基于迁移学习的任务型对话系统研究

--以跨任务、跨领域、跨语言迁移为例

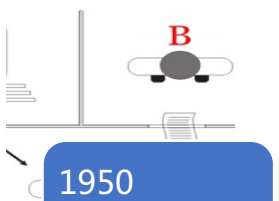
覃立波

导师：车万翔

哈尔滨工业大学

社会计算与信息检索研究中心

人机对话系统简史



1950

- 图灵测试

1966

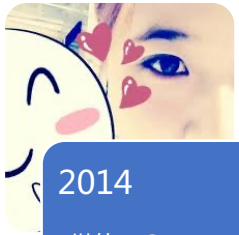
- ELIZA：MIT 开发的“心里医生”

1994

- ALICE：基于模式匹配的聊天机器人

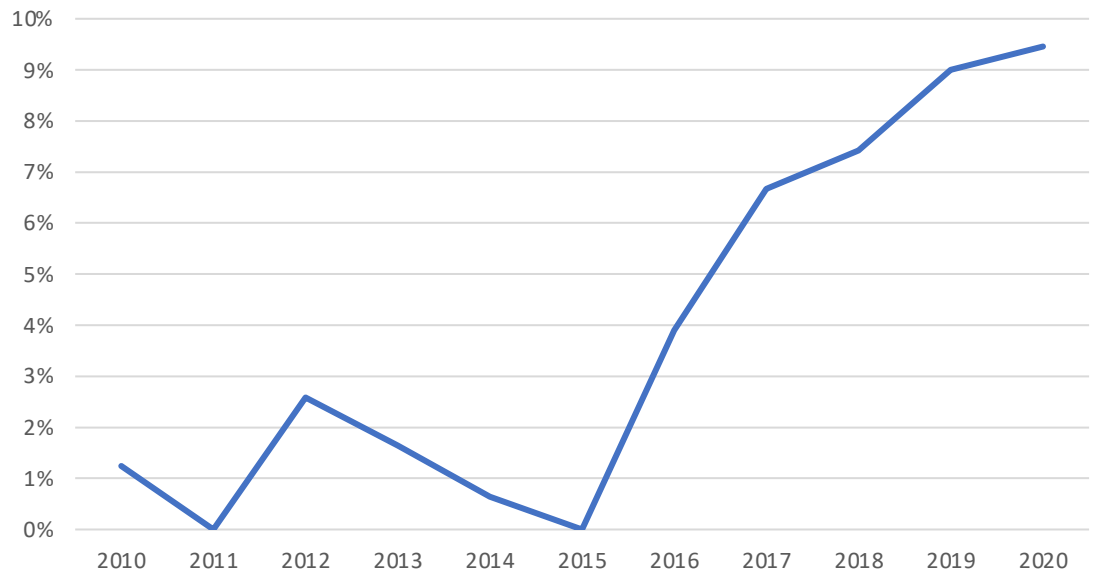
2011

- Siri
- Google Now

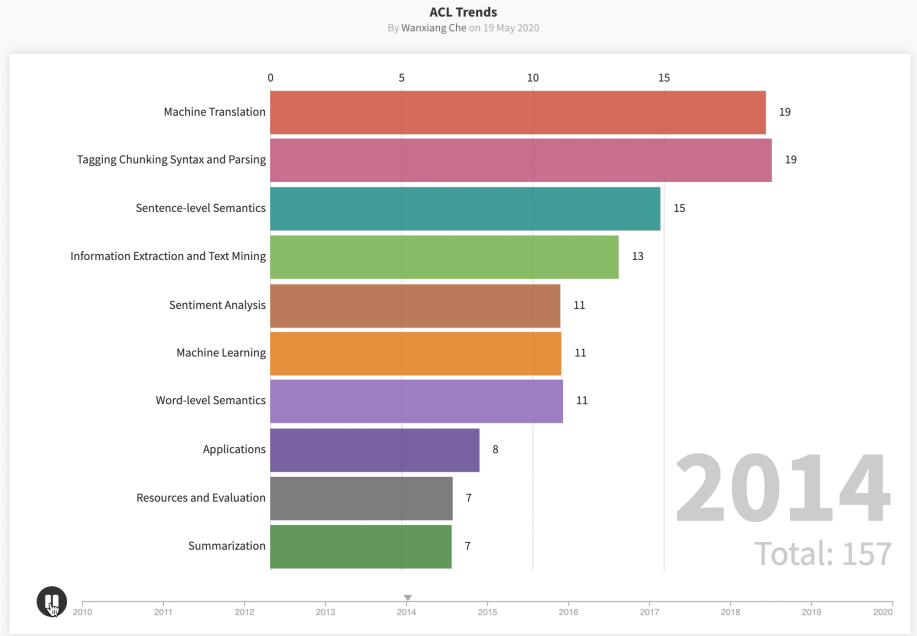


2014

- 微软：Cortana、小冰
- Amazon ALEXA



人机对话在ACL会议上异军突起



https://mp.weixin.qq.com/s/0VJOXstLRuSYFNZbpXsTYA

人机对话系统的四大功能

	任务型 Task	聊天 Chat	知识问答 Knowledge	推荐 Recommen-dation
目的	完成任务或动作	闲聊	知识获取	信息推荐
领域	特定域（垂类）	开放域	开放域	特定域
以话轮数评价	越少越好	话轮越多越好	越少越好	越少越好
应用	虚拟个人助理	娱乐、情感陪护	客服、教育	个性化推荐
典型系统	Siri、Cortana、Google Assistant、度秘	小冰、笨笨	Watson、Wolfram Alpha	阿里小蜜

任务型对话系统具有广泛应用

语义交互无处不在

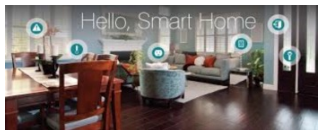
智能硬件 | 汽车 | 个人助理 | 咨询类机器人









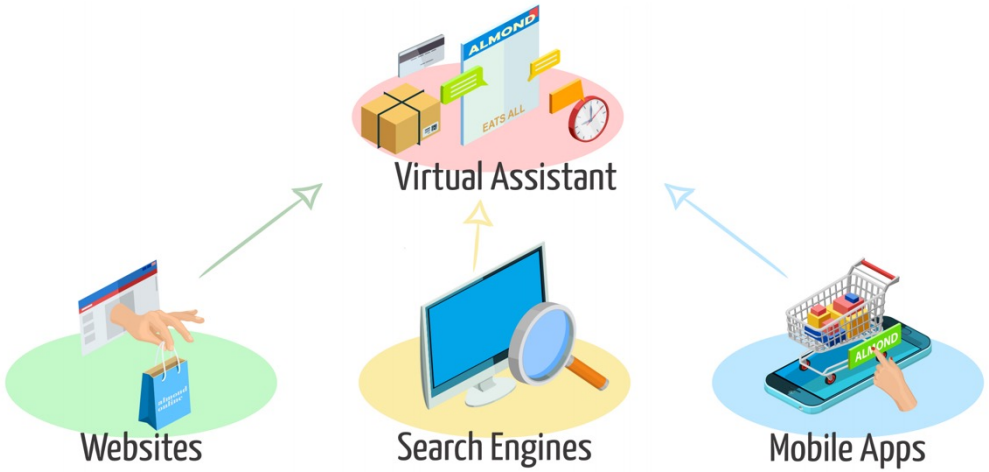



无人机
智能家居

车载语音
智能汽车

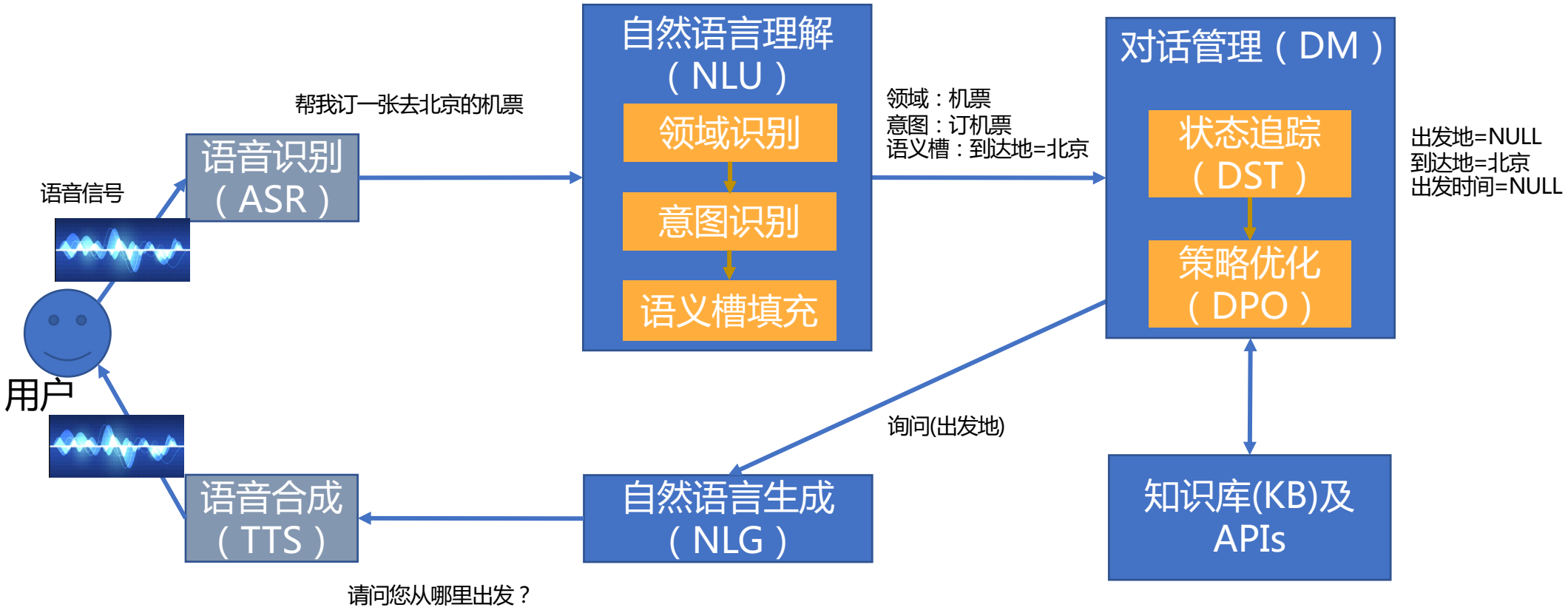
订餐
出行

导诊机器人
电商客服

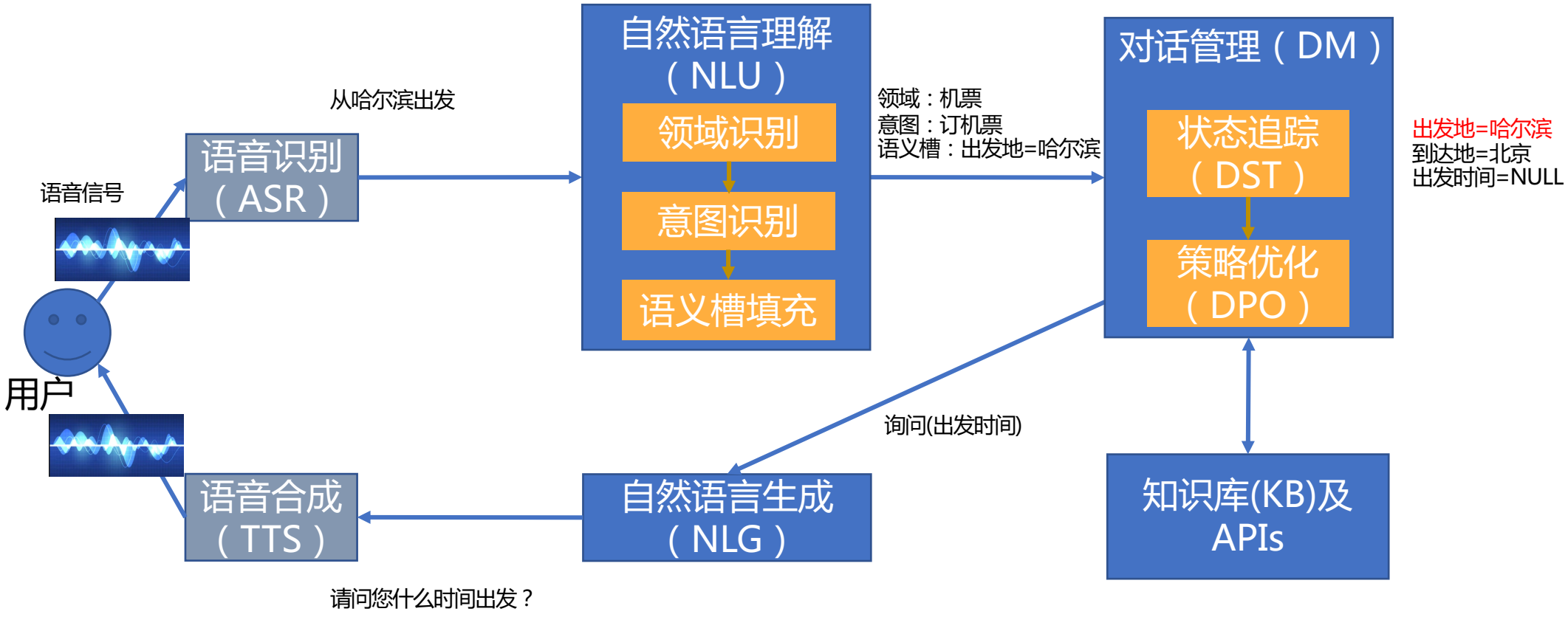


Virtual Assistants Might Eat the Internet.
-- Chris Manning (2020)

任务型对话系统的结构（Pipeline系统）

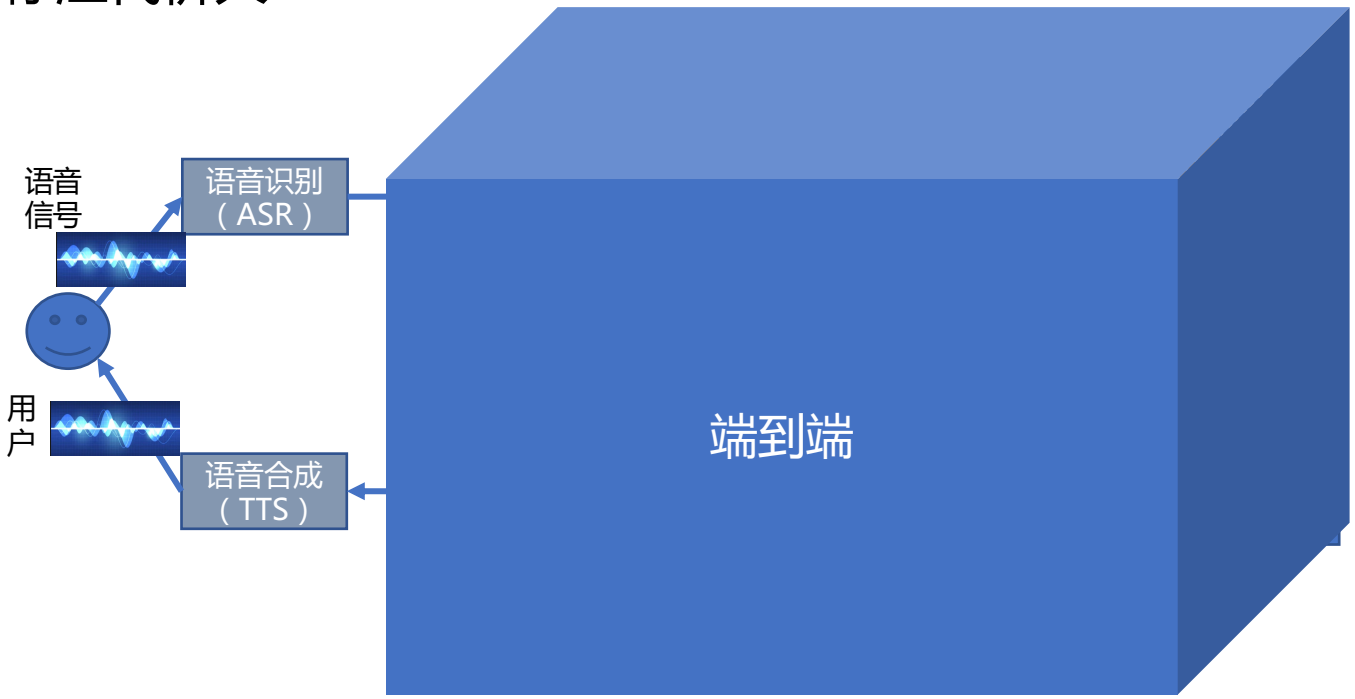


任务型对话系统的结构（Pipeline系统）



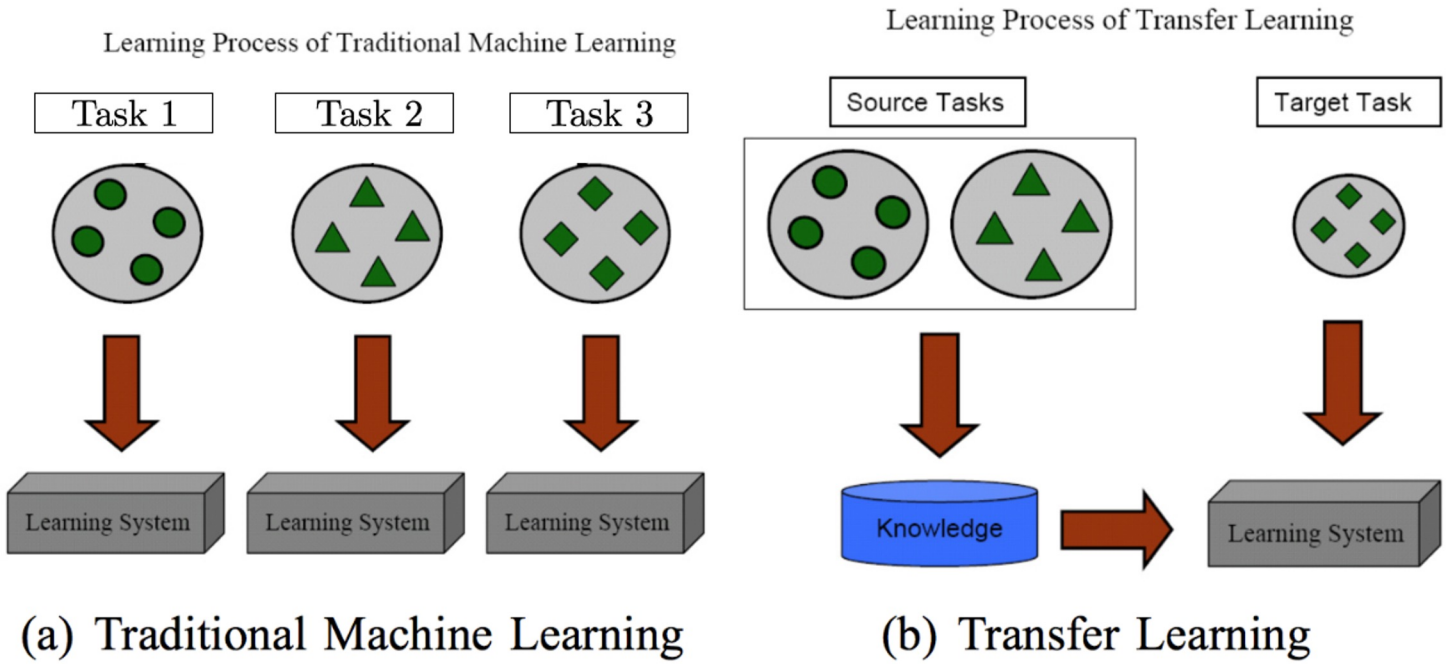
任务型对话系统的结构（端到端系统）

- Pipeline系统的缺点
  - 错误级联
  - 各模块数据标注代价大



面临的主要问题

- 大规模、高质量的标注数据较难获得，尤其在对话领域更为突出
- 不同任务、不同领域、不同语言存在着共享知识，可以充分利用



(a) Traditional Machine Learning (b) Transfer Learning

Pan and Yang (2010)

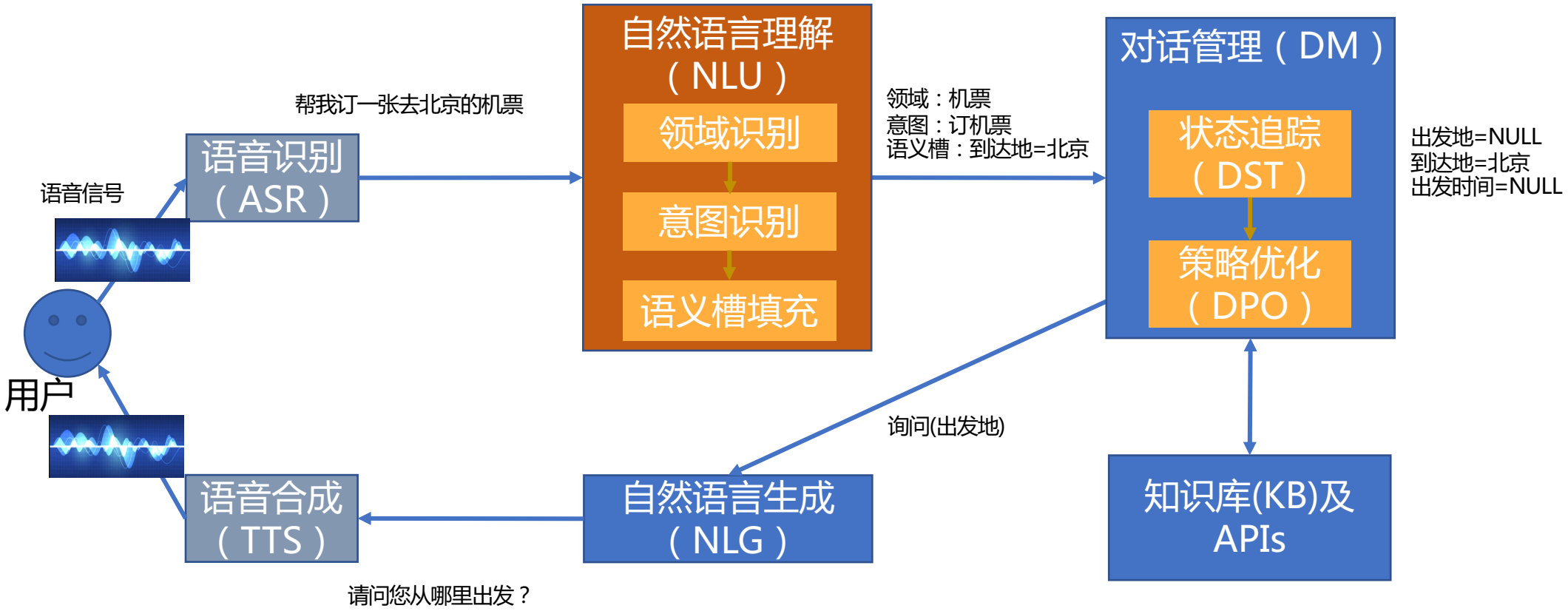
提纲

- 基于迁移学习的任务型对话系统
  - 跨任务迁移
  - 跨领域迁移
  - 跨语言迁移
- 总结及趋势展望

提纲

- 基于迁移学习的任务型对话系统
  - 跨任务迁移
  - 跨领域迁移
  - 跨语言迁移
- 总结及趋势展望

NLU模块多个任务串行执行

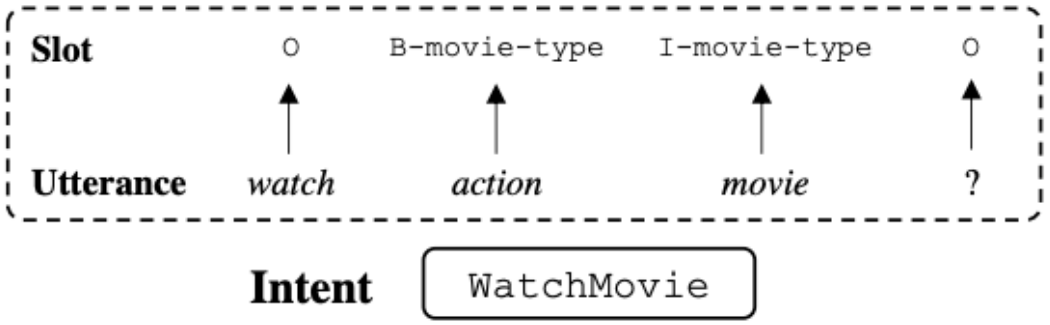


各项任务示例

领域	意图	语义槽	举例
火车	查询	出发地、到达地	厦门到福建建阳的火车是几点呢
	预定	出发地、到达地、出发时间、车次	订一张明天哈尔滨到北京的Z15
地图	路线	出发地、到达地	南山医院到北大医院路线
	位置	POI	北京西站在哪
短信	发信息	接收人、内容	发短信给叶少云晚上要不要出去玩
	发送联系人	接收人、联系人	把哥哥的电话发给殷龙

各项任务互相帮助

- 传统意图识别和槽填充串行，既会引出错误级联，也无法利用共有的知识
- 意图识别与槽填充不是相互独立而是紧密联系的
  - 例如：如果这句话意图是 WatchMovie，那么这句话包含的Slot槽值应该是电影相关而不是音乐相关，反之亦然



多任务联合学习

Task A

Intent Detection

Task B

Slot Filling

前人工作：共享编码的双任务学习

- 首次使用RNN-based (GRU)的方法联合建模意图识别和槽填充任务
- GRU的每一个时刻出来的向量进行槽填充任务
- GRU编码句子后通过max-pooling层得到该句的表示进行意图识别
- 通过共享的GRU层来进行两个任务的联合学习，从而隐式的学习两者的关系

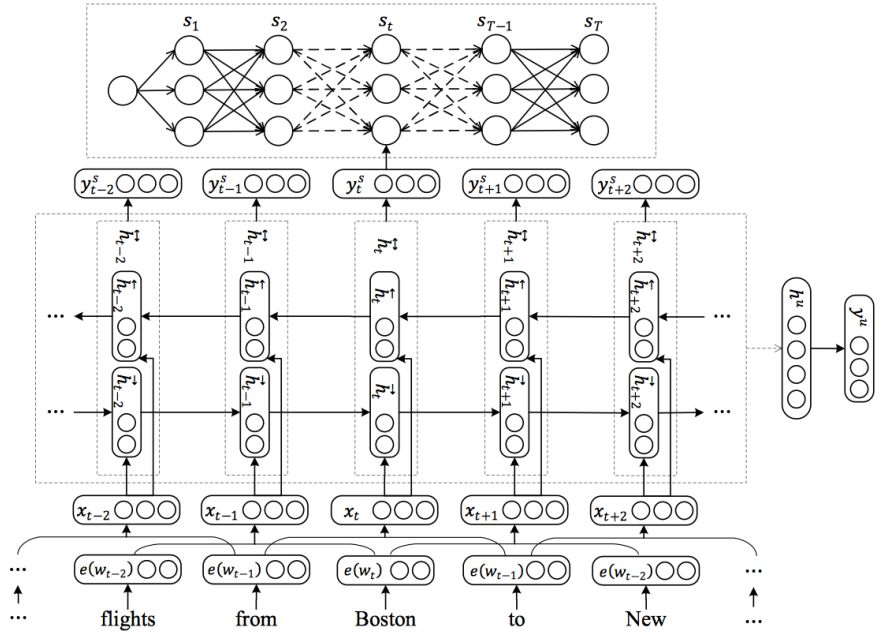
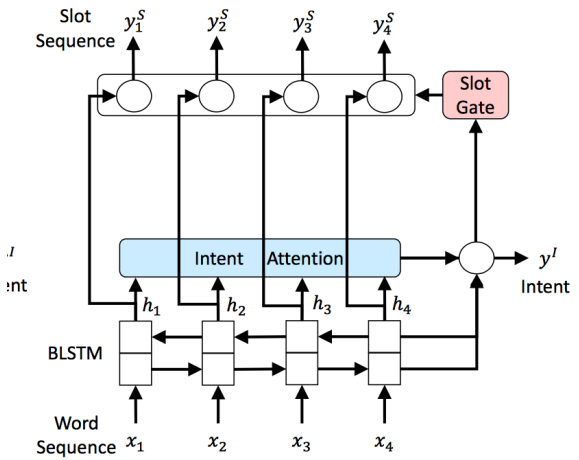


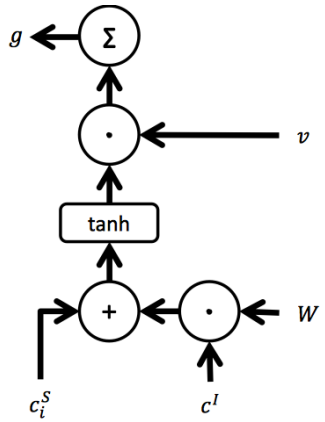
Figure 1: The structure of our model

Xiaodong Zhang and Houfeng Wang. A Joint Model of Intent Determination and Slot Filling for Spoken Language Understanding. IJCAI2016

前人工作：用意图控制槽填充

- 首次利用Slot-gate机制来显式的建模了槽填充任务和意图识别任务之间的关系
  - *g*越大，表示Intent和Slots的关系越大
- 不足
  - 基于gate机制的交互并不充分
  - 难以捕获每个token的信息





Slot Gate $g = \sum v \cdot \tanh(c_i^S + W \cdot c^I)$

W^S: matrix for output layer
b^S : bias for output layer

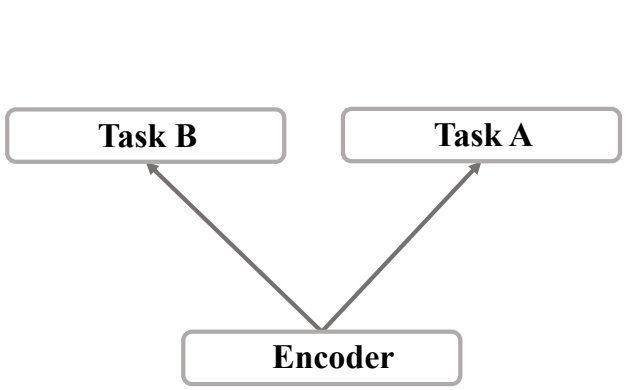
Slot Prediction $y_i^S = softmax(W^S(h_i + c_i^S) + b^S)$ ⟶ $y_i^S = softmax(W^S(h_i + \boldsymbol{g} \cdot c_i^S) + b^S)$

$\boldsymbol{g}$ will be larger if slot and intent are better related

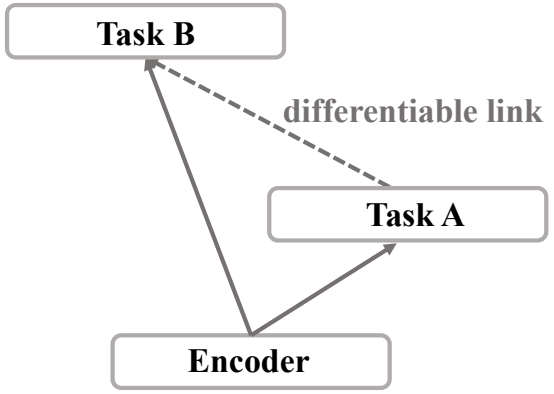
Chih-Wen Goo, Guang Gao. et al. Slot-Gated Modeling for Joint Slot Filling and Intent Prediction. NAACL2018.

基于Stack-propagation的联合学习

- Stack-propagation
  - 一种多任务学习框架
  - 任务之间有层次依赖关系



(a) Multi-task framework



(b) Stack-propagation

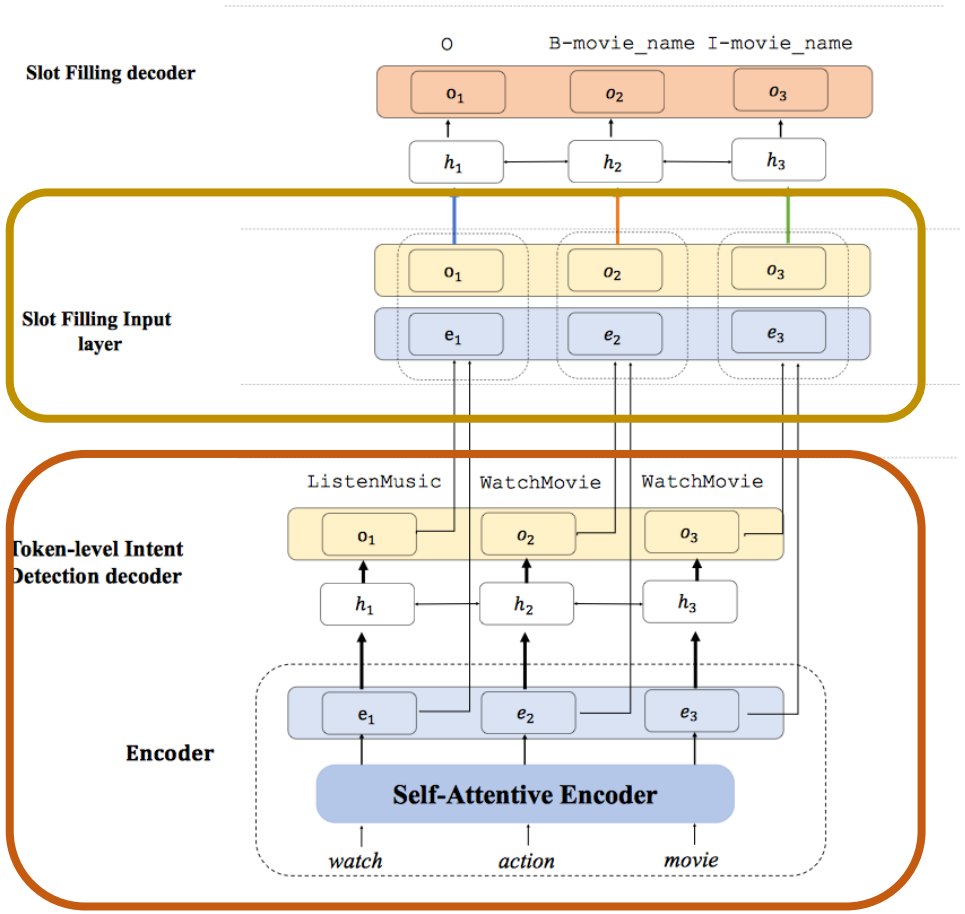
Libo Qin, Wanxiang Che, Yangming Li, Haoyang Wen and Ting Liu. A Stack-Propagation Framework with Token-Level Intent Detection for Spoken Language Understanding. EMNLP 2019.

基于Stack-propagation的联合学习

- 系统架构图

- Stack-propagation引入意图信息
  - 显示的利用intent信息去指导slot预测

- Token-level 意图识别
  - 提供token级别的intent信息
  - 充分利用监督信号，提高性能



Libo Qin, Wanxiang Che, Yangming Li, Haoyang Wen and Ting Liu. A Stack-Propagation Framework with Token-Level Intent Detection for Spoken Language Understanding. EMNLP 2019.

基于Stack-propagation的联合学习

实验结果

Model	SNIPS			ATIS		
	Slot (F1)	Intent (Acc)	Overall (Acc)	Slot (F1)	Intent (Acc)	Overall (Acc)
Joint Seq (Hakkani-Tür et al., 2016)	87.3	96.9	73.2	94.3	92.6	80.7
Attention BiRNN (Liu and Lane, 2016)	87.8	96.7	74.1	94.2	91.1	78.9
Slot-Gated Full Atten (Goo et al., 2018)	88.8	97.0	75.5	94.8	93.6	82.2
Slot-Gated Intent Atten (Goo et al., 2018)	88.3	96.8	74.6	95.2	94.1	82.6
Self-Attentive Model (Li et al., 2018)	90.0	97.5	81.0	95.1	96.8	82.2
Bi-Model (Wang et al., 2018)	93.5	97.2	83.8	95.5	96.4	85.7
CAPSULE-NLU (Zhang et al., 2019)	91.8	97.3	80.9	95.2	95.0	83.4
SF-ID Network (E et al., 2019)	90.5	97.0	78.4	95.6	96.6	86.0
Our model	**94.2***	**98.0***	**86.9***	**95.9***	**96.9***	**86.5***
Oracle (Intent)	96.1	-	-	96.0	-	-

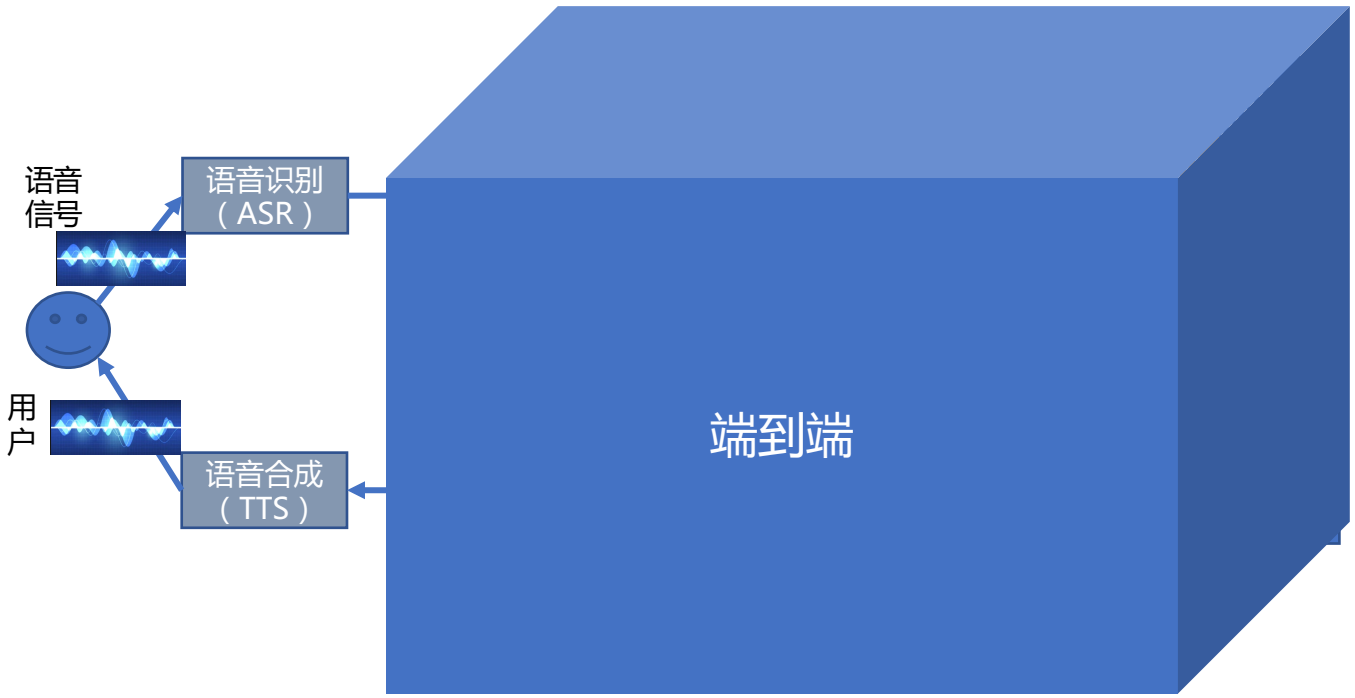
Model	SNIPS			ATIS		
	Slot (F1)	Intent (Acc)	Overall (Acc)	Slot (F1)	Intent (Acc)	Overall (Acc)
Our model	94.2	98.0	86.9	95.9	96.9	86.5
Intent detection (BERT)	-	97.8	-	-	96.5	-
Slot filling (BERT)	95.8	-	-	95.6	-	-
BERT SLU (Chen et al., 2019)	97.0	98.6	92.8	96.1	97.5	88.2
Our model + BERT	97.0	99.0	92.9	96.1	97.5	88.6

Libo Qin, Wanxiang Che, Yangming Li, Haoyang Wen and Ting Liu. A Stack-Propagation Framework with Token-Level Intent Detection for Spoken Language Understanding. EMNLP 2019.

提纲

- 迁移学习在任务型对话系统中的研究
  - 跨任务迁移
  - 跨领域迁移
  - 跨语言迁移
- 总结及趋势展望

端到端任务型对话系统



基于Seq2seq的端到端任务型对话系统

Address	Distance	POI type	POI	Traffic info
638 Amherst St	3 miles	grocery store	Sigona Farmers Market	car collision nearby
269 Alger Dr	1 miles	coffee or tea place	Cafe Venetia	car collision nearby
5672 barringer street	5 miles	certain address	5672 barringer street	no traffic
200 Alester Ave	2 miles	gas station	Valero	road block nearby
899 Ames Ct	5 miles	hospital	Stanford Childrens Health	moderate traffic
481 Amaranta Ave	1 miles	parking garage	Palo Alto Garage R	moderate traffic
145 Amherst St	1 miles	coffee or tea place	Teavana	road block nearby
409 Bollard St	5 miles	grocery store	Willows Market	no traffic

Knowledge Base (KB)

Seq2seq Model

Driver Address to the gas station.
Car Valero is located at 200 Alester Ave.
Driver OK , please give me directions via a route that avoids all heavy traffic.

Dialog History

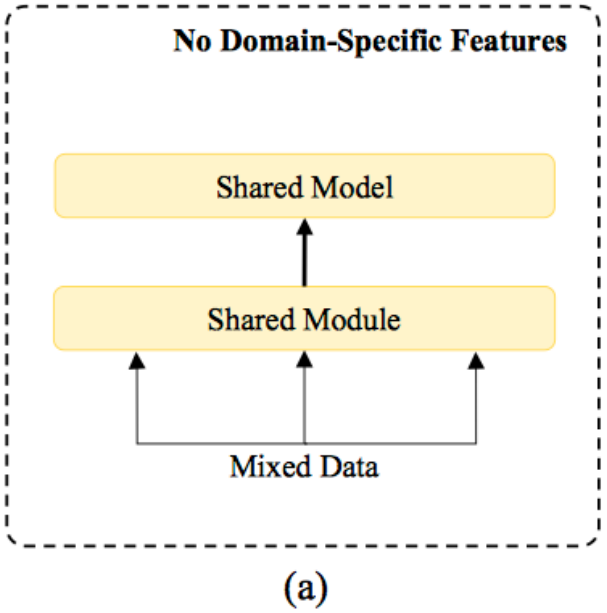
Response

Car Since there is a road block nearby, I found another route for you and I sent it on your screen.

存在多个领域的数据，如何充分利用？

前人工作：直接将多领域数据混合

- 在混合数据集上训练一个通用模型
  - 简单的共享一套参数难以捕获所有领域的特征
  - 忽略了领域特有的特征

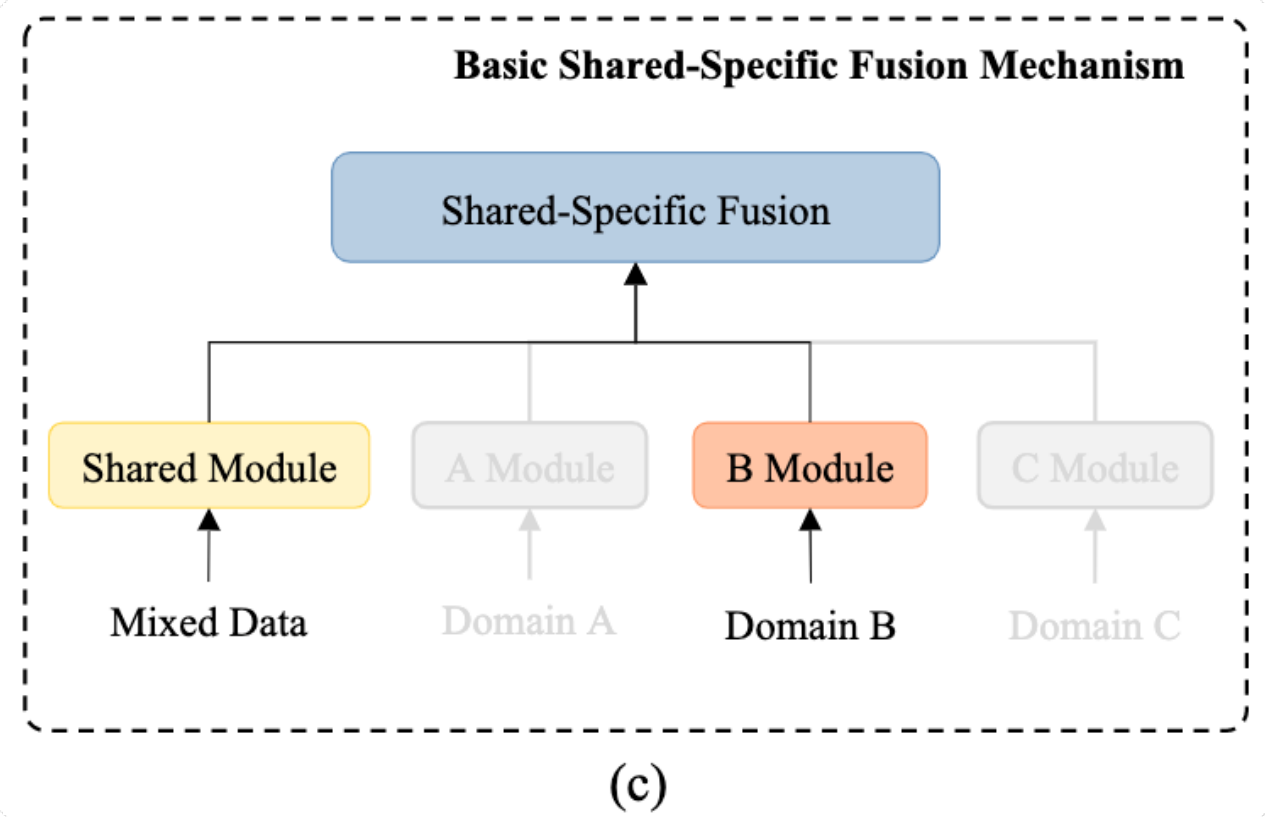


(a)

Key-Value Retrieval Networks for Task-Oriented Dialogue Mihai Eric, Christopher D.Manning. SIGDial 2018
Andrea Madotto , Chien-Sheng Wu , Pascale Fung Mem2Seq: Effectively Incorporating Knowledge Bases into End-to-End Task-Oriented Dialog Systems. ACL2018
Chien-Sheng Wu , Richard Socher, Caiming Xiong. Global-to-local Memory Pointer Networks for Task-Oriented Dialogue. ICLR2019

前人工作：Shared-private框架

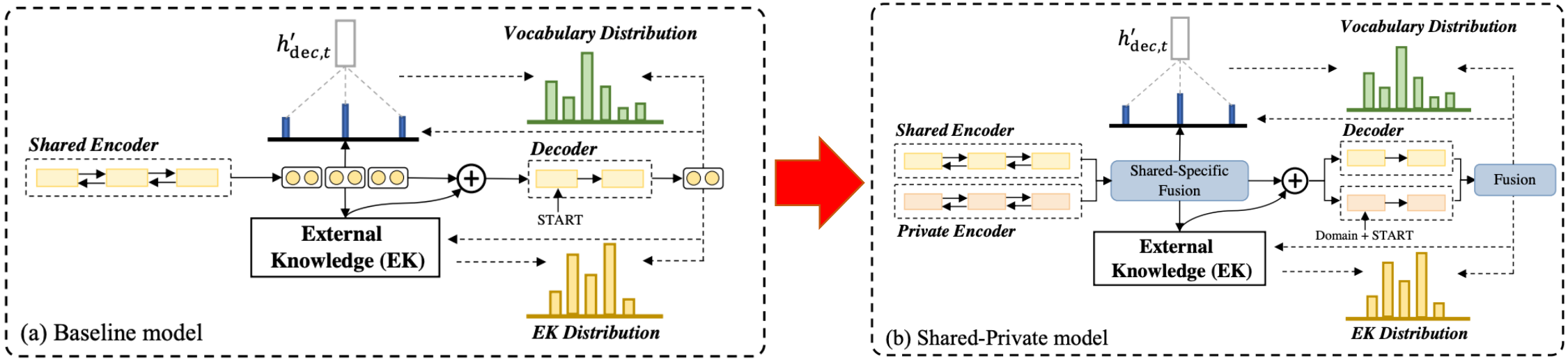
- 显示融入领域共享的知识和领域无关的知识



Pengfei Liu, Xipeng Qiu and Xuanjing Huang. Adversarial Multi-task Learning for Text Classification. ACL2017
Jiang Guo, Darsh J Shah and Regina Barzilay. Multi-Source Domain Adaptation with Mixture of Experts. EMNLP2018

前人工作：Shared-private框架

- 每一个样例X都会过shared和它特有的private的encoder-decoder模型
  - Shared：捕获领域共有的特征
  - Private：捕捉领域特定的特征

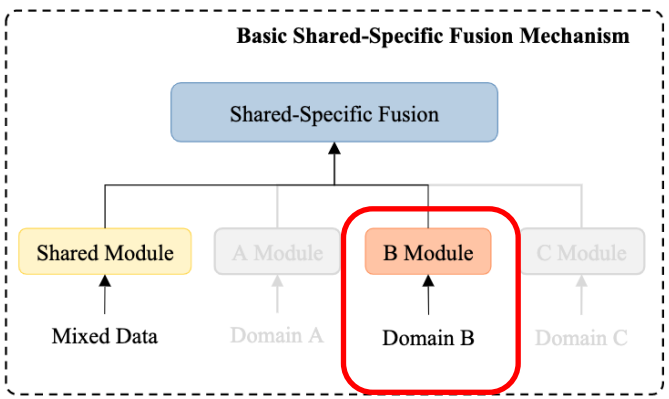


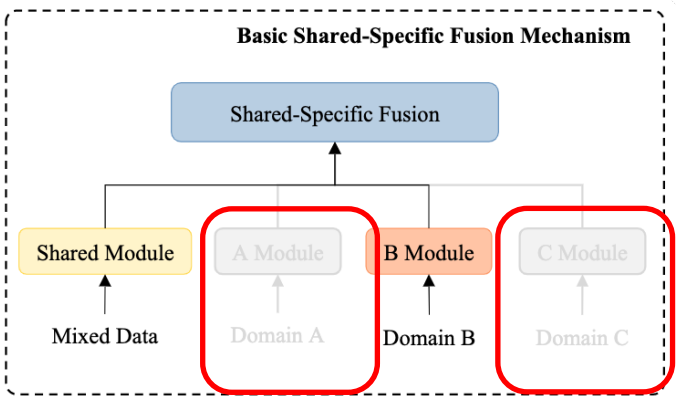
Pengfei Liu, Xipeng Qiu and Xuanjing Huang. Adversarial Multi-task Learning for Text Classification. ACL2017

Jiang Guo, Darsh J Shah and Regina Barzilay. Multi-Source Domain Adaptation with Mixture of Experts. EMNLP2018

基本的Shared-private框架的缺点

- 某一领域可能缺少训练数据
- 忽略领域之间的细粒度关系
  - 如：电影和音乐更近，和订机票比较远

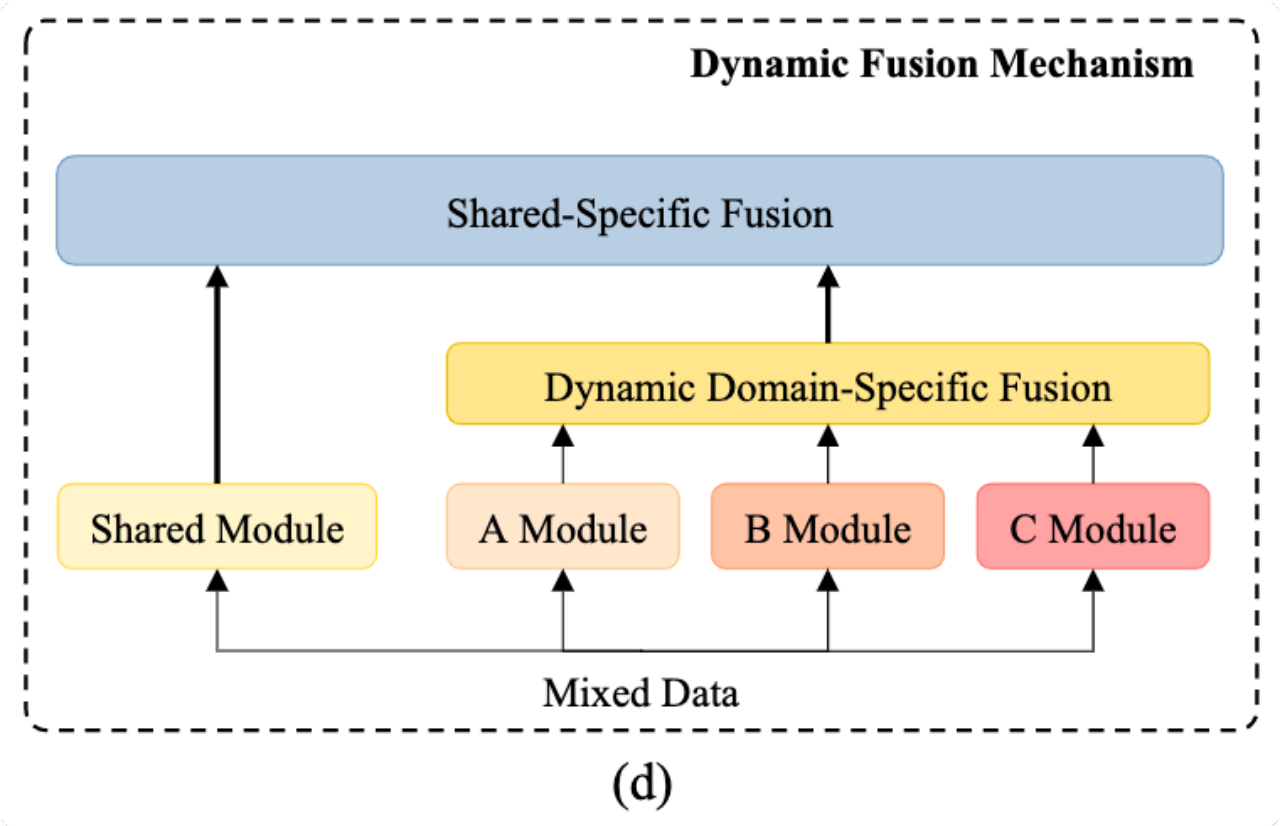




Libo Qin, Xiao Xu, Wanxiang Che, Yue Zhang, Ting Liu. Dynamic Fusion Network for Multi-Domain End-to-end Task-Oriented Dialog. ACL 2020.

动态聚合框架

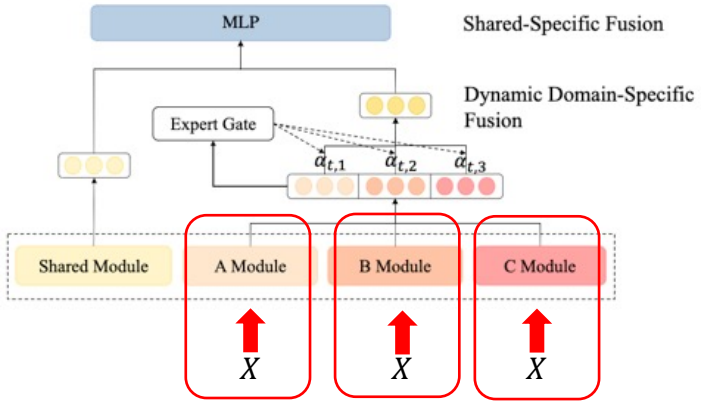
- 提出动态聚合函数显示建模领域之间的联系



Libo Qin, Xiao Xu, Wanxiang Che, Yue Zhang, Ting Liu. Dynamic Fusion Network for Multi-Domain End-to-end Task-Oriented Dialog. ACL 2020.

动态聚合网络 (DF-Net)

- 动态聚合层
  - 考虑领域之间的关系，从而更好地利用所有领域的知识
  - a_t代表当前instance 与每个领域之间的相关度
  - 使用加权来得到最终的领域私有特征



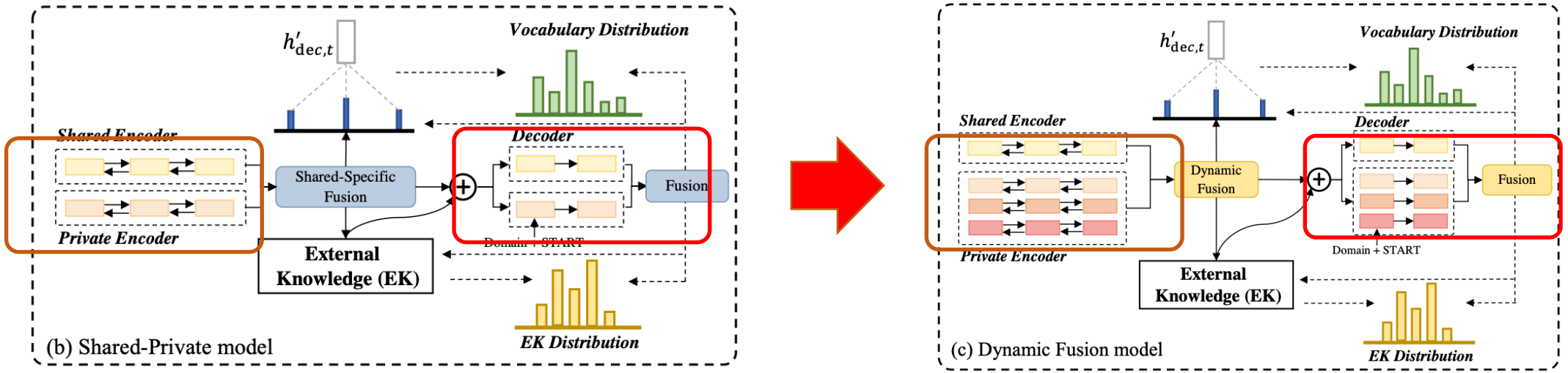
$$\alpha_t = \text{Softmax}(W * h^d_{\text{dec},t} + b).$$

$$h^{d_f}_{\text{dec},t} = \sum_i \alpha_{t,i} h^{d_i}_{\text{dec},t}.$$

Libo Qin, Xiao Xu, Wanxiang Che, Yue Zhang, Ting Liu. Dynamic Fusion Network for Multi-Domain End-to-end Task-Oriented Dialog. ACL 2020.

动态聚合网络 (DF-Net)

动态聚合层



Libo Qin, Xiao Xu, Wanxiang Che, Yue Zhang, Ting Liu. Dynamic Fusion Network for Multi-Domain End-to-end Task-Oriented Dialog. ACL 2020.

实验

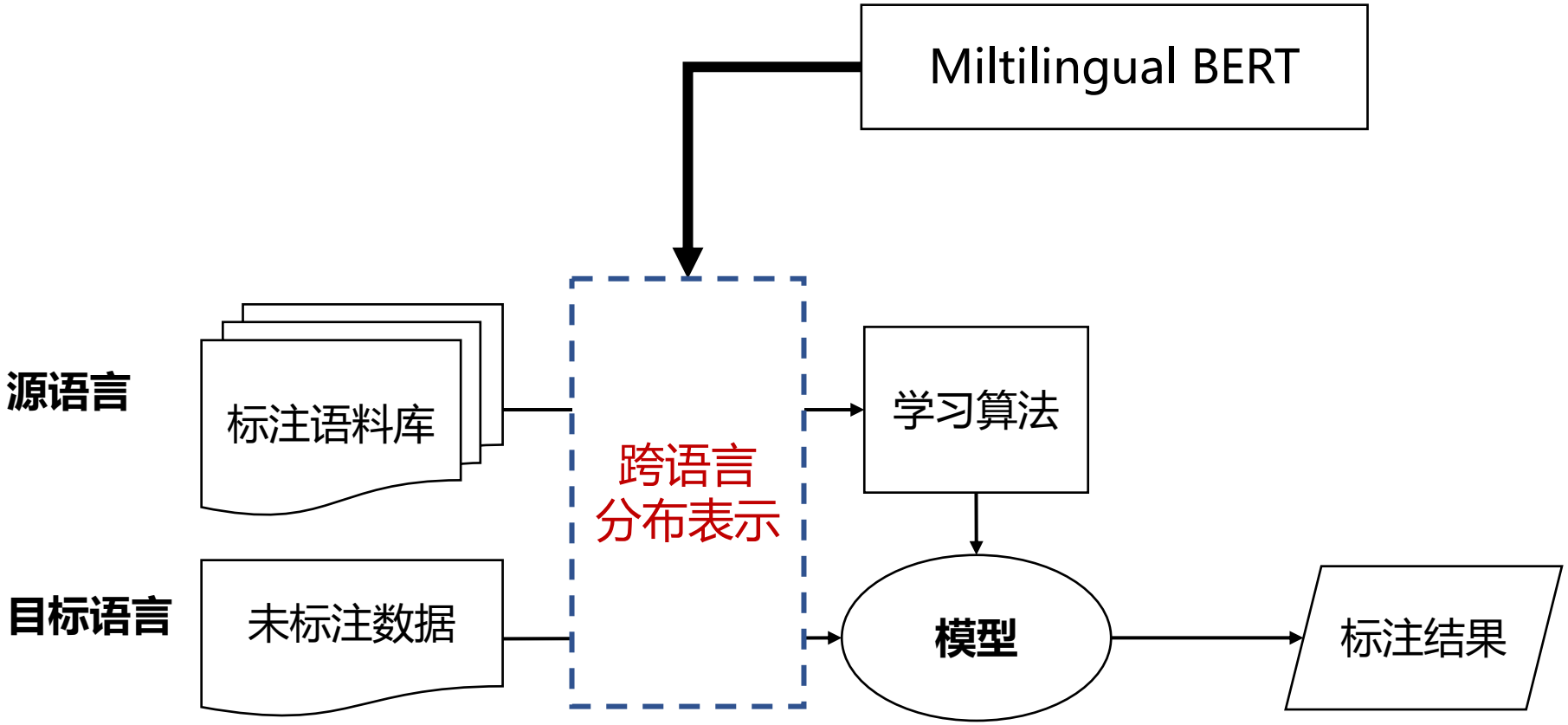
	SMD					Multi-WOZ 2.1				
Model	BLEU	F1	Navigate F1	Weather F1	Calendar F1	BLEU	F1	Restaurant F1	Attraction F1	Hotel F1
Mem2Seq (Madotto et al., 2018)	12.6	33.4	20.0	32.8	49.3	6.6	21.62	22.4	22.0	21.0
DSR (Wen et al., 2018)	12.7	51.9	52.0	50.4	52.1	9.1	30.0	33.4	28.0	27.1
KB-retriever (Qin et al., 2019)	13.9	53.7	54.5	52.2	55.6	-	-	-	-	-
GLMP (Wu et al., 2019a)	13.9	60.7	54.6	56.5	72.5	6.9	32.4	38.4	24.4	28.1
Shared-Private framework (Ours)	13.6	61.7	56.3	56.5	72.8	6.6	33.8	39.8	26.0	28.3
Dynamic Fusion framework (Ours)	**14.4***	**62.7***	**57.9***	**57.6***	**73.1***	**9.4***	**35.1***	**40.9***	**28.1***	**30.6***

Libo Qin, Xiao Xu, Wanxiang Che, Yue Zhang, Ting Liu. Dynamic Fusion Network for Multi-Domain End-to-end Task-Oriented Dialog. ACL 2020.

提纲

- 迁移学习在任务型对话系统中的研究
  - 跨任务迁移
  - 跨领域迁移
  - 跨语言迁移
- 总结及趋势展望

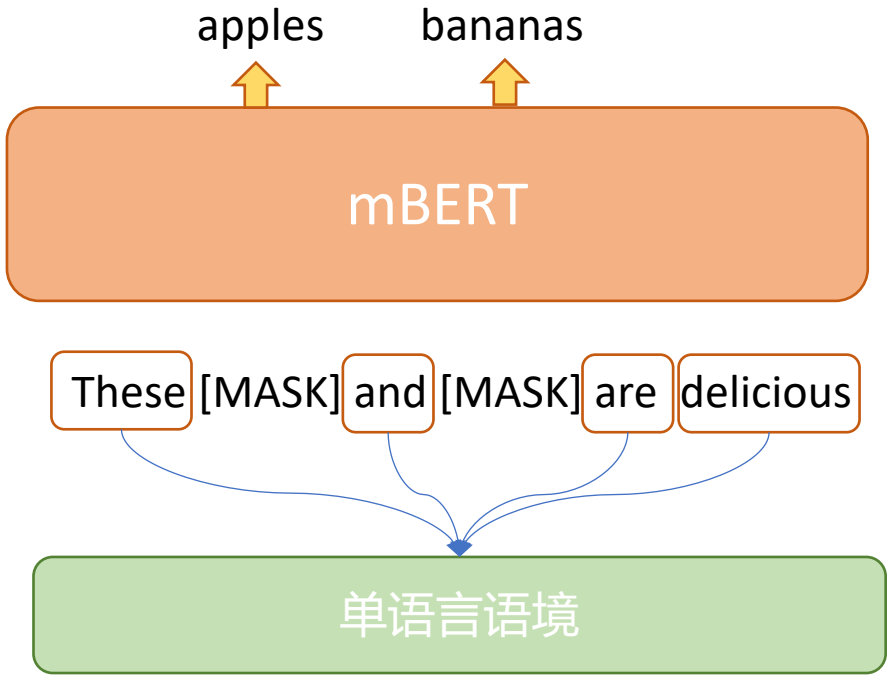
Zero-shot 跨语言框架



Beto, Bentz, Becas: The Surprising Cross-Lingual Effectiveness of BERT. EMNLP 2019

mBERT缺点

- mBERT预训练是多个语言数据混合
- 但对于单个句子来说仍是单语言语境
- 没有任何的跨语言表示对齐信号



前人工作：AIML

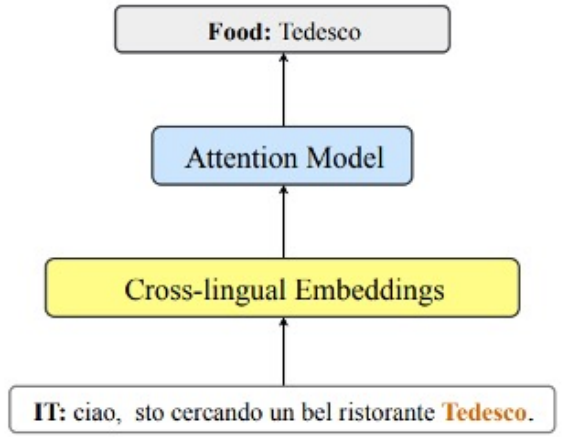
- 提出构造code-switch数据进行fine-tune mBERT，可以使模型隐式对齐两个语言之间的表示空间

原训练句子：hi, im looking for a nice German restaurant.

Attention找到分值最高的单词：German

进行替换后的训练数据：hi, im looking for a nice Tedesco restaurant.

Zero-shot测试过程

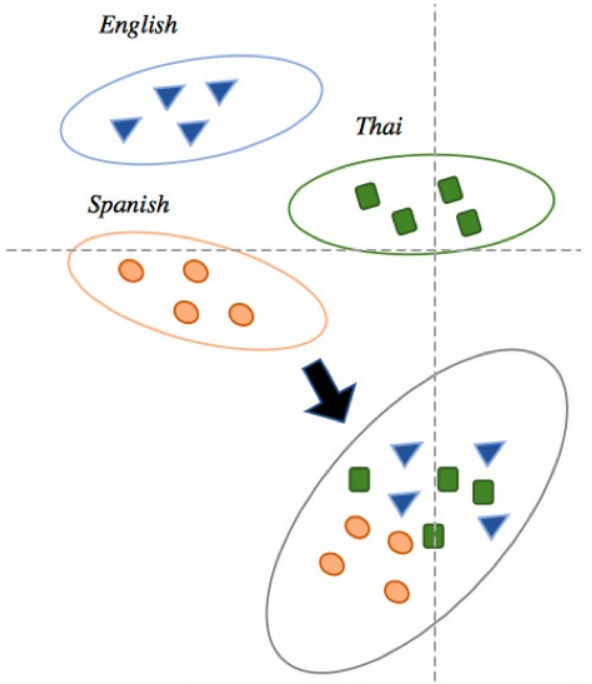


Zero-shot Test Step

Zihan Liu , Genta Indra Winata, Zhaojiang Lin, Peng Xu, Pascale Fung. Attention-Informed Mixed-Language Training for Zero-shot Cross-lingual Task-oriented Dialogue Systems. AAAI2020

我们的方法

- AIML Training 缺点
  - 需要设计启发规则 (Attention)找到该替换的单词，并且只替换一个单词，损失信息
  - 一次训练只能部署到一种目标语言，因为每次训练过程只翻译一种特定语言

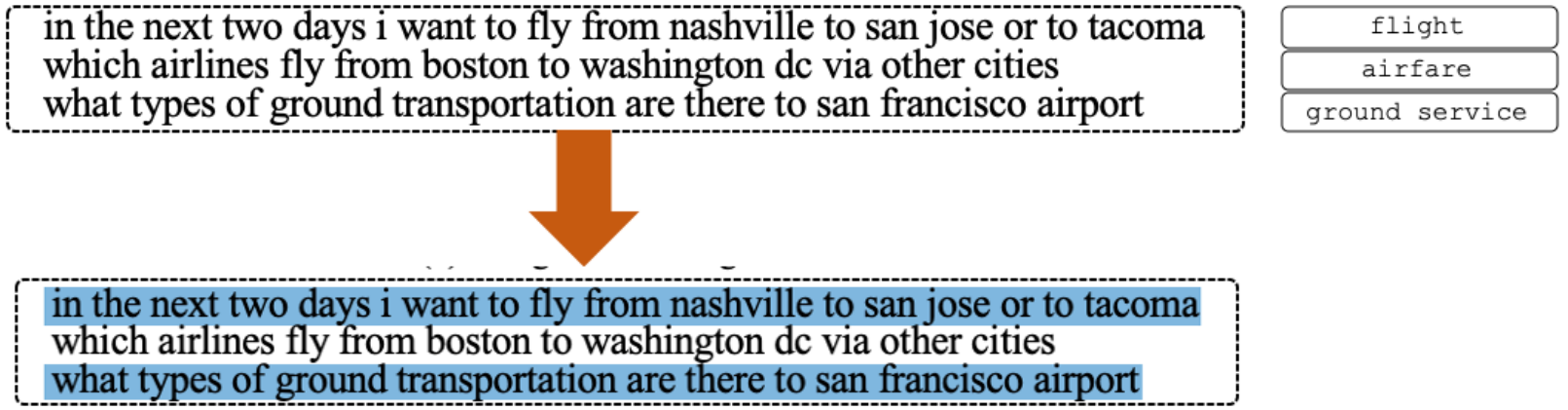


- 我们提出一种Multi-lingual的code-switch数据增广方法更好的fine-tuning mBERT
  - 可以同时对齐多个语言，一次训练，多次部署
  - 采用随机替换，不需要任何的启发规则

Libo Qin, Minheng Ni, Yue Zhang and Wanxiang Che. CoSDA-ML: Multi-Lingual Code-Switching Data Augmentation for Zero-Shot Cross-Lingual NLP. IJCAI 2020.

方法流程：选择句子

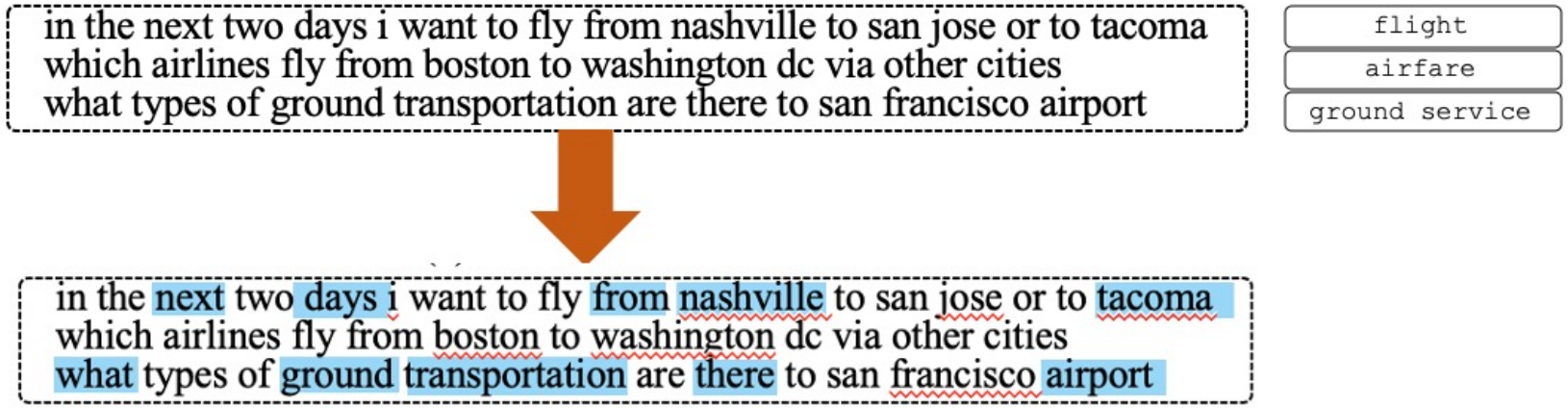
Fine-tuning的
原始训练数据
(e.g., 意图分类)



Libo Qin, Minheng Ni, Yue Zhang and Wanxiang Che. CoSDA-ML: Multi-Lingual Code-Switching Data Augmentation for Zero-Shot Cross-Lingual NLP. IJCAI 2020.

方法流程：选择单词

Fine-tuning的
原始训练数据
(e.g., 意图分类)

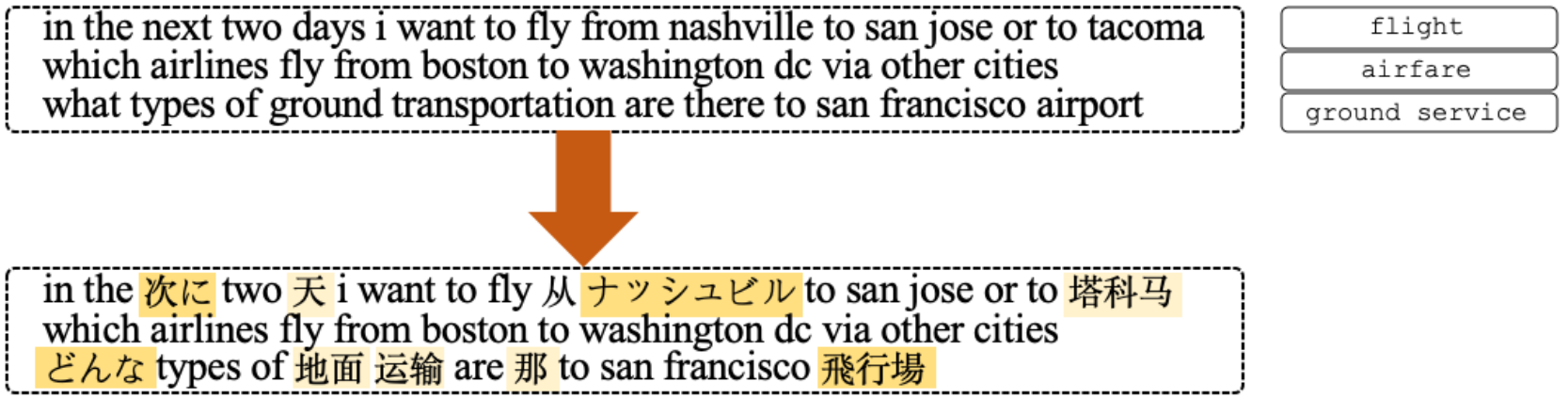


Libo Qin, Minheng Ni, Yue Zhang and Wanxiang Che. CoSDA-ML: Multi-Lingual Code-Switching Data Augmentation for Zero-Shot Cross-Lingual NLP. IJCAI 2020.

方法流程：单词替换

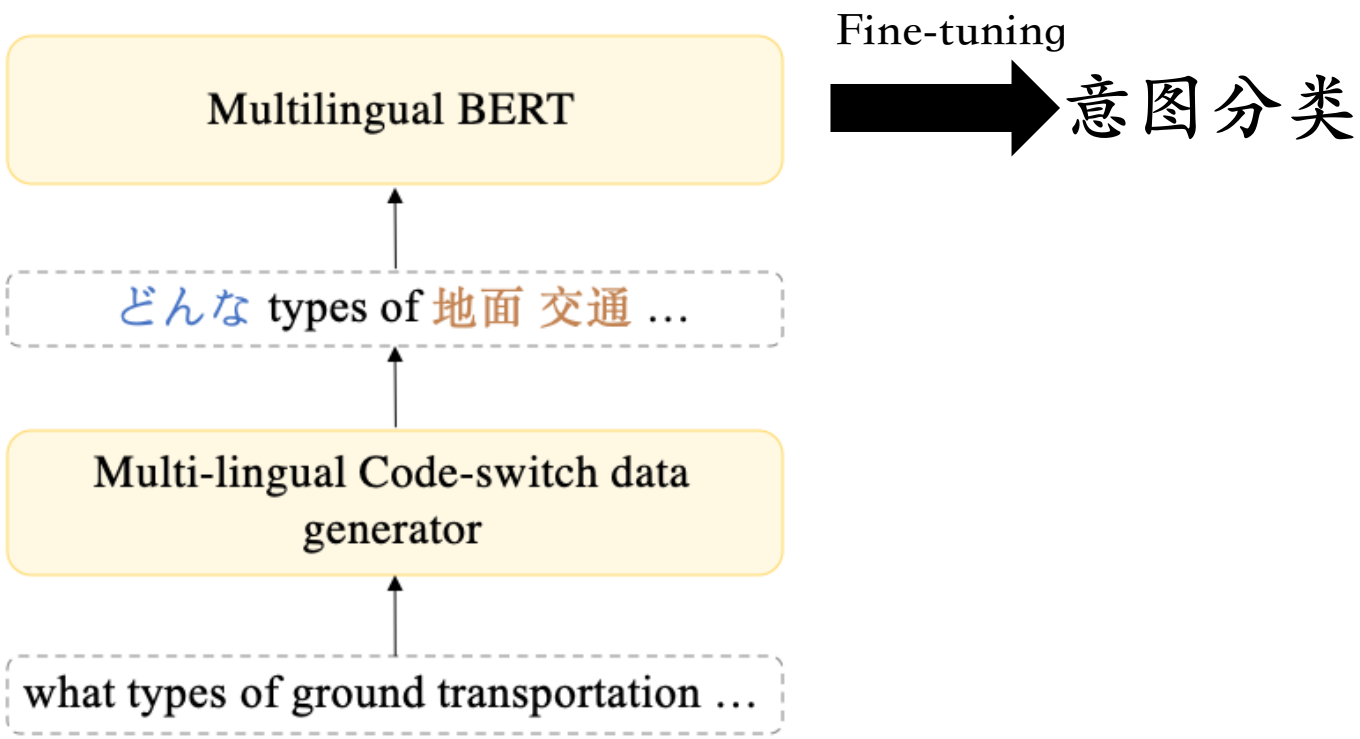
Fine-tuning的原始训练数据(e.g., 意图分类)

增广后的fine-tuning数据(e.g., 意图分类)



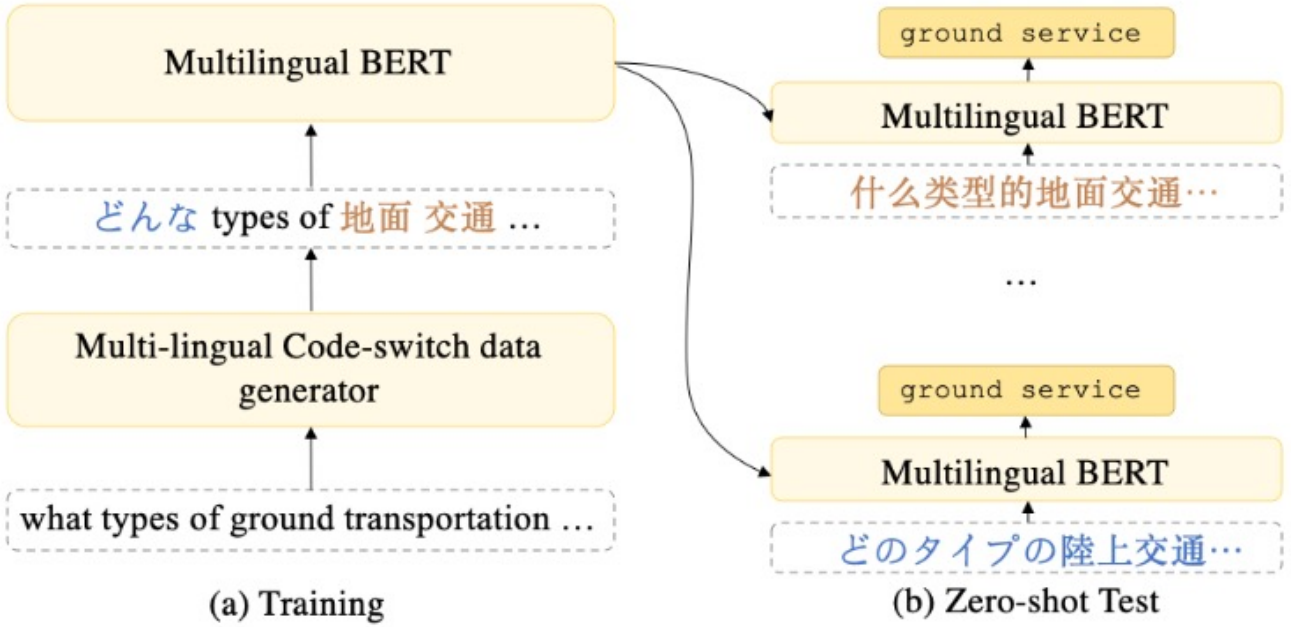
Libo Qin, Minheng Ni, Yue Zhang and Wanxiang Che. CoSDA-ML: Multi-Lingual Code-Switching Data Augmentation for Zero-Shot Cross-Lingual NLP. IJCAI 2020.

方法示例



Libo Qin, Minheng Ni, Yue Zhang and Wanxiang Che. CoSDA-ML: Multi-Lingual Code-Switching Data Augmentation for Zero-Shot Cross-Lingual NLP. IJCAI 2020.

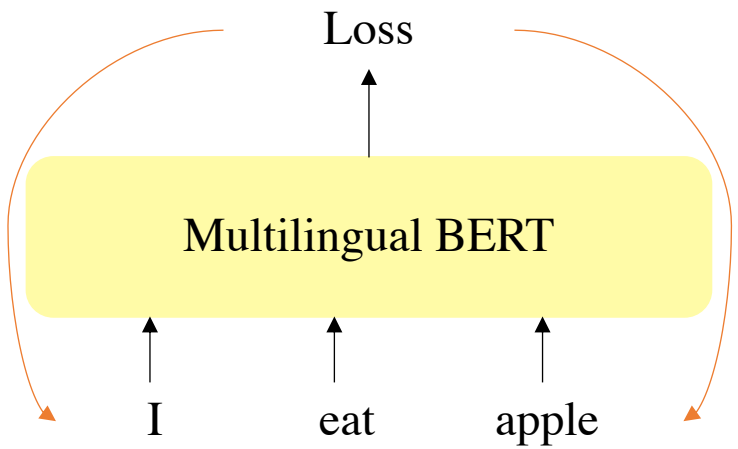
方法示例

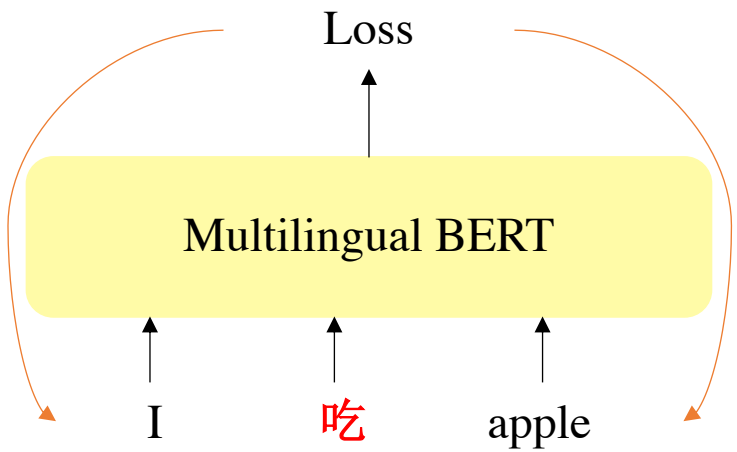


Libo Qin, Minheng Ni, Yue Zhang and Wanxiang Che. CoSDA-ML: Multi-Lingual Code-Switching Data Augmentation for Zero-Shot Cross-Lingual NLP. IJCAI 2020.

方法背后的直觉

- 相同的loss，不同的输入
- 拉近不同语言的表示





Libo Qin, Minheng Ni, Yue Zhang and Wanxiang Che. CoSDA-ML: Multi-Lingual Code-Switching Data Augmentation for Zero-Shot Cross-Lingual NLP. IJCAI 2020.

实验

Model	Spanish		Thai	
	Intent acc.	Slot F1	Intent acc.	Slot F1
Multi. CoVe [Yu *et al.*, 2018]	53.9	19.3	70.7	35.6
Attention-Informed Mixed Training [Liu *et al.*, 2019b]	86.5	74.4	70.6	28.5
XLM from Liu *et al.* [2019b]	62.3	42.3	31.6	7.9
+ CoSDA-ML	94.3	68.8	85.8	36.2
mBERT from Liu *et al.* [2019b]	73.7	51.7	28.2	10.6
+ CoSDA-ML (Static)	92.8	75.2	74.8	28.1
+ CoSDA-ML	**94.8***	**81.3***	**81.7***	**38.4***

Table 4: Slot filling and Intent detection experiments.

Libo Qin, Minheng Ni, Yue Zhang and Wanxiang Che. CoSDA-ML: Multi-Lingual Code-Switching Data Augmentation for Zero-Shot Cross-Lingual NLP. IJCAI 2020.

实验

Model	German			Italian		
	slot acc.	joint goal acc.	request acc.	slot acc.	joint goal acc.	request acc.
XL-NBT [Chen *et al.*, 2018]	55.0	30.8	68.4	72.0	41.2	81.2
Attention-Informed Mixed Training [Liu *et al.*, 2019b]	69.5	32.2	86.3	69.5	31.4	85.2
XLM from Liu *et al.* [2019b]	58.0	16.3	75.7	-	-	-
+CoSDA-ML	77.4	48.7	88.3	-	-	-
mBERT from Liu *et al.* [2019b]	57.6	15.0	75.3	54.6	12.6	77.3
+CoSDA-ML	**82.7***	**62.6***	**95.7***	**83.4***	**67.1***	**93.9***

Table 5: Dialog State Tracking experiments. "-" represents the absence of languages in the XLM models and we cannot report the results.

Libo Qin, Minheng Ni, Yue Zhang and Wanxiang Che. CoSDA-ML: Multi-Lingual Code-Switching Data Augmentation for Zero-Shot Cross-Lingual NLP. IJCAI 2020.

提纲

- 迁移学习在任务型对话系统中的研究
  - 跨任务迁移
  - 跨领域迁移
  - 跨语言迁移
- 总结及趋势展望

总结

- 充分利用多种知识缓解数据不足的问题
- 探索了迁移学习在任务型对话系统中的应用
  - 跨任务 (通过Stack-propagation显示交互提高多项任务的性能)
  - 跨领域 (提出动态聚合shared-private框架提升多领域端到端任务型对话系统性能)
  - 跨语言 (提出multi-lingual code-switching数据增广方法提升Zero-shot跨语言能力)

论文和代码

- A Stack-Propagation Framework with Token-Level Intent Detection for Spoken Language Understanding. EMNLP 2019
  - 论文：https://www.aclweb.org/anthology/D19-1214.pdf
  - 代码：https://github.com/LeePleased/StackPropagation-SLU
- Dynamic Fusion Network for Multi-Domain End-to-end Task-Oriented Dialog. ACL 2020
  - 论文： https://arxiv.org/pdf/2004.11019.pdf
  - 代码：https://github.com/LooperXX/DF-Net
- CoSDA-ML: Multi-Lingual Code-Switching Data Augmentation for Zero-Shot Cross-Lingual NLP. IJCAI 2020
  - 论文： https://arxiv.org/pdf/2006.06402.pdf
  - 代码： https://github.com/kodenii/CoSDA-ML


flickr
facebook
YouTube
Instagram
twitter
LinkedIn
skype

谢谢大家！