Personalized Review Recommendation based on Implicit dimension mining

Bei Xu[1,2,†] **Yifan Xu**[1,†,*]
[1]School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
[2]Jiangsu Key Laboratory of Big Data Security & Intelligent Processing, Nanjing 210023, China
xubei@njupt.edu.cn
xyf-0209@foxmail.com

Abstract

Users usually browse product reviews before buying products from e-commerce websites. Lots of e-commerce websites can recommend reviews. However, existing research on review recommendation mainly focuses on the general usefulness of reviews and ignores personalized and implicit requirements. To address the issue, we propose a Large language model driven Personalized Review Recommendation model based on Implicit dimension mining (*PRR-LI*). The model mines implicit dimensions from reviews and requirements, and encodes them in the form of "text + dimension". The experiments show that our model significantly outperforms other state-of-the-art textual models on the Amazon-MRHP dataset, with some of the metrics outperforming the state-of-the-art multimodal models. And we prove that encoding "*text + dimension*" is better than encoding "*text*" and "*dimension*" separately in review recommendation.

1 Introduction

Online product reviews are referential because they reflect the experience of past users. Some studies (Ventre and Kolbe, 2020) have shown the impact of online reviews on new users' purchase intention. Therefore, recommending useful reviews is helpful for users as well as e-commerce websites.

Current review recommendation techniques focus on review helpfulness prediction, in which a key step is to extract features from reviews and user requirements. Most features are extracted from the textual content (Saumya et al., 2023), which mainly includes: lexical, textual, readability, and others (Hong et al., 2017; Qazi et al., 2016; Malik and Hussain, 2018). Other features include non-textual content (Ghose and Ipeirotis, 2011; Lee et al., 2018), product-related factors (Hu et al., 2014; Lee and Choeh, 2014), and reviewer-related factors (Krishnamoorthy, 2015; Korfiatis et al., 2012; Allahbakhsh et al., 2015). Previous review recommendation methods take the product attributes or user preferences that directly appear in reviews as features (Liu et al., 2005), such as appearance, size, price, or components of products. However, some implicit features are ignored. For example, in the review of a computer: "*My game runs very smoothly*", "*performance*" is implicit because "*performance*" does not appear in the review. And a requirement "*I want to buy a computer to run my 3D game*" also implicitly indicate a request for performance.

Semantic enhancement is an approach to enhance semantic information of data. Related studies mainly use knowledge graphs or external knowledge to extend input or enrich knowledge facts (Zhang et al., 2019; Bhatt et al., 2020; Lyu et al., 2023). But current semantic enhancement methods are hard to enhance reviews because reviews are often unprofessional and casual. They are also hard to mine the implicit features from requirements because of the lack of context.

We propose a Large language model driven Personalized Review Recommendation based on Implicit dimension mining (*PRR-LI*). The model only uses textual content of reviews and

† The authors have contributed equally to this work.

* Yifan Xu is the corresponding author.

Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers), pages 86–91
June 16-21, 2024 ©2024 Association for Computational Linguistics

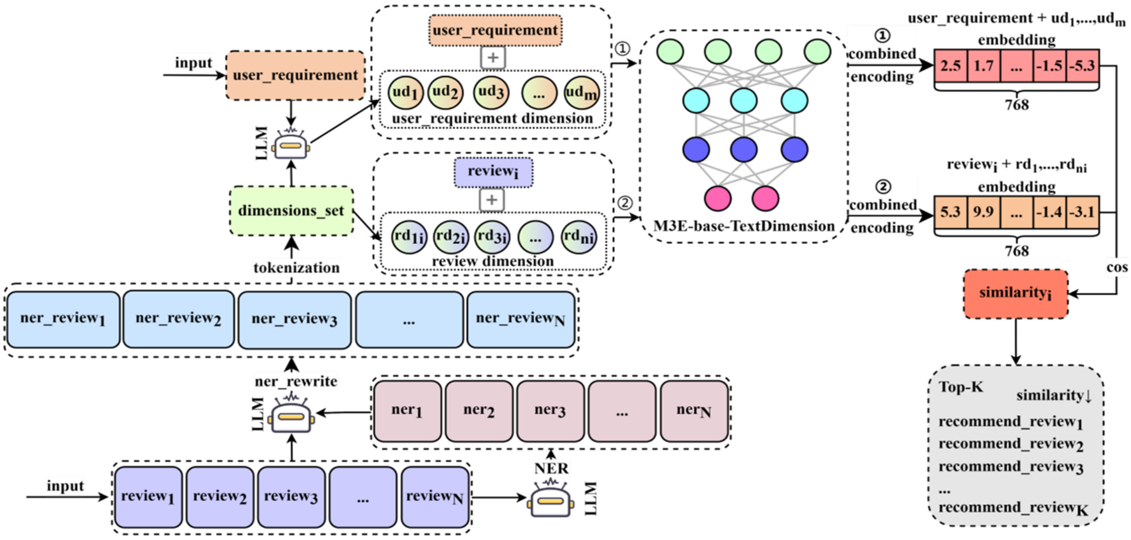

Figure 1: Framework of *PRR-LI*

requirements. The implicit dimensions of reviews and requirements are mined by using a large language model (*LLM*). We design prompts to guide the *LLM* to rewrite review text while keeping the original meaning, and then mine the implicit dimensions in reviews. At the same time, implicit dimensions are also mined from requirements. Finally, *PRR-LI* encodes enhanced reviews and requirements together by combined encoding. The experiments show that our model significantly outperforms other state-of-the-art text-only models, and some of the metrics exceed nearly 10% or are close to the performance of state-of-the-art multimodal models.

2 Review Dimension

We define review dimension as any entity or attribute expressed by a review that can reflect an explicit or implicit requirement. We classify the dimensions as explicit or implicit depending on whether the dimensions are directly mentioned in the review. Let R represent a review, the dimension D of R is denoted as $\{d_1, d_2, \ldots, d_n\}$. If R literally contains d_i, d_i is an explicit dimension of R. If R does not literally contain d_i, d_i is an implicit dimension of R. For example, "*gift*" is an explicit dimension in the review "*The packaging is perfect for a gift*". In the reviews "*The phone is easy to hold in one hand*" and "*This monitor is too big for my desk*", "*size*" does not appear directly, but is implied in the reviews. So "*size*" is an implicit dimension.

3 Model

The framework of *PRR-LI* is shown in Figure 1. The model takes reviews as input, acquires explicit and implicit entities by *LLM*, then inputs the reviews and the entities into the *LLM* again to obtain the rewritten reviews, and finally uses the tool (He and Choi, 2021) to tokenize the rewritten reviews and preserve words with parts of speech[1] *n*, *nz*, *nx* as review dimensions. The acquired review dimensions include both explicit and implicit dimensions expressed in the original reviews. We use the *API* version of the basic *LLM*, ChatGLM-Pro, with temperature and top_p set to 0.9 and 0.7 respectively. Then, the requirement and the acquired review dimensions are fed into the *LLM* to find the dimensions that meet the requirements. The prompts are shown in Table 1.

We design a text combined encoding module based on *M3E-Base*. *M3E-Base-TextDimension* is a version of *M3E-Base* after fine-tuning. The data "*review*" and "*review dimension*" are combined and then input into the module to be transformed into enhanced review embedding. The data "*requirement*" and "*requirement dimension*" are combined and input into the module to be

[1] https://hanlp.hankcs.com/docs/annotations/pos/pku.html

Name	Prompt templates
Entity recognize	*NER* Task: You need to perform fine-grained entity recognition on the text of a user's review of product. Please perform fine-grained entity recognition on the following reviews:\n{*content*}
Text rewrite	Text rewriting task, you need to rewrite the text of the user's review of the product.\n{*entity*}\nPlease rewrite the following reviews in conjunction with the entity recognition results, and output the rewritten text without any other explanatory notes.\n{*content*}
Check dimension	{*content*}\nIf there is any direct or indirect reference to <{*dimension*}> in the text above, please answer <yes> or <no>. No further explanation is required.
User requirement	I will give you a paragraph of text describing the user's requirements and a dimension word and ask you to judge whether the user is likely to be interested in this dimension.\nPlease make a judgement on the following, if the user is likely to be interested, answer 'yes', otherwise answer 'no', do not add any other irrelevant explanatory notes.\nText:\n{*content*}\nWords:\n{*dimension*}

Table 1: The prompt templates.

Type	Method	Clothing			Electronics			Home		
		M@5	*N@3*	*N@5*	*M@5*	*N@3*	*N@5*	*M@5*	*N@3*	*N@5*
Text-only	*BiMPN*	57.7	41.8	46.0	52.3	40.5	44.1	56.6	43.6	47.6
	EG-CNN	56.4	40.6	44.7	51.5	39.4	42.1	55.3	42.4	46.7
	Conv-KNRM	57.2	41.2	45.6	52.6	40.5	44.2	57.4	44.5	48.4
	PRHNet	58.3	42.2	46.5	52.4	40.1	43.9	57.1	44.3	48.1
Multimodal	*SSE-Cross*	65	56	59.1	53.7	43.8	47.2	60.8	51	54
	D&R Net	65.2	56.1	59.2	53.9	44.2	47.5	61.2	51.8	54.6
	MCR	67	**58.1**	61.1	56	**56.5**	49.7	63.2	**54.2**	57.3
Ours	*PRR-LI*	62.7	44.4	54.2	59.6	44.1	53.1	**66.6**	46.3	**57.9**
	PRR-LI_FT	**71.1**	51.5	**62.1**	**68.8**	54	**61.2**	64.6	50.1	57.1

Table 2: Results on the Amazon-MRHP dataset.

transformed into enhanced requirement embedding. Then we use cosine distance to calculate the semantic similarity between enhanced review embedding and enhanced requirement embedding. The model recommends the Top-N reviews in descending order.

4 Experiments

4.1 Dataset

We compare our model with others on the benchmark dataset *Amazon-MRHP* (Ni et al., 2019; Liu et al., 2021), which contains 87,492 reviews for clothing, 79,570 reviews for electronics, and 111,193 reviews for home. Under the premise of not violating relevant laws and regulations, as well as the website's robot exclusion protocol, we built a dataset JDDataset from the JingDong website for other experiments. JDDataset is available at https://www.modelscope.cn/datasets/Jerry0/JDDataset. It contains 437,646 reviews, of which 90,000 were used for training, 2,000 for validation, and 880 for testing.

4.2 Experimental setups

We use the v2.1 native version of HanLP (He and Choi, 2021). The stop words contain both Chinese and English. The Adam optimizer is chosen for fine-tuning, batch_size is 16, the learning rate is $5e^{-5}$, weight_decay is $1e^{-3}$, and epoch is 4.

We use the metrics commonly used in the recommendation: (1) Recall@N, denoted as *R@N*; (2) MAP@N, denoted as *M@N*; (3) NDCG@N (Järvelin and Kekäläinen, 2017), denoted as *N@N*.

We compare our model with two types of state-of-the-art review recommendation models. One is the models that only use textual content: BiMPN (Wang et al., 2017), EG-CNN (Chen et al., 2018), Conv-KNRM (Dai et al., 2018), and PRHNet (Fan et al., 2019). The other is the multimodal models: SSE-Cross (Abavisani et al., 2020), D&R Net (Xu et al., 2020), and MCR (Liu et al., 2021).

PRR-LI_FT is a version of *PRR-LI* after fine-tuning. The two models are text-only models.

4.3 Results on *Amazon-MRHP*

We conduct comparative experiments on the benchmark dataset *Amazon-MRHP*. The results are shown in Table 2. *PRR-LI_FT* and *PRR-LI*

		R@5	*R@10*	*R@15*	*M@5*	*M@10*	*M@15*	*N@5*	*N@10*	*N@15*
M3E-base	separated	72	66.44	72.83	63.57	53.56	49.97	88.49	87.42	87.11
	combined	76	74	74.92	68.67	62.8	59.22	93.48	92.33	91.66
M3E-base-TD	separated	68	71	82.9	69.83	70.93	70.69	79.6	81.82	82.85
	combined	**96**	**93**	**89.9**	**98.38**	**97.09**	**95.23**	**99.39**	**98.95**	**98.46**

Table 3: Results on separated and combined encoding. *M3E-base-TD* refers to *M3E-base-TextDimension*.

significantly outperform the text-only models. After fine-tuning, *PRR-LI_FT* continues to improve significantly on most metrics because *PRR-LI_FT* can encode the type of data "text + dimension" better than *PRR-LI*. And *PRR-LI_FT* is better than the multimodal models on *MAP@5*.

The performance of *PRR-LI* and *PRR-LI_FT* is not as good as the multimodal models in *N@3* and *N@5* for home data, while the performance of *PRR-LI* and *PRR-LI_FT* is close to the multimodal models for clothing data. One reason is that the images of home and clothing products help reflect the requirements more visually. For electronics data, *PRR-LI* and *PRR-LI_FT* outperform the multimodal model by almost 10% in both *MAP@5* and *N@5*. One reason is that the images of electronic products do not reflect the requirements as much as the images of home and clothing.

4.4 Ablation experiment

Figure 2 shows that adding different parts of *PRR-LI* can effectively optimize recommendation. The dataset is JDDataset. The performance decreases significantly without *rewrite*, *review dimension,* or *require dimension*. And *rewrite with NER* is better than *rewrite*.

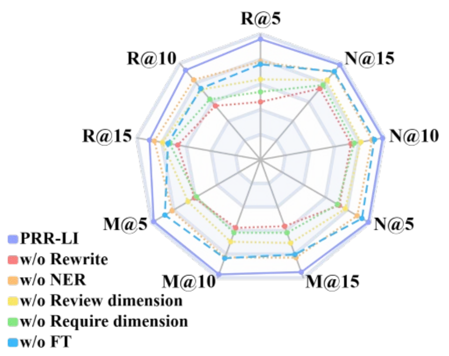

Figure 2: Ablation experiment

We further test other *LLMs*' abilities to rewrite with NER as shown in Table 4. "*Rewrite*" and "*NER_rewrite*" respectively means rewrite text without and with *NER*. The values are average *proffer*. *Proffer* reflects the implicit dimension mining effect, and refers to the proportion of acquired dimensions to the total dimensions as shown in equation 1,

LLMs	Rewrite	NER_rewrite
ChatGLM2-6B v1.0.12	35.5	**37.1**
Qwen-7B-Chat v1.1.5	**40.7**	34.7
Baichuan2-7B-Chat v1.0.4	**39.7**	31.9
internlm-chat-7b v1.0.1	**13.3**	3.5
Llama2-Chinese-7b-Chat-ms v1.0.0	20.3	**23.8**
ChatGLM-Pro	29.2	**33.6**

Table 4: Rewrite with NER. The *LLMs* with parameters 6b and 7b are from https://www.modelscope.cn.

$$Proffer = \frac{id}{id+ed} \quad (1),$$

Where *id* is the number of implicit dimensions and *ed* is the number of explicit dimensions.

We can see that some *LLMs* are not suitable for rewriting with *NER*.

4.5 Experiments on encoding models

We test other encoding models in *PRR-LI* on JDDataset as shown in Figure 3. "*dimension*" refers to vectorizing the text using the dimensions of the review. *M3E-base* and *text2vec-bge-large* series are from https://huggingface.co. We can see that the *M3E-base-TextDimension* reaches the best. The results on "*dimension*" show that ignoring the text content weakens the ranking and the recall.

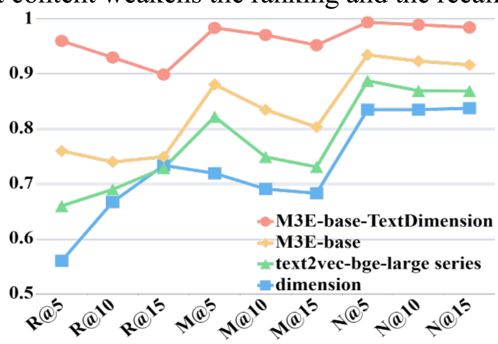

Figure 3: Results on encoding models

4.6 Experiments on the encoding method

We test separated encoding, which encodes text and dimension separately, and combined encoding, which encodes text and dimension in the form of "text + dimension". Table 3 shows that the combined encoding achieves better results on both *M3E* models, and *M3E-base* can handle the type of "text + dimension" data better after fine-tuning.

5 Conclusion

PRR-LI and the fine-tuned version, *PRR-LI_FT*, significantly outperform the text-only review recommendation models, and even outperform the multimodal models in some metrics. This reveals that review text may contain a wealth of implicitly semantic information that has yet to be fully exploited. Furthermore, the results on electronics are better than those on clothing and home products. This suggests that review text can reveal more information about objects that lack intuitive visual information, compared to objects that possess abundant visual representations.

We also demonstrate that, in review recommendation, encoding "*text + dimension*" together is better than encoding "*text*" and "*dimension*" separately. It indicates that "*text + dimension*" may serve as a more effective input for language models compared to plain text.

In conclusion, our model offers a method to extract implicit dimension from review text and integrate them with the text itself. Our model has the potential to be utilized in other applications that involve processing the semantics of short text.

6 Limitations

Although this model achieves competitive performance, its efficiency has a bottleneck caused by acquiring requirement dimensions one by one. It is crucial to find a way to acquire all requirement dimensions at once to improve efficiency and expand the model's applicability. And the performance of our model on long text has not yet been tested and validated.

In addition, considering that *PRR-LI* and *PRR-LI_FT* do not use data other than text, it is very likely that the models' performance can be further improved by using multimodal data, such as images.

References

Mahdi Abavisani, Liwei Wu, Shengli Hu, Joel Tetreault, and Alejandro Jaimes. 2020. Multimodal Categorization of Crisis Events in Social Media. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14667–14677, Seattle, WA, USA. IEEE.

Mohammad Allahbakhsh, Aleksandar Ignjatovic, Hamid Reza Motahari-Nezhad, and Boualem Benatallah. 2015. Robust evaluation of products and reviewers in social rating systems. *World Wide Web*, 18(1):73–109.

Shreyansh Bhatt, Amit Sheth, Valerie Shalin, and Jinjin Zhao. 2020. Knowledge Graph Semantic Enhancement of Input Data for Improving AI. *IEEE Internet Computing*, 24(2):66–72.

Cen Chen, Yinfei Yang, Jun Zhou, Xiaolong Li, and Forrest Sheng Bao. 2018. Cross-Domain Review Helpfulness Prediction Based on Convolutional Neural Networks with Auxiliary Domain Discriminators. In Marilyn Walker, Heng Ji, and Amanda Stent, editors, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 602–607, New Orleans, Louisiana. Association for Computational Linguistics.

Zhuyun Dai, Chenyan Xiong, Jamie Callan, and Zhiyuan Liu. 2018. Convolutional Neural Networks for Soft-Matching N-Grams in Ad-hoc Search. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 126–134, New York, NY, USA. Association for Computing Machinery.

Miao Fan, Chao Feng, Lin Guo, Mingming Sun, and Ping Li. 2019. Product-Aware Helpfulness Prediction of Online Reviews. In *The World Wide Web Conference*, pages 2715–2721, New York, NY, USA. Association for Computing Machinery.

Anindya Ghose and Panagiotis G. Ipeirotis. 2011. Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23(10):1498–1512.

Han He and Jinho D. Choi. 2021. The Stem Cell Hypothesis: Dilemma behind Multi-Task Learning with Transformer Encoders. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 5555–5577, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Hong Hong, Di Xu, G. Alan Wang, and Weiguo Fan. 2017. Understanding the determinants of online review helpfulness: A meta-analytic investigation. *Decision Support Systems*, 102:1–11.

Nan Hu, Noi Sian Koh, and Srinivas K. Reddy. 2014. Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. *Decision Support Systems*, 57:42–53.

Kalervo Järvelin and Jaana Kekäläinen. 2017. IR evaluation methods for retrieving highly relevant documents. *ACM SIGIR Forum*, 51(2):243–250.

Nikolaos Korfiatis, Elena García-Bariocanal, and Salvador Sánchez-Alonso. 2012. Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, 11(3):205–217.

Srikumar Krishnamoorthy. 2015. Linguistic features for review helpfulness prediction. *Expert Systems with Applications*, 42(7):3751–3759.

Pei-Ju Lee, Ya-Han Hu, and Kuan-Ting Lu. 2018. Assessing the helpfulness of online hotel reviews: A classification-based approach. *Telematics and Informatics*, 35(2):436–445.

Sangjae Lee and Joon Yeon Choeh. 2014. Predicting the helpfulness of online reviews using multilayer perceptron neural networks. *Expert Systems with Applications*, 41(6):3041–3046.

Bing Liu, Minqing Hu, and Junsheng Cheng. 2005. Opinion observer: analyzing and comparing opinions on the Web. In *Proceedings of the 14th International Conference on World Wide Web*, pages 342–351, New York, NY, USA. Association for Computing Machinery.

Junhao Liu, Zhen Hai, Min Yang, and Lidong Bing. 2021. Multi-perspective Coherent Reasoning for Helpfulness Prediction of Multimodal Reviews. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 5927–5936, Online. Association for Computational Linguistics.

Ziyu Lyu, Yue Wu, Junjie Lai, Min Yang, Chengming Li, and Wei Zhou. 2023. Knowledge Enhanced Graph Neural Networks for Explainable Recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 35(5):4954–4968.

Msi Malik and Ayyaz Hussain. 2018. An analysis of review content and reviewer variables that contribute to review helpfulness. *Information Processing & Management*, 54(1):88–104.

Jianmo Ni, Jiacheng Li, and Julian McAuley. 2019. Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 188–197, Hong Kong, China. Association for Computational Linguistics.

Aika Qazi, Karim Bux Shah Syed, Ram Gopal Raj, Erik Cambria, Muhammad Tahir, and Daniyal Alghazzawi. 2016. A concept-level approach to the analysis of online review helpfulness. *Computers in Human Behavior*, 58:75–81.

Sunil Saumya, Pradeep Kumar Roy, and Jyoti Prakash Singh. 2023. Review helpfulness prediction on e-commerce websites: A comprehensive survey. *Engineering Applications of Artificial Intelligence*, 126:107075.

Ivan Ventre and Diana Kolbe. 2020. The Impact of Perceived Usefulness of Online Reviews, Trust and Perceived Risk on Online Purchase Intention in Emerging Markets: A Mexican Perspective. *Journal of International Consumer Marketing*, 32(4):287–299.

Zhiguo Wang, Wael Hamza, and Radu Florian. 2017. Bilateral Multi-Perspective Matching for Natural Language Sentences. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pages 4144–4150, Melbourne, Australia. International Joint Conferences on Artificial Intelligence Organization.

Nan Xu, Zhixiong Zeng, and Wenji Mao. 2020. Reasoning with Multimodal Sarcastic Tweets via Modeling Cross-Modality Contrast and Semantic Association. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3777–3786, Online. Association for Computational Linguistics.

Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu. 2019. ERNIE: Enhanced Language Representation with Informative Entities. In Anna Korhonen, David Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1441–1451, Florence, Italy. Association for Computational Linguistics.