

国際会議
EMNLP-IJCNLP2019
参加報告

NHK放送技術研究所
宮崎太郎

自己紹介

- 宮崎太郎(みやざきたろう)
- 2011年よりNHK技研でNLP周りの研究
 - Twitterからの情報抽出
 - 日本語→手話の翻訳
- 今回が初の有名会議での発表！



本日の内容

- EMNLP-IJCNLP 2019概観
 - 会議概要
 - 発表タイトルに見るトレンド
- Knowledgeを活用した学習まわりの論文紹介
 - Knowledge-Enriched Transformer for Emotion Detection in Textual Conversations
 - Machine Reading Comprehension Using Structural Knowledge Graph-aware Network
 - Language Models as Knowledge Bases?

EMNLP-IJCNLP 2019



2019年11月3-7日 @香港 Asia World Expo



本会議683件+18ワークショップ433件の発表など



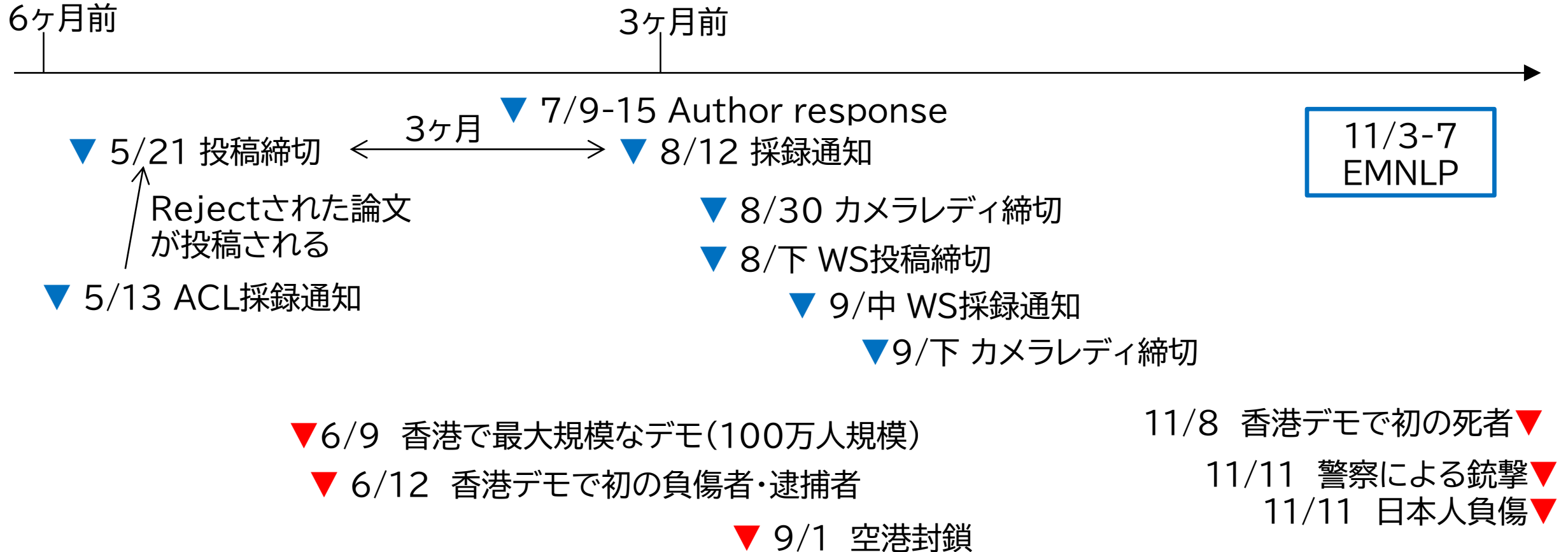
参加者数1,957人

広い! 人多い!

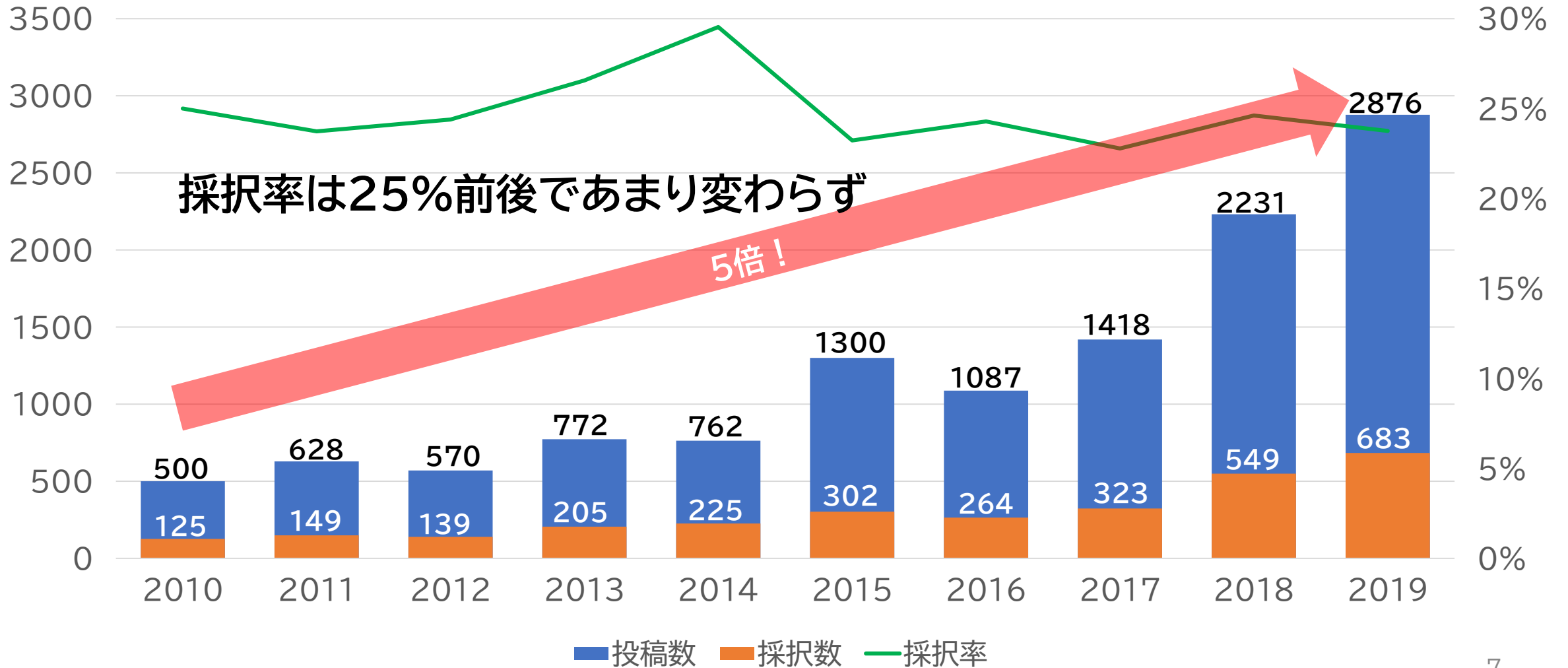
ソーシャルイベント @香港ディズニーランド!



スケジュール感



投稿数, 採択数の推移



Reviewスコアと採択の関係

Decisions by Score



EMNLP 2018

ACCEPTED PAPERS BY AVERAGE SCORE



EMNLP 2019
NAIST 須藤さまのTweetから拝借しました

今年のレビューの特徴

- スコアは”Overall Recommendation”のみ
 - Author response時にはスコア一切不明
 - 査読者は“Reviewer Confidence“などを入力しているようですが・・・
- 結構頑張ってAuthor responseしたけど, コメントの変化なし
 - ACからのMeta-Reviewには反映されてそうなのでやはり重要そう

論文タイトルから見る 研究のトレンド

論文タイトルから見る研究トレンド

- これだけではトレンドはよくわからん...
- これでは何も面白くないので以下のFeatureを加味
 - $Score(w) = \log(TF(w)) * IDF^*(w^w)$
 - $IDF^*(w) = \log\left(\frac{|D|}{\{d:d\exists w\}+1}\right)$, $d \in EMNLP2018$ の論文タイトル集合
 - 要は, 前年と比較して増えたものを強調

単語ごとの論文数推移

単語	EMNLP 2018	EMNLP 2019
understanding	6	16
approach	8	19
analysis	13	30
text	30	68
generative	6	13
transfer	7	15
supervised	7	15
latent	7	14
cross-lingual	15	30
generation	32	63
graph	14	27
dataset	13	24
knowledge	24	44
dialogue	11	20
extraction	17	30

※EMNLP 2018で6回(1%)以上出現の単語のみを対象
増加率上位15件を表示

単語	EMNLP 2018	EMNLP 2019
bert	0	14
stance	0	6
recommendation	0	5
ner	0	5
fact	0	5
task-oriented	1	10
global	1	10
style	1	8
response	2	11
reasoning	4	19
low-resource	4	15
conversation	4	13
questions	4	12
event	5	15
dialog	5	14

※EMNLP 2018で出現が少なかったものから
気になったもの

単語ごとの論文数推移

単語	EMNLP 2018	EMNLP 2019
understanding	6	16
approach	8	19
analysis	13	30
text	30	68
generative	6	13
transfer	7	15
supervised	7	15
latent	7	14
cross-lingual	15	30
generation	32	63
graph	14	27
dataset	13	24
knowledge	24	44
dialogue	11	20
extraction	17	30

単語	EMNLP 2018	EMNLP 2019
bert	0	14
stance	0	6
recommendation	0	5
ner	0	5
fact	0	5
task-oriented	1	10
global	1	10
style	1	8
response	2	11
reasoning	4	19
low-resource	4	15
conversation	4	13
questions	4	12
event	5	15
dialog	5	14

タスク:赤, 手法:緑で色分け

流行りの手法は・・・

- **BERT**はとにかく多くの論文で使われている！
 - 論文投稿締め切り半年前に公開されたことを考えると驚異的！
- テキスト生成以外のタスクでも**Generative**なアプローチ
 - 文書分類, イベント抽出などに使われているようです
- **Knowledge, Graph**など, 外部知識を取り入れる手法
 - 文脈を取り入れた翻訳なども広い意味でこのカテゴリ？

流行りのタスクは・・・

- **対話生成, 対話分析**の復興

- 文章理解や生成など, 複雑なことを組み合わせても性能がでるようになった？

- **Fact check**がとにかく注目されている

- Keynote, Workshopなどもたくさん, が難しい・・・

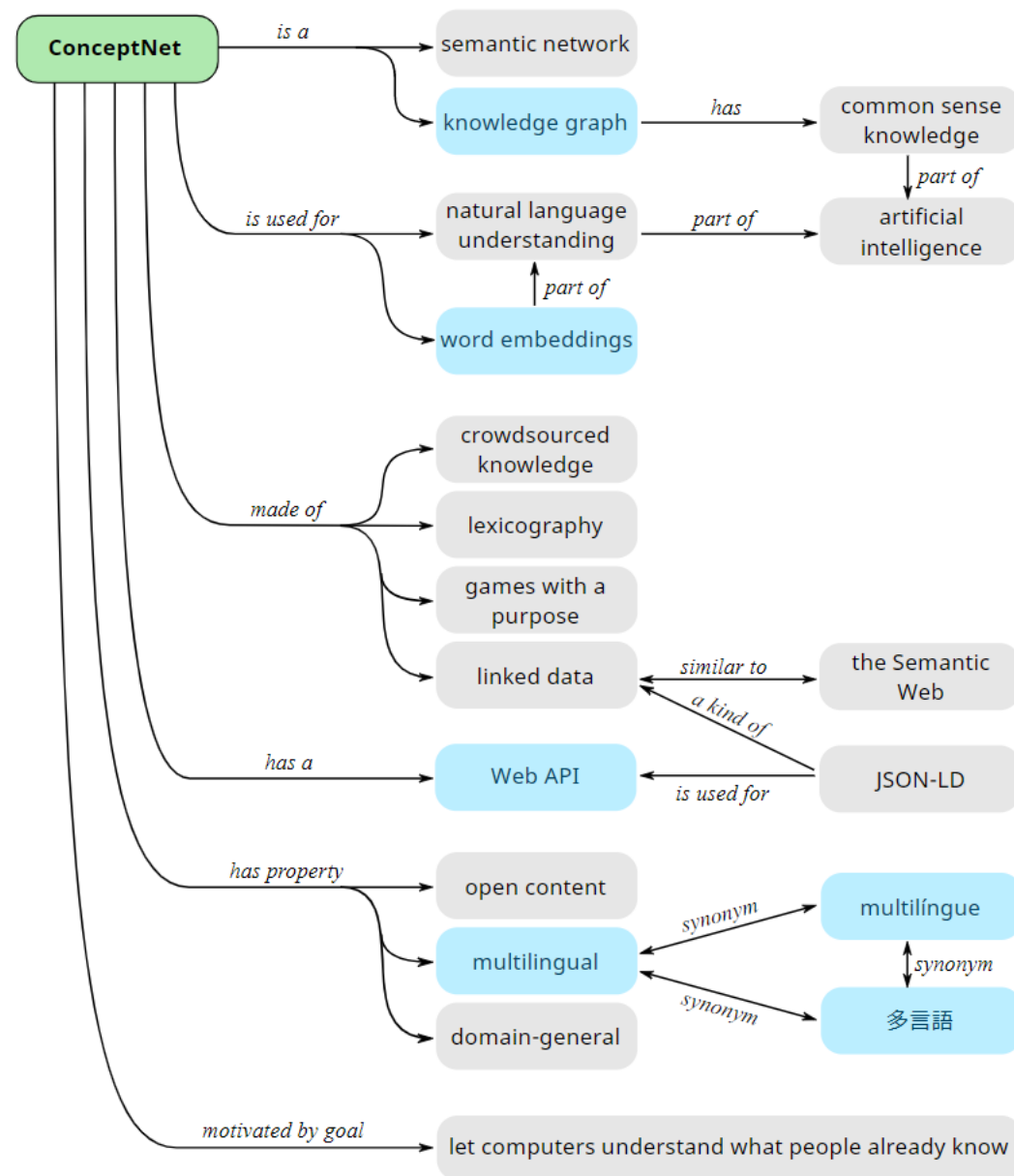
- **多言語**を同時に扱う手法タスク・・・

- BERTがmulti-lingualなことの影響な気も

Knowledgeまわりの論文紹介

Knowledge... ?

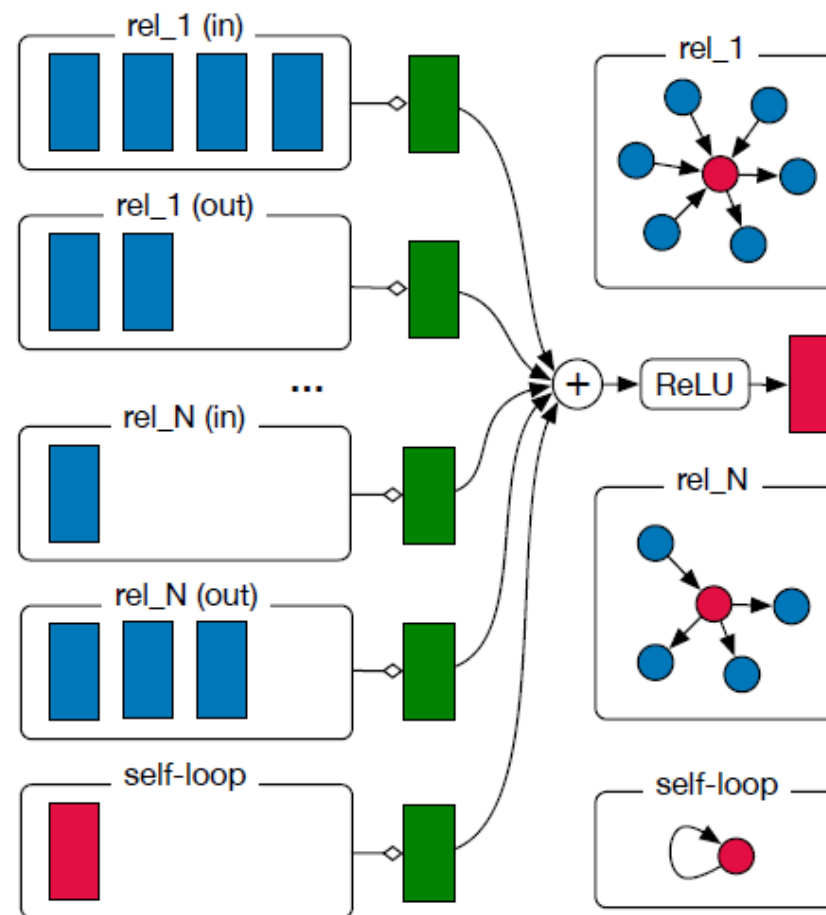
- 単語と単語の関係をつなぐKB (Knowledge Base)の利用が主流
 - ConceptNetやYagoなどが公開
 - 「外部知識」として活用



出典: ConceptNet (<http://conceptnet.io/>)

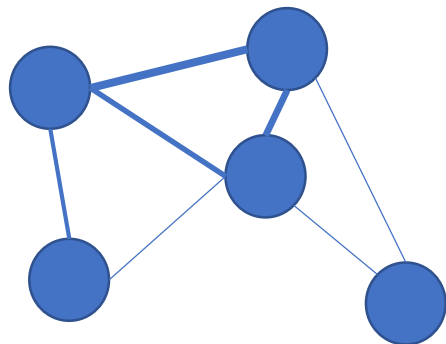
Graph... ?

- KBのようなGraphも入力するNN
 - 基本的には, 注目している単語とそれにKB上でつながる単語のベクトルを足し合わせたベクトルを利用
 - 足し合わせ方, 重みの決定方法が違う
いろいろな手法が提案



【参考】GCN(Graph Convolutional Network) [Kipf et al., 2016]

- 対象単語とその周辺単語のベクトルを足し合わせる重みを**事前情報**として与える(学習しない)
 - エッジの**重みを固定**で与える
 - 性能が良くて計算は速いが, KBが変わったときに対応不可能



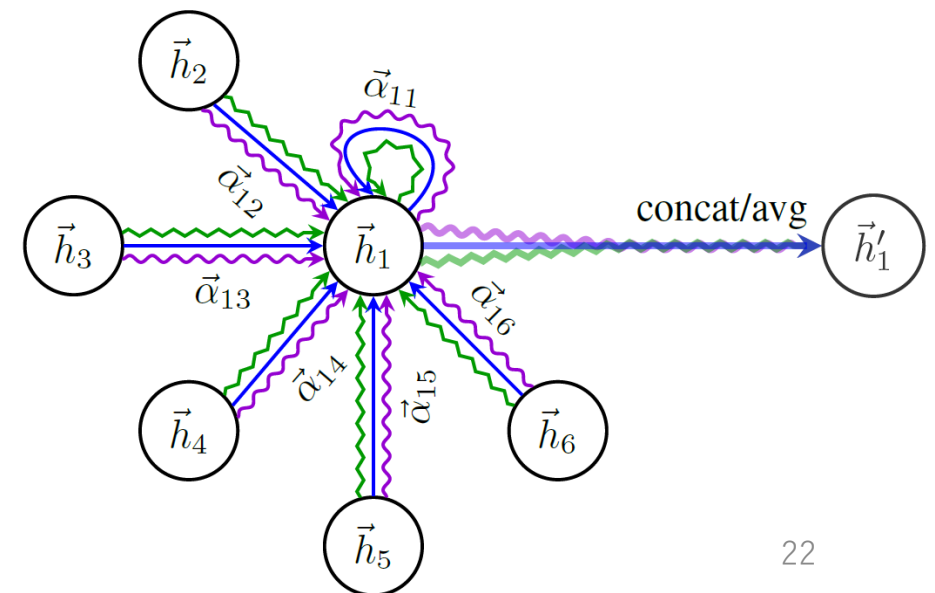
$$A = \begin{pmatrix} 1 & \dots & 0.3 \\ \vdots & \ddots & \vdots \\ 0.7 & \dots & 1 \end{pmatrix}$$

足し合わせ重み = 隣接行列(固定) 21

【参考】GAT (Graph Attention network) [Veličković et al., 2018]

- 対象単語とその周辺単語のベクトルを足し合わせる重みを「アテンションの値」として学習

- それぞれのエッジの重みを持たないため、メモリ削減+学習が早い
- KBに新たなエッジが追加されても大丈夫



Knowledge-Enriched Transformer for Emotion Detection in Textual Conversations

Peixiang Zhong^{1,2}, Di Wang¹, Chunyan Miao^{1,2,3}

¹Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly

²Alibaba-NTU Singapore Joint Research Institute

³School of Computer Science and Engineering

Nanyang Technological University, Singapore

peixiang001@e.ntu.edu.sg, {wangdi, ascymiao}@ntu.edu.sg

概要

- 対話テキストの感情推定
 - Transformer-basedなモデルに文脈とKBを使う手法(KET)の提案
- Ablation testにより, 文脈, KBの双方が性能向上に貢献していることを示した

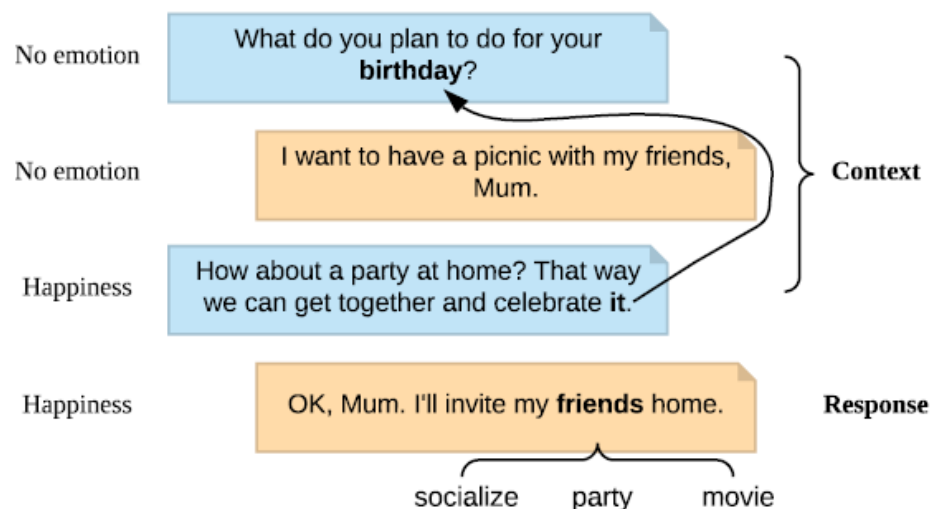
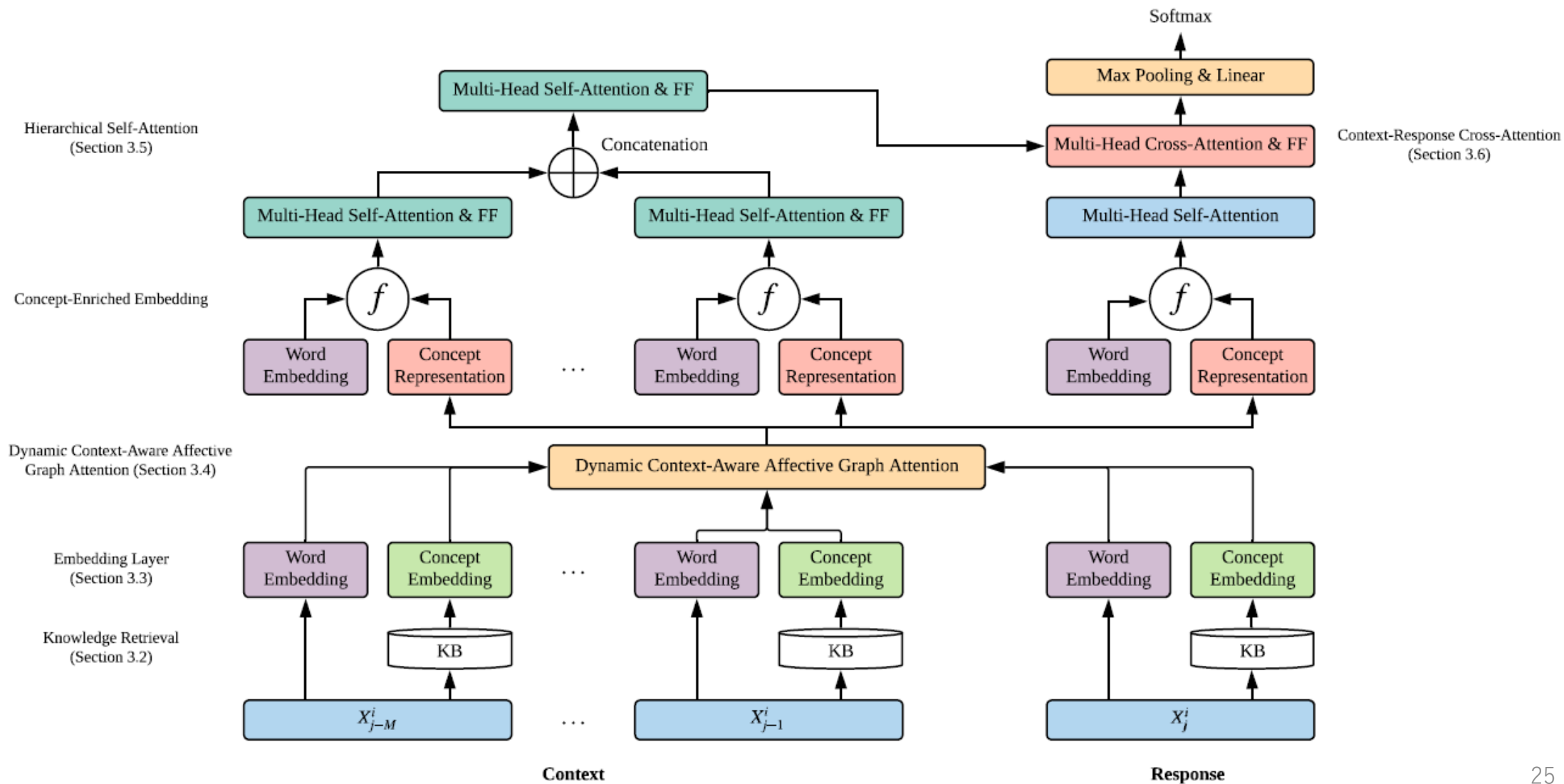
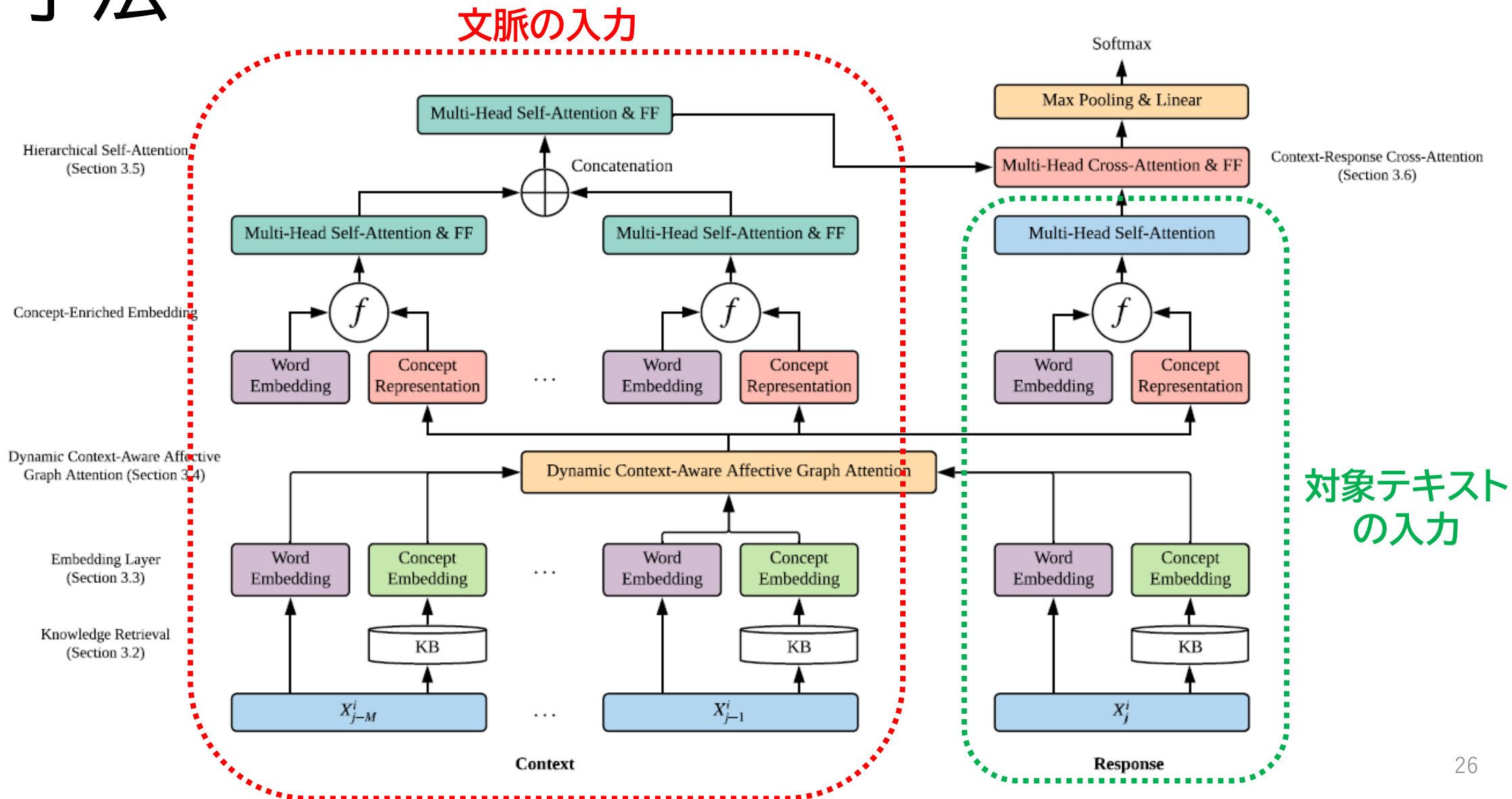


Figure 1: An example conversation with annotated labels from the DailyDialog dataset (Li et al., 2017). By referring to the context, “it” in the third utterance is linked to “birthday” in the first utterance. By leveraging an external knowledge base, the meaning of “friends” in the fourth utterance is enriched by associated knowledge entities, namely “socialize”, “party”, and “movie”. Thus, the implicit “happiness” emotion in the fourth utterance can be inferred more easily via its enriched meaning.

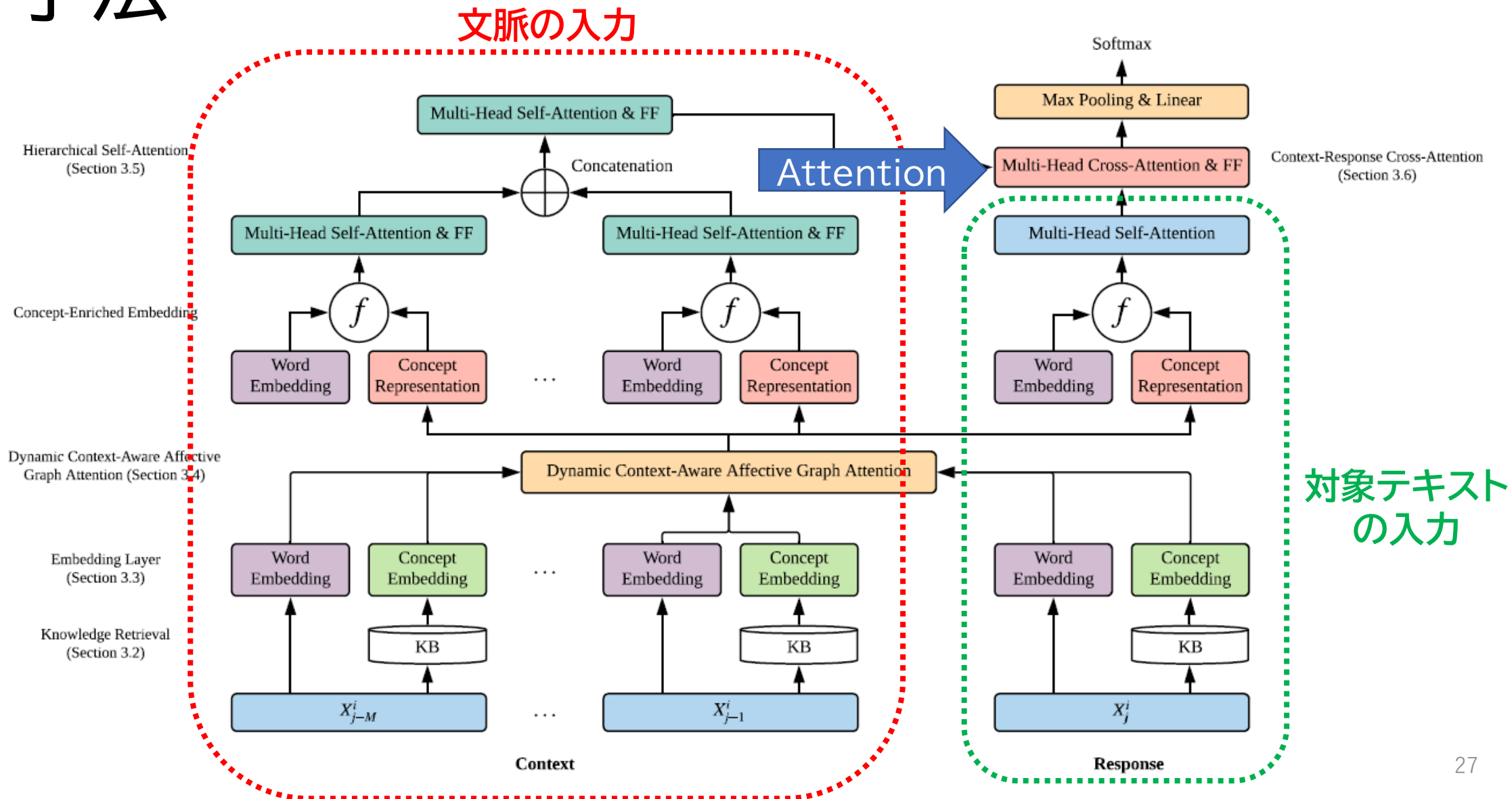
手法



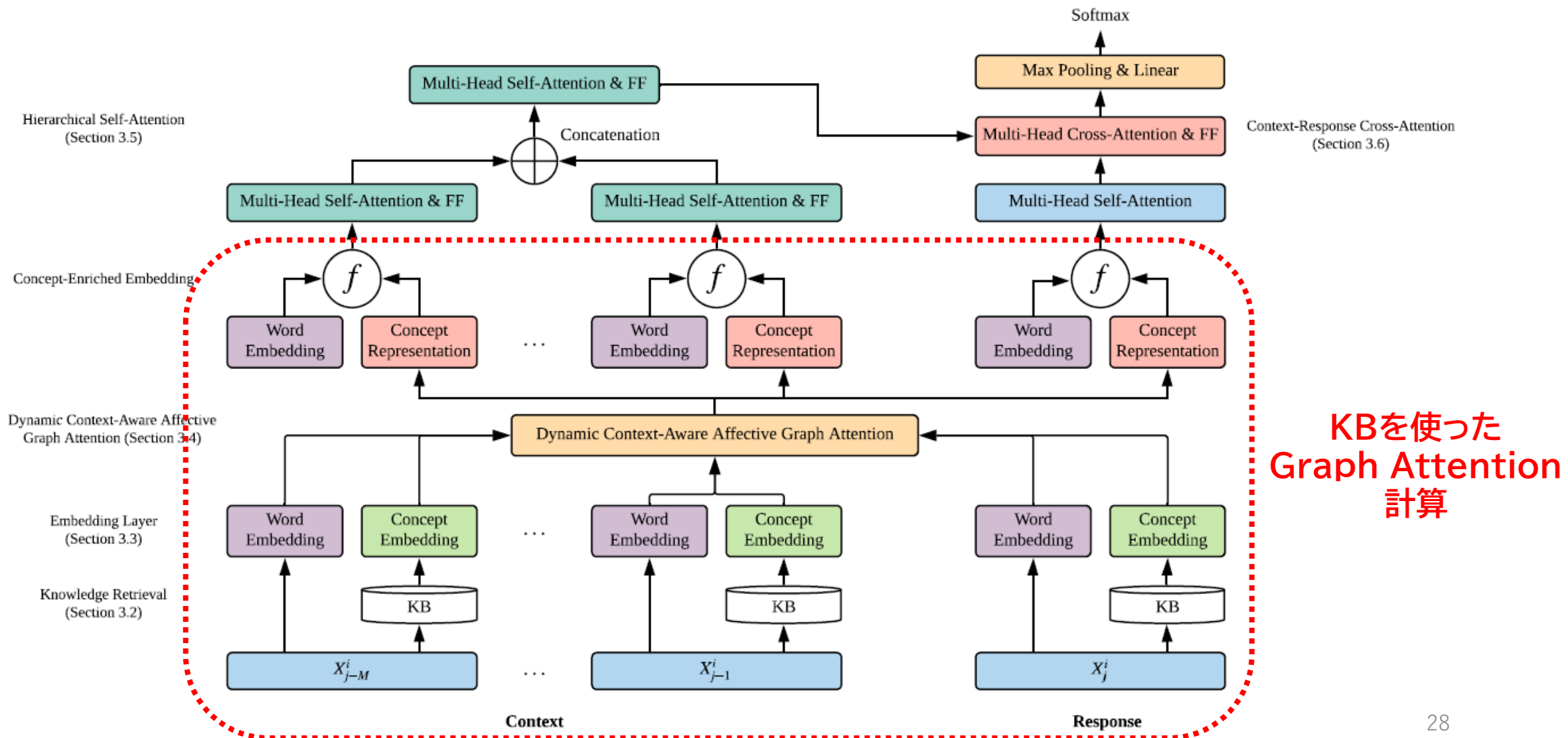
手法



手法



手法



実験結果

Model	EC	DailyDialog	MELD	EmoryNLP	IEMOCAP
cLSTM	0.6913	0.4990	0.4972	0.2601	0.3484
CNN (Kim, 2014)	0.7056	0.4934	0.5586	0.3259	0.5218
CNN+cLSTM (Poria et al., 2017)	0.7262	0.5024	0.5687	0.3289	0.5587
BERT_BASE (Devlin et al., 2018)	0.6946	0.5312	0.5621	0.3315	0.6119
DialogueRNN (Majumder et al., 2019)	0.7405	0.5065	0.5627	0.3170	0.6121
KET_SingleSelfAttn (ours)	0.7285	0.5192	0.5624	0.3251	0.5810
KET_StdAttn (ours)	0.7413	0.5254	0.5682	0.3353	0.5861
KET (ours)	0.7348	0.5337	0.5818	0.3439	0.5956

Table 2: Performance comparisons on the five test sets. Best values are highlighted in bold.

- 提案手法が概ね良好
 - Attentionの計算方法によって若干のブレはある

実験結果

- 文脈, KBはそれぞれ使ったほうが良い結果

Dataset	KET	-context	-knowledge
EC	0.7451	0.7343	0.7359
DailyDialog	0.5544	0.5282	0.5402
MELD	0.5401	0.5177	0.5248
EmoryNLP	0.3712	0.3564	0.3553
IEMOCAP	0.5389	0.4976	0.5217

Table 5: Ablation study for KET on the validation sets.

Machine Reading Comprehension Using Structural Knowledge Graph-aware Network

**Delai Qiu¹, Yuanzhe Zhang², Xinwei Feng⁴, Xiangwen Liao¹,
Wenbin Jiang⁴, Yajuan Lyu⁴, Kang Liu^{2,3}, Jun Zhao^{2,3}**

¹ College of Mathematics and Computer Science, Fuzhou University, Fuzhou, China

² Institute of Automation, Chinese Academy of Sciences, Beijing, China

³ University of Chinese Academy of Sciences, Beijing, China

⁴ Baidu Inc., Beijing, China

noneqdl@gmail.com, {yzzhang, kliu, jzhao}@nlpr.ia.ac.cn
lxw@fzu.edu.cn, {fengxinwei, jiangwenbin, lvyaajuan}@baidu.com

概要

- Machine Reading Comprehension(MRC)タスクでKBを活用
 - MRC: テキストを読んでそれに関する質問に答えるタスク
- Graph Attention Network (GAT) を応用したモデルの提案
- いくつかのデータセットを使って**えらい高い**性能を発揮

手法

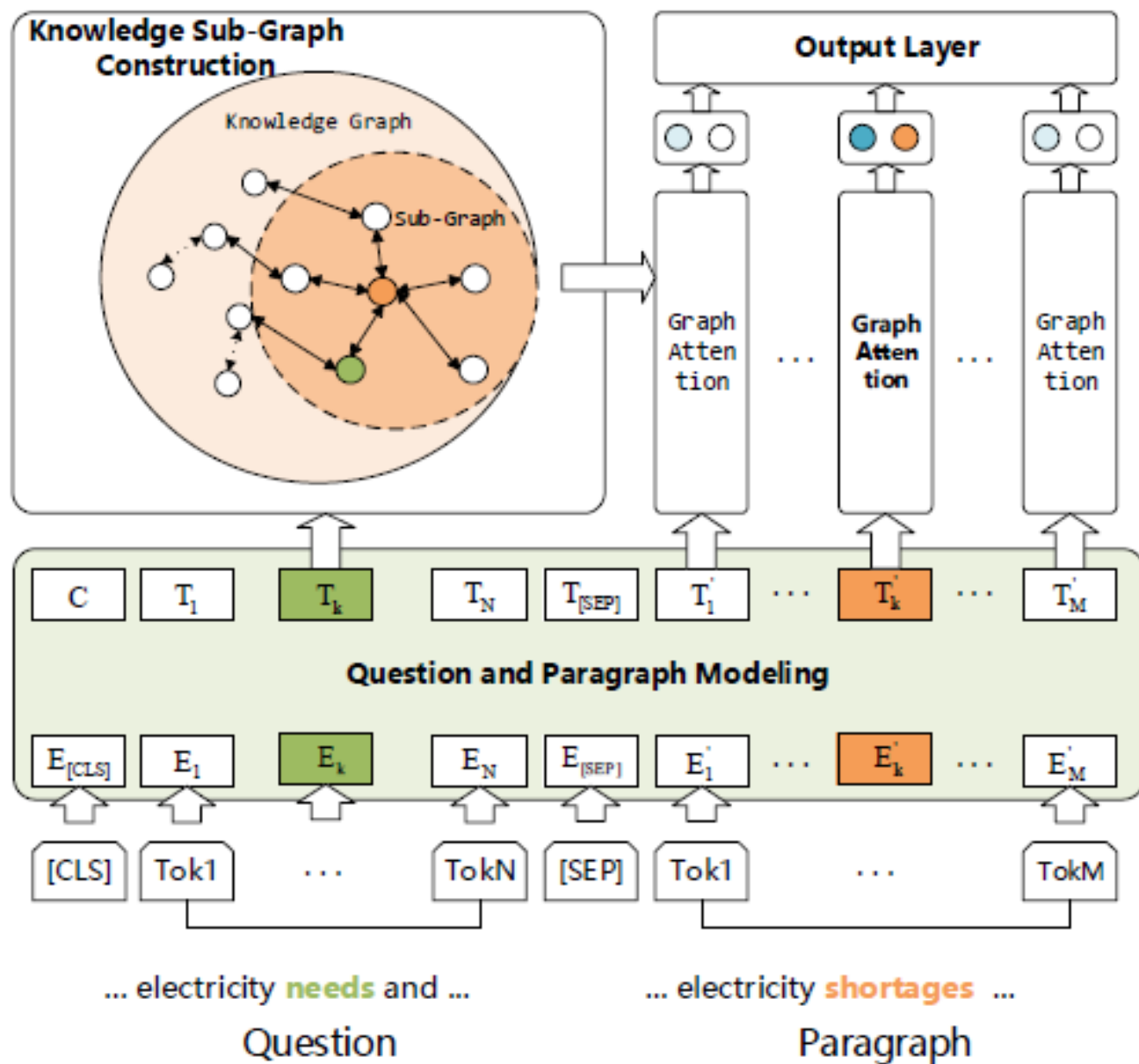
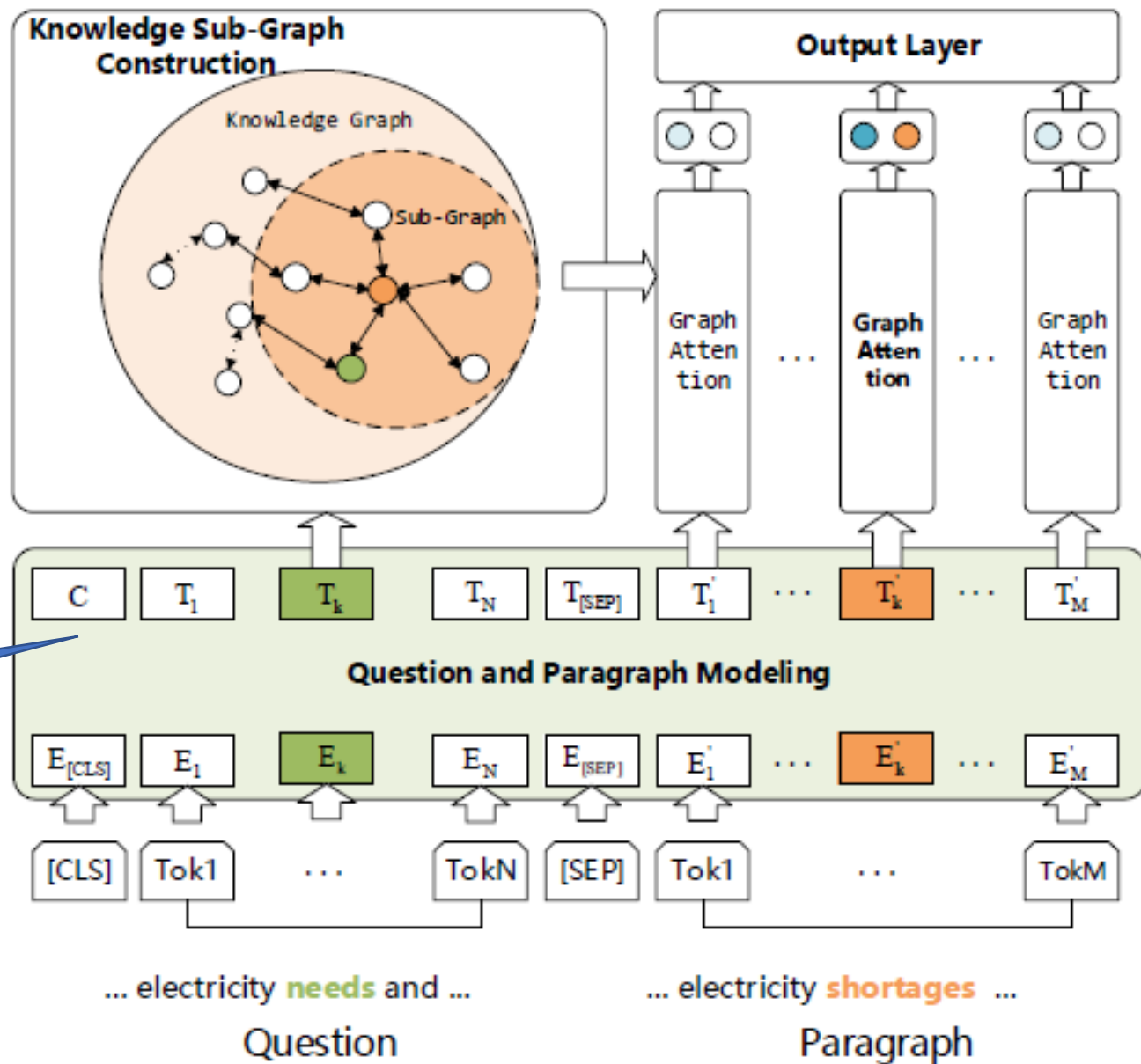


Figure 1: Framework of our SKG model.

手法



①BERTに入力

Figure 1: Framework of our SKG model.

手法

②質問文に出現する
単語とKBから
サブグラフ生成

①BERTに入力

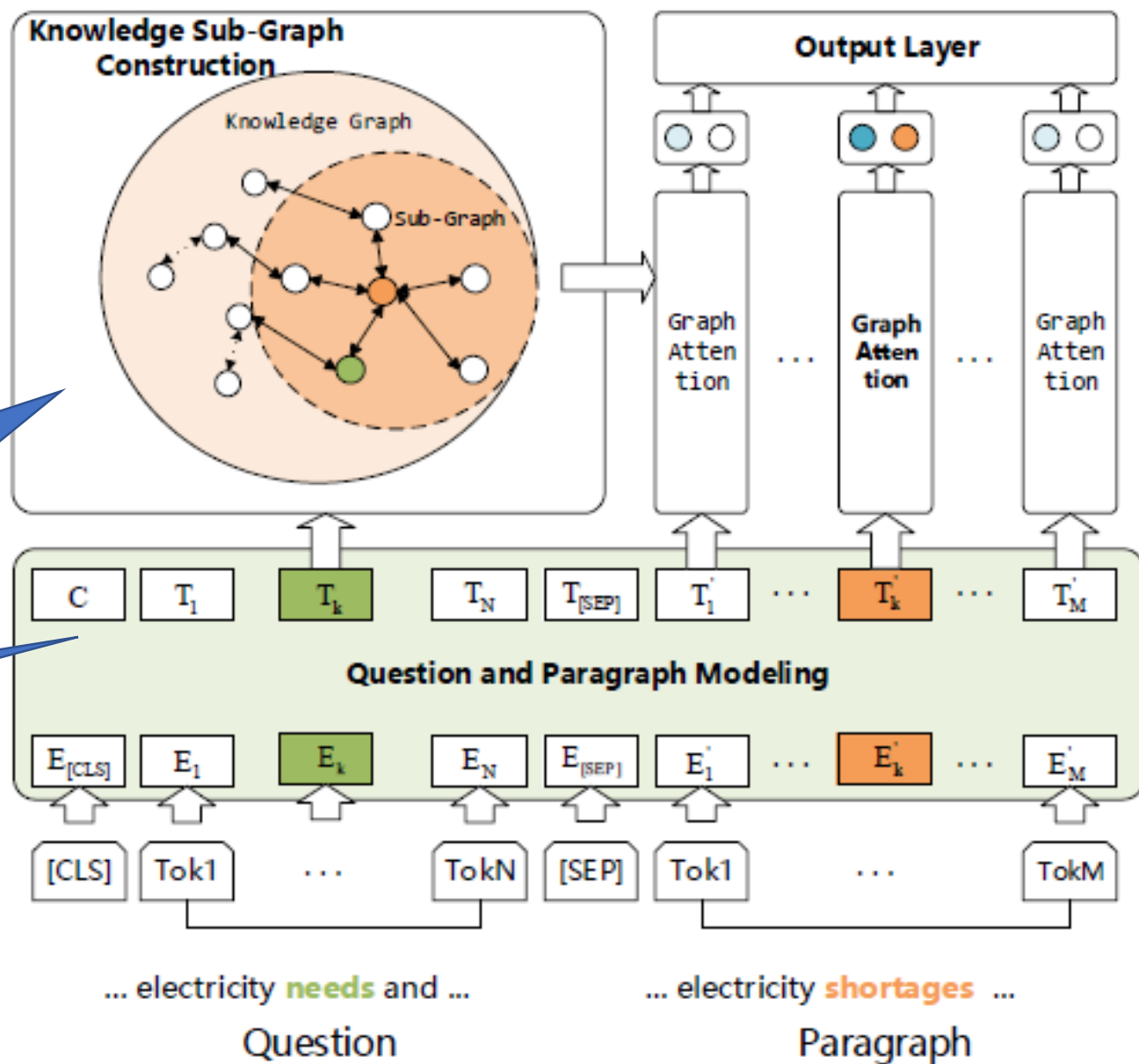


Figure 1: Framework of our SKG model.

手法

② 質問文に出現する
単語とKBから
サブグラフ生成

① BERTに入力

③ GATの要領で
Attentionを計算

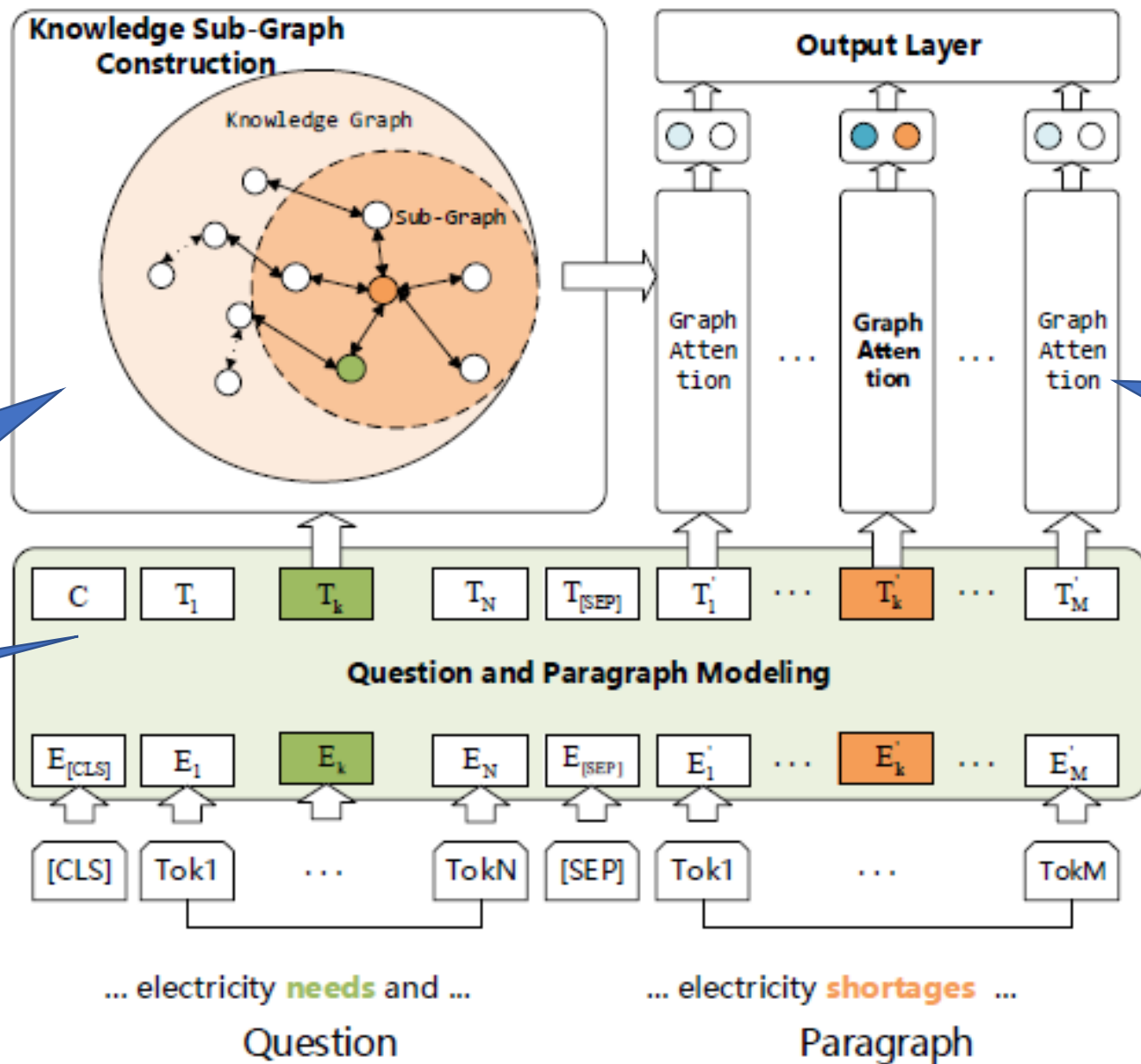
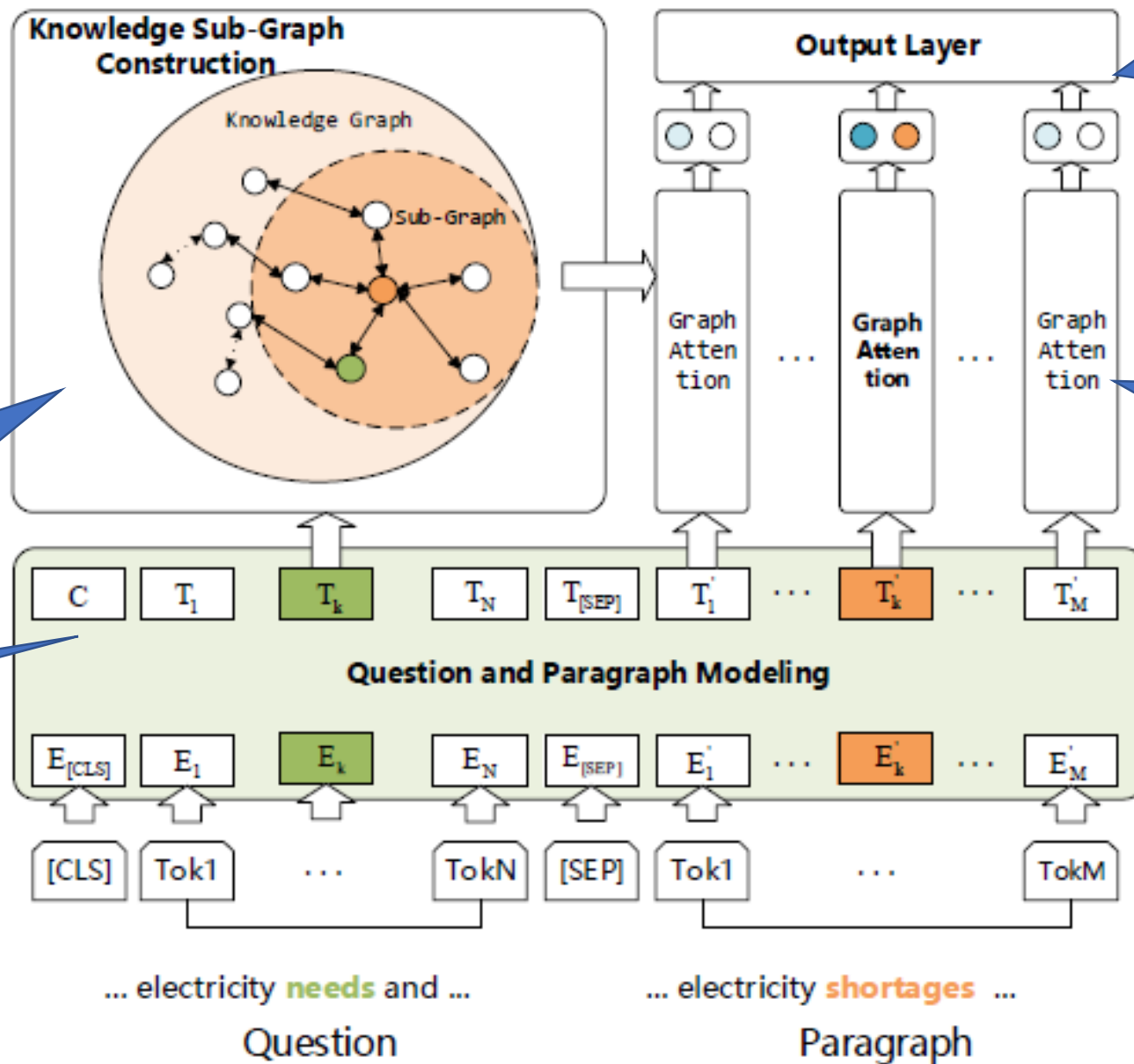


Figure 1: Framework of our SKG model.

手法

②質問文に出現する
単語とKBから
サブグラフ生成

①BERTに入力



④Decoderで出力

③GATの要領で
Attentionを計算

Figure 1: Framework of our SKG model.

結果

Model	EM		F1	
	Dev	Test	Dev	Test
QANet (Yu et al., 2018)	35.38	36.51	36.75	37.79
SAN (Liu et al., 2018)	38.14	39.77	39.09	40.72
DocQA w/o ELMo (Clark and Gardner, 2018)	36.59	38.52	37.89	39.76
DocQA w/ ELMo (Clark and Gardner, 2018)	44.13	45.44	45.39	46.65
SKG+BERT-Large(ours)	70.94	72.24	71.55	72.78

Table 1: The performance of different models on ReCoRD dataset.

Language Models as Knowledge Bases?

**Fabio Petroni¹ Tim Rocktäschel^{1,2} Patrick Lewis^{1,2} Anton Bakhtin¹
Yuxiang Wu^{1,2} Alexander H. Miller¹ Sebastian Riedel^{1,2}**

¹Facebook AI Research

²University College London

{fabiopetroni, rockt, plewis, yolo, yuxiangwu, ahm, sriedel}@fb.com

概要

- BERT等の学習済みモデルをKBとして使えるんじゃないか？



- BERT-LargeはQAタスクなどを精度良く解ける！

- KBを明示的に使う手法より良い場合も！

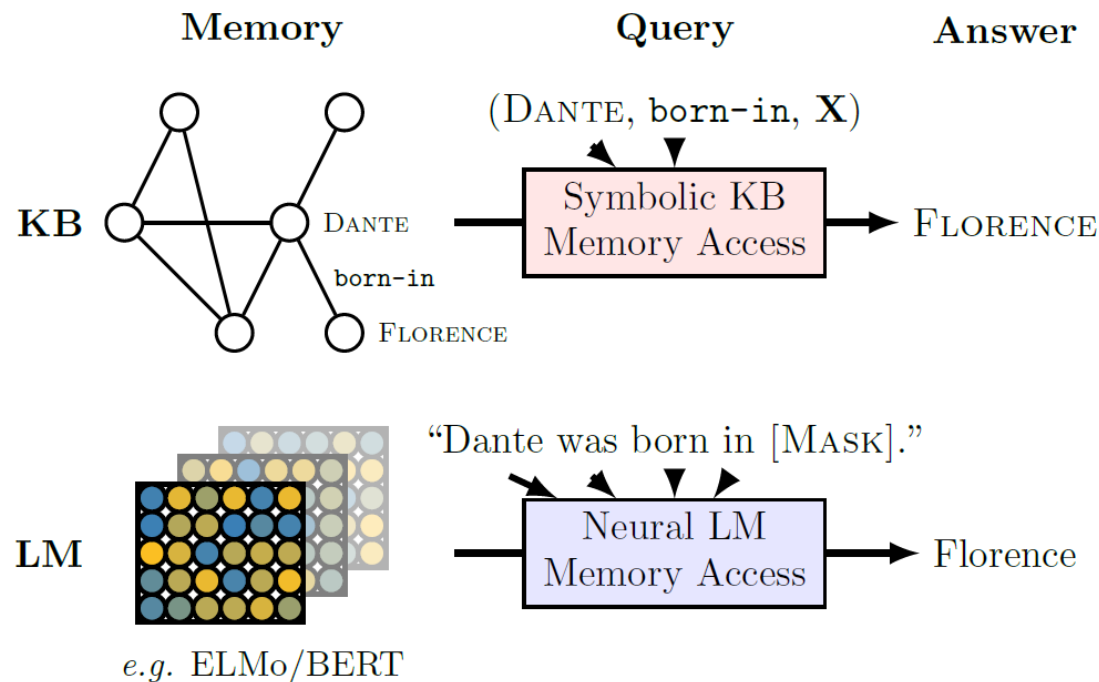


Figure 1: Querying knowledge bases (KB) and language models (LM) for factual knowledge.

概要

- BERT等の学習済みモデルをKBとして使えるんじゃないか？



- BERT-LargeはQAタスクなどを精度良く解ける！

- KBを明示的に使う手法より良い場合も！

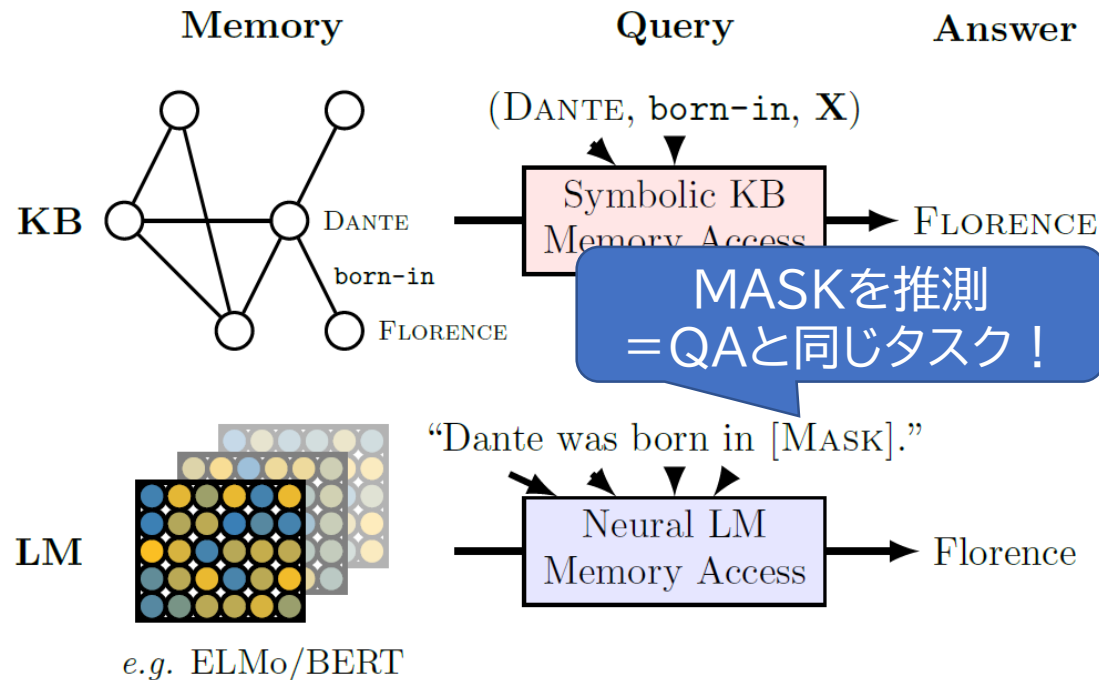


Figure 1: Querying knowledge bases (KB) and language models (LM) for factual knowledge.

結果

Corpus	Relation	Statistics		Baselines		KB		LM					
		#Facts	#Rel	Freq	DrQA	RE _n	RE _o	Fs	Txl	Eb	E5B	Bb	Bl
Google-RE	birth-place	2937	1	4.6	-	3.5	13.8	4.4	2.7	5.5	7.5	14.9	16.1
	birth-date	1825	1	1.9	-	0.0	1.9	0.3	1.1	0.1	0.1	1.5	1.4
	death-place	765	1	6.8	-	0.1	7.2	3.0	0.9	0.3	1.3	13.1	14.0
	Total	5527	3	4.4	-	1.2	7.6	2.6	1.6	2.0	3.0	9.8	10.5
T-REx	1-1	937	2	1.78	-	0.6	10.0	17.0	36.5	10.1	13.1	68.0	74.5
	<i>N</i> -1	20006	23	23.85	-	5.4	33.8	6.1	18.0	3.6	6.5	32.4	34.2
	<i>N</i> - <i>M</i>	13096	16	21.95	-	7.7	36.7	12.0	16.5	5.7	7.4	24.7	24.3
	Total	34039	41	22.03	-	6.1	33.8	8.9	18.3	4.7	7.1	31.1	32.3
ConceptNet	Total	11458	16	4.8	-	-	-	3.6	5.7	6.1	6.2	15.6	19.2
SQuAD	Total	305	-	-	37.5	-	-	3.6	3.9	1.6	4.3	14.1	17.4

KB自動獲得
手法

BERT-
Large

Table 2: Mean precision at one (P@1) for a frequency baseline (Freq), DrQA, a relation extraction with naïve entity linking (RE_n), oracle entity linking (RE_o), fairseq-fconv (Fs), Transformer-XL large (Txl), ELMo original (Eb), ELMo 5.5B (E5B), BERT-base (Bb) and BERT-large (Bl) across the set of evaluation corpora.

ただし！

- すべてのタスクはCloze-Style
 - SQuADのようなQAタスクは手作業でCloze-Styleに変換
 - “Who developed the theory of relatively?”
→”The theory of relatively was developed by ____”
- もう少し複雑なタスクの場合どうなるかは不明
 - BERTなどで回答文生成は難しいと思うので・・・

論文紹介パートまとめ

- Knowledgeを使うことで性能が向上！ することも多い
 - Knowledge自体の適用化, タスクに合わせた拡張なども必要な場合も
 - それ自体が一つの研究テーマ(→KB Completionなど)
- Knowledgeを使うやり方はまだ皆さん試行錯誤中？
 - Graph Attention Network[Veličković et al, 2018] basedな手法が流行？
 - NeurIPS 2019でも80件近い「Graph」をタイトルに含む論文が