

北京航空航天大学博士学位论文答辩

弱标注下的自然语言语义解析技术研究

博士生：刘 乾

指导老师：赵沁平、周彬



虚拟现实技术与系统国家重点实验室

STATE KEY LABORATORY OF VIRTUAL REALITY TECHNOLOGY AND SYSTEMS

汇报提纲

1

研究背景与意义

2

研究现状与问题

3

研究目标与内容

4

研究成果与创新

1. 研究背景与意义：语义解析

语义解析旨在将自然语言解析成形式化的中间表示。中间表示既可以是受语言学启发设计的**语义表征**，也可以是面向任务可执行的**计算机程序**。



1. 研究背景与意义：语义解析

语义解析允许没有编程背景的用户通过自然语言完成复杂任务，提高了用户完成任务的效率，具有**很高的研究意义与应用价值**。

智能服务

Google 谷歌的ceo是谁

全部 新闻 图片 地图 视频 更多

找到约 27,900,000 条结果 (用时 0.44 秒)

谷歌 / 首席执行官

孙达尔·皮柴

2015年10月2日-

用户还搜索了

安佳丽·皮查伊 拉里·佩奇 萨蒂亚·纳德拉

自然语言驱动的开放问答

智能金融

"measles in the uk"

Disease Outbreaks Around the World

Measles in the UK (2008-2014)

Year	Measles	Mumps	Rubella	Whooping Cough
2008	10	5	2	1
2009	15	5	2	1
2010	10	5	2	1
2011	15	5	2	1
2012	35	5	2	1
2013	40	5	2	1
2014	10	5	2	1

Impact of Diseases

Disease	Impact
Measles	10.0
Whooping Cough	4.0
Mumps	1.0
Rubella	1.0

自然语言驱动的数据分析

智能家居

Sure, I can move "Take a walk outside" to tomorrow at 8:00 AM. Is that right?

Take a walk outside
Tomorrow • 8:00 – 9:00 AM
Bellevue Downtown Park

Yeah, that sounds great.

Okay, I've moved it to tomorrow.

Nicholas and his manager from 9:00 to 1:30 today. And uh, we're gonna need a room in Bellevue.

自然语言驱动的虚拟助手

1. 研究背景与意义：语义解析与代码生成

语义解析与代码生成关联密切，本文主要关注在**面向任务的语义解析**。



语义解析

Are there any jobs requiring NLP experience?

```
SELECT Jobs WHERE  
Experience = NLP
```

Text-to-SQL
[HLT 1990]

```
answer(job(experience,  
NLP))
```

Text-to-FunQL
[AAAI 1996]

```
(lambda $0:e(and_(job  
$0)(skill $0 NLP)))
```

Text-to-Prolog
[EMCL 2001]



代码生成

Return the k-th largest value in my_list.

```
sorted(my_list, reverse=False)[-k]
```

Text-to-Python
[ACL 2005]

```
Collections.sort(my_list,  
Collections.reverseOrder());  
my_list.get(k)
```

Text-to-Java
[ACL 2015]

1. 研究背景与意义：text-to-SQL语义解析

年份	城市	国家	参赛国家数量
1896	雅典	希腊	14
...
2008	北京	中国	204

国家	区域	面积
希腊	欧洲	13万
...
中国	亚洲	960万

知识库



奥运会总共举行过多少次？
按参赛国家数量排序城市。
希腊在哪一年举行了其最后一届夏季奥运会？

```
SELECT COUNT (*)
```

```
SELECT 城市 ORDER BY 参赛国家数量
```

```
SELECT 年份 WHERE 国家 = 希腊 ORDER BY 年份 DESC LIMIT 1
```

32
雅典, 巴黎, ...
2004

汇报提纲

1

研究背景与意义

2

研究现状与问题

3

研究目标与内容

4

研究成果与创新

2. 研究现状及问题

围绕text-to-SQL语义解析，从**规则驱动语义解析**，**强监督语义解析**，**弱监督语义解析**和**对话式语义解析**展开本领域的调研与分析。

基于规则的语义解析

规则驱动 语义解析

- ❑ 标注输入：无
- ❑ 标注输出：无

基于机器学习的语义解析

2.1 强监督 语义解析

- ❑ 标注输入：单轮语句
- ❑ 标注输出：程序

2.2 弱监督 语义解析

- ❑ 标注输入：单轮语句
- ❑ 标注输出：结果

2.3 对话式 语义解析

- ❑ 标注输入：多轮对话
- ❑ 标注输出：程序

2.1 强监督语义解析

强监督语义解析方法可以分为基于**组合文法**的方法，基于**机器翻译**的方法和基于**程序语法**的方法。

基于组合文法的方法 [UAI 2005]

希腊在哪一年举行了其
最后一届夏季奥运会？

语言标注

基于机器翻译的方法 [ACL 2015]

```
SELECT 年份 WHERE 国  
家 = 希腊 ORDER BY  
年份 DESC LIMIT 1
```

程序标注

基于程序语法的方法 [ACL 2017]

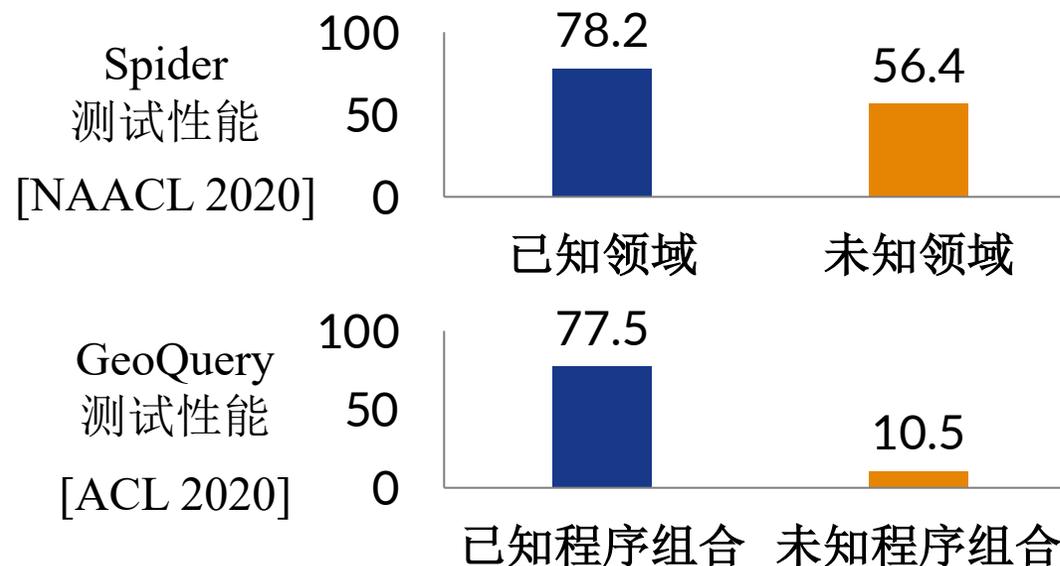
强监督语义解析方法需要程序标注

2.1 强监督语义解析

强监督语义解析数据集规模小，覆盖的知识库领域和程序组合有限，模型容易过拟合到**特定知识库领域**和**特定程序组合**上。

数据集	样本数量(个)
GeoQuery [AAAI 1996]	880
Jobs [ECML 2001]	640
ATIS [UAI 2005]	5,418
Overnight [ACL 2015]	13,682
Spider [EMNLP 2018]	10,181

最先进的基于程序语法的语义解析



数据集规模小，模型过拟合现象严重

2.2 弱监督语义解析

2015年，为降低语义解析模型的标注成本，扩展数据集规模，研究者们开始探索仅标注自然语言问题**答案**的**弱监督语义解析** [ACL 2015]。

数据集	样本数量(个)
WikiTableQuestions [ACL 2015]	22,033
SQA [ACL 2017]	17,553
WikiSQL [2017]	87,673

希腊在哪一年举行了其
最后一届夏季奥运会?

语言标注

2004

答案标注

弱监督语义解析数据集更容易规模化扩展

2.2 弱监督语义解析

弱监督语义解析方法可以分为基于**强化学习**的方法，基于**边缘似然**的方法和基于**可微分**的方法。

基于强化学习的方法 [NIPS 2018] **43.8**

↓ **10.2%**

基于边缘似然的方法 [ACL 2015] **37.1**

↓ **24.0%**

基于可微分的方法 [ICLR 2017] **34.2**

↓ **30.1%**

方法在WikiTableQuestions上的准确率

与程序标注模型准确率相比

答案标注监督弱，模型学习较困难

2.3 对话式语义解析

对话式语义解析是个新兴研究领域，已有方法主要有基于**分层编码**的、基于**复制操作**的和基于**数据流**的方法。

基于分层编码的方法 [NAACL 2018]

基于复制操作的方法 [EMNLP 2019]

基于数据流的方法 [TACL 2020]

希腊在哪一年举行了其
最后一届夏季奥运会？

对话标注

```
SELECT 年份 WHERE 国  
家 = 希腊 ORDER BY  
年份 DESC LIMIT 1
```

程序标注

那法国呢？

对话标注

```
SELECT 年份 WHERE 国  
家 = 法国 ORDER BY  
年份 DESC LIMIT 1
```

程序标注

对话式语义解析方法需要每轮自然语言的程序标注

2.3 对话式语义解析

对话式语义解析是个新兴研究领域，相关的数据集不多。目前主要有SParC [EMNLP 2018] 和 CoSQL [EMNLP 2019] 两个数据集。

自然语言 What are the name and budget of the departments with average instructor salary greater than the overall average?

SQL程序

```
SELECT T2.name, T2.budget
FROM instructor as T1 JOIN department
as T2 ON T1.department_id = T2.id
GROUP BY T1.department_id
HAVING avg(T1.salary) >
(SELECT avg(salary) FROM instructor)
```

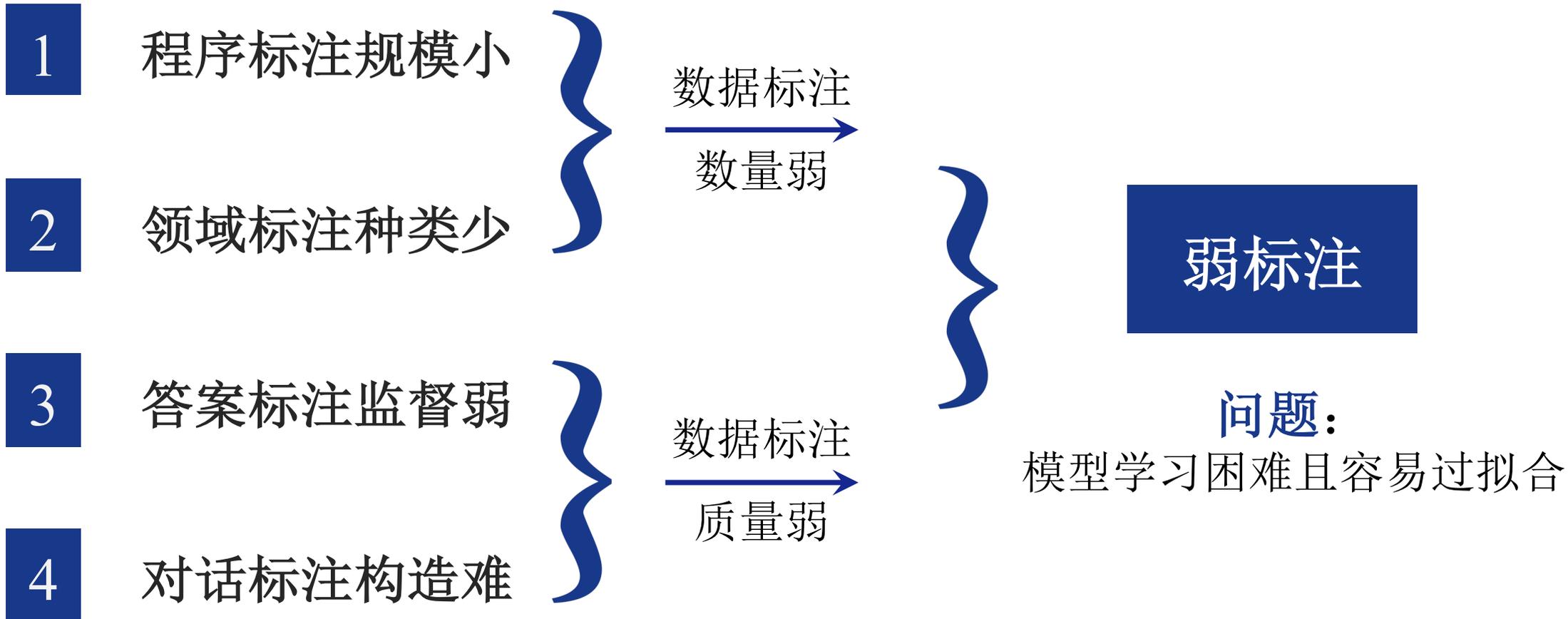
数据集	自然语言种类	
	语境无关	语境相关
SParC	47.5%	52.5%
CoSQL	68.2%	31.8%

每个程序标注需 **6** 分钟
[EMNLP 2018]

有**20%**数据偏离真实分布
[ACL 2021]

对话式语义解析数据集构造的成本与难度高

2. 研究现状与问题



汇报提纲

1

研究背景与意义

2

研究现状与问题

3

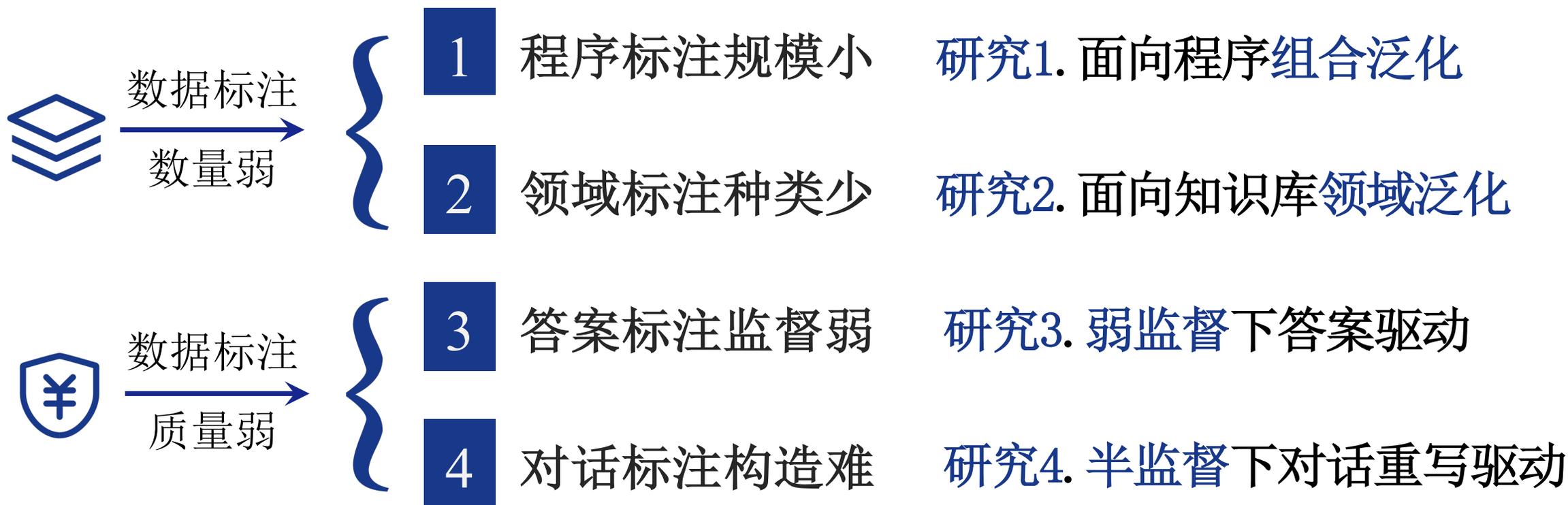
研究目标与内容

4

研究成果与创新

3. 研究目标与内容

研究目标: 在弱标注下提升自然语言语义解析模型的性能与泛化能力



汇报提纲

1 研究背景与意义

2 研究现状与问题

3 研究目标与内容

3.1 面向程序组合泛化的自然语言语义解析方法

3.2 面向知识库领域泛化的自然语言语义解析方法

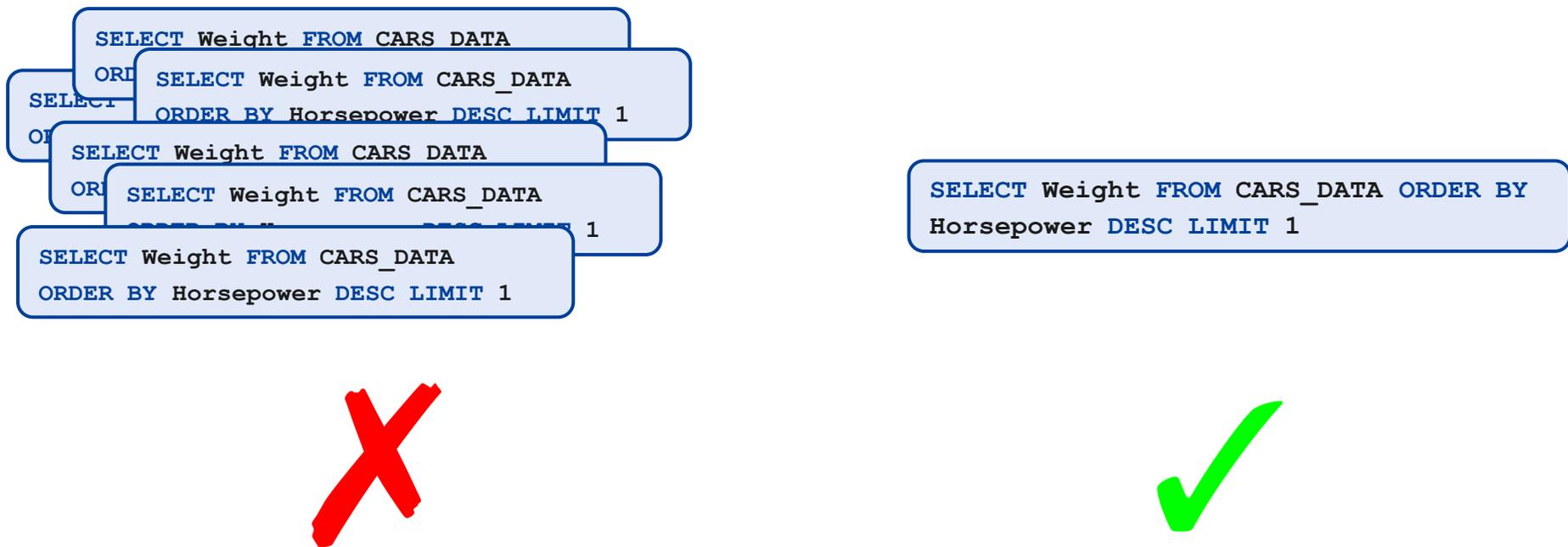
3.3 弱监督下答案驱动的自然语言语义解析方法

3.4 半监督下对话重写驱动的对话式自然语言语义解析方法

4 研究成果与创新

3.1 面向程序组合泛化的自然语言语义解析方法：研究目标

立足于强监督语义解析场景，面向语义解析模型**程序标注数量弱**的情况，本研究内容旨在提升语义解析模型**面向程序组合泛化**的能力。



3.1 面向程序组合泛化的自然语言语义解析方法：相关工作

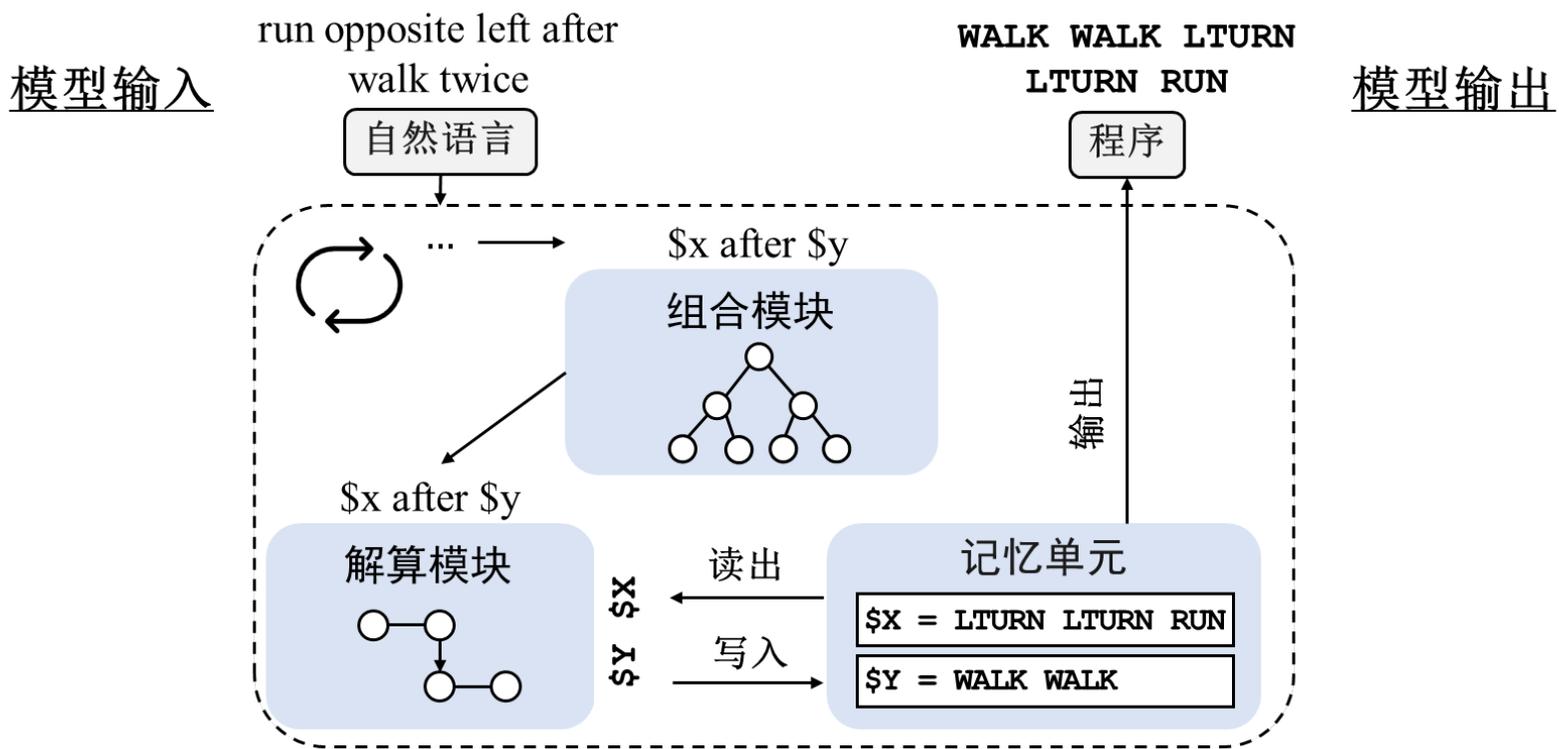
为达成该目标，前人工作关注在设计拥有组合泛化能力的神经网络架构。

模型	泛化测试1	泛化测试2	泛化测试3
<i>Seq2Seq</i>	1.2	2.5	13.8
<i>CNN</i>	69.2	56.7	0.0
<i>Syntactic Attention (MILA)</i>	91.0	28.9	15.2
<i>CGPS (Baidu Research)</i>	98.8	83.2	20.3
<i>GECA (MIT)</i>	86.0	82.0	-
<i>Meta Seq2Seq (FAIR)</i>	99.9	99.9	16.6
<i>Equivariant Seq2Seq (Cambridge)</i>	99.1	92.0	15.9

尚无模型能够全方位应对已有的组合泛化挑战

3.1 面向程序组合泛化的自然语言语义解析方法：总体结构

针对**程序标注规模小**的难点，受启发于人类的层次化抽象思维，本文提出**带记忆单元**的模型架构，通过多步迭代增强模型面向程序组合泛化的能力。



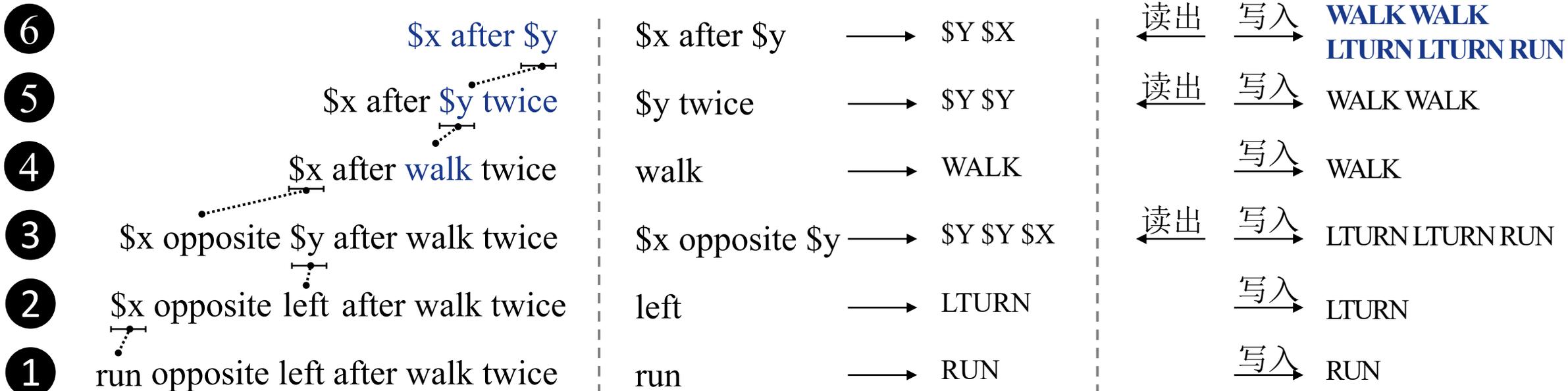
3.1 面向程序组合泛化的自然语言语义解析方法：多步迭代

本方法将程序的生成过程分解，通过模块多步迭代得到程序。

组合模块

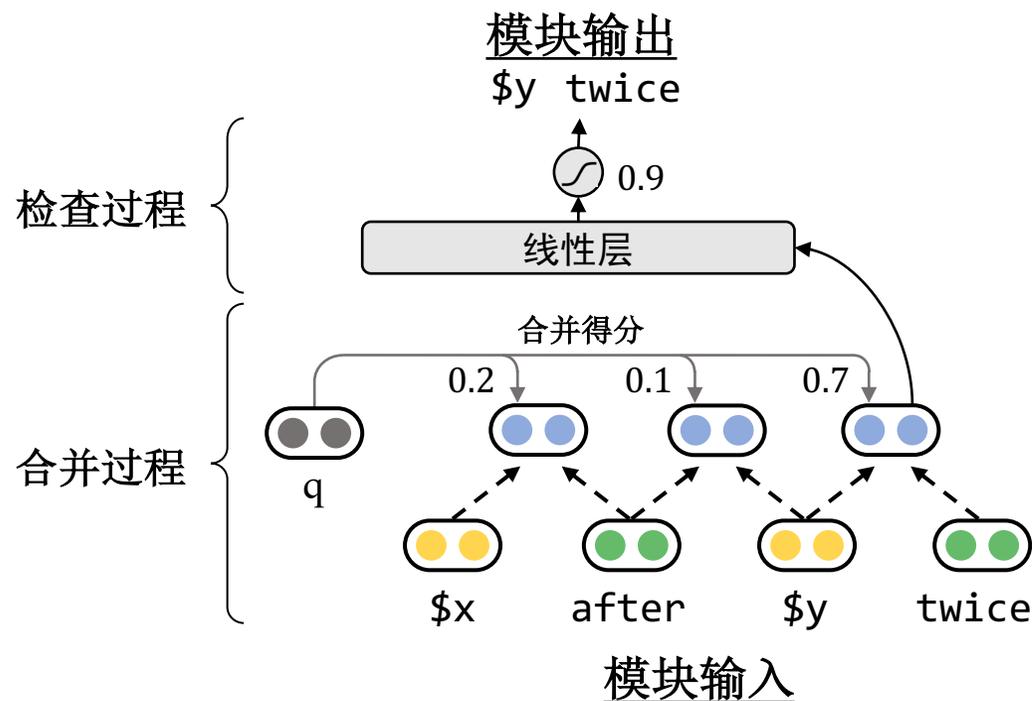
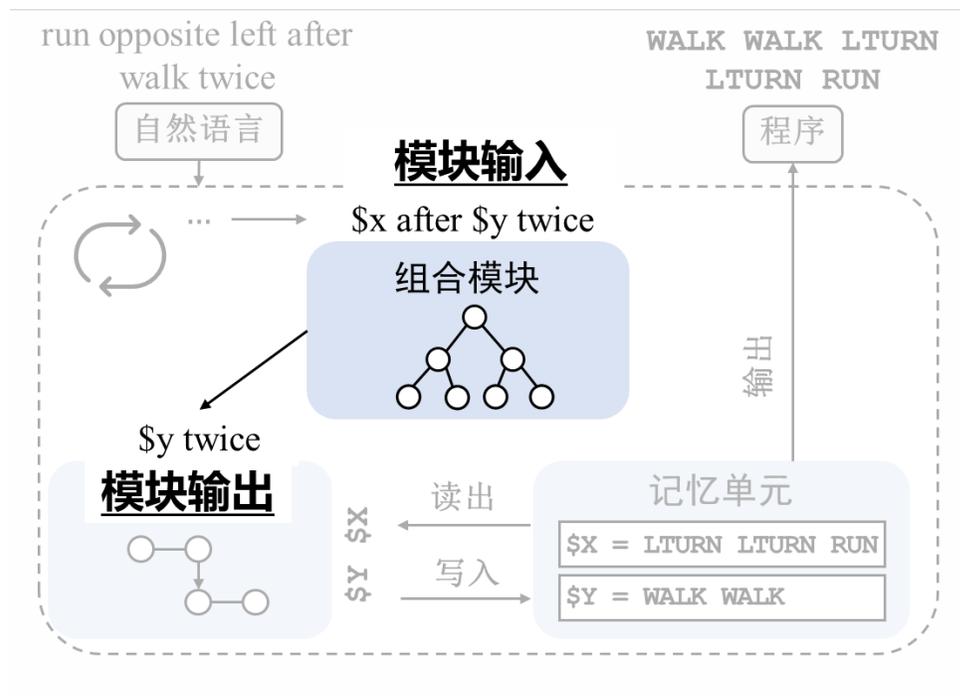
解算模块

记忆单元



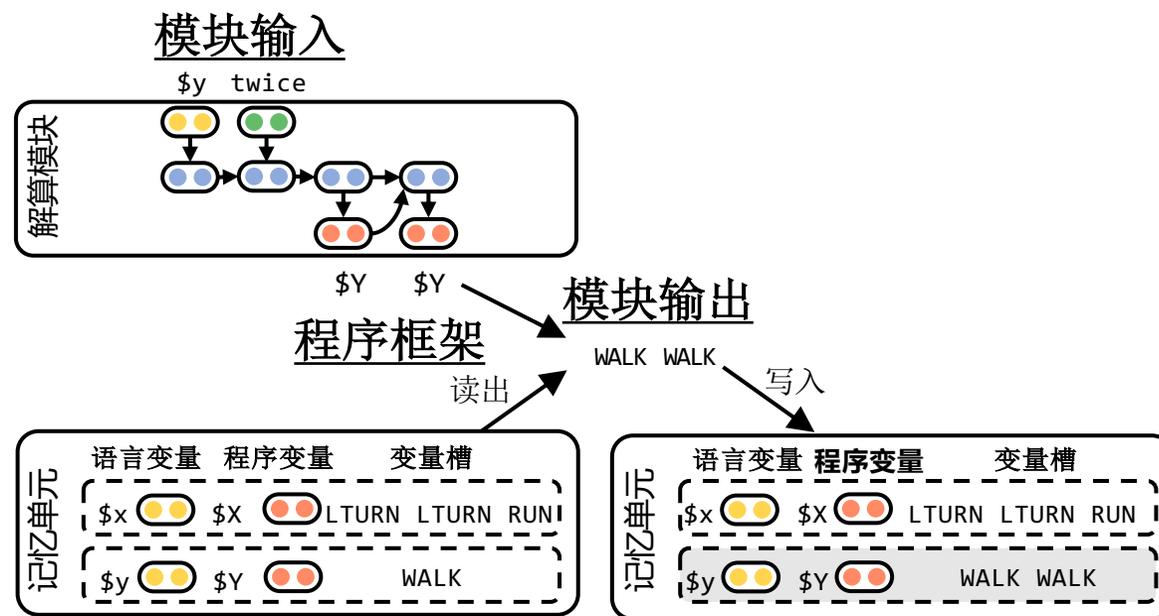
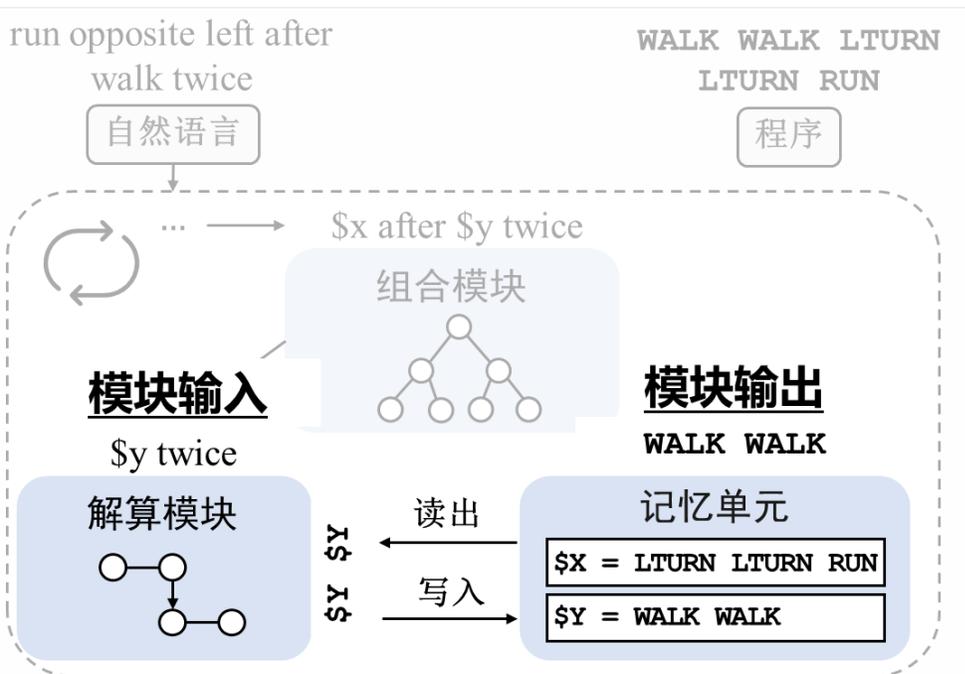
3.1 面向程序组合泛化的自然语言语义解析方法：组合模块

组合模块使用树状长短期序列网络实现，通过**自底向上**地合并语言表达式中的每一对相邻单词，直到找到一个合理的局部语言表达式。



3.1 面向程序组合泛化的自然语言语义解析方法：解算模块

解算模块使用序列到序列模型实现，接收组合模块的输出作为输入，通过**解码生成程序框架与读写记忆单元**完成对输入的解析。

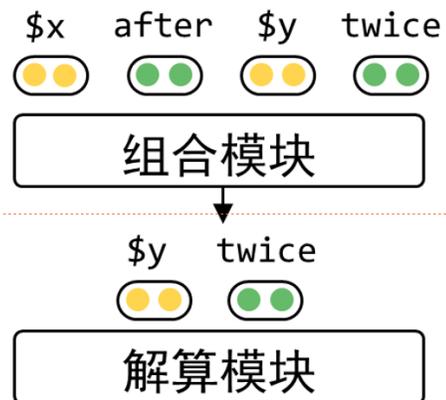


3.1 面向程序组合泛化的自然语言语义解析方法：模型训练



训练挑战

(i) 组合模块和解算模块之间的信息是离散的，模型**无法反向传播**直接优化。

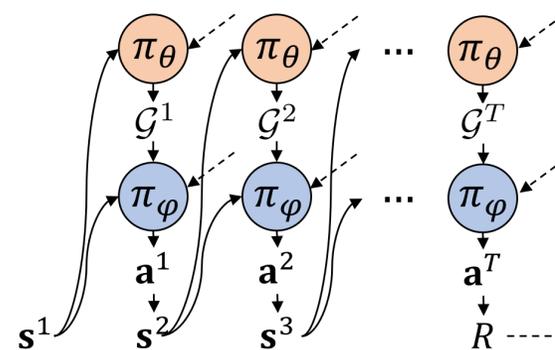


(ii) 强化学习的**奖励稀疏**，较难训练。



优化策略

(i) 采用**分层强化学习**，将组合模块和解算模块视为不同层级的代理。



(ii) 使用**课程学习**，从简单到复杂组织训练过程。

3.1 面向程序组合泛化的自然语言语义解析方法：奖励函数

- 基于**匹配度**的奖励函数设计：使用交并比来衡量动作序列的相似性。

预测：JUMP **LTURN RUN**

监督：WALK WALK **LTURN RUN**

奖励： $2 / (3 + 4 - 2) = 0.4$

- 基于**复杂度**的奖励函数设计：复用已有的表达式，使学到的表达式更简单。

预测1：“jump twice” → JUMP JUMP



预测2：“jump” → JUMP, “\$x twice” → \$X \$X



3.1 面向程序组合泛化的自然语言语义解析方法：实验设置

SCAN数据集主要包括 Add Jump, Around Right 和 Length Generalization 三大组合泛化挑战，后续工作扩展了MCD(1/2/3) 和小样本条件下的MiniSCAN。

训练

Add Jump

jump
walk twice
walk around left

测试

jump around left

训练时没有关于
*jump*的复杂语句

Around Right

turn around left
turn opposite right
walk around left

turn around right

训练时未出现
*around right*短语

Length Generalization

look around left
look around left twice
look around left twice after look

look around left twice after look around left

训练集对应的导航动作序列长度不超过24，测试时的导航动作序列都超过24。

3.1 面向程序组合泛化的自然语言语义解析方法：实验结果

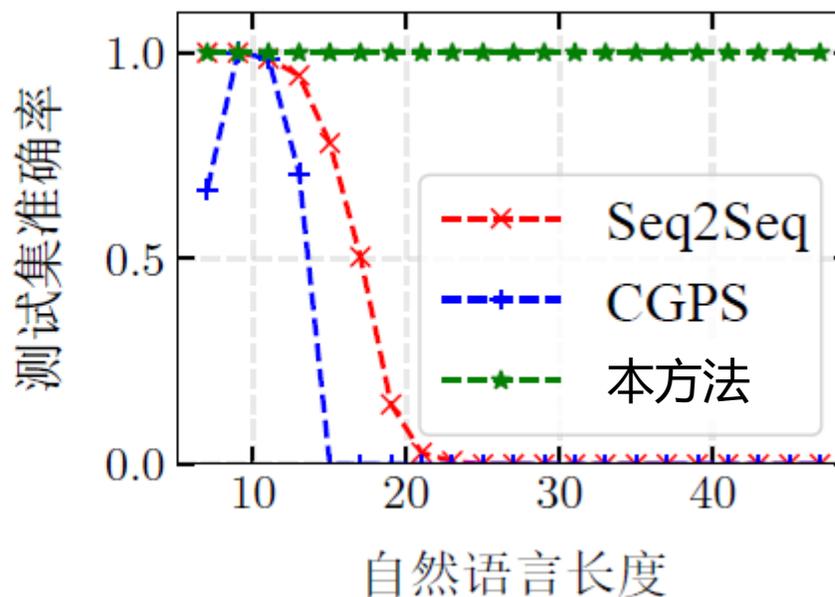
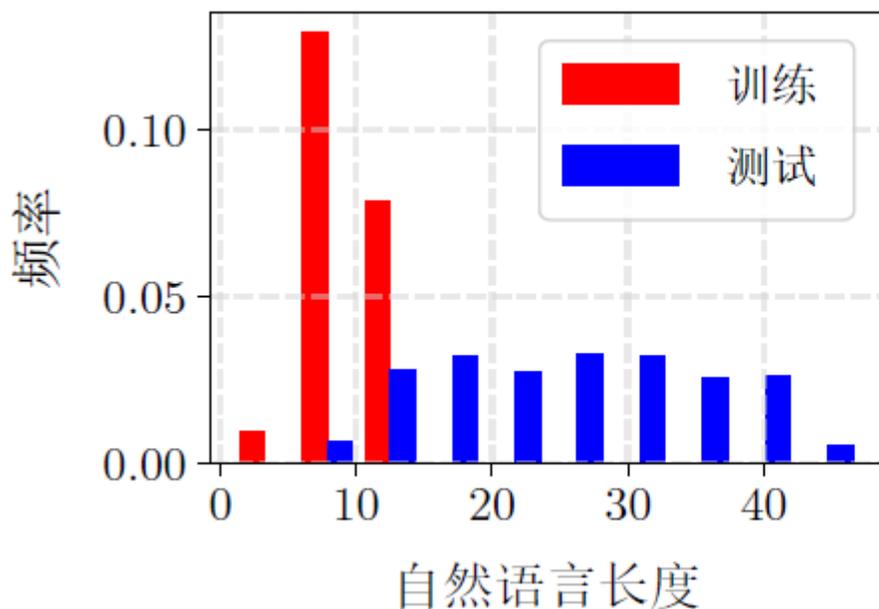
本方法以**100%准确率**解决了SCAN相关测试基准上的全部组合泛化任务。

组合泛化挑战（提出机构）	性能
<i>Add Jump (Facebook)</i>	100.0%
<i>Around Right (Facebook)</i>	100.0%
<i>Length Generalization (Facebook)</i>	100.0%
<i>MCD1/2/3 (Google Brain)</i>	100.0%

本方法无需任何额外标注即可解决前人提出的全部组合泛化任务

3.1 面向程序组合泛化的自然语言语义解析方法：实验结果

基线模型的测试准确率主要由训练集中**输入长度所占的频率**所决定。相反，随着输入长度的增加，本方法的性能保持着很好的趋势。



本方法对输入自然语言的长度鲁棒，表现出较好的组合泛化性

3.1 面向程序组合泛化的自然语言语义解析方法：实验结果

在**程序数据标注弱**的情况下，本方法显著提升了模型的组合泛化能力。

方法	MiniSCAN 测试集性能
Seq2Seq	2.5%
CGPS	76.0%
人类水平	84.3%
本方法	100.0%

达成研究目标，在弱数据数量下增强了模型的泛化能力

3.1 面向程序组合泛化的自然语言语义解析方法：实验分析

详尽的消融实验结果验证了本方法中每个组件的有效性，包括**层次化抽象设计思想**、**课程学习训练策略**与**基于复杂度的奖励**。

消融组件	Simple	Add Jump	Length	Around Right	MCD1	MCD2	MCD3
层次化抽象	98.5 ± 0.6	0.0	11.1 ± 13.1	0.0	5.3 ± 2.4	0.7 ± 0.3	2.6 ± 0.9
课程学习	0.0	0.0	0.0	0.0	0.0	0.0	0.0
复杂度奖励	100.0	100.0	100.0	0.0	100.0	100.0	78.8 ± 4.2

Qian Liu, Shengnan An, Jian-Guang Lou, Bei Chen, Zeqi Lin, Yan Gao, Bin Zhou, Nanning Zheng, Dongmei Zhang. Compositional Generalization by Learning Analytical Expressions. *NeurIPS 2020 (CCF-A)*.

汇报提纲

1 研究背景与意义

2 研究现状与问题

3 研究目标与内容

3.1 面向程序组合泛化的自然语言语义解析方法

3.2 面向知识库领域泛化的自然语言语义解析方法

3.3 弱监督下答案驱动的自然语言语义解析方法

3.4 半监督下对话重写驱动的对话式自然语言语义解析方法

4 研究成果与创新

3.2 面向知识库领域泛化的语义解析方法：研究目标

立足于强监督语义解析场景，面向语义解析模型**领域标注数量弱**的情况，本研究内容旨在提升语义解析模型**面向知识库的领域泛化**能力。



database: concert singer

Train



Show all *countries* and the number of *singers* in each *country*.



```
SELECT Country, count(*) FROM Singer GROUP BY Country
```



database: farm

Test



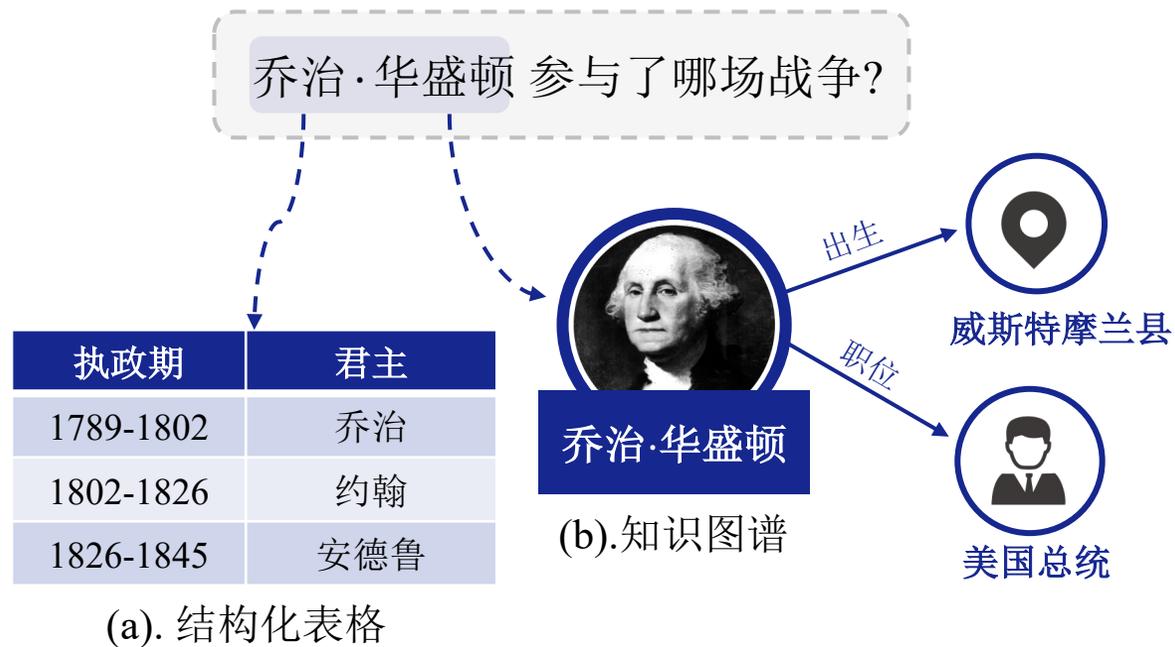
Please show the different *statuses* of *cities* and the average *population* of cities with each *status*.



```
SELECT Status, avg(Population) FROM City GROUP BY Status
```

3.2 面向知识库领域泛化的语义解析方法：相关工作

为达成该目标，研究者们一般使用**实体链接机制**来提升模型的领域泛化能力。

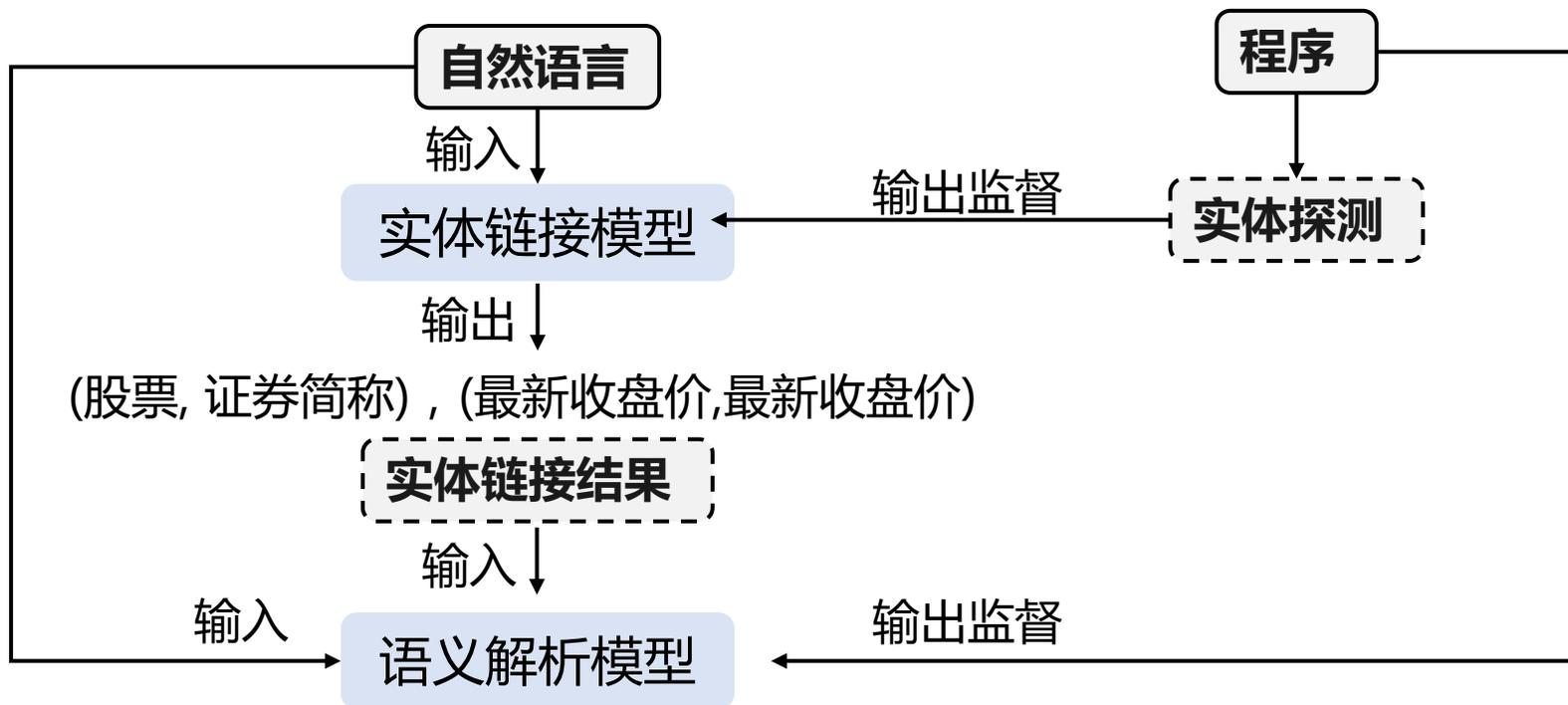


已有的实体链接模型需要额外的数据标注

3.2 面向知识库领域泛化的语义解析方法：总体结构

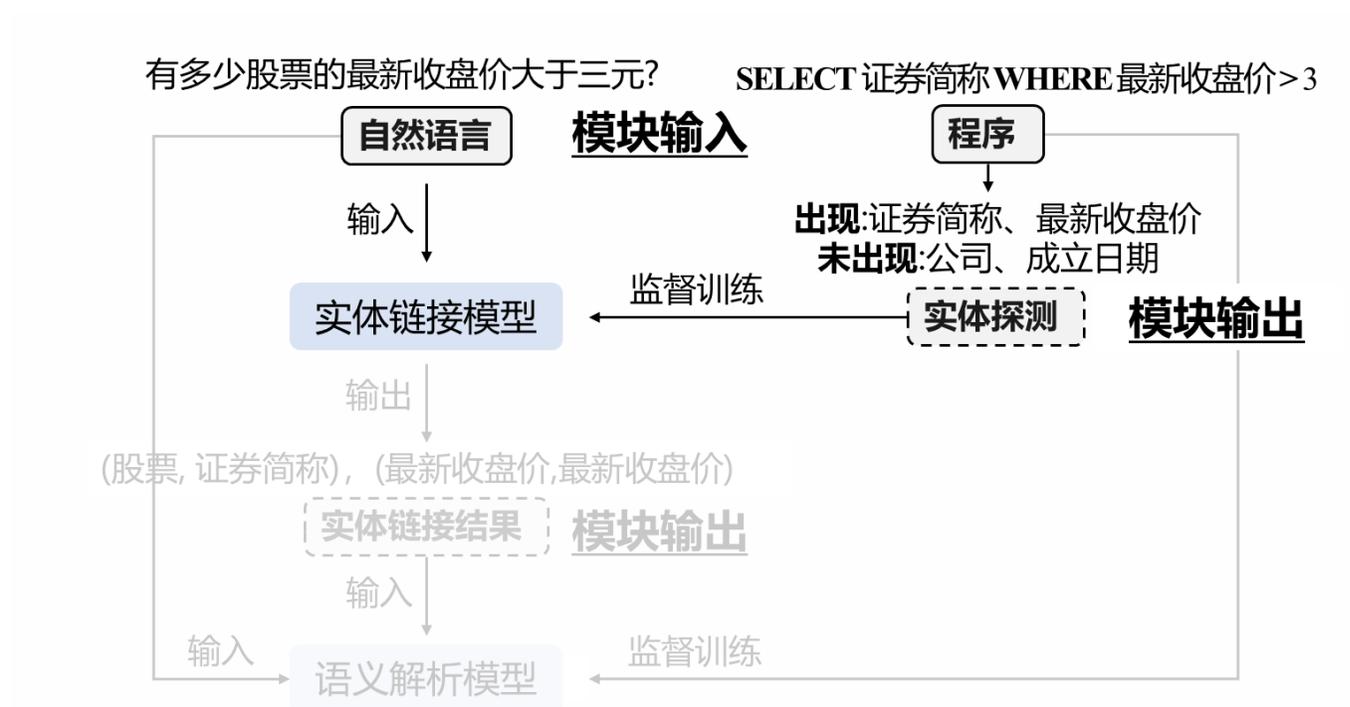
针对**领域标注种类少**的难点，本文提出一种无需额外标注数据即可训练实体链接模型，再用其增强语义解析模型领域泛化能力的方法。

模型输入 有多少股票的最新收盘价大于三元？ `SELECT 证券简称 WHERE 最新收盘价 > 3` 模型输出



3.2 面向知识库领域泛化的语义解析方法：实体链接

与前人方法需要用到额外的实体链接数据不同，本方法通过将语义解析数据确定性地转换成**实体探测**信号，并通过该信号训练实体链接模型。



3.2 面向知识库领域泛化的语义解析方法：实体链接

给定问句和实体集合，**实体探测**信号指示每个实体在问句中是否被提及。

年份	城市	国家	参赛国家数量
1896	雅典	希腊	14
1900	巴黎	法国	24
1904	路易斯	美国	12
...
2004	雅典	希腊	201
2008	北京	中国	204
2012	伦敦	英国	204



希腊在哪一年举行了其最后一届夏季奥运会？

```
SELECT 年份 WHERE 国家 = 希腊  
ORDER BY 年份 DESC LIMIT 1
```



导出实体探测监督



🍏 年份, 🍏 国家, 🍏 希腊
🍏 城市, 🍏 参赛国家数量

3.2 面向知识库领域泛化的语义解析方法：实体链接

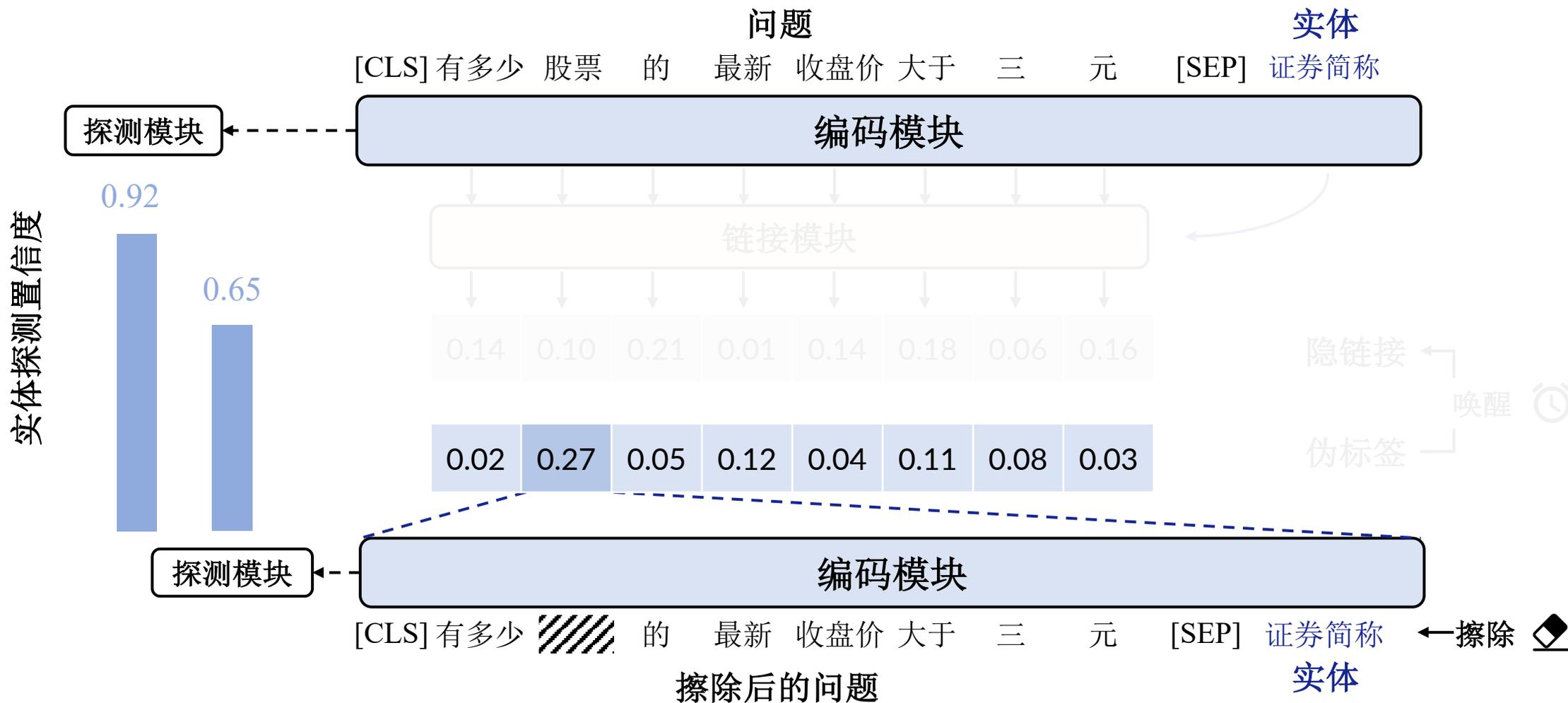


3.2 面向知识库领域泛化的语义解析方法：实体链接

当擦除问句中某个词之后，如果一个训练好的实体探测模块在某个实体上的探测置信度有**明显下降**，可以认为该词与给定的实体密切相关。

问题	实体	置信度
有多少 股票 的 最新 收盘价 大于 三元？	证券简称	0.92
问题	实体	置信度
有多少  的 最新 收盘价 大于 三元？	证券简称	0.65
		差值
		0.27

3.2 面向知识库领域泛化的语义解析方法：实体链接



3.2 面向知识库领域泛化的语义解析方法：实体链接

通过对若干候选实体执行上述过程，得到的链接伪标签可以作为链接模块训练使用的权重，从而指导链接模块的训练。

问题

有多少 股票 的 最新 收盘价 大于 三元？

0.02 0.27 0.05 0.12 0.04 0.11 0.08 0.03

0.01 0.13 0.02 0.00 0.00 0.02 0.01 0.04

0.01 0.02 0.01 0.02 0.74 0.32 0.03 0.02

实体

证券简称

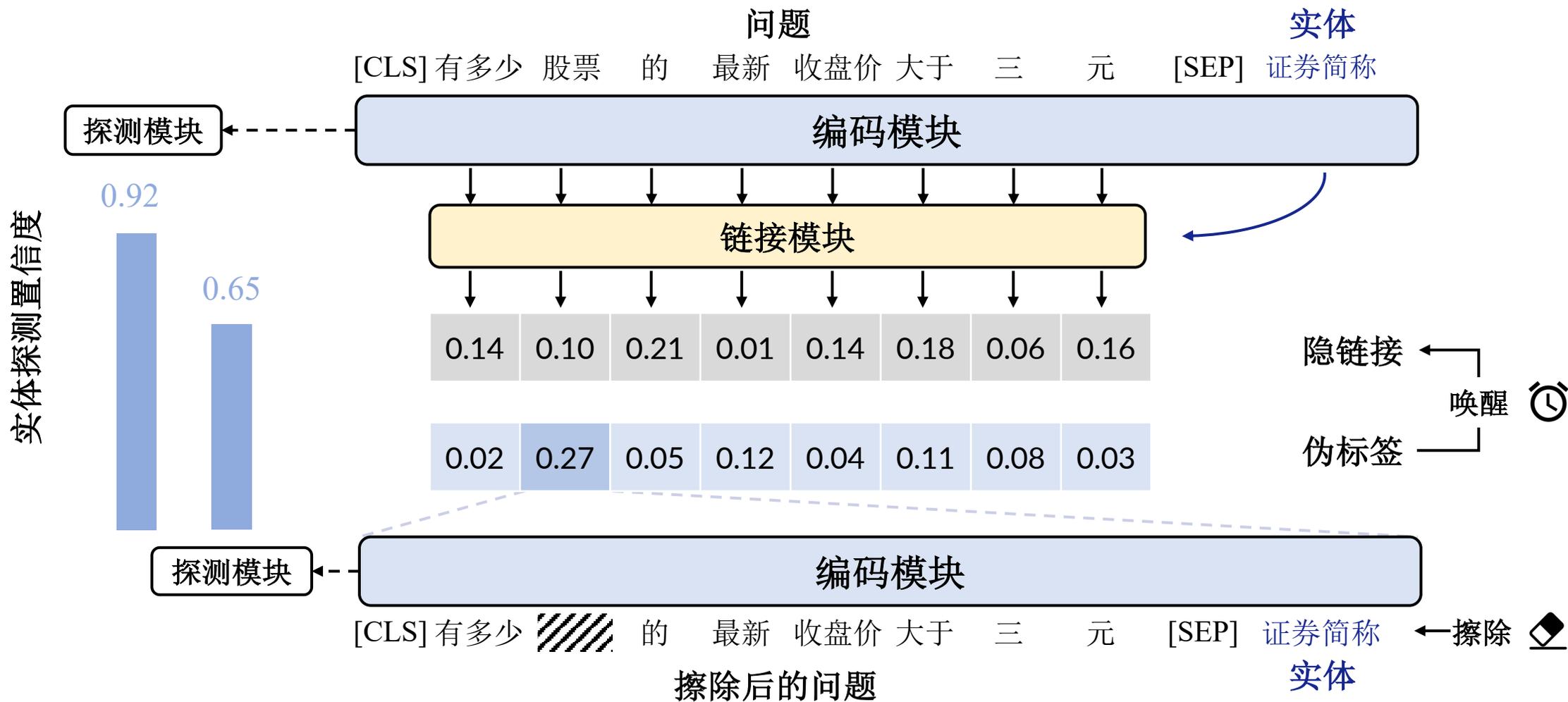
公司

最新收盘价

训练

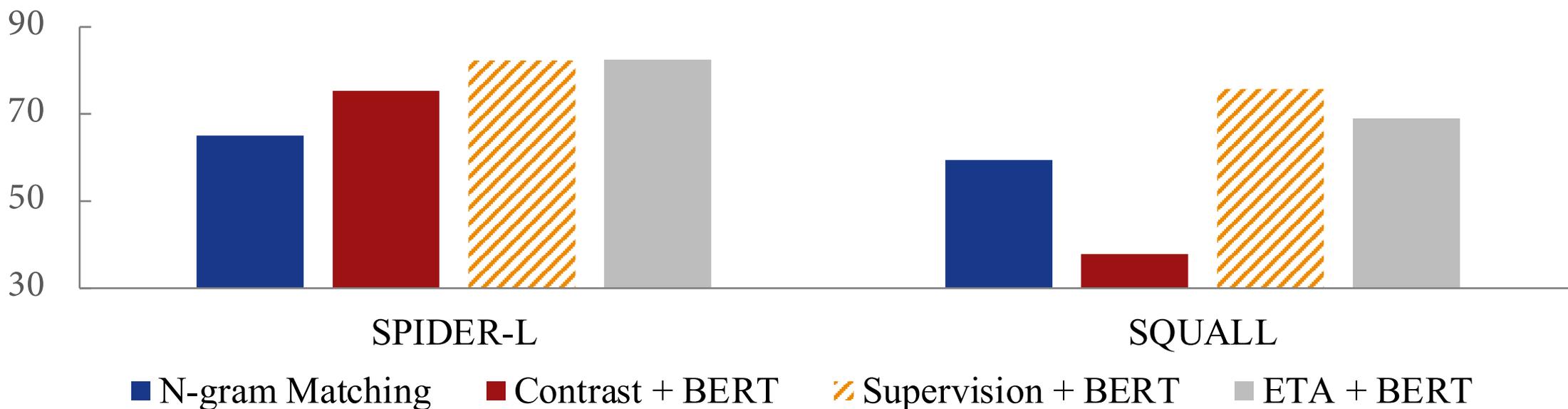
链接模块

3.2 面向知识库领域泛化的语义解析方法：实体链接



3.2 面向知识库领域泛化的语义解析方法：实体链接实验结果

虽然没有在细粒度的实体链接监督上进行训练，但我们的模型（■）与全监督下训练的模型（▨）性能相当。



本方法所学习到的实体链接模型与人类专家的标注表现出了高度一致性

3.2 面向知识库领域泛化的语义解析方法：实体链接实验分析

在实体链接数据集上的分析表明，本方法的错误主要有四种类型：链接遗漏，语义正确，部分正确，错误链接。

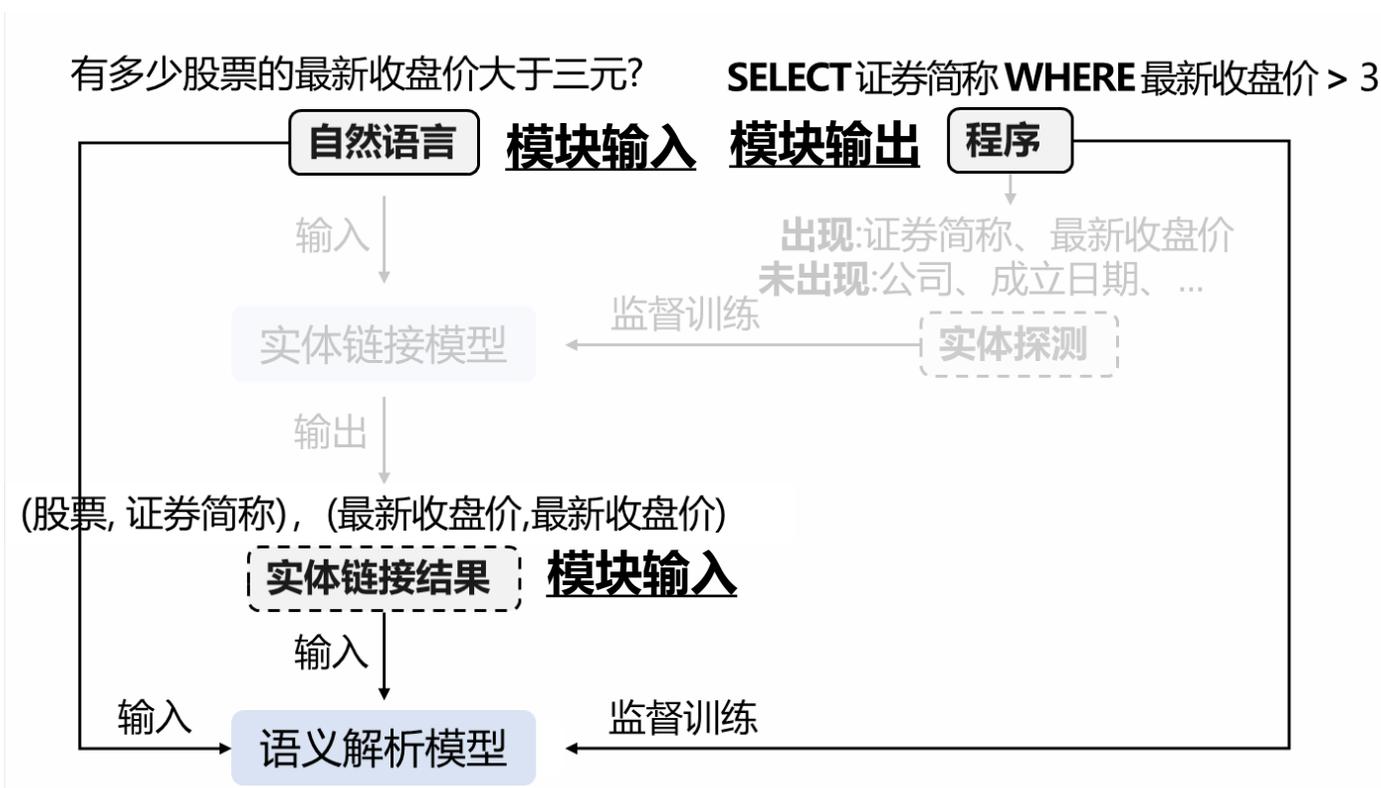
错误类型（占比）	错误样例
遗漏链接（43.1%）	How many points did arnaud demare receive? 翻译：阿尔诺·德马雷得到多少分？ 标注： points →“UCI world tour points” 预测：空
语义正确（21.0%）	Total population of millbrook first nation ? 翻译：米尔布鲁克最大的民族有多少人？ 标注： population →“Population” 预测： population →“Population”； nation →“Community”

错误类型（占比）	错误样例
部分正确（15.8%）	Who was the first winning captain ? 翻译：谁是首个冠军队长？ 标注： the first →“Year”； winning captain →“Winning Captain” 预测： first →“Year”； winning captain →“Winning Captain”
错误链接（10.1%）	Were the matinee and evening performances held earlier than the 8th anniversary? 翻译：日场和夜场演出是否在早于 8 周年纪念日时举行？ 标注： earlier →“Date” 预测： matinee →“Performance”； earlier →“Date”

本方法所学习到的实体链接模型的主要挑战在于召回率较低

3.2 面向知识库领域泛化的语义解析方法：语义解析

本方法可以适用于任何基于知识库的语义解析模型，下面本文以text-to-SQL为例展示**实体链接强化的语义解析模型架构**。



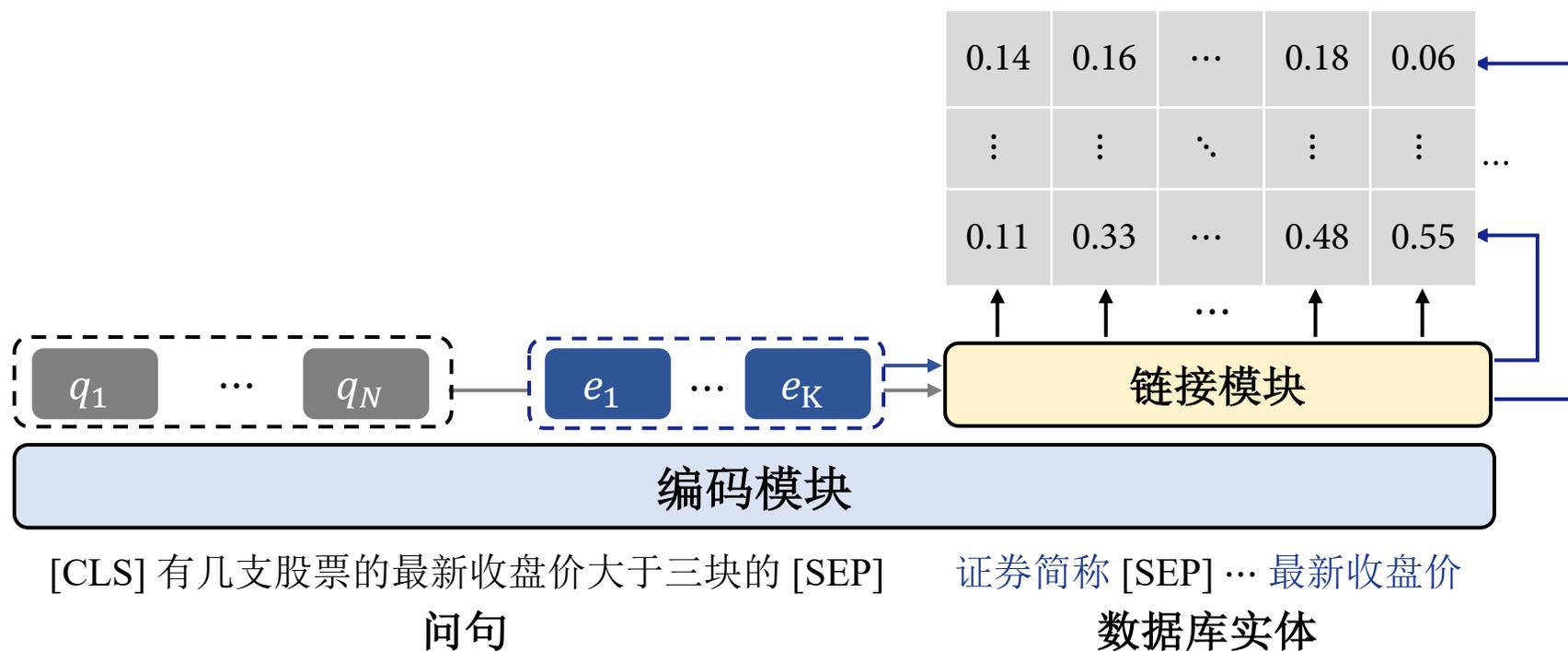
3.2 面向知识库领域泛化的语义解析方法：语义解析

给定一个自然语言问句和数据库中所有实体，模型首先将它们拼接起来构成一个输入序列，得到问题表示 $\langle q_1, \dots, q_N \rangle$ 和实体表示 $\langle e_1, \dots, e_K \rangle$ 。



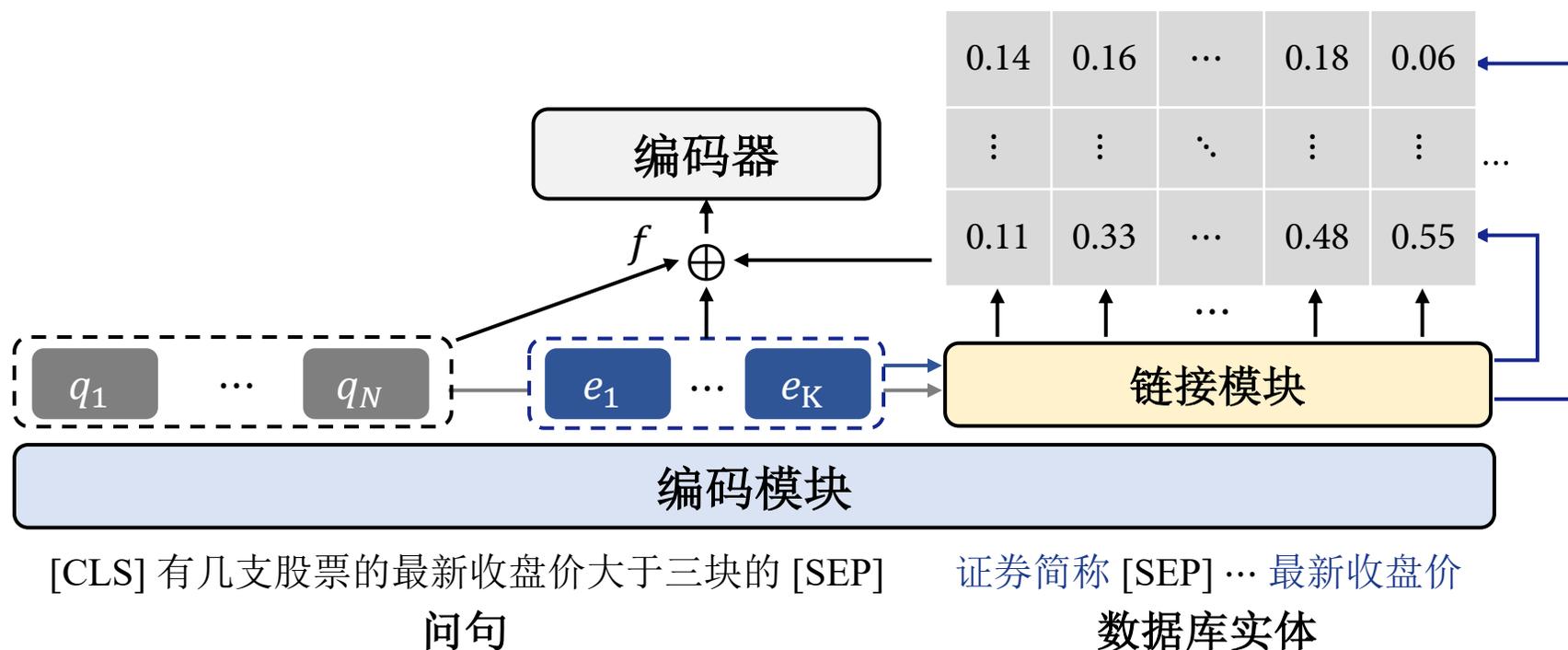
3.2 面向知识库领域泛化的语义解析方法：语义解析

将这些表示送入已经学习好的链接模块，从而得到实体链接结果。



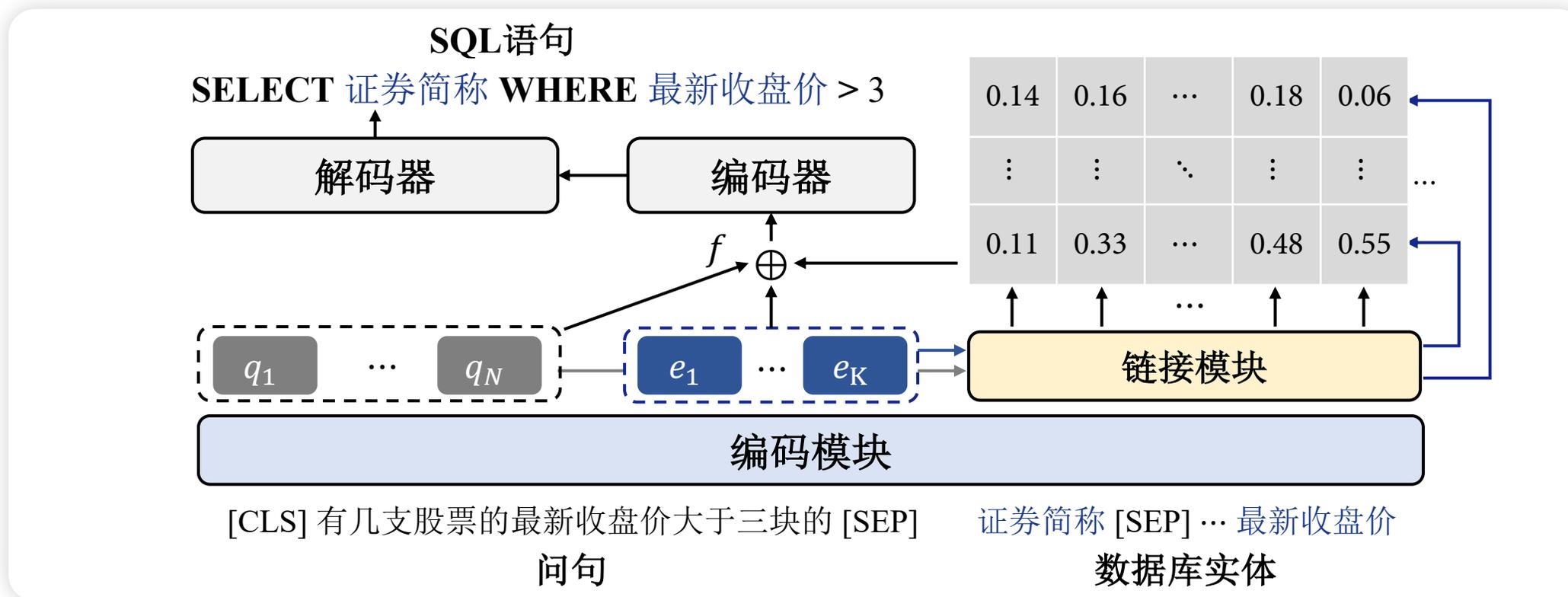
3.2 面向知识库领域泛化的语义解析方法：语义解析

将实体链接结果作为**实体先验**，被用于为自然语言问句中每个单词计算一个**实体可感知**的表征。该表征与原始单词表征拼接送入编码器。



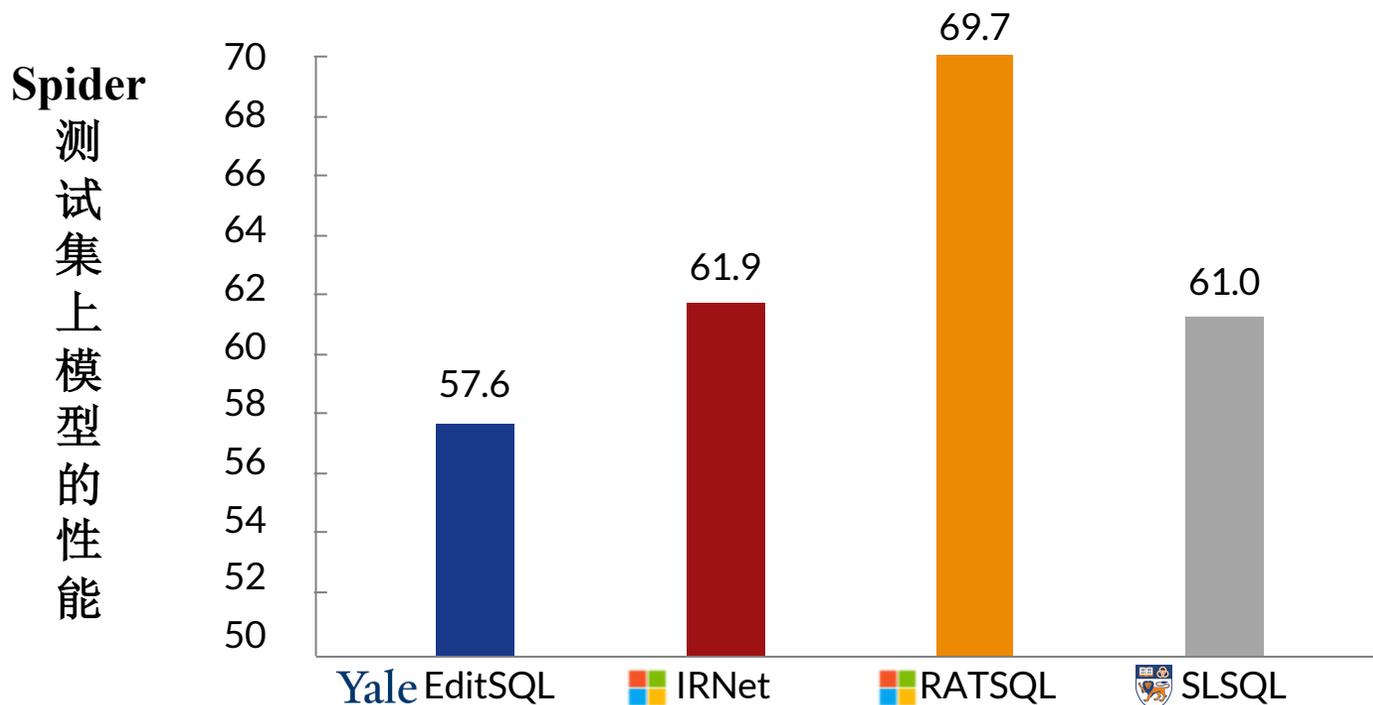
3.2 面向知识库领域泛化的语义解析方法：语义解析

最终，解码器负责生成SQL语句。通过增加实体感知表征强化编码器，本方法促使解码器更好地预测SQL语句中的实体。



3.2 面向知识库领域泛化的语义解析方法：语义解析实验结果

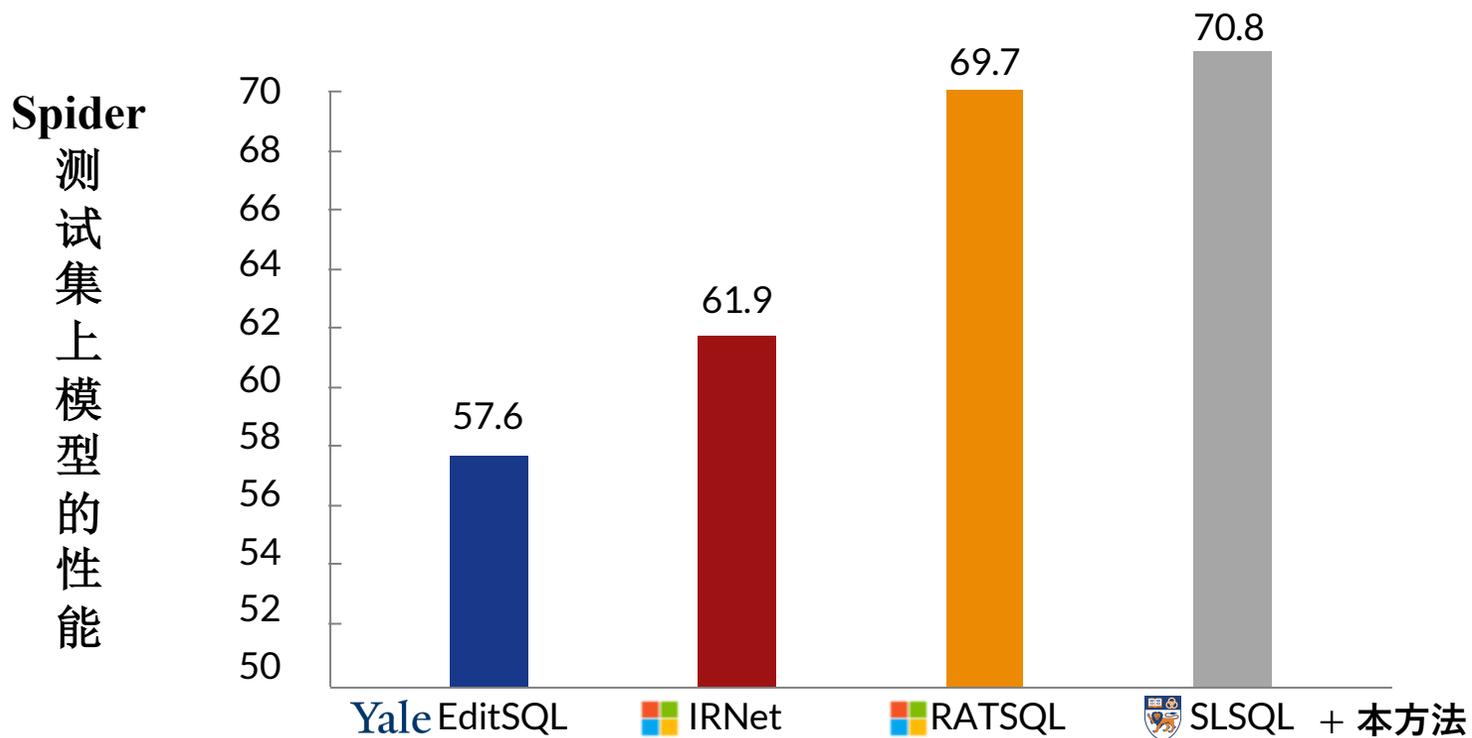
在**知识库领域数据标注有限**时，本方法显著提升了模型的领域泛化能力。



达成研究目标，在弱数据数量下增强了模型的泛化能力

3.2 面向知识库领域泛化的语义解析方法：语义解析实验结果

在**知识库领域数据标注有限**时，本方法显著提升了模型的领域泛化能力。



达成研究目标，在弱数据数量下增强了模型的泛化能力

3.2 面向知识库领域泛化的语义解析方法：语义解析实验分析

实体链接的结果为用户理解SQL语句的生成过程提供了一些**可解释性**，可以帮助他们更好地理解语义解析模型的预测过程。

自然语言

Show **name**₁, **country**₂, **age**₃ for all **singers**₄ ordered by **age**₃ from the **oldest**₃ to the youngest.

翻译：按年龄从大到小显示所有歌手的名字、国家和年龄。

SQL语句

```
SELECT name1, country2, age3 FROM singer4  
ORDER BY age3 DESC
```

Qian Liu, Dejian Yang, Jiahui Zhang, Jiaqi Guo, Bin Zhou, Jian-Guang Lou. Awakening Latent Grounding from Pretrained Language Models for Semantic Parsing. *ACL-Findings 2021*.

汇报提纲

1 研究背景与意义

2 研究现状与问题

3 研究目标与内容

3.1 面向程序组合泛化的自然语言语义解析方法

3.2 面向知识库领域泛化的自然语言语义解析方法

3.3 弱监督下答案驱动的自然语言语义解析方法

3.4 半监督下对话重写驱动的对话式自然语言语义解析方法

4 研究成果与创新

3.3 弱监督下答案驱动的语义解析方法：研究目标

立足于弱监督语义场景，面向语义解析模型**数据标注质量弱化**的情况，本研究内容旨在提升语义解析模型**弱监督下的模型性能**。

年份	城市	国家	参赛国家数量
1896	雅典	希腊	14
1900	巴黎	法国	24
1904	路易斯	美国	12
...
2004	雅典	希腊	201
2008	北京	中国	204
2012	伦敦	英国	204

问题：希腊在哪一年举行了其最后一届夏季奥运会？

答案：2004

问题：哪个城市举办的奥运会的参赛国家数量首次超过20？

答案：巴黎

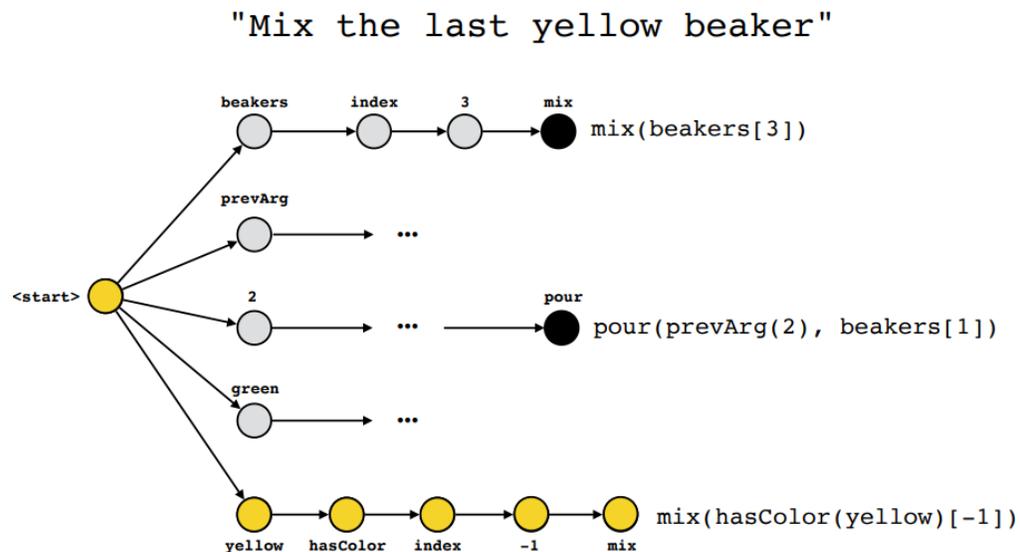
问题：哪几年参赛的国家数量最多？

答案：2008, 2012

3.3 弱监督下答案驱动的自然语言语义解析方法：相关工作

为达成该目标，前人工作一般使用强化学习或监督学习范式来解决。

强化学习

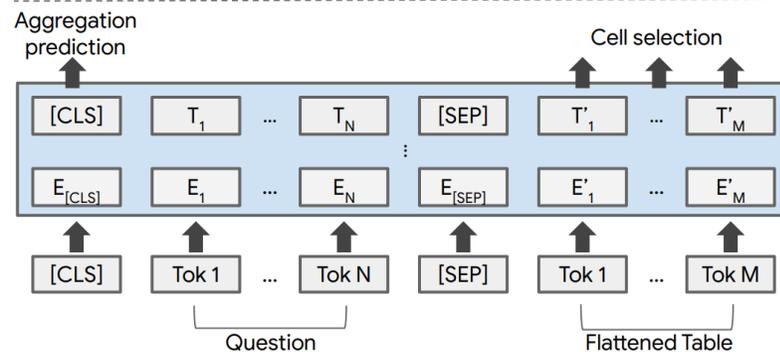


监督学习

op	$P_a(op)$	compute(op, P_s, T)
NONE	0	-
COUNT	0.1	.9 + .9 + .2 = 2
SUM	0.8	.9×37 + .9×31 + .2×15 = 64.2
AVG	0.1	64.2 ÷ 2 = 32.1

Rank	...	Days	P_s
1	...	37	0.9
2	...	31	0.9
3	...	17	0
4	...	15	0.2
...	0

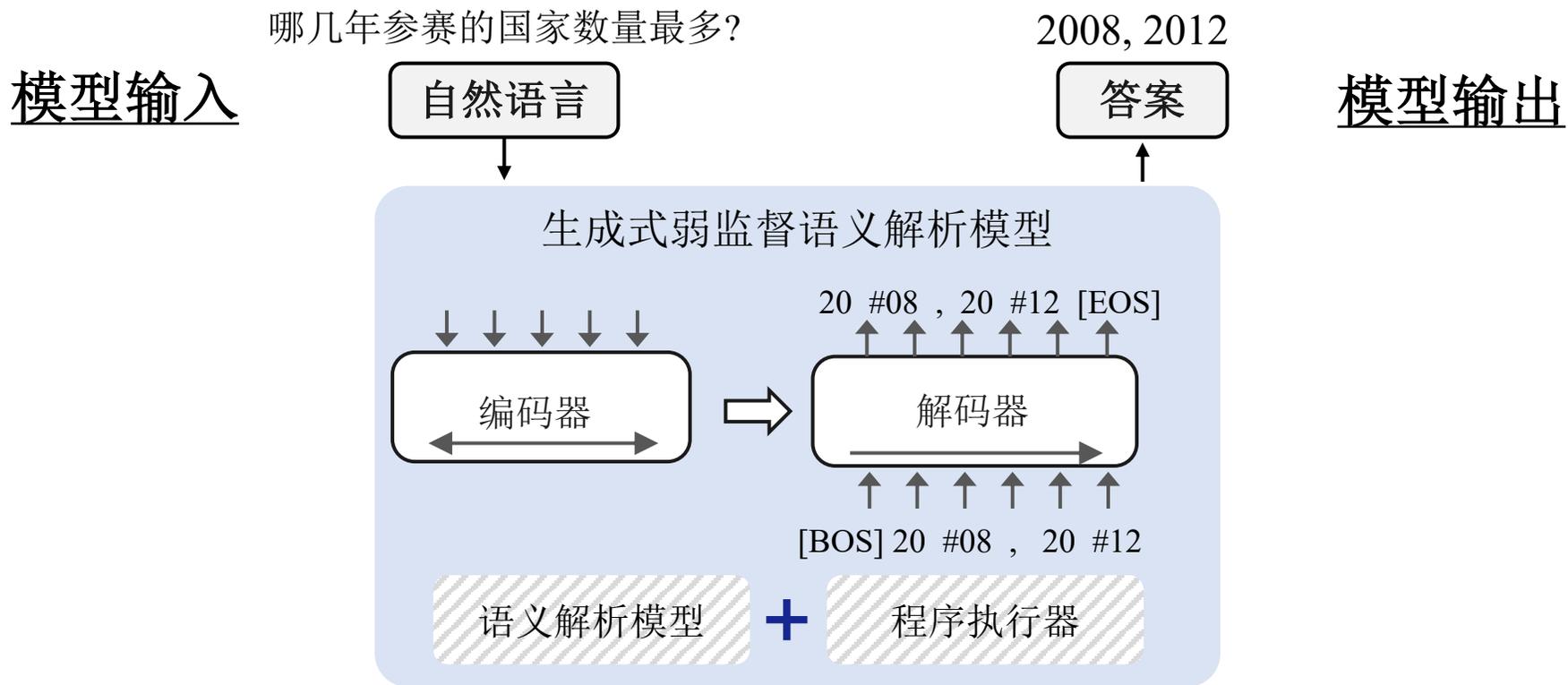
$$s_{pred} = .1 \times 2 + .8 \times 64.2 + .1 \times 32.1 = 54.8$$



强化学习模型训练不稳定，监督学习牺牲了模型的灵活性

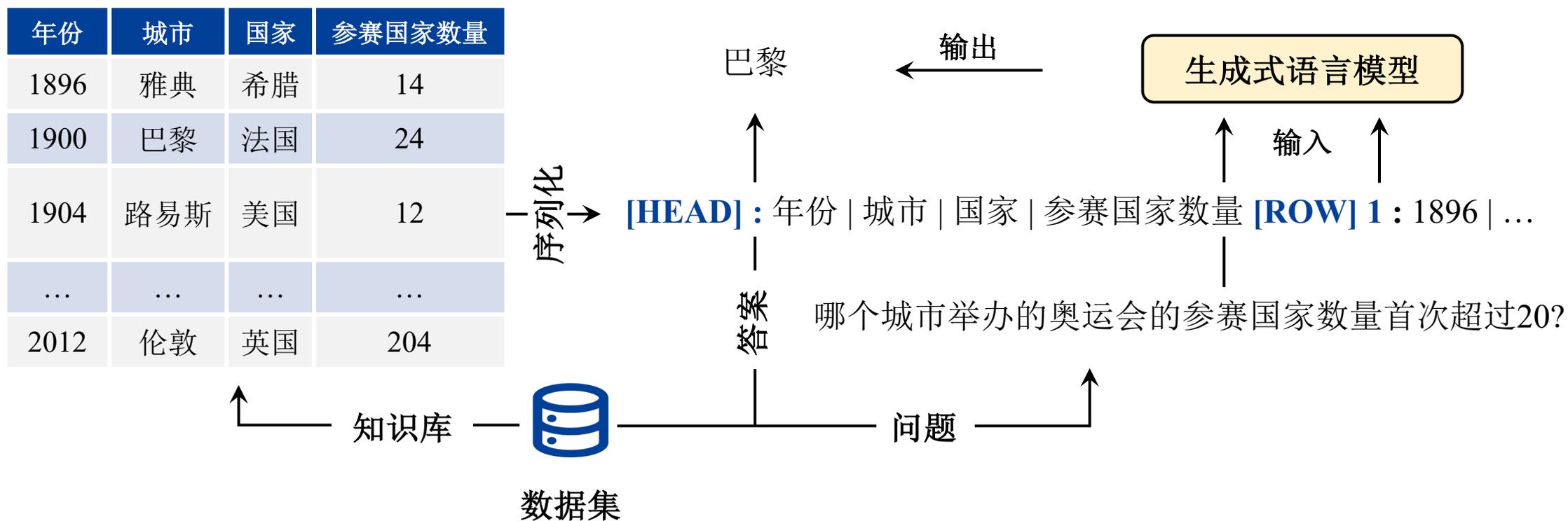
3.3 弱监督下答案驱动的自然语言语义解析方法：总体结构

针对**答案标注监督弱**的难点，提出使用**生成式语言模型**来可微地解决该任务，并利用**继续预训练**进一步提升模型性能。



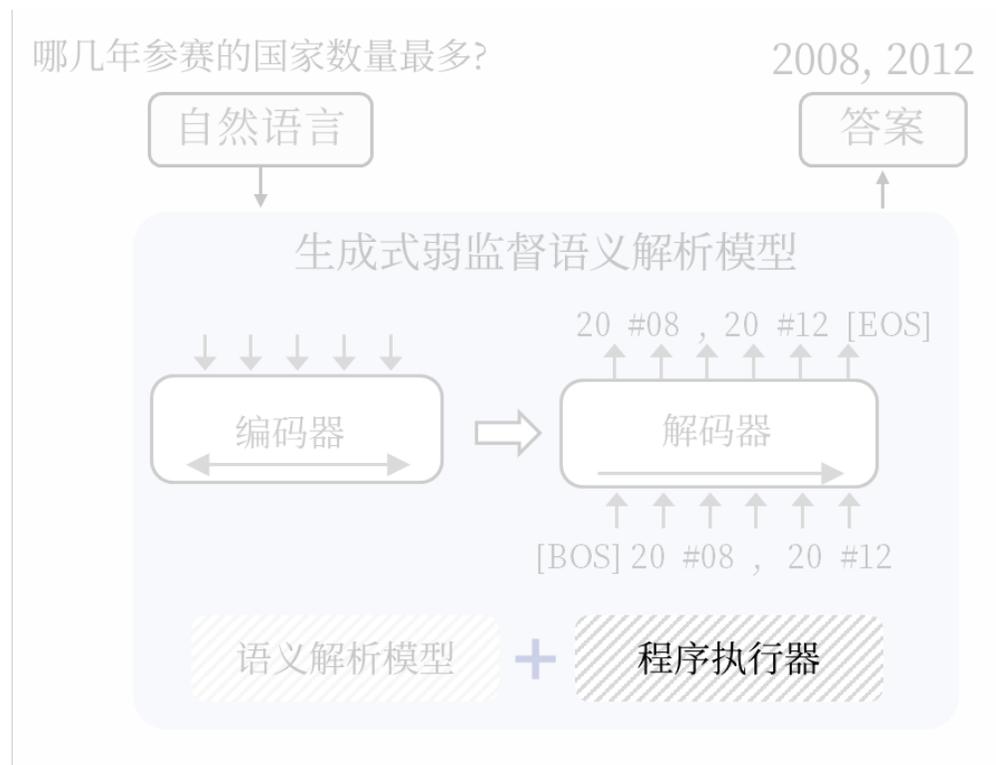
3.3 弱监督下答案驱动的语义解析方法：生成式弱监督语义解析

本文将弱监督语义解析建模为序列生成任务，并利用**生成式预训练语言模型**（例如BART）直接生成答案。



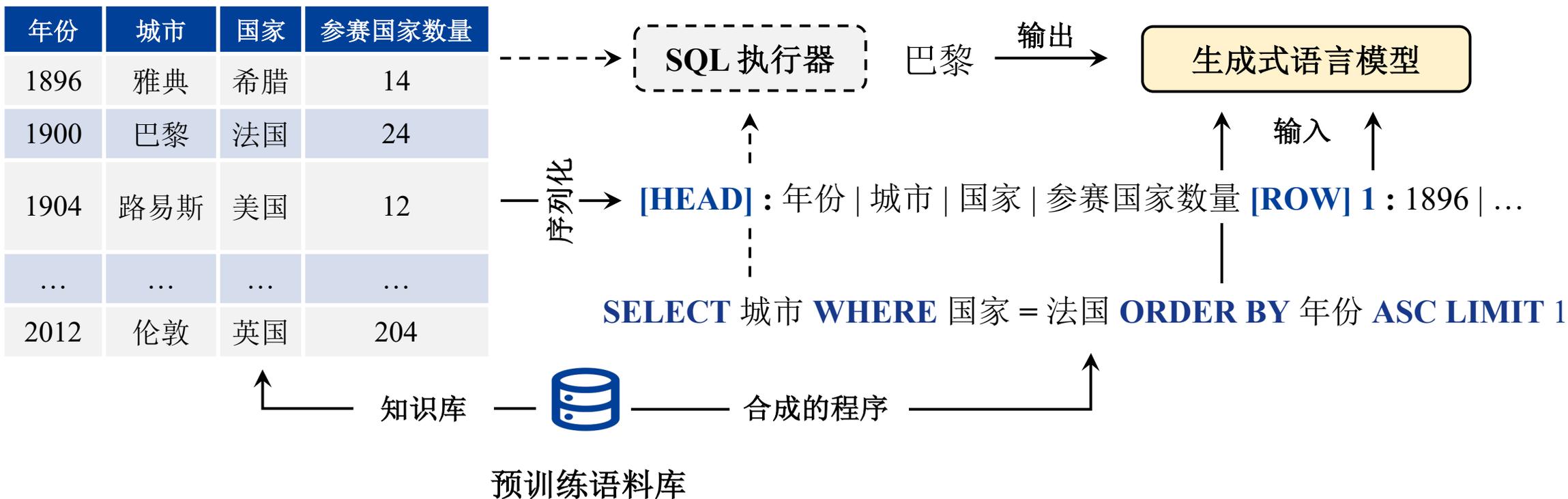
3.3 弱监督下答案驱动的自然语言语义解析方法：执行引导的预训练

将生成式语言模型视作语义解析器与程序执行器的结合，本方法提出通过预训练强化模型内部**程序执行器**，以达到更好的任务性能。



3.3 弱监督下答案驱动的自然语言语义解析方法：执行引导的预训练

本方法通过预训练方法 TAPEX **模仿生成知识库上程序的执行结果**，让模型强化潜在的程序执行器，并以此更好地“执行”自然语言。



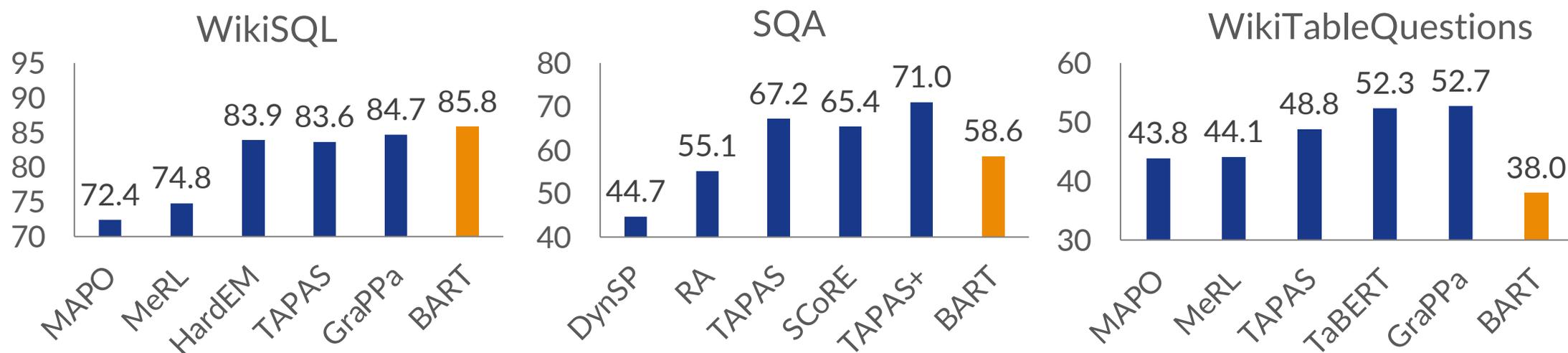
3.3 弱监督下答案驱动的语义解析方法：实验设置

三个经典弱监督语义解析数据集 WikiSQL, WikiTableQuestions 和 SQA。

WikiSQL	WikiTableQuestions	SQA
How many CFL teams are from York College?	Which album released by the band schnell fenster produced the most singles appearing on the Australian peak chart?	Where are the players from?
简单问题	复杂问题	简单对话

3.3 弱监督下答案驱动的语义解析方法：实验结果

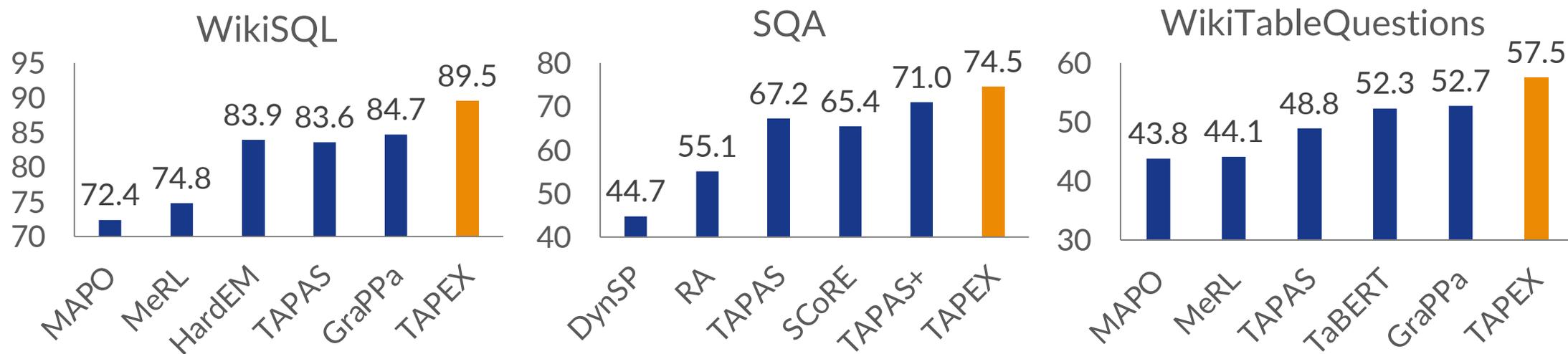
- 即使BART并没有进行过弱监督语义解析相关的预训练，但它单独就可以在WikiSQL上取得高达**85.8%**的答案准确率。
- 然而，在**数据量小**的SQA和WikiTableQuestions上，BART的效果并不好。



生成式语言模型在资源丰富的弱监督语义解析场景效果较好

3.3 弱监督下答案驱动的语义解析方法：实验结果

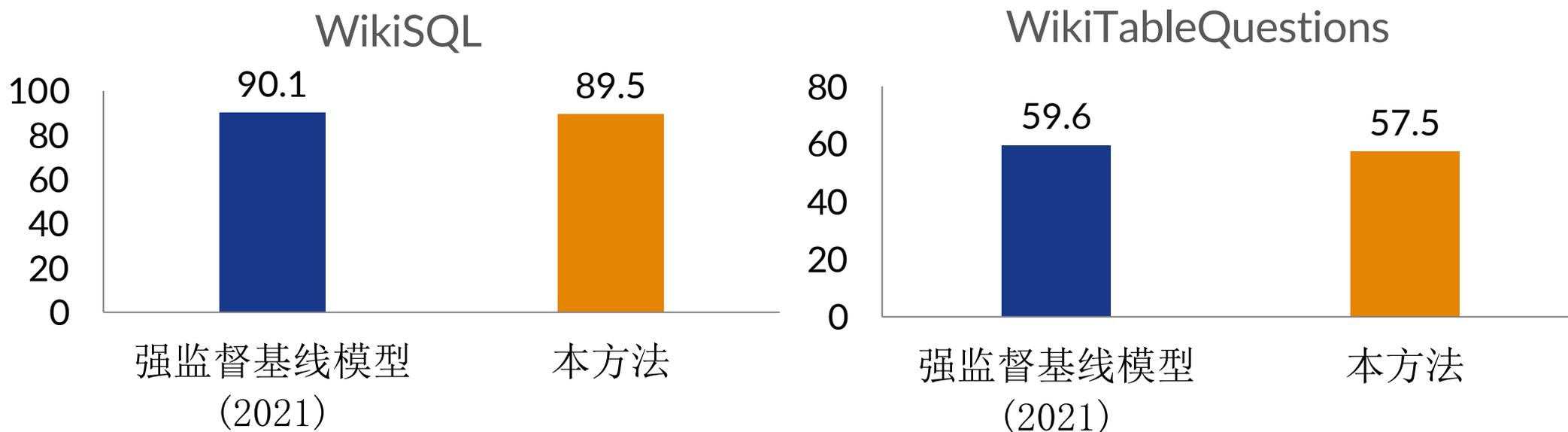
- TAPEX在所有弱监督语义解析基准数据集上都取得了**最先进的性能**。
- 在**数据量小**的SQA和WikiTableQuestions上，TAPEX的绝对提升高达**19.5%**。



本文的预训练方法可以极大地缓解低资源下的数据饥饿问题

3.3 弱监督下答案驱动的语义解析方法：实验结果

在**数据标注质量弱**时，本方法极大地缩小了弱监督与强监督的性能差异。



达成研究目标，在弱数据质量下增强了模型的性能

3.3 弱监督下答案驱动的自然语言语义解析方法：实验分析

为了理解预训练在哪些问题上会对模型有明显的帮助，人工分析了 WikiTableQuestions 开发集上 **8 种问题** 的答案准确率。

问题类别	示例问题	BART	TAPEX	问题类别	示例问题	BART	TAPEX
选择 (Select)	What is the years won for each team? 翻译: 每支队伍分别是在哪年赢球的?	41.3%	64.8% (+23.5%)	算术 (Arithmetic)	What is the difference between White voters and Black voters in 1948? 翻译: 1948 年, 白人选民和黑人选民的选票差了多少?	33.1 %	53.5% (+20.4%)
过滤 (Filter)	How long did Taiki Tsuchiya last? 翻译: Taiki Tsuchiya 活了多久?	40.1%	65.7% (+25.6%)	比较 (Comparative)	Besides Tiger Woods, what other player won between 2007 and 2009 ? 翻译: 除了泰格伍兹, 还有哪位球员在 2007 年和 2009 年之间曾赢过比赛?	30.0 %	55.9% (+25.9%)
聚合 (Aggregate)	What is the amount of matches drawn? 翻译: 有多少场比赛?	26.9 %	57.4% (+30.5%)	分组 (Group)	What was score for each winning game? 翻译: 每场获胜的比赛的得分是多少?	49.5 %	66.7% (+17.2%)
极值 (Superlative)	What was the last Baekje Temple? 翻译: 哪个是最后一个百济寺?	46.3 %	64.3% (+18.0%)				

Qian Liu, Bei Chen, Jiaqi Guo, Morteza Ziyadi, Zeqi Lin, Weizhu Chen, Jian-Guang Lou. TAPEX: Table Pre-training via Learning a Neural SQL Executor. *ICLR 2022*.

汇报提纲

1 研究背景与意义

2 研究现状与问题

3 研究目标与内容

3.1 面向程序组合泛化的自然语言语义解析方法

3.2 面向知识库领域泛化的自然语言语义解析方法

3.3 弱监督下答案驱动的自然语言语义解析方法

3.4 半监督下对话重写驱动的对话式自然语言语义解析方法

4 研究成果与创新

3.4 半监督下对话重写驱动的对话式语义解析方法：研究目标

立足于对话式语义场景，面向语义解析模型**数据标注质量弱化**的情况，本研究内容旨在提升语义解析模型**半监督下的模型性能**。

年份	城市	国家	参赛国家数量
1896	雅典	希腊	14
1900	巴黎	法国	24
1904	路易斯	美国	12
...
2004	雅典	希腊	201
2008	北京	中国	204
2012	伦敦	英国	204



希腊在哪一年举行了其最后一届夏季奥运会？

```
SELECT 年份 WHERE 国家 = 希腊 ORDER BY 年份 DESC LIMIT 1
```



那法国呢？

3.4 半监督下对话重写驱动的对话式语义解析方法：研究目标

已有的对话式语义解析数据集中，上下文无关的问题占将近50%甚至更多。40%以上的SQL都非常简单，与真实场景的**数据分布差距大**。

数据集	上下文关系		SQL难度			
	上下文无关	上下文相关	简单	中等	困难	极其困难
SParC	47.5%	52.5%	40.1%	36.7%	12.1%	11.1%
CoSQL	68.2%	31.8%	41.4%	31.8%	16.2%	10.5%

现有的数据集数据分布远离真实场景

3.4 半监督下对话重写驱动的对话式语义解析方法：总体结构

针对**对话标注构造难**的难点，提出将对话式语义解析任务解耦成单轮语义解析和**对话重写**两个子任务，并使用**基于编辑的方法**完成对话重写任务。

模型输入

How much money has Smith earned

翻译: 史密斯赚了多少钱

How about Bill Collins

翻译: 比尔柯林斯呢



对话

输入

对话重写模型

输出

How much money has Bill Collins earned

翻译: 比尔柯林斯赚了多少钱

语境无关自然语言

输入 ↓

单轮语义解析模型

输出

程序

对话式语义解析

模型输出

3.4 半监督下对话重写驱动的对话式语义解析方法：数据集构建

为了使用半监督数据完成对话式语义解析场景，本文贡献了**第一个语义解析领域的对话重写数据集** FollowUp。



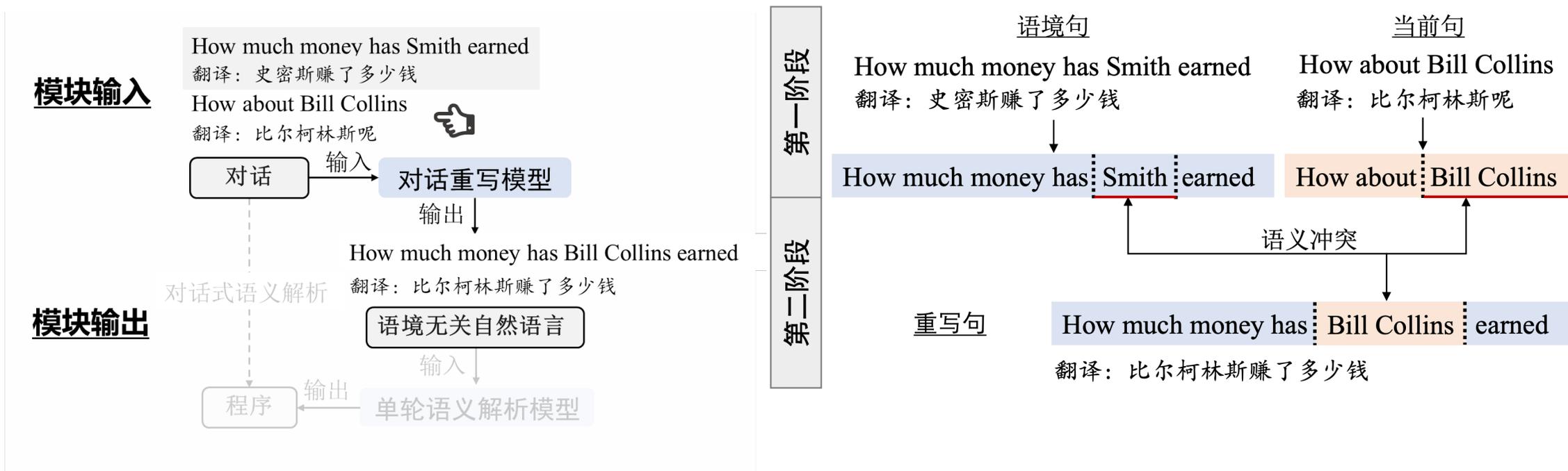
3.4 半监督下对话重写驱动的对话式语义解析方法：数据集构建

FollowUp数据集包括了**1,000**个通过众包标注的对话，这些对话覆盖了**120**个不同的知识库表格，并涵盖了多个典型的对话场景。

场景	样例
选择	语境句: In 1995, is there any network named CBC? 当前句: Any TSN? 重写句: In 1995, is there any network named TSN?
运算	语境句: List all universities founded before 1855. 当前句: Show their number. 重写句: Show the number of all universities founded before 1855.
比较	语境句: How much money has Smith earned? 当前句: How about Bill Collins? 重写句: How much money has Bill Collins earned?

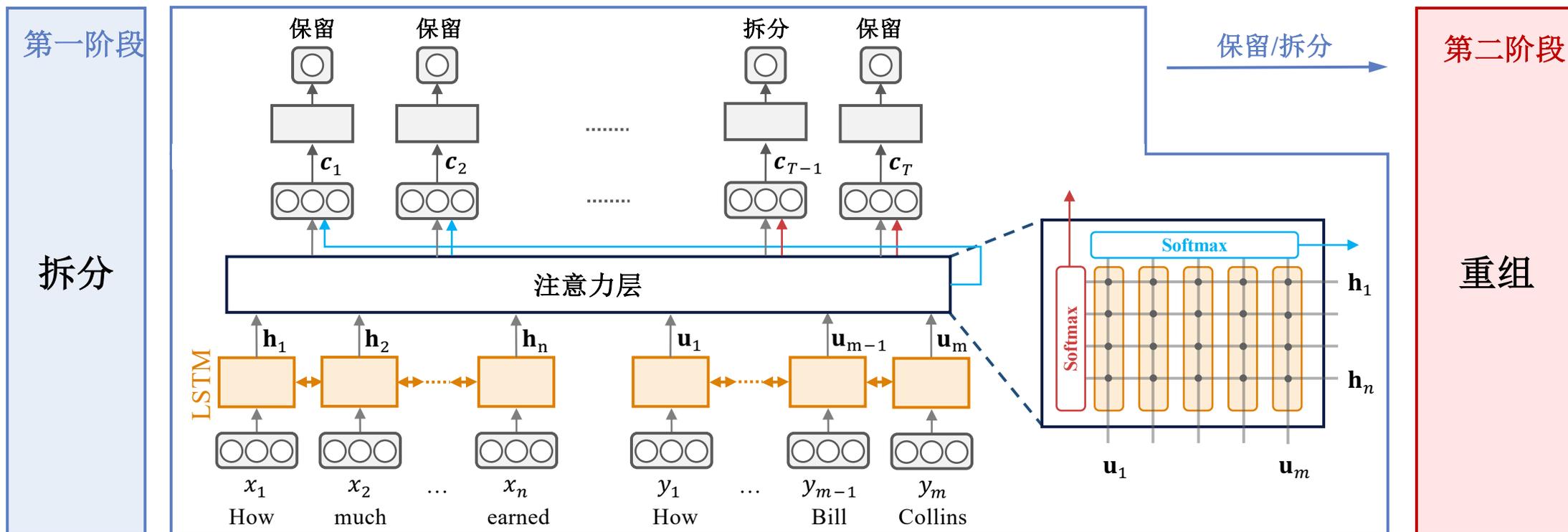
3.4 半监督下对话重写驱动的对话式语义解析方法：编辑对话

不同于逐词解码生成重写句的序列生成模型，本文提出一个以**片段**为单位，通过**编辑对话**来生成重写句的方法STAR。



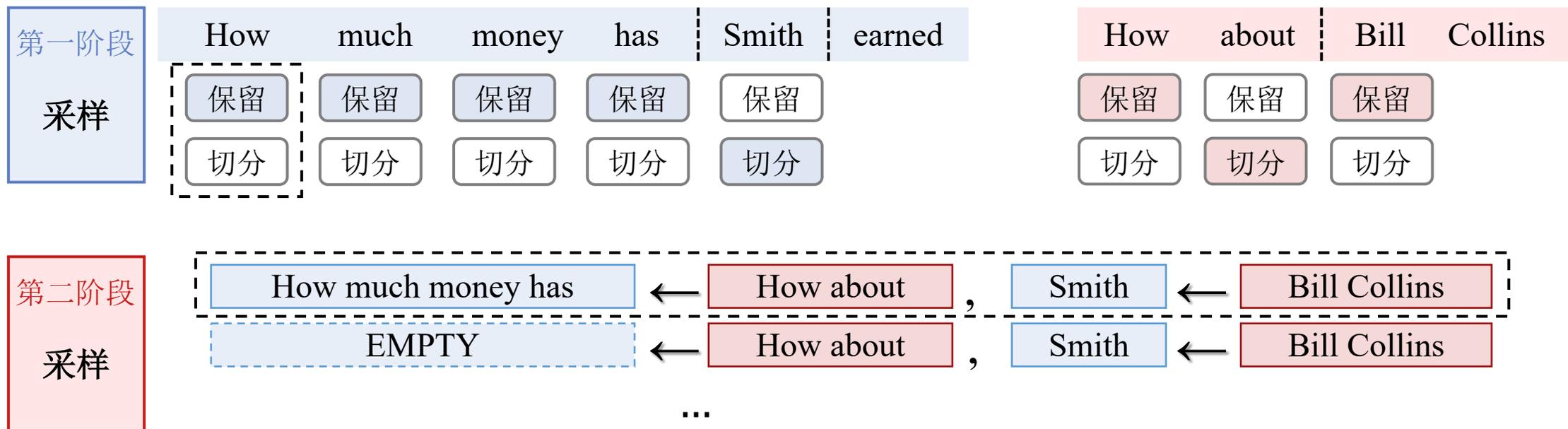
3.4 半监督下对话重写驱动的对话式语义解析方法：拆分重组

- 第一阶段**拆分**，语境句和当前句都被拆分若干个片段。
- 第二阶段**重组**，模型找出最可能的语义冲突对，通过重写过程产生重写句。



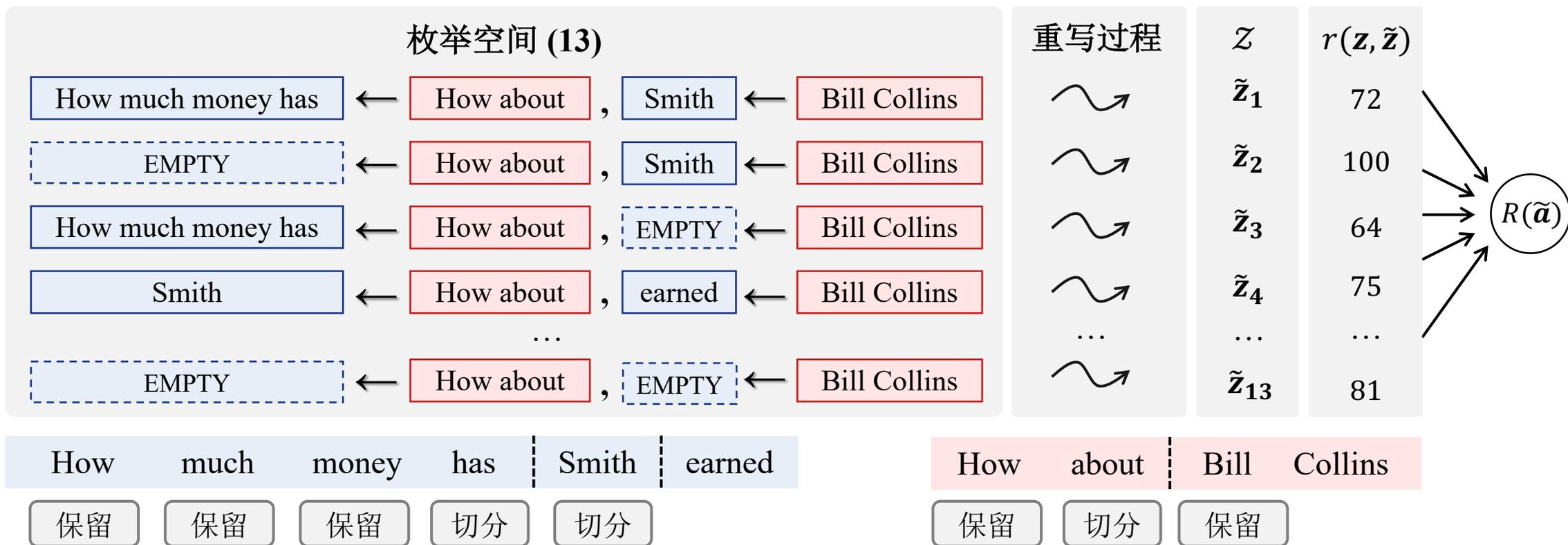
3.4 半监督下对话重写驱动的对话式语义解析方法：模型训练

推理时模型以片段为单位进行编辑，但训练数据并没有相关的监督信息，所以无法使用监督学习范式进行训练。本文使用**强化学习**联合优化两阶段。



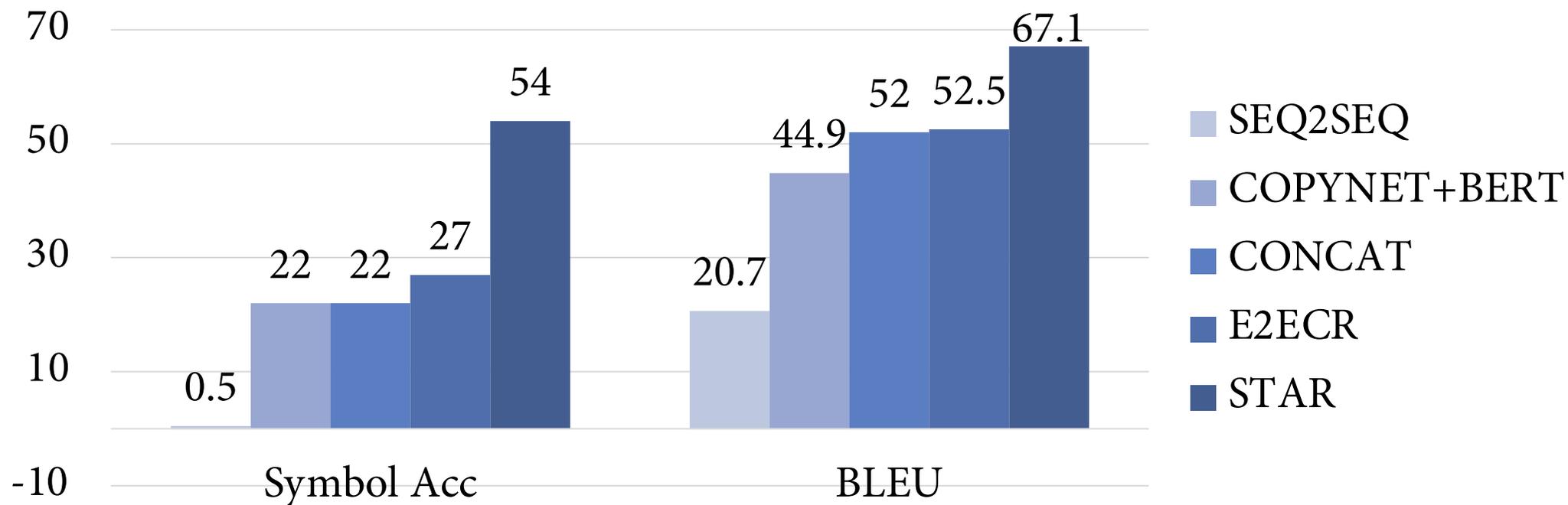
3.4 半监督下对话重写驱动的对话式语义解析方法：模型训练

若对两部分同时进行采样，采样效率较低，模型学习较为困难。为了缓解该问题，本节提出**枚举第二阶段**以提高采样效率。



3.4 半监督下对话重写驱动的对话式语义解析方法：实验结果

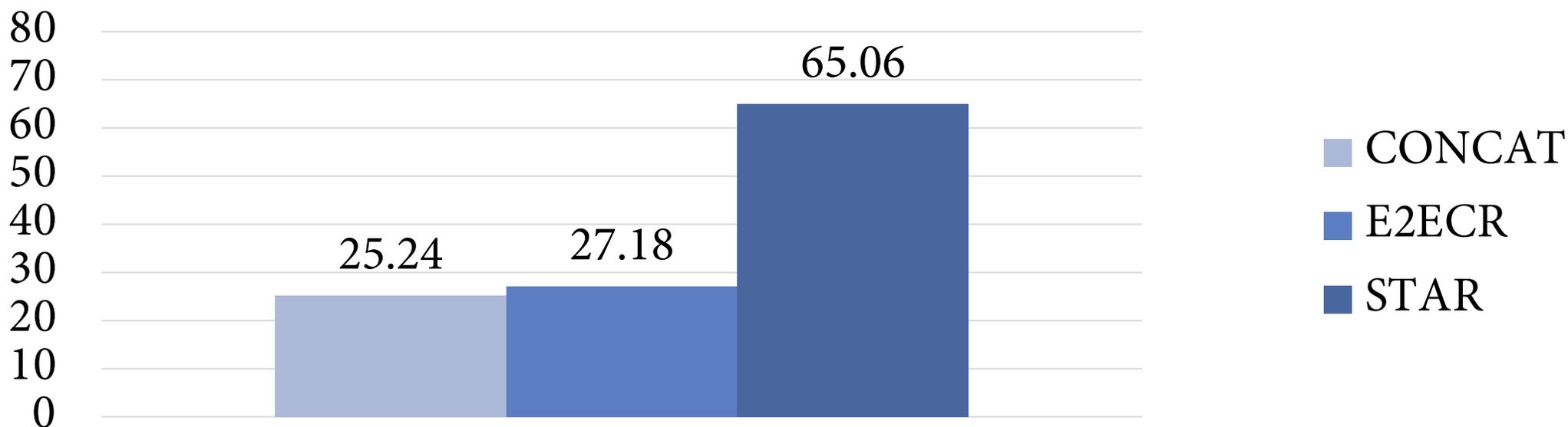
在对话重写任务上，本方法相比最好的基线模型在BLEU上获得 **14.6%** 改进。



本方法通过编辑对话可以更好地完成对话重写任务

3.4 半监督下对话重写驱动的对话式语义解析方法：实验结果

在**数据标注质量弱**时，本方法使用半监督取得全监督下 **65%** 的性能。



达成研究目标，在弱数据质量下增强了模型的性能

3.4 半监督下对话重写驱动的对话式语义解析方法：实验分析

样例 1

语境句	: [compared to Glebe park] [, does] [Hampden park] [holds more attendances at capacity ?]
翻译	: 与Glebe公园相比, Hampden公园是不是能容纳更多的游客?
当前句	: [how about] [compared to Balmoor]
翻译	: 与Balmoor相比如何?
预测重写句	: compared to Balmoor , does Hampden park holds more attendances at capacity ?
翻译	: 与Balmoor公园相比, Hampden公园是不是能容纳更多的游客?

样例 2

语境句	: [Is there any book which belongs to] [Nancy miller]
翻译	: 有没有属于南希米勒的书?
当前句	: [I mean] [the writer Nancy miller]
翻译	: 我是说作家南希米勒。
预测重写句	: Is there any book which belongs to the writer Nancy miller
翻译	: 有没有属于作家南希米勒的书?

Qian Liu, Bei Chen, Jian-Guang Lou, Ge Jin, Dongmei Zhang. FANDA: A Novel Approach to Perform Follow-up Query Analysis. *AAAI 2019 (CCF-A)*.

Qian Liu, Bei Chen, Haoyan Liu, Lei Fang, Jian-Guang Lou, Bin Zhou, Dongmei Zhang. A Split-and-Recombine Approach for Follow-up Query Analysis. *EMNLP 2019 (CCF-B)*.

汇报提纲

1

研究背景与意义

2

研究现状与难点

3

研究目标与内容

4

研究成果与创新

4. 研究成果与创新：论文创新点

- 提出一个记忆单元增强的神经网络架构，同时提出分层强化学习算法和课程学习的训练策略，显著提升了语义解析模型面向程序组合泛化的能力。
- 提出一种无需额外数据训练实体链接模型，并增强语义解析模型的方法，显著提升了语义解析模型面向知识库领域泛化能力。
- 提出一种使用生成式模型可微地解决弱监督语义解析的方法，并提出一种执行引导的预训练方法增强其性能，方法在弱监督下效果与强监督基线模型持平。
- 提出一种将对话式语义解析任务解耦为对话重写和单轮语义解析的思路，并构建对话重写数据集，方法在半监督下性能可以保持全监督下65%的性能。

4. 研究成果与创新：已发表或已录用的论文

15 篇 发表论文 **12** 篇 中国计算机学会推荐A/B类国际会议 **8** 篇 一作论文

1. **Qian Liu**, Bei Chen, Jian-Guang Lou, Ge Jin, Dongmei Zhang. FANDA: A Novel Approach to Perform Follow-up Query Analysis. AAAI 2019. (中国计算机学会推荐的A类国际会议)
2. **Qian Liu**, Yihong Chen, Bei Chen, Jian-Guang Lou, Zixuan Chen, Bin Zhou, Dongmei Zhang. You Impress Me: Dialogue Generation via Mutual Persona Perception. ACL 2020. (中国计算机学会推荐的A类国际会议)
3. **Qian Liu**, Bei Chen, Jiaqi Guo, Jian-Guang Lou, Bin Zhou and Dongmei Zhang. How Far are We from Effective Context Modeling? An Exploratory Study on Semantic Parsing in Context. IJCAI 2020. (中国计算机学会推荐的A类国际会议)
4. **Qian Liu**, Shengnan An, Jian-Guang Lou, Bei Chen, Zeqi Lin, Yan Gao, Bin Zhou, Nanning Zheng, Dongmei Zhang. Compositionality Generalization by Learning Analytical Expressions. NeurIPS 2020. (中国计算机学会推荐的A类国际会议)
5. **Qian Liu**, Dejian Yang, Jiahui Zhang, Jiaqi Guo, Bin Zhou, Jian-Guang Lou. Awakening Latent Grounding from Pretrained Language Models for Semantic Parsing. ACL 2021 Findings. (中国计算机学会推荐的A类国际会议)
6. **Qian Liu**, Bei Chen, Jiaqi Guo, Morteza Ziyadi, Zeqi Lin, Weizhu Chen, Jian-Guang Lou. TAPEX: Table Pre-training via Learning a Neural SQL Executor. ICLR 2022.
7. **Qian Liu**, Bei Chen, Haoyan Liu, Lei Fang, Jian-Guang Lou, Bin Zhou, Dongmei Zhang. A Split-and-Recombine Approach for Follow-up Query Analysis. EMNLP 2019. (中国计算机学会推荐的B类国际会议)
8. **Qian Liu**, Bei Chen, Jian-Guang Lou, Bin Zhou and Dongmei Zhang. Incomplete Utterance Rewriting as Semantic Segmentation. EMNLP, 2020. (中国计算机学会推荐的B类国际会议)

4. 研究成果与创新：在读期间所获奖励

1. 2019年度北京航空航天大学博士研究生国家奖学金
2. 2020年度百度奖学金提名奖（**全球20强**）
3. 2021年度北京航空航天大学博士研究生国家奖学金
4. 2021年度北京航空航天大学博士研究生卓越学术基金

请各位专家批评指正!

博士生：刘 乾

指导老师：赵沁平、周彬

虚拟现实技术与系统国家重点实验室

STATE KEY LABORATORY OF VIRTUAL REALITY TECHNOLOGY AND SYSTEMS