UOM-FJ at the NTCIR-15 SHINRA2020-ML Task

Hiyori Yoshikawa [1] Chunpeng Ma[1] Aili Shen [2] Qian Sun [2] Chenbang Huang [2]
Guillaume Pelat [3] Akiba Miura [1] Daniel Beck [2] Timothy Baldwin [2] Tomoya Iwakura [1]

[1]Fujitsu Laboratories Ltd., Japan [2]University of Melbourne, Australia [3]École Polytechnique, France

FUJITSU

Summary

- We participated in the **SHINRA2020-ML Task**: mapping Wikipedia entities into Extended Named Entity (ENE) categories.
- Our model was trained to capture multiple aspects of Wikipedia articles: **text, structured knowledge, images, page layout and the ENE class hierarchy**.
- Our system **ranked first in four languages**, achieving an F1-score of 82.73 on the English subtask.

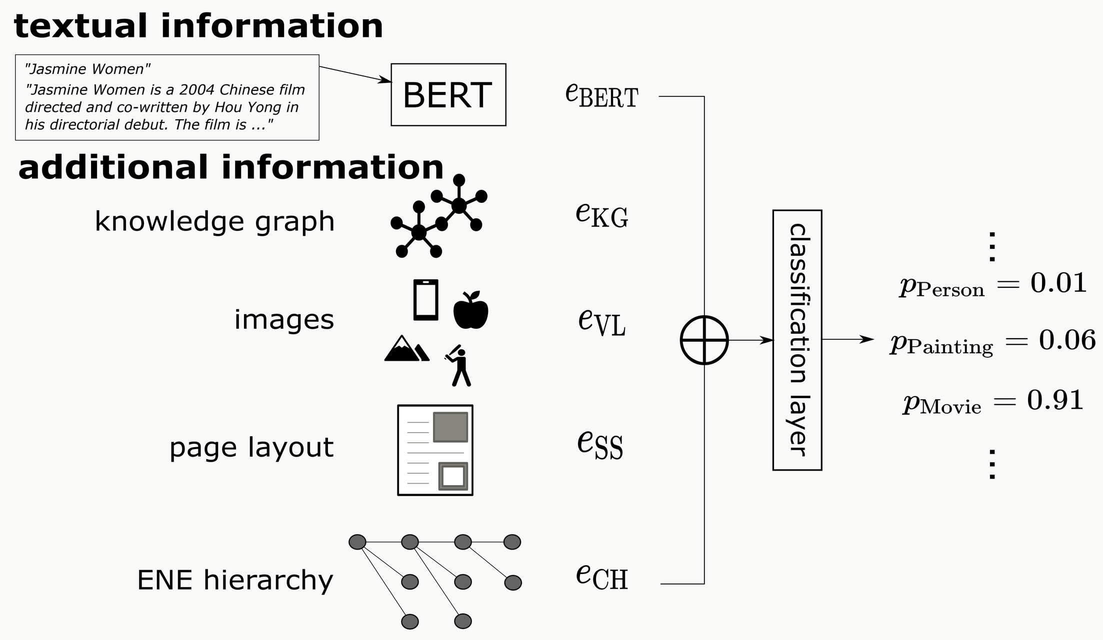
Our System Overview

Document representation: $e = \mathrm{FFNN}(e_{\mathrm{BERT}} \oplus e_{\mathrm{KG}} \oplus e_{\mathrm{VL}} \oplus e_{\mathrm{SS}} \oplus e_{\mathrm{CH}})$

- BERT encoder [1] for **pure textual information** (e_{BERT})
- A set of separately-trained document representations that encodes different aspects of a given document: **knowledge graph features (e_{KG}), text and images (e_{VL}), page screenshot layout (e_{SS}), and ENE class hierarchy (e_{CH})**

Label probability: $p(c, e) = \sigma(w_c^\top e + b_c)$

- BERT encoder is fine-tuned while training the classifier
- Other document representations are separately trained, and fixed during training of the final classifier



Multi-aspect Document Representations

Knowledge graph (e_{KG})

- We use pre-trained embeddings of a Wikidata graph [3]
- `wikibase_item` field of the Wikipedia dump is used to identify the Wikidata entity corresponding to a Wikipedia article
- 98.3% of English Wikipedia articles have corresponding Wikidata entries

Text–image representation (e_{VL})

- VL-BERT-based [4] encoding of text and images in a given Wikipedia article
- Multiple images are rescaled and concatenated to compose a single large image; individual images are treated as regions-of-interest (ROIs)
- The VL-BERT model is fine-tuned on the SHINRA2020-ML task and used to generate the text–image representation

Page screenshot layout (e_{SS})

- We obtain visual renderings of Wikipedia articles from a Wikipedia dump and generate their screenshots
- The screenshots are resized to a fixed size and fed into INCEPTION (Inception V3 with different convolution filters) [5] to obtain visual representations

ENE class hierarchy (e_{CH})

- We employ a hierarchy-aware global model (HiGAM) [6] to capture the label hierarchy and correlations between ENE categories
- Textual representation of a document is fed into a Hierarchy-GCN [2], where each node of the hierarchy graph represents an ENE category, and each directed edge represents either hierarchical relational information or correlation information, to obtain hierarchy-aware embeddings
- The encoder is trained with binary cross-entropy loss over the hierarchical label space

Results

Official evaluation results (English)

Submission name	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}_1$	Rank
`jointrep`	**81.77**	83.71	**82.73**	1
`jointrepPostprocess`	81.46	83.71	82.57	3
`jointrepUnionPostprocess`	80.66	84.80	82.68	2
Other teams best	79.65	**85.00**	82.23	

- `jointrep`: output the categories whose estimated probability exceeds the threshold $\theta = 0.5$
- `jointrepPostprocess`: `jointrep` + assign the label `CONCEPT` if the probability scores of all candidate categories are below the threshold
- `jointrepUnionPostprocess`: take the union of different predictions made by five models trained on different subsets of the training data, and then apply the post-processing step as in `jointrepPostprocess`

Official evaluation results (Other languages*, F1) * Non-English results are based on VL-BERT model

	ar	bg	ca	cs	da	de	es	fa	fi	fr	he	hi	hu	id
Ours	64.55	83.07	**79.82**	81.29	80.56	81.03	**81.39**	80.38	80.91	78.21	81.09	66.67	85.02	78.51
Others best	76.27	83.77	76.28	84.47	82.30	81.86	80.94	81.70	83.62	81.01	83.79	76.43	85.46	81.93

	it	ko	nl	no	pl	pt	ro	ru	th	tr	uk	vi	zh
Ours	**82.02**	82.51	81.64	78.79	84.52	80.87	80.83	82.90	65.02	84.85	81.61	77.06	78.58
Others best	81.92	83.67	83.29	80.53	84.53	83.23	84.60	84.08	81.26	86.50	83.12	80.34	81.25

Analysis

Ablation study (leaderboard results, `jointrep`, English)

	Full	Full $-e_{\mathrm{X}}$				$e_{\mathrm{BERT}} \oplus e_{\mathrm{X}}$			
		KG	VL	SS	CH	KG	VL	SS	CH
$\mathcal{F}_1$	75.7	75.2	74.8	76.0	75.1	71.0	73.9	70.9	74.5

Common error patterns

- Confusion between `CONCEPT` and other classes, e.g. `1.7.21:Doctrine_Method`
- Informatin mismatch: some information in the corresponding Japanese article is missing in the target article, and *vice versa*
- Page redirects to a different entity, e.g. Tailoring → Tailor

References

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, June 2019.

[2] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[3] A. Lerer, L. Wu, J. Shen, T. Lacroix, L. Wehrstedt, A. Bose, and A. Peysakhovich. PyTorch-BigGraph: A large-scale graph embedding system. In *Proceedings of the 2nd SysML Conference*, 2019.

[4] W. Su, X. Zhu, Y. Cao, B. Li, L. Lu, F. Wei, and J. Dai. Vl-bert: Pre-training of generic visual-linguistic representations. In *International Conference on Learning Representations*, 2020.

[5] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016.

[6] J. Zhou, C. Ma, D. Long, G. Xu, N. Ding, H. Zhang, P. Xie, and G. Liu. Hierarchy-aware global model for hierarchical text classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1106–1117, 2020.