A Multi-label Multi-hop Relation Detection Model based on Relation-aware Sequence Generation

Linhai Zhang[1], Deyu Zhou[1]*, Chao Lin[1], Yulan He[2]

[1]School of Computer Science and Engineering, Key Laboratory of Computer Network and Information Integration, Ministry of Education, Southeast University, China
[2]Department of Computer Science, University of Warwick, UK
{lzhang472,d.zhou,c.lin}@seu.edu.cn,
yulan.he@warwick.ac.uk

Abstract

Multi-hop relation detection in Knowledge Base Question Answering (KBQA) aims at retrieving the relation path starting from the topic entity to the answer node based on a given question, where the relation path may comprise multiple relations. Most of the existing methods treat it as a single-label learning problem while ignoring the fact that for some complex questions, there exist multiple correct relation paths in knowledge bases. Therefore, in this paper, multi-hop relation detection is considered as a multi-label learning problem. However, performing multi-label multi-hop relation detection is challenging since the numbers of both the labels and the hops are unknown. To tackle this challenge, multi-label multi-hop relation detection is formulated as a sequence generation task. A relation-aware sequence relation generation model is proposed to solve the problem in an end-to-end manner. Experimental results show the effectiveness of the proposed method for relation detection and KBQA.

1 Introduction

With the development of Knowledge Bases (KBs) such as DBpedia, Freebase, and WikiData, Knowledge Base Question Answering (KBQA) system (Berant et al., 2013; Bordes et al., 2015; Yin et al., 2016; Hao et al., 2018) is attracting more and more attention. The KBQA system often contains two core components: (1) entity linking, which identifies the topic entity mentioned in the question; (2) relation detection, which detects the relation paths starting from the topic entity to the answer node.

Relation detection in KBQA can be categorized into singe-relation (one-hop) detection and multi-relation (multi-hop) detection. Most existing single-relation detection methods (Yin et al., 2016; Yu et al., 2017; Lukovnikov et al., 2017; Yu et al.,
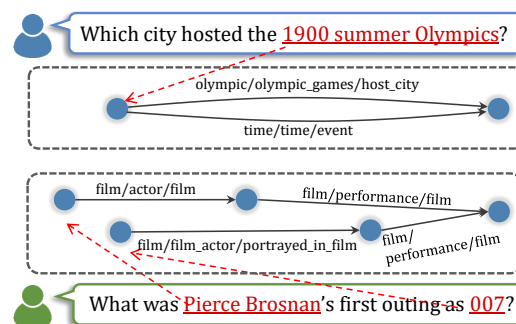

Figure 1: Examples of ground-truth relation paths corresponding to the given questions.

2018) rely on measuring the semantic similarity of questions and candidate relations. He and Golub (2016) proposed an encoder-decoder based generative framework for single-relation extraction. For multi-relation detection, some approaches (Yih et al., 2015; Yu et al., 2017, 2018) proposed to tackle two or three-relation detection by applying some constraint which makes the number of hops fixed. Xiong et al. (2017) and Das et al. (2017) modeled the relation reasoning problem as a Markov decision process. Chen et al. (2019) exploited a transition-based search framework to select the relation dynamically. Recently, some researchers attempt to model prediction uncertainties in the simple question answering task with Bayesian neural network (Zhang et al., 2021). Generally, most of existing methods focus on detecting one optimal relation path, considering the task a single-label learning problem.

However, for some question, there may exist multiple relation paths to the correct answer. For example, as shown in the upper part of Figure 1, there are two distinct relation paths ***time/time/event*** and ***olympic/olymipic_games/host_city*** with the same meaning, making the instance multi-label. Moreover, as shown in the lower part of Figure 1, there are two relation paths starting from different topic entities ***Pierce Brosnan*** and ***007***. A robust

---

*Corresponding author.

KBQA system should be able to infer the final answer based on multiple relation paths. We therefore consider multi-label multi-hop relation detection in this paper.

Nevertheless, it is challenging to perform multi-label multi-hop relation detection since both the number of relation paths and the number of hops in each relation path are unknown. To deal with such a challenge, in this paper, we formulate it as a sequence generation task in the following form $\{r_1^1, \ldots, r_{n_1}^1, \texttt{[SEP]}, r_1^2, \ldots, r_{n_2}^2, \texttt{[SEP]}, r_1^m, \ldots, r_{n_m}^m, \texttt{[END]}\}$, where r_i^j denotes the i-th relation in j-th path, the comma splits the relation chains and the $\texttt{[SEP]}$ indicates the division of different relation paths. A relation-aware sequence generation model (RSGM) is proposed to learn the sequence generation task end-to-end, without the need of knowing the number of labels and hops beforehand. In specific, a pre-trained Bidirectional Encoder Representations from Transformers (BERT) model is employed as the encoder of RSGM while a Gated Recurrent Unit (GRU) with relation-aware attention is design as the decoder to incorporating the semantic information of the relations. Moreover, a constraint-learning strategy is proposed to mitigate the exposure bias and label repetition problem in sequence generation.

The main contributions of this paper are:

- An end-to-end relation-aware sequence generation model, RSGM, is proposed to deal with the multi-label multi-hop relation detection problem.

- Experimental results show the effectiveness of proposed model both on relation detection and KBQA end-task.

2 Methodology

2.1 Problem Setting

Let $G = (\mathcal{S}, \mathcal{R}, \mathcal{O})$ be the KB, where $\mathcal{S}$ represents the set of subject entities, $\mathcal{O}$ represents the set of object entities, and $\mathcal{R}$ represents the set of relations between the subject entities and the object entities. Assume a set of questions $Q = \{q_1, q_2, \ldots q_{|Q|}\}$, where each instance $q_i = \{w_1, w_2, \ldots w_{N_i}\}$ has N_i words.

Given question q and knowledge base G, traditional methods treat multi-hop relation detection as a single-label learning problem, aiming at find the optimal relation path p, where

$$p = (r_1, r_2, \ldots, r_n | r_i \in \mathcal{R}) \qquad (1)$$

In this paper, multi-label multi-hop relation detection is considered. Therefore, the objective is to find a set of relation paths P based on question q and knowledge base G, where

$$\begin{aligned} P &= \{p_1, p_2, \ldots, p_m\} \\ p_j &= (r_{1j}, r_{2j}, \ldots, r_{nj} | r_{ij} \in \mathcal{R}) \end{aligned} \qquad (2)$$

As both the number of labels and the number of hops are unknown, the task is formulated as a sequence generation problem. The objective is to generate a token sequence Y given question q and knowledge base G, where

$$\begin{aligned} Y &= (y_1, y_2, \ldots, y_M) \\ y_i &\in \mathcal{R} \cup \texttt{[SEP]} \cup \texttt{[END]} \end{aligned} \qquad (3)$$

where the $\texttt{[SEP]}$ indicates the division of different relation paths and $\texttt{[END]}$ indicates the end of the sequence.

2.2 The Proposed Model

The overall architecture of the proposed relation-aware sequence generation model (RSGM) is presented in Figure 2. It consists of two components which will be discussed in more details: (1) the question encoder, where the question is transformed into a context-aware representation using BERT; (2) the relation decoder, where a GRU is employed to generate the relation sequentially with a relation-aware attention.

Question Encoder To utilize the abundant semantic information of the large pre-trained model, BERT is employed as the encoder, which takes the question as the input and learns the context-aware representation for each token.

Specifically, for the given question q_i, the token $\texttt{[CLS]}$ is inserted as the first token to obtain the representation of the whole question, i.e.,

$$q_i = \{\texttt{[CLS]}, w_1, w_2, \ldots w_{N_i}\} \qquad (4)$$

The representation of the word w_i is computed as the sum of three embeddings: token embedding h_i^{tok}, segmentation embedding h_i^{seg}, and the position embedding h_i^{pos}, which is denoted as h_i^w,

$$h_i^w = h_i^{tok} + h_i^{seg} + h_i^{pos} \qquad (5)$$

As a result, a list of token embeddings $\mathbf{W}_0 = \{h^q, h_1^w, h_2^w \ldots h_N^w\}$ are obtained and then fed into a series of L pre-trained transformer blocks,

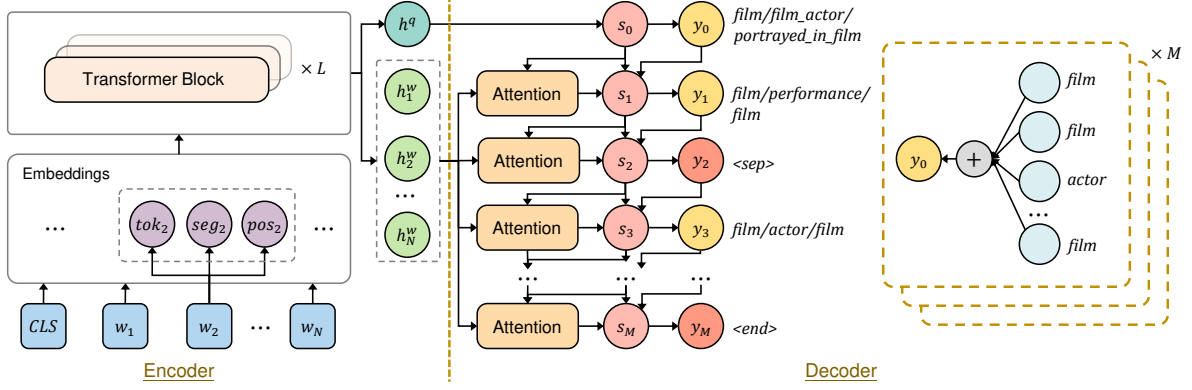$$\mathbf{W}_i = \text{Transformer}(\mathbf{W}_{i-1}), i = 1, 2, \cdots, L \quad (6)$$

Figure 2: The architecture of the proposed RSGM model. Each relation can be split into multiple word tokens and its representations can be obtained by adding such word token embeddings to mitigate the zero-shot learning problem.

Relation Decoder To generate relation sequentially, a GRU is employed. The prediction y_i at time i is affected by three factors, the hidden state of time $i-1$, the token predicted at time $i-1$ and the relation-aware question representation, formally,

$$
\begin{aligned}
s_i &= \text{GRU}(s_{i-1}, [h_{y_{i-1}}; c_i]) \\
y_i &= \text{softmax}(\tanh(\mathbf{W}_o s_i + \mathbf{b}_o))
\end{aligned}
\tag{7}
$$

where s_{i-1} denotes the hidden state of time $i-1$, $h_{y_{i-1}}$ denotes the embedding of token y_{i-1} predicted at time $i-1$, c_i denotes the relation-aware question representation.

Assuming the relation r consisting of N_r words, its embedding is defined as the sum of its word embeddings to encode its semantic information:

$$
h_r = \sum_{k=1}^{N_r} e_k
\tag{8}
$$

The token `[SEP]` is initialized randomly and updated during training.

The relation-aware question representation c_i is calculated by taking the word embedding h_i^w as the key and the value, and the hidden state s_{i-1} as the query:

$$
\begin{aligned}
c_i &= \sum_{j=1}^{N} \alpha_{ij} h_j^w \\
\alpha_{ij} &= \frac{exp(e_{ij})}{\Sigma_{k=1}^{N} exp(e_{ik})} \\
e_{ik} &= \mathbf{v}_a^T tanh(\mathbf{W}_a[s_{i-1}; h_j^w])
\end{aligned}
\tag{9}
$$

2.3 Training and Inference

The proposed model is trained under regular sequence-to-sequence loss by maximizing the likelihood of the ground-truth token sequence. At the training stage, to bridge the gap between training and inference, the scheduled sampling policy (Bengio et al., 2015) is employed, which exploits part of the ground-truth to guide model learning. At the testing stage, the beam search optimization approach (Wiseman and Rush, 2016) is used to mitigate the exposure bias problem. Additionally, to avoid the repetition problem, a constraint mechanism is added. Unlike text generation or machine translation tasks, the multiple paths to the same question are usually mutually exclusive. Therefore, when a relation path is generated, an infinite penalty is added for that relation path, in order to avoid it being generated again.

3 Experiments

We conduct experiments on a large KBQA benchmark dataset FreebaseQA (Jiang et al., 2019) to evaluate the effectiveness of the proposed RSGM model.

3.1 Dataset

FreebaseQA (Jiang et al., 2019) is a novel KBQA dataset generated by matching trivia-type question-answer pairs with facts existed in FreeBase. In particular, for each question in the dataset, there may often exists multiple multi-hop relation paths that can give rise to the correct answer. Here multi-hop means it should takes multiple hops in the knowledge base to reach the correct answer node. Compared with the existing well-known KBQA datasets SimpleQuestions (Bordes et al., 2015) and WebQuestion (Berant et al., 2013), it has the following characteristics: (1) for the give question, it provides multiple annotated relation paths to the

Methods	Presion($\uparrow$)	Recall($\uparrow$)	Micro F1($\uparrow$)	HL($\times 10^{-4}$)($\downarrow$)
CNN-multichannel (Kim, 2014)	0.5158	0.3952	0.4475	1.4285
MLKNN (Zhang and Zhou, 2014)	0.5327	0.3287	0.4066	1.4049
HAN (Yang et al., 2016)	0.4965	0.4254	0.4582	1.4728
SGM (Yang et al., 2018)	0.5039	0.3976	0.4445	1.4549
SGM-BERT	0.5992	0.4372	0.5056	1.2437
RSGM	**0.6795**	**0.5285**	**0.5945**	**1.0552**

Table 1: Performance comparision of the proposed approach with other approaches on FreebaseQA test set. "HL" represents the metric of Hamming Loss.

Dateset	FreebaseQA
Train	20358
Dev	3994
Test	3996

Table 2: Question numbers of the subsets of FreebaseQA dataset

correct answer as ground-truth. (2) the linguistic structure of the question is more sophisticated. (3) more training instances is provided enabling effectively training for neural networks. The detailed statistics are shown in Table 2. The FreebaseQA dataset is publicly available[1]. Since FreebaseQA is the only KBQA dataset that is annotated with multiple relations, we mainly conduct our experiments on this dataset.

3.2 Baselines

To our best knowledge, there is no other KBQA method consider multi-label multi-hop relation detection, so to demonstrate the effectiveness of the proposed model, we mainly include the following baselines and modify them to perform a multi-label prediction task:

- **CNN-multichannel** (Kim, 2014): multiple filters are employed to extract features of sentences and a full connected layer with the sigmoid function is utilized to obtain the probability of each label.

- **MLKNN** (Zhang and Zhou, 2014): the maximum a posterior (MAP) rule is exploited to make prediction by reasoning with the labeling information implied in the k-nearest neighbors, which ignores utilizing label correlations.

---

[1] http://github.com/infinitecold/FreebaseQA

- **HAN** (Yang et al., 2016): a hierarchical attention network is employed to obtain sentence representations and then generate document representations based on sentence representations.

- **SGM** (Yang et al., 2018): a novel sequence-to-sequence structure with global embedding is proposed to capture the correlations between labels.

- **SGM-BERT**: a variant of SGM by replacing the encoder of SGM with BERT.

3.3 Evaluation Metrics

To evaluate the performance of different approaches, several evaluation metrics are employed including Precision, Recall, Micro-F1 score and Hamming Loss as suggested in (Zhang and Zhou, 2007).

3.4 Model Setup

The uncased BERT$_{base}$ is employed as text encoder, with the parameters fine-tuned during training. For decoder GRU, the hidden state dimension is set to 128 and beam size is 5. The whole model is trained by the Adam optimizer (Kingma and Ba, 2014) with a learning rate of 1e-4 and a dropout rate of 0.3. The number of epochs is 10 and the mini-batch size of the input is set at 20. The parameters are chosen based on the evaluation results from dev subset.

3.5 Relation Detection Results

Experimental results on the FreebaseQA benchmark are listed in Table 1. It can be observed that: (1) SGM-BERT outperforms SGM, demonstrating the effectiveness of BERT encoder; (2) The proposed RSGM outperforms SGM by large margin on all metrics. The reasons can be summarized as followings: (1) RSGM employs a relation-aware

attention mechanism, which provide more informative sentence representation compared with vanilla self-attention; (2) RSGM formulated multi-label multi-hop relation detection as a sequence generation task, which gets a much smaller search space compared with other methods that perform multi-label perdiction task.

3.6 KBQA End-Task Results

To investigate the effectiveness of the proposed relation detection method for the KBQA end-task, we perform entity linking and retrieve the final answer with the relation path detected by RSGM.

K	Recall
50	0.80
100	0.88
200	0.94

Table 3: Entity Linking Results

Table 3 gives the entity linking results with different sizes of candidate entities, where recall is calculated as the fraction of the questions for which the candidate topic entities collection includes the correct subject entity.

Methods	Accuracy%
FOFE-net (Jiang et al., 2019)	37.0
RSGM-50	36.1
RSGM-100	**38.1**
RSGM-200	38.0

Table 4: KBQA Results on FreeBaseQA test set. RSGM-x represent top x of candidate entities are used from entity linking.

The experimental results of KBQA end-task are shown in Table 4. The FOFE-net (Jiang et al., 2019) is a pipeline KBQA system built based on FOFE-net (Xu et al., 2017) , which achieves outstanding results on SimpleQuestions and WebQSP datasets. The RSGM result is obtained by performing entity linking and relation detection with proposed model. Based on the multiple relation paths generated by RSGM, a majority vote strategy is employed to get the final answer. The results show that RSGM outperforms FOFE-net in KBQA end-task on FreeBaseQA dataset.
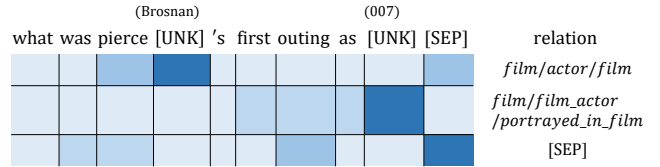

Figure 3: Visualization of relation-aware attention. Deeper color indicate larger attention weight.

3.7 Visualization of relation-aware attention

Let us refer back to the bottom example described in Figure 1, it can be observed that when predicting different relations, the different range of the question plays a different role. At the same time, attention mechanism between relations and questions can be utilized to select the most meaningful words in the given question. To demonstrate the above observation, the weights in attention layer are extracted and further visualized in different kinds of colors that reflect the contributions of different words. The results are shown in Figure 3. From the Figure, it can be observed that the attention is captured properly and indicates which parts in the question make more contribution. For instance, "pierce [UNK]" has been paid more attention when detecting the relation "*film/film_character/portrayed_in_films*".

4 Conclusion

In this paper, we frame multi-hop relation detection as a multi-label learning problem. To solve the challenge of multi-label multi-hop relation detection, we cast it as a sequence generation problem. A relation-aware sequence relation generation model is proposed to learn the problem in an end-to-end manner. Experimental results show that our approach not only achieves better relation detection performance, but also improves the results of the state-of-the-art KBQA system.

References

Samy Bengio, Oriol Vinyals, Navdeep Jaitly, and Noam Shazeer. 2015. Scheduled sampling for sequence prediction with recurrent neural networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, page 1171–1179, Cambridge, MA, USA.

Jonathan Berant, Andrew Chou, Roy Frostig, and Percy Liang. 2013. Semantic parsing on Freebase from question-answer pairs. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1533–1544, Seattle, Washington, USA.

Antoine Bordes, Nicolas Usunier, Sumit Chopra, and Jason Weston. 2015. Large-scale simple question answering with memory networks. *ArXiv*, abs/1506.02075.

Zi-Yuan Chen, Chih-Hung Chang, Yi-Pei Chen, Jijnasa Nayak, and Lun-Wei Ku. 2019. UHop: An unrestricted-hop relation extraction framework for knowledge-based question answering. 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 345–356, Minneapolis, Minnesota.

Rajarshi Das, Shehzaad Dhuliawala, Manzil Zaheer, Luke Vilnis, Ishan Durugkar, Akshay Krishnamurthy, Alex Smola, and Andrew McCallum. 2017. Go for a walk and arrive at the answer: Reasoning over paths in knowledge bases using reinforcement learning. *arXiv preprint arXiv:1711.05851*.

Yanchao Hao, Hao Liu, Shizhu He, Kang Liu, and Jun Zhao. 2018. Pattern-revising enhanced simple question answering over knowledge bases. In *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018*, pages 3272–3282, Santa Fe, New Mexico, USA.

Xiaodong He and David Golub. 2016. Character-level question answering with attention. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1598–1607, Austin, Texas.

Kelvin Jiang, Dekun Wu, and Hui Jiang. 2019. FreebaseQA: A new factoid QA data set matching trivia-style question-answer pairs with Freebase. 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 318–323, Minneapolis, Minnesota.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980.*

Denis Lukovnikov, Asja Fischer, Jens Lehmann, and Sören Auer. 2017. Neural network-based question answering over knowledge graphs on word and character level. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017*, pages 1211–1220, Republic and Canton of Geneva, Switzerland.

Sam Wiseman and Alexander M. Rush. 2016. Sequence-to-sequence learning as beam-search optimization. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1296–1306, Austin, Texas.

Wenhan Xiong, Thien Hoang, and William Yang Wang. 2017. DeepPath: A reinforcement learning method for knowledge graph reasoning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 564–573, Copenhagen, Denmark.

Mingbin Xu, Hui Jiang, and Sedtawut Watcharawittayakul. 2017. A local detection approach for named entity recognition and mention detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1237–1247, Vancouver, Canada. Association for Computational Linguistics.

Pengcheng Yang, Xu Sun, Wei Li, Shuming Ma, Wei Wu, and Houfeng Wang. 2018. SGM: Sequence generation model for multi-label classification. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3915–3926, Santa Fe, New Mexico, USA.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 1480–1489.

Wen-tau Yih, Ming-Wei Chang, Xiaodong He, and Jianfeng Gao. 2015. Semantic parsing via staged query graph generation: Question answering with knowledge base. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, ACL 2015*, pages 1321–1331, Beijing, China.

Wenpeng Yin, Mo Yu, Bing Xiang, Bowen Zhou, and Hinrich Schütze. 2016. Simple question answering by attentive convolutional neural network. In *Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers, COLING 2016*, pages 1746–1756, Osaka, Japan.

Mo Yu, Wenpeng Yin, Kazi Saidul Hasan, Cicero dos Santos, Bing Xiang, and Bowen Zhou. 2017. Improved neural relation detection for knowledge base question answering. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017*, pages 571–581, Vancouver, Canada.

Yang Yu, Kazi Saidul Hasan, Mo Yu, Wei Zhang, and Zhiguo Wang. 2018. Knowledge base relation detection via multi-view matching. *CoRR*, abs/1803.00612.

Linhai Zhang, Chao Lin, Deyu Zhou, Yulan He, and Meng Zhang. 2021. A bayesian end-to-end model with estimated uncertainties for simple question answering over knowledge bases. *Computer Speech & Language*, 66:101167.

M. Zhang and Z. Zhou. 2014. A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 26(8):1819–1837.

Min-Ling Zhang and Zhi-Hua Zhou. 2007. Ml-knn: A lazy learning approach to multi-label learning. *Pattern recognition*, 40(7):2038–2048.