25TH ACM SIGKDD CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING
KDD2019

中国科学技术大学 University of Science and Technology of China
科大讯飞 iFLYTEK

QuesNet: A Unified Representation for Heterogeneous Test Questions

Yu Yin[1], Qi Liu[1], Zhenya Huang[1], Enhong Chen[1], Wei Tong[1] , Shijin Wang[2,3], Yu Su[2,3]*

[1]Anhui Province Key Laboratory of Big Data Analysis and Application, University of Science and Technology of China,
{yxonic,huangzhy}@mail.ustc.edu.cn, {cheneh,qiliuql}@ustc.edu.cn
[2]iFLYTEK Research, iFLYTEK CO., LTD., [3]State Key Laboratory of Cognitive Intelligence, {sjwang3,yusu}@iflytek.com

Abstract

Understanding learning materials (e.g. test questions) is a crucial issue in online learning systems. Unfortunately, many supervised approaches suffer from the problem of scarce human labeled data, whereas abundant unlabeled resources are highly underutilized. While pretraining has the ability to alleviate this problem, existing pretraining methods in NLP area are infeasible to learn test question representations due to several problems. First, questions usually comprise of heterogeneous data including content text, images and side information. Second, there exists both basic linguistic information as well as domain logic and knowledge. To this end, here we propose a novel pre-training method, namely *QuesNet*. We first design a unified framework to aggregate question information with its heterogeneous inputs into a comprehensive vector. Then we propose a two-level hierarchical pre-training algorithm consisting of a novel holed language model objective and a domain-oriented objective, to learn better understanding of test questions. We conduct extensive experiments on large-scale real-world question data, where the experimental results clearly demonstrate the effectiveness of QuesNet for question understanding as well as its superior applicability.

Introduction

Background

How to get better question understanding is the fundamental issue promoting many questionbased applications, such as difficulty estimation, knowledge mapping and score prediction.

Existing Approaches

- Rules for syntactic p:atterns or semantic encodings
- Representation networks in end-to-end frameworks

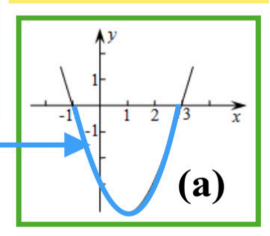
Pre-training Take full advantage of large-scale unlabeled question corpus

Challenges

- Test questions contain coherent heterogeneous data
- We need to carefully consider the advanced logic information aside from the basic linguistic context
- The learned question representations should have great accessibility and be easy to apply to downstream tasks

Eg1. The image of the quadratic function $y = x^2 - 2x - 3$ is shown in figure (a). Find the range of independent variable x when $y < 0$.

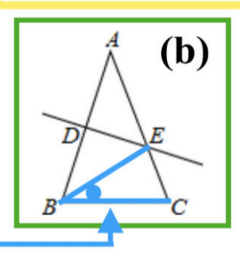
Function



A. x < - 1   B. x > 3   C. - 1 < x < 3   D. x < - 3 or x > 3

Eg2. In the isosceles triangle ABC shown in figure (b), AB = AC, angle A = 20 degrees. Now let DE be the perpendicular bisector of the line segment AB that intersects AB at D and AC at E. Connect B and E, then how many degrees does angle CBE equal?

Geometry



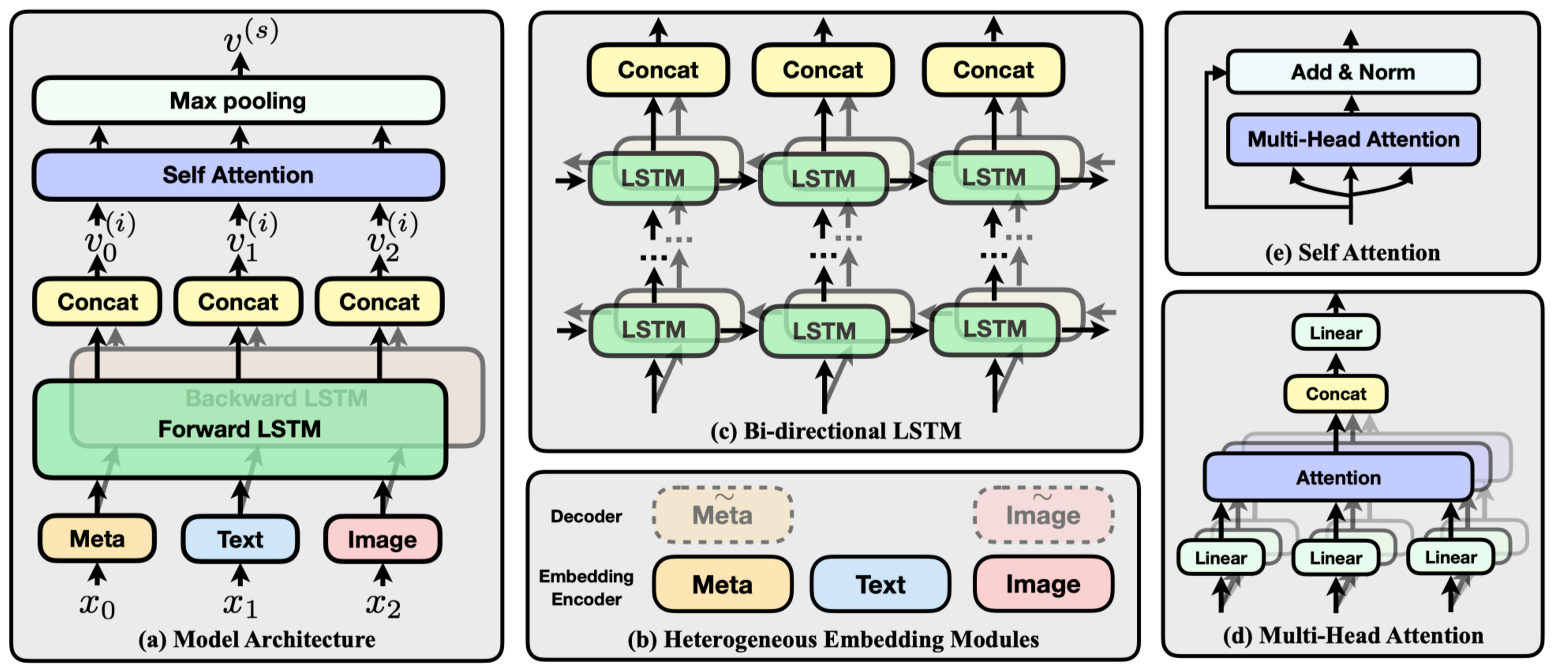
A. 50 degrees   B. 60 degrees   C. 70 degrees   D. 80 degrees

QuesNet: Modeling and Pre-training

QuesNet Architecture

The overall architecture of QuesNet consists of three hierarchical layers:

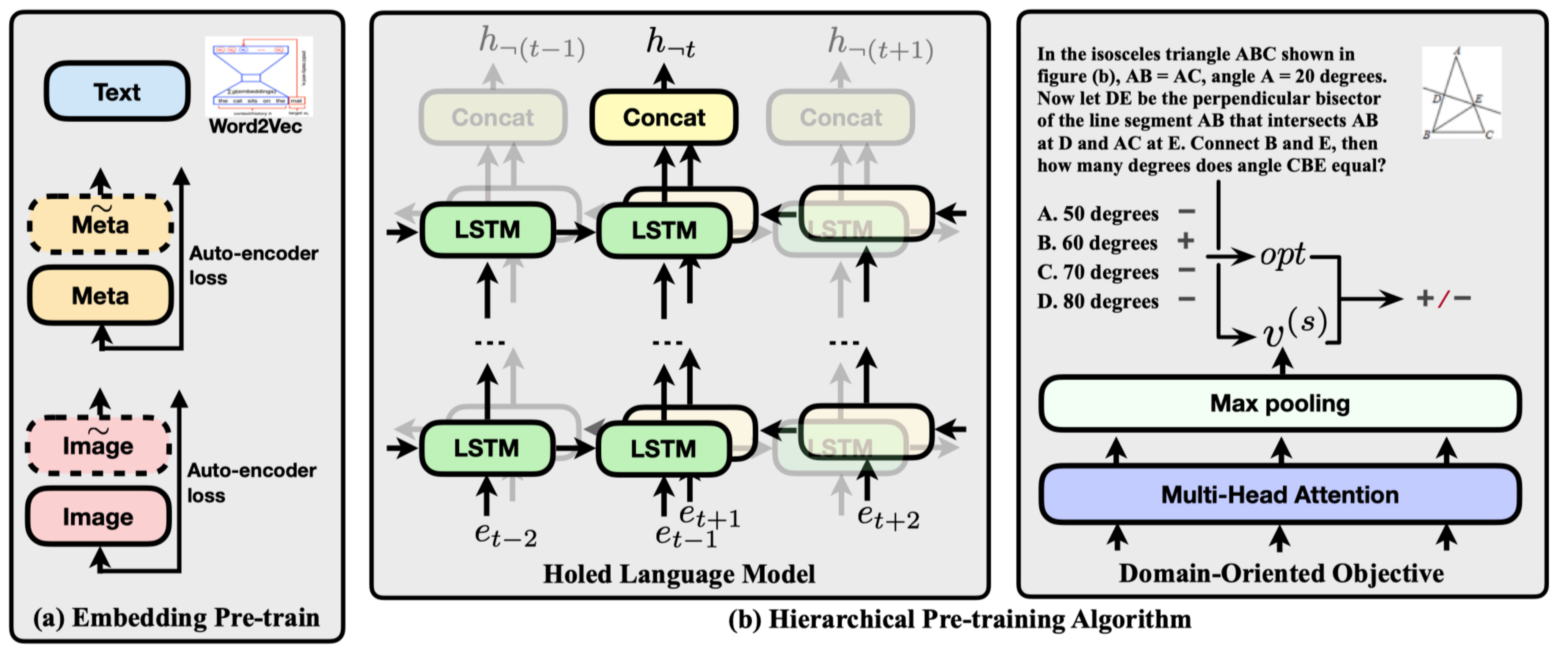
1. **Embedding layer**: encodes heterogeneous information
2. **Content layer**: models linguistic relation and context
3. **Sentence layer**: aggregate linguistic information focusing on long-term and global complex relations



Pre-training

For solving previously mentioned challenges, we design a novel hierarchical pre-training algorithm:

1. **Embedding pretraining**: pre-train separately
2. **Holed Language Model (HLM)**: like LM, but the probability of an input is conditioned by its context from both sides
3. **Domain-Oriented Objective**: use answers and options of a test question as a natural guidance



Datasets

The dataset we used, along with the large question corpus, are supplied by iFLYTEK Co., Ltd., from their online education system called Zhixue.

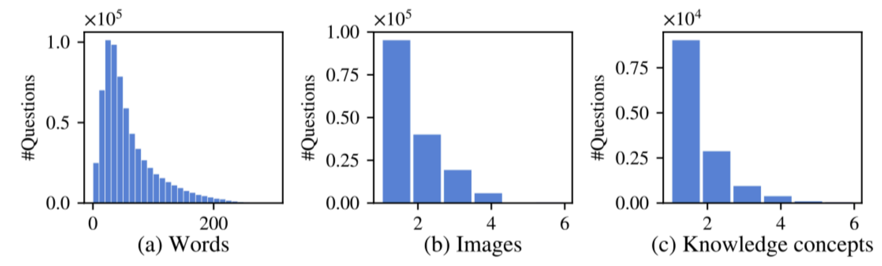


Figure 4: Distribution of question inputs and labels.

Table 1: Statistics of datasets.

	All	KM	DE	SP
#Questions	675,264	13,372	2,465	15,045
#Questions with image	165,859	3,318	299	2,952
#Questions with meta	488,352	8,030	1,896	5,948
#Questions with option	242,960	4840	1,389	4,364
Avg. Words per question	59.10	58.43	60.25	51.94
#Students	-	-	-	50,945
#Records	-	-	-	3,358,111
Label sparsity	-	1.98%	0.37%	2.22%

Evaluation Tasks

Knowledge mapping: a multi-label task
Difficulty estimation: estimate a numerical
Score prediction: a sophisticated domain-specific application

Comparison Methods

Method	Text	Image	Meta	Low level	High level
Original	✔			-	-
ELMo	✔			-	-
BERT	✔			-	-
H-BERT	✔	✔	✔	-	-
QN-T	✔			✔	✔
QN-I		✔		✔	✔
QN-M			✔	✔	✔
QN-TI	✔	✔		✔	✔
QN-TM	✔		✔	✔	✔
QN-IM		✔	✔	✔	✔
QN (no pre)	✔	✔	✔		
QN-L	✔	✔	✔	✔	
QN-H	✔	✔	✔		✔
QuesNet	✔	✔	✔	✔	✔

Experiments

Performance

- Pre-training based methods outperforms original ones in general;
- QuesNet has the best performance guarantee among all the pre-training based methods.

Ablation Experiments

The result shows the importance of heterogeneous information, and also how each purposed techniques contribute to question understanding.

Table 3: Performance of comparison methods on different tasks.

Methods	Knowledge mapping				Difficulty estimation				Student performance prediction			
	ACC	Precision	Recall	F-1	MAE	RMSE	DOA	PCC	MAE	RMSE	ACC	AUC
Original	0.5744	0.4147	0.7872	0.5432	0.2200	0.2665	0.6064	0.3050	0.4245	0.4589	0.7459	0.5400
ELMo	0.6942	0.7960	0.7685	0.7820	0.2250	0.2655	0.5561	0.4299	0.3569	**0.4361**	0.7866	0.5773
BERT	0.6224	0.7326	0.6711	0.7005	0.2265	0.2975	0.6258	0.3600	0.4009	0.4630	0.7390	0.5279
H-BERT	0.6261	0.7608	0.6911	0.7243	0.2097	0.2698	**0.6597**	0.3713	0.3925	0.4528	0.7784	0.5838
QuesNet	**0.7749**	**0.8659**	**0.8075**	**0.8357**	**0.2029**	**0.2530**	0.6137	**0.4499**	**0.3445**	0.4403	**0.7999**	**0.6354**

Table 4: Ablation experiments.

Methods	Knowledge mapping				Difficulty estimation				Student performance prediction			
	ACC	Precision	Recall	F-1	MAE	RMSE	DOA	PCC	MAE	RMSE	ACC	AUC
QN-T	0.7050	0.8264	0.7436	0.7829	0.2166	0.2733	0.6123	0.3040	0.4488	0.4713	0.7454	0.6052
QN-I	0.1136	0.2232	0.3195	0.2628	0.2265	0.2713	0.5961	0.2178	0.4711	0.4899	0.7400	0.5921
QN-M	0.0355	0.1396	0.2853	0.1875	0.2251	0.2737	0.5549	0.2205	0.4719	0.4908	0.7410	0.5502
QN-TI	0.7207	0.8307	0.7595	0.7935	0.2110	0.2647	0.6029	0.3333	0.4279	0.4678	0.7523	0.6221
QN-TM	0.7196	0.8428	0.7523	0.7950	0.2114	0.2664	0.6151	0.3315	0.4353	0.4803	0.7456	0.6156
QN-IM	0.1428	0.2323	0.2818	0.2547	0.2277	0.2707	0.5766	0.2279	0.4710	0.4906	0.7411	0.5513
QN (no pre)	0.5659	0.6816	0.7091	0.6951	0.2225	0.2657	0.5750	0.3087	0.4349	0.4759	0.7488	0.5891
QN-L	0.7185	0.8352	0.7457	0.7879	0.2193	0.2630	0.5721	0.3359	0.3843	0.4561	0.7747	0.6237
QN-H	0.6807	0.8052	0.7271	0.7642	0.2161	0.2665	**0.6291**	0.3328	0.3946	0.4475	0.7814	0.6058
QuesNet	**0.7749**	**0.8659**	**0.8075**	**0.8357**	**0.2029**	**0.2530**	0.6137	**0.4499**	**0.3445**	**0.4403**	**0.7999**	**0.6354**