CKG: Dynamic Representation Based on Context and Knowledge Graph

Xunzhu Tang*
National Engineering Research Center for Big Data Technology and System, Services Computing Technology and System Lab, Cluster and Grid Computing Lab, School of Computer Science and Technology, Huazhong University of Science and Technology
Wuhan, China
tangxz@hust.edu.cn

Tiezhu Sun
Momenta
Suzhou, China
suntiezhu@momenta.ai

Rujie Zhu
Department of Electrical and Computer Engineering
University of Central Florida
Orlando, FL, USA
rujie.zhu@ucf.edu

Shi Wang*
Institute of Computing Technology, Chinese Academy
beijing, China
wangshi@ict.ac.cn

***Abstract*—Recently, neural language representation models pre-trained on large corpus can capture rich co-occurrence information and be fine-tuned in downstream tasks to improve the performance. As a result, they have achieved state-of-the-art results in a large range of language tasks. However, there exists other valuable semantic information such as similar, opposite, or other possible meanings in external knowledge graphs (KGs). We argue that entities in KGs could be used to enhance the correct semantic meaning of language sentences. In this paper, we propose a new method CKG: Dynamic Representation Based on Context and Knowledge Graph. On the one side, CKG can extract rich semantic information of large corpus. On the other side, it can make full use of inside information such as co-occurrence in large corpus and outside information such as similar entities in KGs. We conduct extensive experiments on a wide range of tasks, including QQP, MRPC, SST-5, SQuAD, CoNLL 2003, and SNLI. The experiment results show that CKG achieves SOTA 89.2 on SQuAD compared with SAN (84.4), ELMo (85.8), and BERT$_{Base}$ (88.5).**

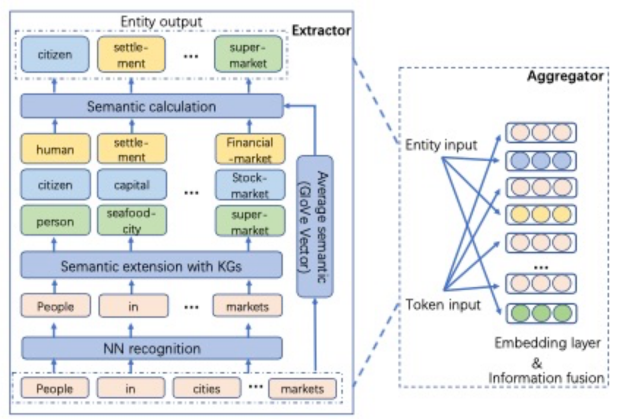


Fig. 1: Architecture of CKG, which consists of two main parts (*i.e.,* Extractor and Aggregator).

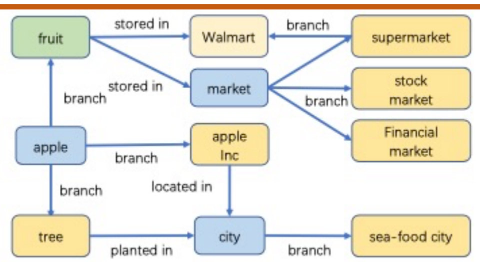


Fig. 2: Partial Knowledge Graph from Dbpedia. The figure shows that 'apple' could be a kind of food, as well as a company. We can get the precise meaning of 'apple' by the neighbours around it.

TABLE I: Results of various models on FIGER(%)

Model	NFGEC (Attentive)	NFGEC (LSTM)	ERNIE (tsinghua)	CKG
Acc.	54.53	55.60	57.19	58.84
Macro	74.76	75.15	76.51	76.23
Micro	71.58	71.73	73.39	75.24

TABLE II: Comparison among CKG+ELMo and other models in QQP, SST-5, and MRPC tasks

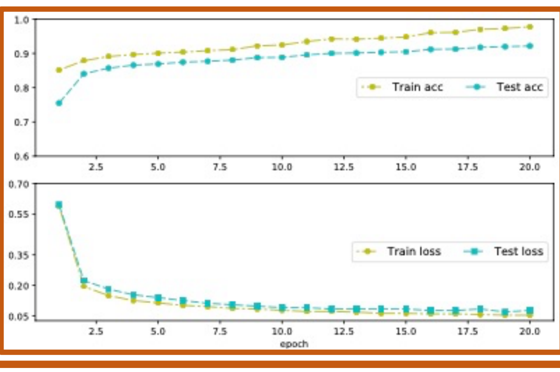
Baseline	Pre-OpenAI SOTA	BiLSTM+ ELMo+Attn	ERNIE (tsinghua)
QQP	93.2	90.4	93.5
MRPC	86.0	84.9	88.9
SST-5	66.1	64.8	71.2

	ERNIE2.0 (baidu)	CKG	CKG+ELMo
QQP	90.4	92.7	93.2
MRPC	88.9	88.2	86.3
SST-5	70.4	70.2	72.3

TABLE III: Comparison of models in word similarity with rank of sepearman. And the comparison of word analogy in semantic, syntactic, average.

Baseline	Semantic	Syntactic	Average	Rank of Spearman
CBOW	73.58	65.95	69.5	73.25
SG	65.62	56.61	60.64	68.69
GloVe	71.39	53.72	61.57	68.93
CKG+GloVe	78.34	69.32	73.83	79.37

Fig. 4: The detailed information on the dataset of SNLI. The upper figure represents the accuracy results of training and test in different epochs, and the lower figure indicates the losses of them.



Task	GloVe	ELMo	CKG+GloVe	CKG
SQuAD	80.8	85.8	85.6	88.7
SNLI	88.1	89.1	90.2	91.1
NER	87.7	91.9	-	92.56

Analysis for SQuAD, SNLI, and NER, comparing different choices. We compares systems with pure GloVe, ELMo, the combination of GloVe and CKG shown as Figure 1 and only CKG. The set of comparison of the three models. ELMo has achieved a great performance in NER task, while CKG performs better.

TABLE V: The set of comparison of the five models. We compare SAN, ELMo, BERT$_{BASE}$, with CKG and CKG+ELMo across SQuAD task. The "INCREASE" column lists improvement over our baseline.

Model	SAN	ELMo	BERT (base)	CKG	CKG + ELMo
SOTA	84.4	85.8	88.5	88.7	89.2
INCREASE	baseline	1.66%	4.86%	5.09%	5.69%