859	0.3148	0.1743	0.2429	0.2625

Experiments

Table 7: The knowledge graph (KG) completion accuracy of ModE [71] on KG-lib and KG-con.

Metric	KG-lib $d = 128$	KG-lib $d = 256$	KG-lib $d = 512$	KG-con $d = 128$	KG-con $d = 256$	KG-con $d = 512$
MR	690.14	622.40	645.23	777.11	740.88	723.90
MRR	0.1312	0.1700	0.1859	0.1128	0.1354	0.1501
HITS@1	0.0842	0.1089	0.1209	0.0685	0.0801	0.0913
HITS@3	0.1316	0.1801	0.1985	0.1127	0.1404	0.1567
HITS@10	0.2133	0.2859	0.3148	0.1981	0.2458	0.2648

Table 8: The knowledge graph (KG) completion accuracy of HAKE [71] on KG-lib and KG-con.

Metric	KG-lib $d = 128$	KG-lib $d = 256$	KG-lib $d = 512$	KG-con $d = 128$	KG-con $d = 256$	KG-con $d = 512$
MR	593.76	597.58	606.03	694.30	684.55	685.92
MRR	0.1474	0.1688	0.1787	0.1311	0.1498	0.1639
HITS@1	0.0904	0.1102	0.1167	0.0831	0.0992	0.1120
HITS@3	0.1541	0.1761	0.1895	0.1348	0.1550	0.1704
HITS@10	0.2595	0.2844	0.3013	0.2205	0.2434	0.2612

Experiments

Table 9: Comparison of our experimental results with the reported results [69] on SemEval (The bold font indicates the results better than the reported results).

Method	SemEval	
	Validation Acc.	Reported Acc.
Word2Vec	**0.7076** ± 0.0104	0.7027
GloVe	**0.8077** ± 0.0251	0.8071
ELMo	0.8666 ± 0.0197	0.8678
BERT	**0.8769** ± 0.0156	0.8692
RoBERTa	**0.8923** ± 0.0112	0.8708
KGAP	N/A	0.8956
KCD	N/A	0.9087
KHAN-RotatE	0.9426 ± 0.0258	N/A
KHAN-HAKE	0.9395 ± 0.0290	N/A
KHAN-ModE	0.9521 ± 0.0183	N/A

Table 10: Comparison of our experimental results with the reported results [69] on AllSides-S (The bold font indicates the results better than the reported results).

Method	AllSides-S	
	Validation Acc.	Reported Acc.
Word2Vec	**0.4977** ± 0.0082	0.4858
GloVe	0.6978 ± 0.0204	0.7101
ELMo	0.8085 ± 0.0178	0.8197
BERT	0.8201 ± 0.0101	0.8246
RoBERTa	**0.8682** ± 0.0081	0.8535
KGAP	N/A	0.8602
KCD	N/A	0.8738
KHAN-RotatE	0.9151 ± 0.0105	N/A
KHAN-HAKE	0.9216 ± 0.0041	N/A
KHAN-ModE	0.9256 ± 0.0098	N/A

Thank you！