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

The Web Conference 2023

Yunyong Ko[†], Seongeun Ryu[§], Soeun Han[§], Youngseung Jeon[*], Jahoon Kim[§]
Sohyun Park[**], Kyungsik Han[§], Hanghang Tong[†], Sang-Wook Kim[§]

[†]University of Illinois at Urbana Champaign, [§] Hanyang University,
[*]University of California Los Angeles, [**]Ajou University

Roadmap

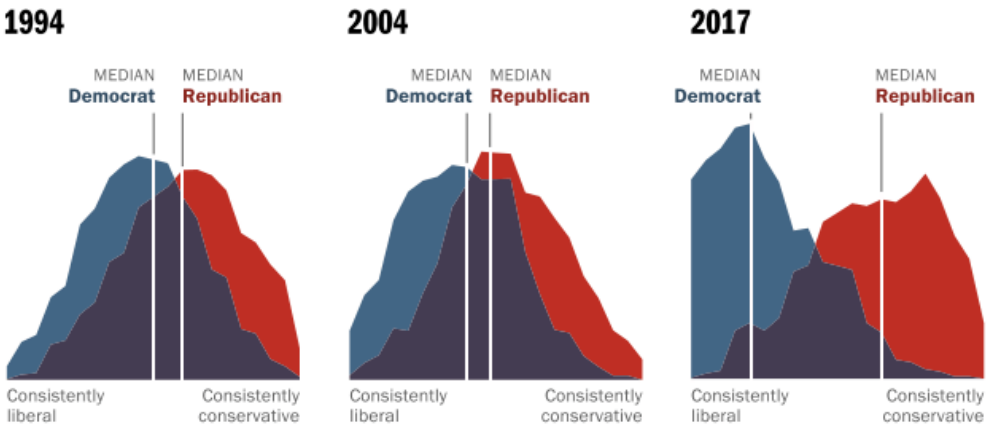
- **Background & Problem Definition**
  - Echo chamber effect
  - Political stance prediction problem
- **Motivation**
  - Limitations of existing works
  - User study
- **Proposed Method: KHAN**
  - Hierarchical attention networks (HAN)
  - Knowledge encoding (KE)
- **Evaluation and Results**
  - Experimental results
  - Conclusion

Background

- **Web-based news platforms**
  - Giving people more chances to access a variety of high-quality news articles
- **Echo chamber effect**
  - Reinforcing people's pre-existing beliefs by limiting exposure to different opinions
  - Leading to serious ***political polarization***

Web-based News Platforms

Trapped in Echo Chambers

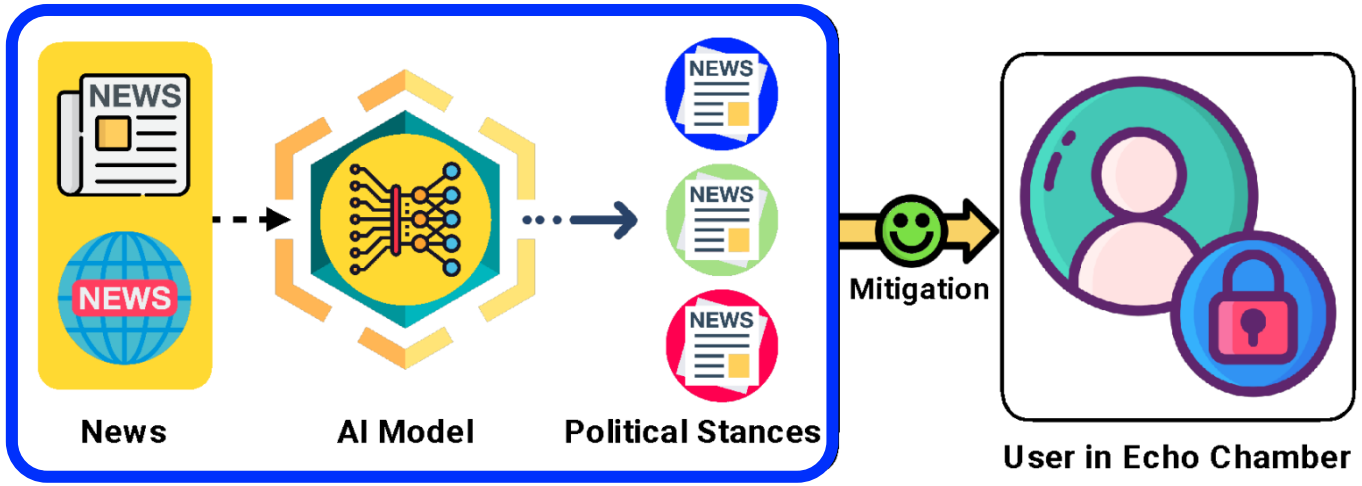


Political Polarization in the US

- Flaxman et al., Filter bubbles, echo chambers, and online news consumption. Public opinion quarterly, 2016.
- Garimella et al., Political Discourse on Social Media: Echo chambers, Gatekeepers, and the Price of Bipartisanship. In Proceedings of the ACM Web Conference (WWW), 2018.
- Cinelli et al., The echo chamber effect on social media. In Proceedings of the National Academy of Sciences, 2021.

Background

- **How to mitigate the echo chamber effect?**
  - Exposing ***diverse opinions*** to people in order to prevent them falling into their existing beliefs
- **Required process: *Accurate political stance prediction***
  - Predicting the political stance of a given news article accurately



Accurate Provision of Diverse Political Stances
to Mitigate the Echo Chamber Effect

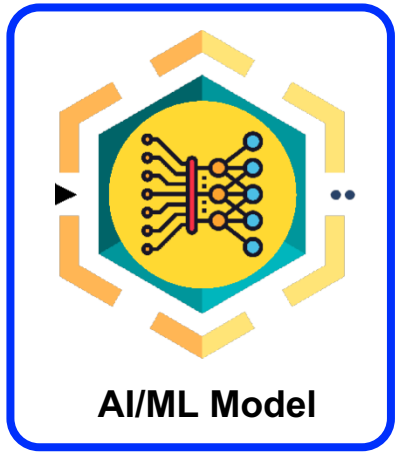
- Bail et al., Exposure to Opposing Views on Social Media Can Increase Political Polarization. The National Academy of Sciences , 2018.
- Gillani et al., Me, My Echo Chamber, and I: Introspection on Social Media Polarization. In Proceedings of the ACM Web Conference (WWW), 2018.

Problem Definition

Political stance prediction

- Given a news article a_i, predict its political stance (e.g., [1,5])

News article a_i



Stance	Probability
Left	0.30
Lean Left	0.45
Center	0.15
Lean Right	0.07
Right	0.03

- To learn the learnable parameters minimizing the loss function
  - A: a given news article datasets
  - $F(\cdot)$: loss function
  - X: the learnable parameters

$$\min \frac{1}{|A|} \sum_{a \in A} F(X, a) + ||X||^2$$

Roadmap

- **Background & Problem Definition**
  - Echo chamber effect
  - Political stance prediction problem
- **Motivation**
  - Limitations of existing works
  - User study
- **Proposed Method: KHAN**
  - Hierarchical attention networks (HAN)
  - Knowledge encoding (KE)
- **Evaluation and Results**
  - Experimental results
  - Conclusion

Limitations of Existing Works

- **General language models (e.g., GloVe, BERT, RoBERTa, …)**
  - *Not designed specially for the political stance prediction*

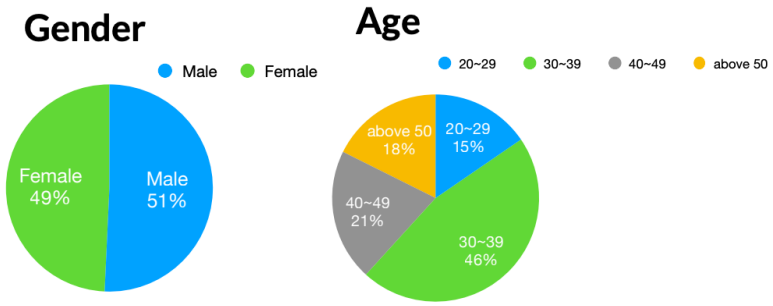
- **DL-based approaches for political stance prediction**
  - Designed to capture *political-related* factors
    - E.g., social, linguistic, and political factors
  - *Not fully justified in terms of the interpretability*
    - Why and how effective are the identified factors in the political stance prediction?

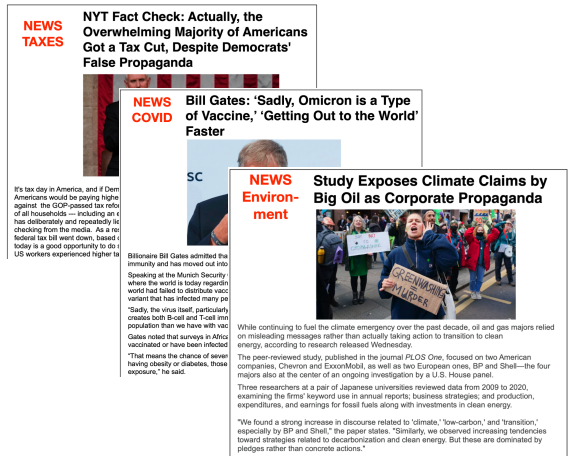
(1) What factors do people really consider and (2) How important are the factors in determining the political stance of a news article?

- Feng et al., KGAP: Knowledge Graph Augmented Political Perspective Detection in News Media, arXiv:2108.03861, 2021.
- Chang Li and Dan Goldwasser, Using social and linguistic information to adapt pretrained representations for political perspective identification. In Findings of the ACL, 2021.
- Zhang et al., KCD: Knowledge Walks and Textual Cues Enhanced Political Perspective Detection in News Media. In Proceedings of the NAACL, 2022.

User Study

- **Goal**
  - To investigate ***the important factors that people really consider*** in political stance prediction
- **Summary**
  - 136 Respondents (recruited via Amazon Mechanical Turk)
    - Gender: male/female
    - Age: 20 - above 50
    - Education: high school or less – university grad above
    - Ethnicity: Caucasian, African American, Asian, Hispanic, …
    - Political stance: very liberal – very conservative
  - 6 news articles with different political stances
    - 3 political-related topic: (1) tax, (2) health, and (3) environment
  - 13 candidate factors
    - E.g., context, tone, keywords, person name, topic, title, imagesm, …





- Lin et al., Which Side Are You On? Identifying Perspectives at the Document and Sentence Levels. In Proceedings of the Computational Natural Language Learning, 2006
- News articles used in our user study are available at: https://github.com/yy-ko/khan-www23/tree/master/user-study

User Study

User study protocol

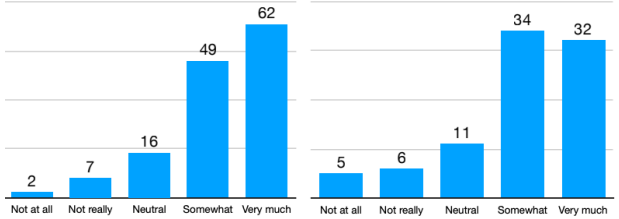
- (1) Provide 6 news articles and 13 factors to the respondents
- (2) Ask them to respond how important each factor with a scale [1: not at all – 5: very much]
- (3) Assess each factor by averaging the importance scores

Results and observations

<Top-5 important factors>

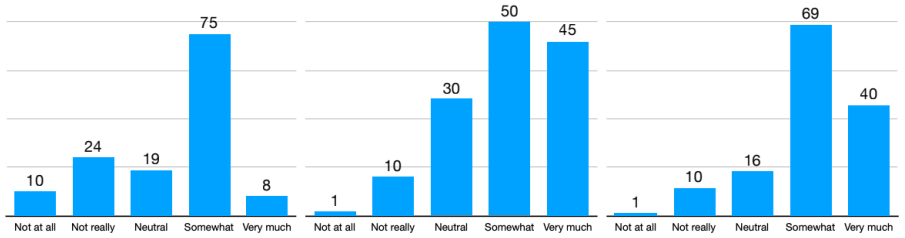
Rank	Factor name	Importance score (1-5)
1	Context	4.19 ± 0.94
2	Keyword	4.01 ± 0.88
3	Person	3.94 ± 0.96
4	Tone	3.93 ± 1.13
5	Freq. used word	3.35 ± 1.07

Both implicit and explicit factors are important in determining the political stance of a news article



<Context> **<Tone>**

Context/tone are ***implicitly*** reflected in a news article



<Freq. used word> **<Person>** **<Keyword>**

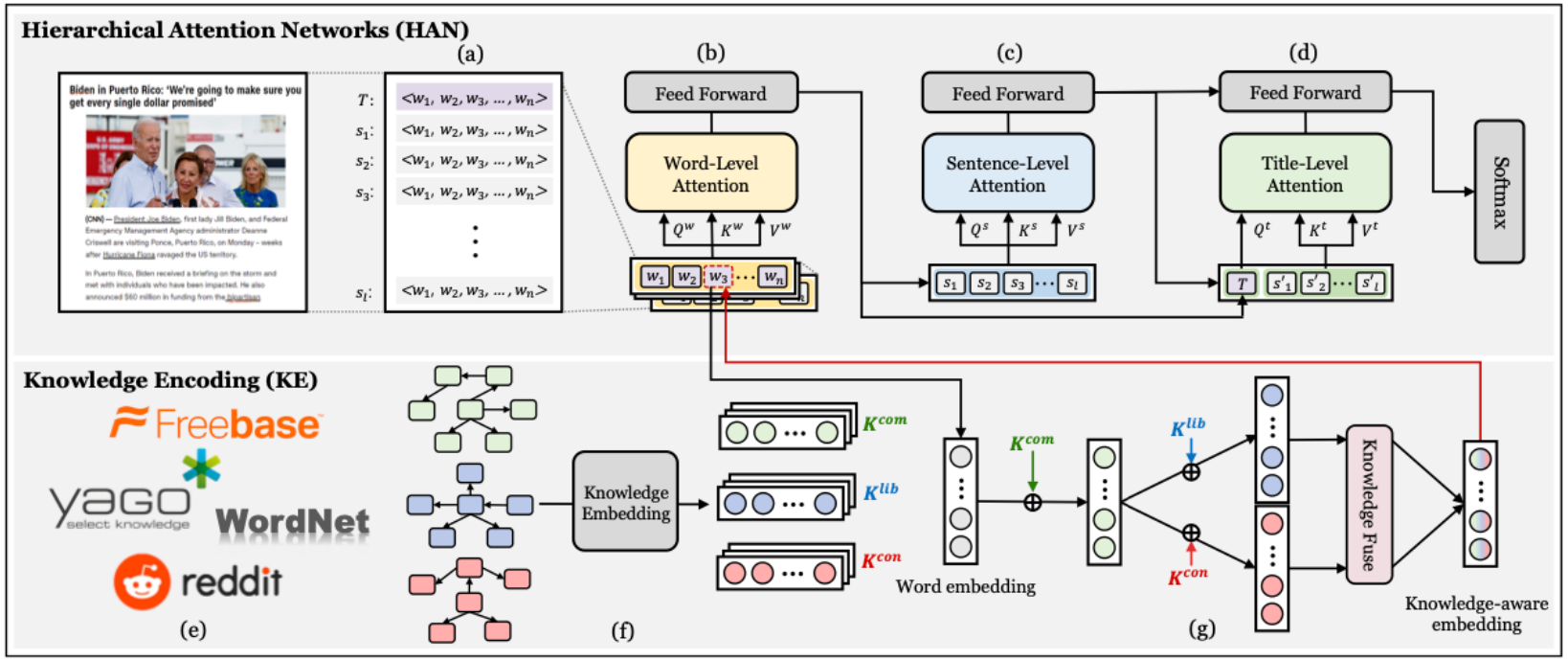
Real-world entities ***explicitly*** appear in a news article, but they are not directly explained

Roadmap

- **Background & Problem Definition**
  - Echo chamber effect
  - Political stance prediction problem
- **Motivation**
  - Limitations of existing works
  - User study
- **Proposed Method: KHAN**
  - Hierarchical attention networks (HAN)
  - Knowledge encoding (KE)
- **Evaluation and Results**
  - Experimental results
  - Conclusion

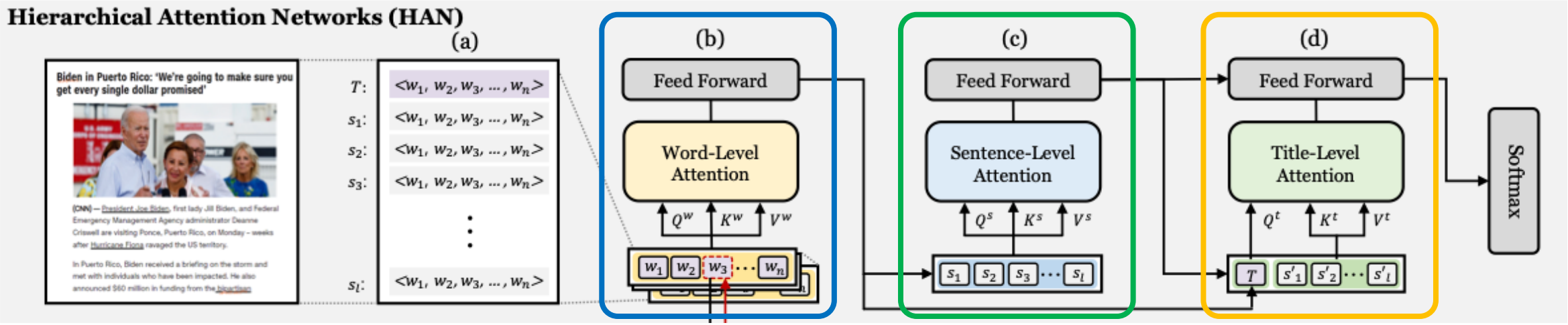
Proposed Method: KHAN

- **Knowledge-aware Hierarchical Attention Networks (KHAN)**
  - Hierarchical attention networks (HAN)
    - To capture the context/tone of a news article with 3-level hierarchy (for ***implicit*** factors)
  - Knowledge encoding (KE)
    - To incorporate external knowledge for real-world entities (for ***explicit*** factors)



Hierarchical Attention Networks (HAN)

- To capture the *implicit* factors of a news article via the *3-level hierarchy*
  - **Word-level attention**
    - To learn the relationships among words in the same sentence (*local context*)
  - **Sentence-level attention**
    - To learn the relationships among sentences (*global context*)
  - **Title-level attention**
    - To reinforce the key idea included in the title (*key idea*)



Hierarchical Attention Networks (HAN)

Word-level Attention $\tilde{W}_j = MultiHead(Q^w, K^w, V^w),$ ⇒ ***Local-context-aware*** word embeddings

Sentence-level Attention $\tilde{S}_k = MultiHead(Q^s, K^s, V^s),$ ⇒ ***Global-context-aware*** sentence embeddings

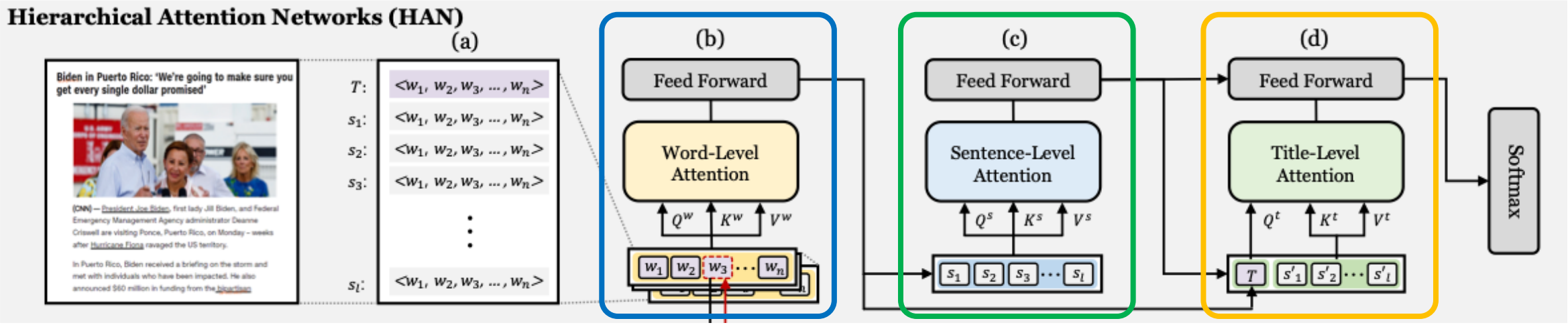
Title-level Attention $\tilde{S}_k^T = MultiHead(Q^t, K^t, V^t),$ ⇒ ***Key-idea-reinforced*** sentence embeddings

Final representation $\tilde{S}_k^* = \tilde{S}_k^T + \tilde{S}_k, \quad a_k = Avg(\tilde{S}_k^*),$ ⇒ ***Implicit-factor-captured*** article embedding

Average

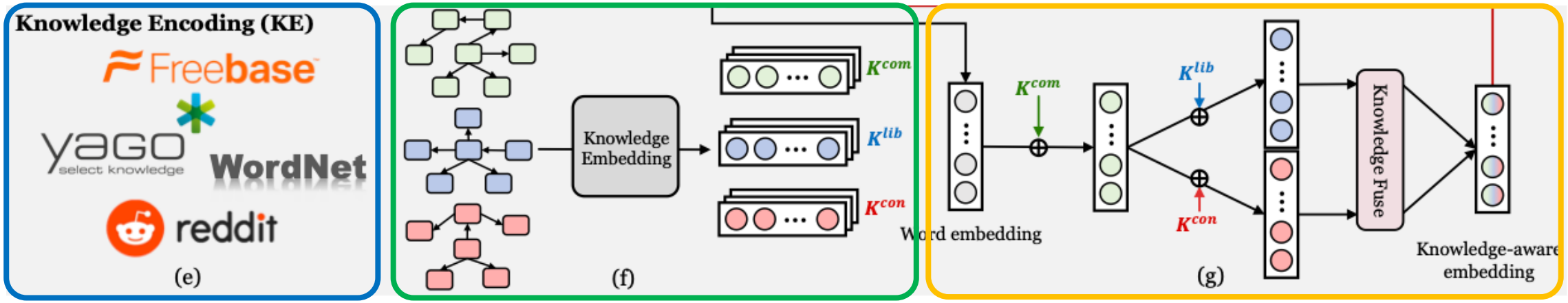
Residual connect & Average

HAN effectively captures not only ***the local and global context*** and ***the key idea*** of a news article



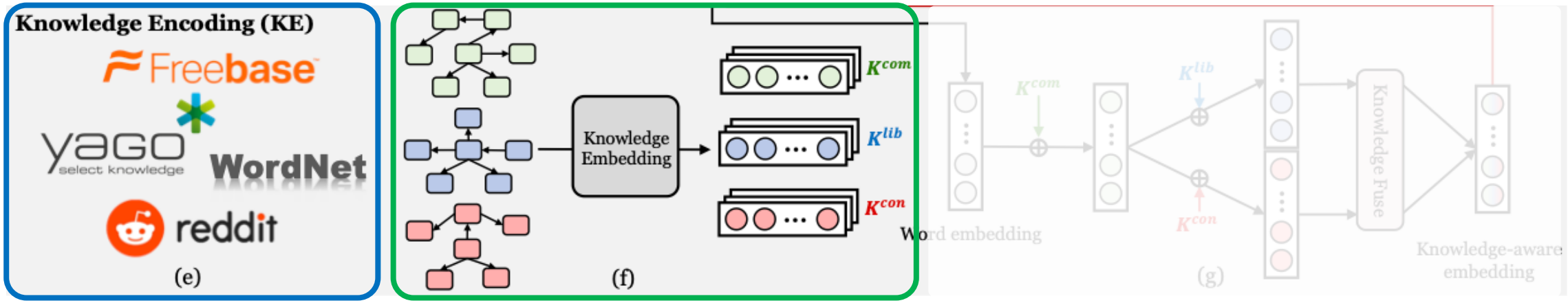
Knowledge Encoding (KE)

- To exploit knowledge for better interpretation of real-world entities (*explicit* factors)
  - Knowledge preparation
    - Common knowledge (YAGO)
    - Political knowledge (liberal and conservative)
  - Knowledge embedding
    - To learn entity embeddings for three different types of knowledge
  - Knowledge injection
    - To fuse three knowledge embeddings and inject the fused knowledge into the corresponding words



Knowledge Encoding (KE)

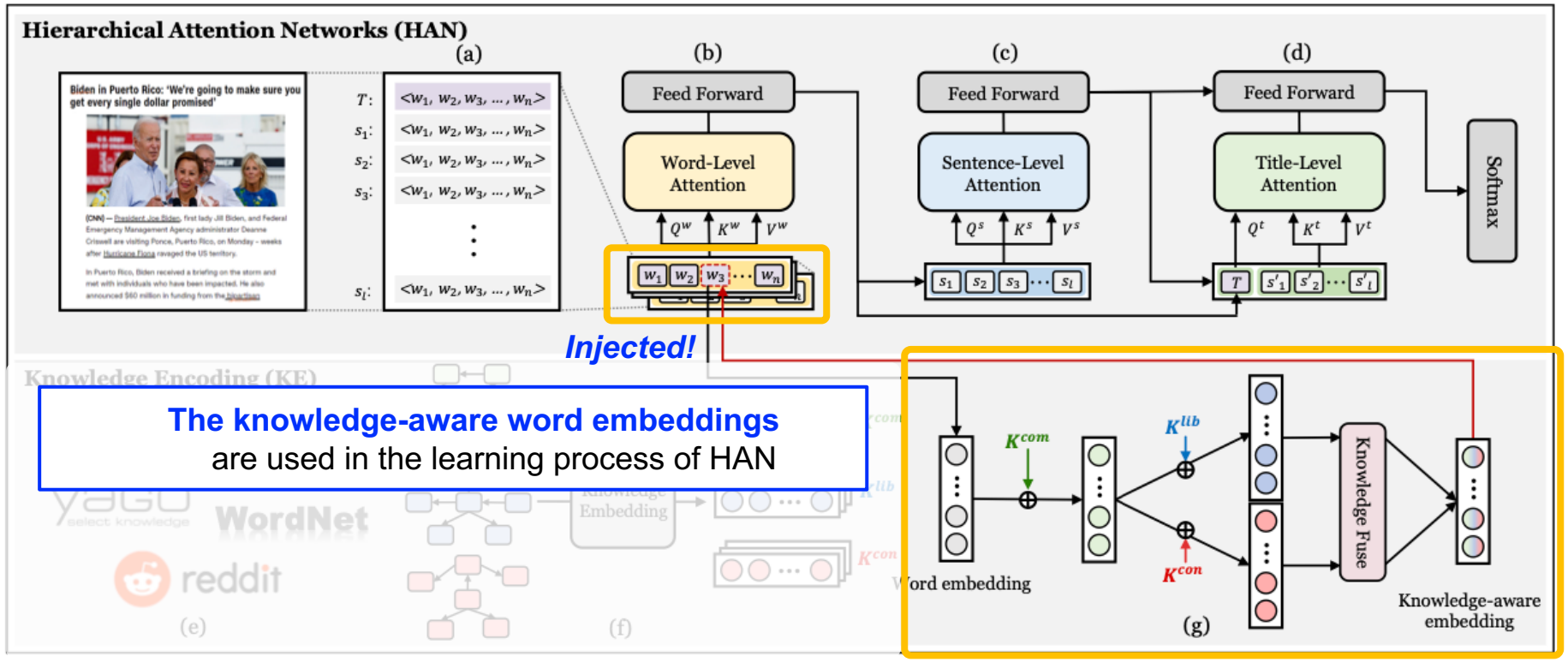
- **Knowledge preparation**
  - Common knowledge graph (YAGO)
    - To exploit *general knowledge* about real-world entities (e.g., person, city, organization, …)
  - Political knowledge graphs (KG-lib and KG-con)
    - To capture *subtle but important difference* in political knowledge for real-world entities
- **Knowledge embedding**
  - To learn the relations among entities in a news article
    - Methods: RotatE, ModE, HAKE



Knowledge Encoding (KE)

Knowledge injection

- How to incorporate the three different knowledge into the political stance prediction?
  - To fuse the three different knowledge embeddings
  - To inject the fused knowledge into the corresponding words in a news article



Roadmap

- **Background & Problem Definition**
  - Echo chamber effect
  - Political stance prediction problem
- **Motivation**
  - Limitations of existing works
  - User study
- **Proposed Method: KHAN**
  - Hierarchical attention networks (HAN)
  - Knowledge encoding (KE)
- **Evaluation and Results**
  - Experimental results
  - Conclusion

Evaluation Questions to Answer

- **EQ1: Model accuracy**
  - To what extent does KHAN improve existing methods in terms of the model accuracy in the political stance prediction?
- **EQ2: Training efficiency**
  - How fast does KHAN converge in terms of training time and epochs?
- **EQ3: Ablation study**
  - How effective are the key components of KHAN in terms of improving the model accuracy in political stance prediction?
- **EQ4: Sensitivity analysis**
  - How sensitive is the model accuracy of KHAN to the hyperparameters?

Experimental Setup

- **Political news datasets**
  - SemEval (2 classes)
    - 10-fold cross validation
  - AllSides-S (3 classes)
    - 3-fold cross validation
  - AllSides-L (5 classes)
    - # of training samples: # of valididation samples = 9:1
- **Seven baseline methods**
  - Five general language models
    - Word2Vec, GloVe, ELMo, BERT, RoBERTa
  - Two political stance prediction models
    - KGAP, KCD (state-of-the-art)

Table 4: 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

- Datasets and codes are available at: https://github.com/yy-ko/khan-www23
- Zhang et al., KCD: Knowledge Walks and Textual Cues Enhanced Political Perspective Detection in News Media. In Proceedings of the NAACL, 2022.

EQ1. Model Accuracy

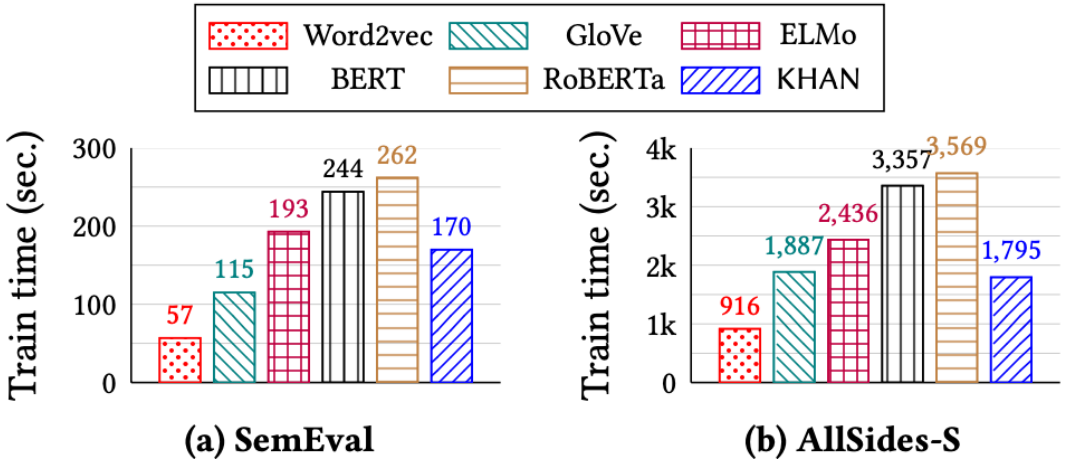
- ***KHAN consistently outperforms all baseline methods regardless of KGE methods***
  - Up to **5.92%** higher than the SOTA method in AllSides-S
  - The improvement of KHAN over all baselines are statistically significant (i.e., *p-value < 0.05*)

Table 5: Comparison of the model accuracy on three real-world datasets (The bold font indicates the best results).

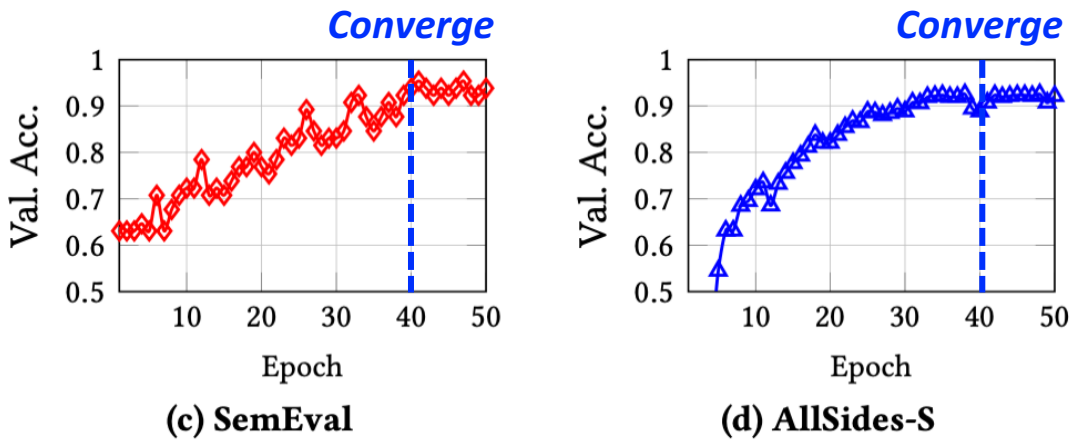
Method	Dataset		
	SemEval	**AllSides-S**	**AllSides-L**
Word2Vec [36]	0.7027	0.4858	0.4851
GloVe [41]	0.8071	0.7101	0.6354
ELMo [42]	0.8678	0.8197	0.7483
BERT [23]	0.8692	0.8246	0.7812
RoBERTa [34]	0.8708	0.8535	0.8222
KGAP [13]	0.8956	0.8602	N/A
KCD [55]	0.9087	0.8738	N/A
KHAN-RotatE	0.9426	0.9151	0.8584
KHAN-HAKE	0.9395	0.9216	0.8563
KHAN-ModE	**0.9521**	**0.9256**	**0.8617**

EQ2. Training Efficiency

- **Training time (for 50 epochs)**
  - KHAN finishes its training in shorter time than ELMo, BERT, and RoBERTa
    - While *achieving the highest accuracy* among all methods
- **Convergence rate w.r.t training epochs**
  - KHAN converges to high accuracy within only around 40 epochs



<Training time>



<Convergence rate with respect to training epochs>

EQ3. Ablation Study

- **Effectiveness of HAN**
  - Each layer of HAN is effective in capturing the implicit factors of a news article
    - KHAN-W (*local context only*) < KHAN-WS (*+ global context*) < KHAN-WST (*+ key idea*)
- **Effectiveness of KE**
  - KHAN-ALL always achieves the highest accuracy
    - External knowledge for real-world entities is important in predicting the political stance

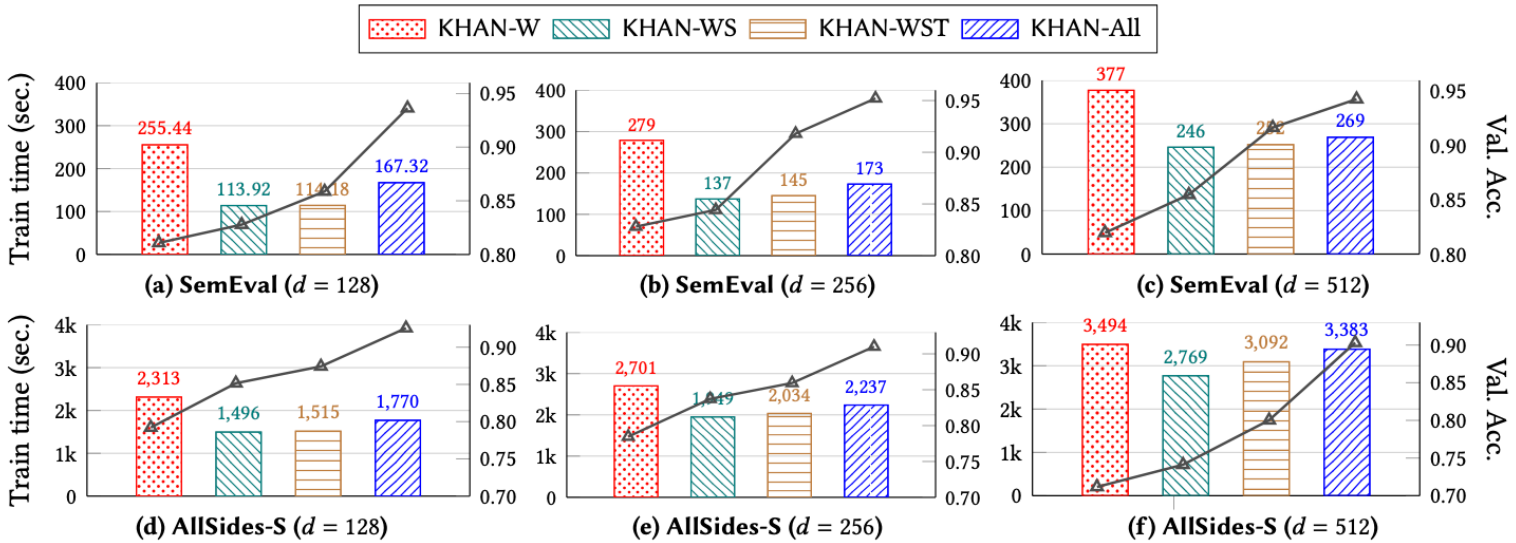
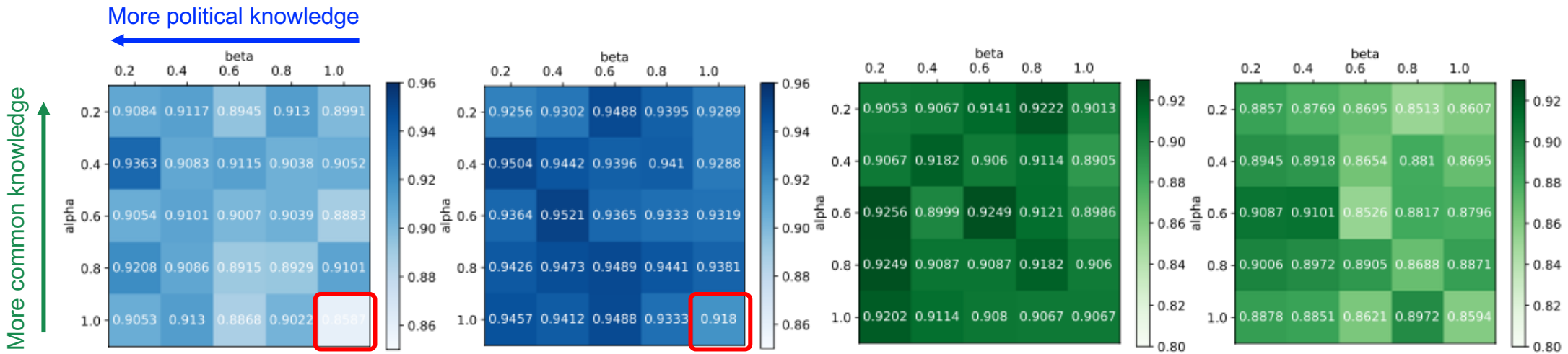


Figure 4: Effectiveness of the main components of KHAN in terms of the training time (bar) and model accuracy (line).

- **KHAN-W**: with only the word-level layer
- **KHAN-WS**: with the word and sentence layers
- **KHAN-WST**: with **HAN** (word, sentence, title)
- **KHAN-ALL**: with **HAN** and **KE** (proposed)

EQ4. Sensitivity Analysis

- **Impact of common (α) and political (β) knowledge**
  - KHAN with smaller β tends to achieve higher accuracy than KHAN with larger β
    - *Political knowledge is useful* in predicting the political stance
  - KHAN with less external knowledge tends to perform poorly
    - e.g., when $\alpha = \beta = 1$, KHAN shows the worst performance in SemEval



(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.

Conclusion

- **Observation**
  - Both implicit and explict factors are important in predicting the political stance of news articles
    - Implicit factors: context and tone
    - Explicit factors: real-world entities (e.g., keyword, person, and frequently used word)
- **Algorithm: KHAN**
  - **K**nowledge-aware **H**ierarchical **A**ttention **N**etworks
    - **HAN**: to capture the *implicit* factors via *3-level hierarchy* (i.e., word-, sentence-, and title-level)
    - **KE**: to incorporate *external knowledge* for real-world entities (*explicit* factors)
- **Evaluation**
  - Demonstrating the superiority of KHAN through extensive evaluations
    - **Accuracy**: KHAN achieves the highest prediction accuracy
    - **Efficiency**: KHAN converges within comparable training time/epochs
    - **Effectiveness**: each of the key components is effective in the political stance prediction

Contact: Yunyong Ko (Email: yyko@illinois.edu)
Code & Datasets: https://github.com/yy-ko/khan-www23