



中国科学院大学
University of Chinese Academy of Sciences

深度学习应用（自然语言处理）





目录



AI DISCOVERY

1 自然语言处理概述

基本概念、技术概览

2 自然语言处理进阶

情感分析、机器阅读、自动问答、文本生成

3 多模态融合

多模态分类、多模态检索、命名实体识别

4 应用与实践

优化方法、调参方法、可视化方法
实践：机器翻译



AI DISCOVERY



目录



AI DISCOVERY

1 自然语言处理概述

基本概念、技术概览

2 自然语言处理进阶

情感分析、机器阅读、自动问答、文本生成

3 多模态融合

多模态分类、多模态检索、命名实体识别

4 应用与实践

优化方法、调参方法、可视化方法
实践：机器翻译

AI DISCOVERY



自然语言处理流程



- ◆ 自然语言处理的两大基本步骤是**自然语言理解**和**自然语言生成**

自然语言理解

理解给定文本的含义，将人类的自然语言输入进行分析、计算，产生计算机可以理解和处理的统一形式，需要解决以下几个歧义性

- 词法歧义性：单词有多重含义
- 句法歧义性：句子有多重解析树
- 语义歧义性：句子有多重含义
- 回指歧义性：前后相同的词不同义

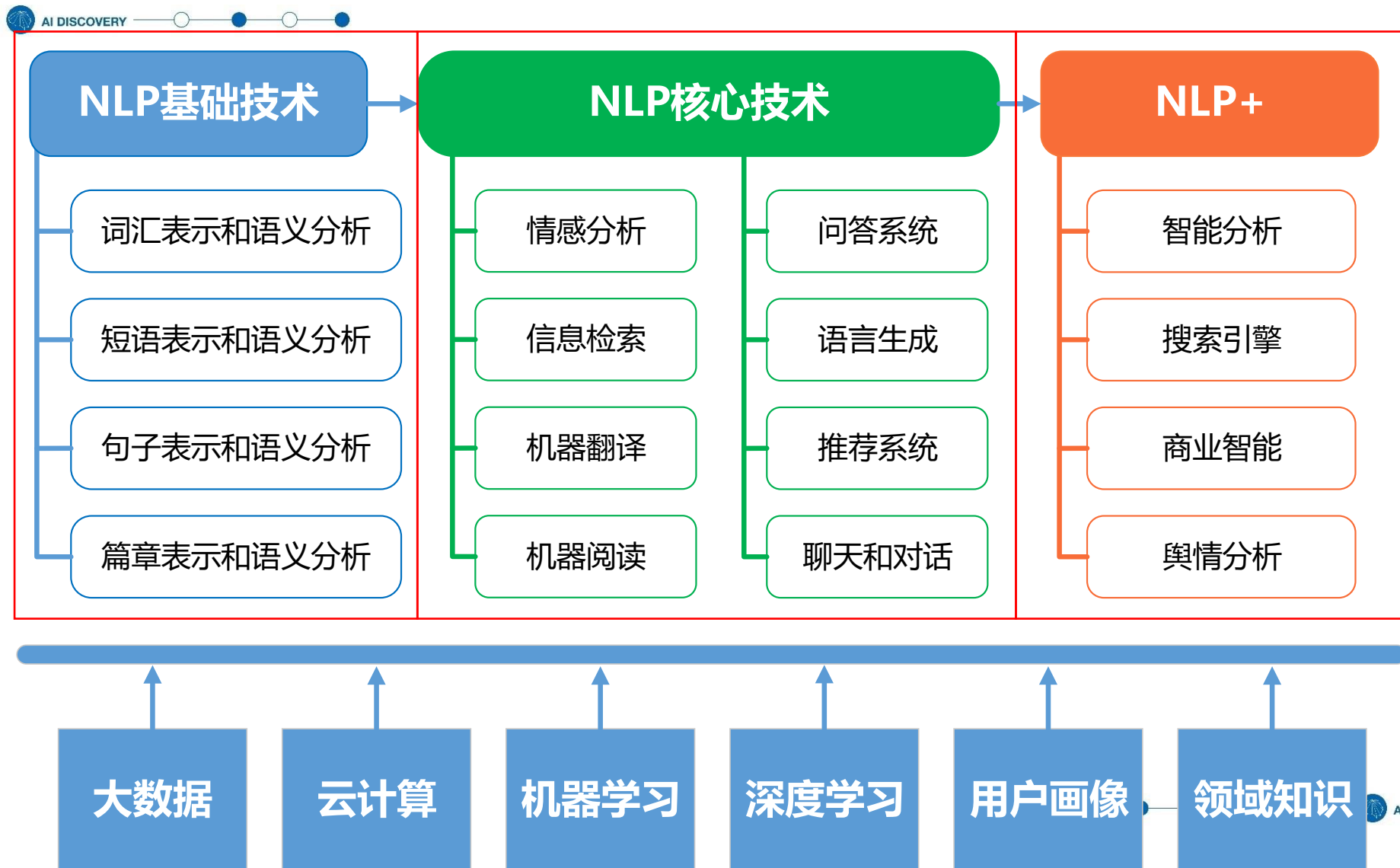
自然语言生成

将计算机产生的语义表达结果转化为人类可以读懂的自然语言的过程，大致可以分为三个阶段

- 文本规划：完成结构化数据中基础内容的规划
- 语句规划：从结构化数据中组合语句，来表达信息流
- 实现：产生语法通顺的语句



自然语言处理技术概览





目录



AI DISCOVERY

1 自然语言处理概述

基本概念、技术概览

2 自然语言处理进阶

情感分析、机器阅读、自动问答、文本生成

3 多模态融合

多模态分类、多模态检索、命名实体识别

4 应用与实践

优化方法、调参方法、可视化方法
实践：机器翻译



AI DISCOVERY



自然语言处理



AI DISCOVERY

情感分析

机器阅读

自动问答

文本生成



AI DISCOVERY



情感分析与人工智能



Marvin Lee Minsky
“人工智能之父”

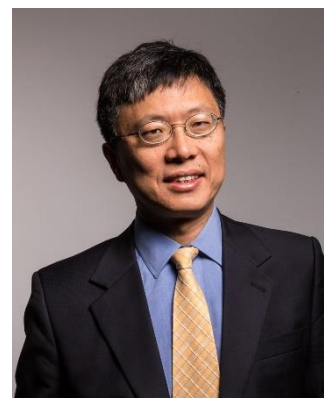
The question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without any emotions.

我认为，智能不光是 IQ，更重要的还要有情感



李飞飞
斯坦福人工智能实验室主任

人类的情感是非常丰富的。
未来，从情绪到情感，将是人工智能未来前进的方向。



沈向洋
微软全球执行副总裁



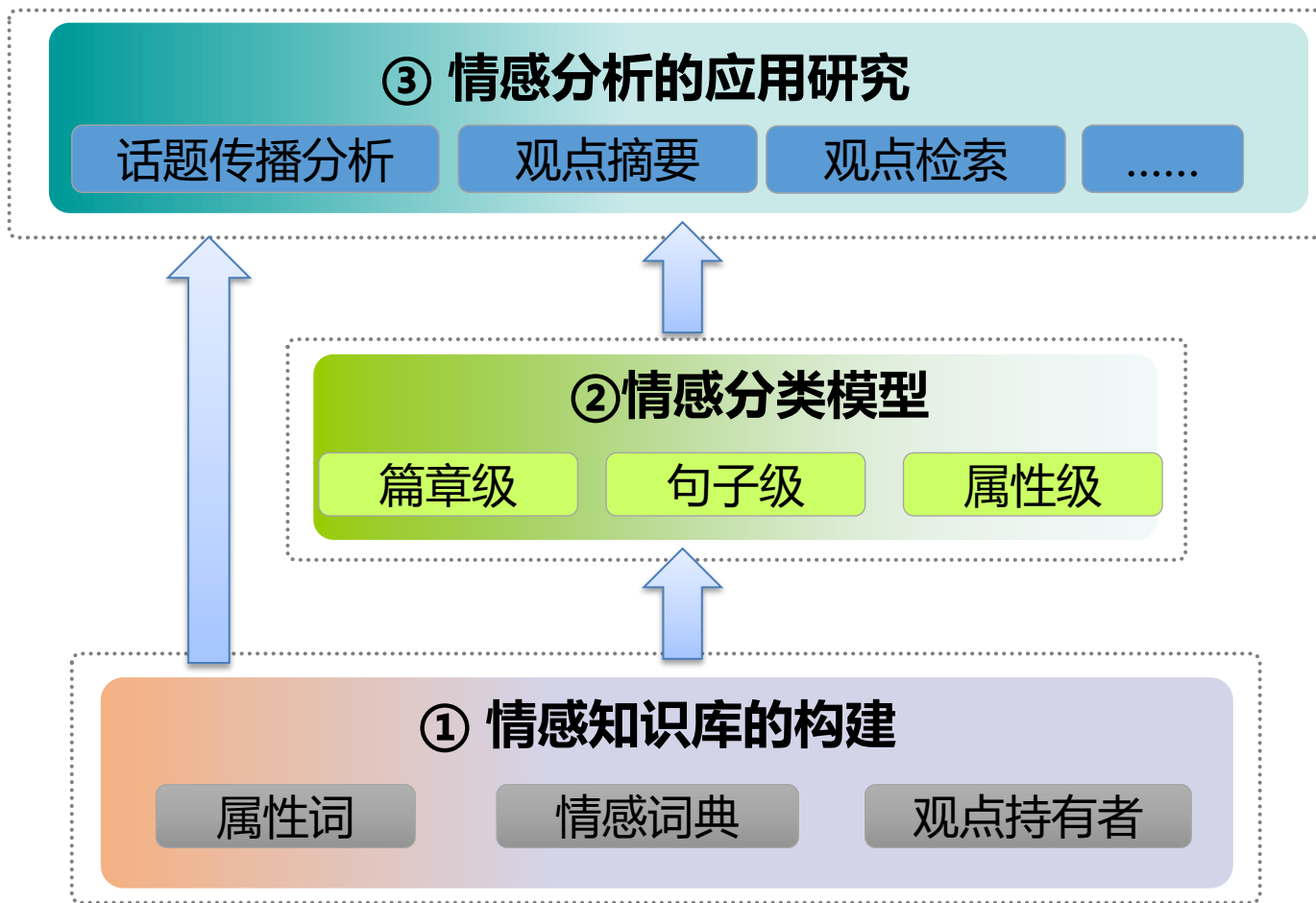
情感分析技术体系

AI DISCOVERY

应用需求

情感分析

基础资源



AI DISCOVERY

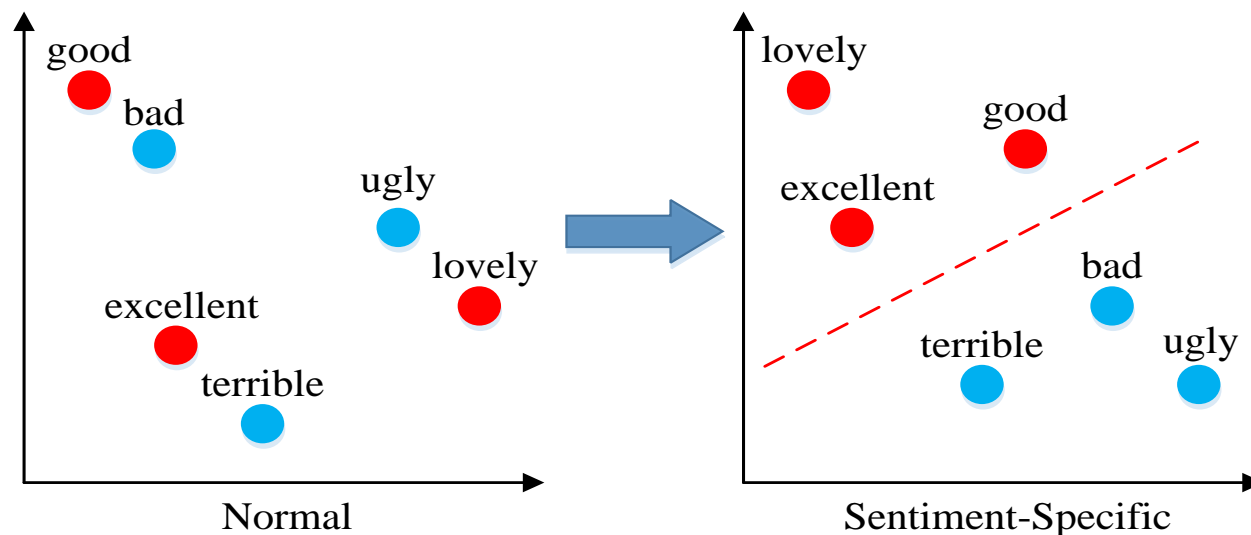


情感词向量意义

◆ 问题分析

- 传统词向量学习方法根据目标词的上下文，建立基于上下文窗口的语言模型，通过最大化目标词的概率来学习词向量
- 具有相似语法和语义的词，在词向量空间中的距离比较近
- 但与此同时，也会将具有相似上下文，但是情感极性相反的词映射到相近的向量空间，导致很难直接用于情感分析

*I feel **good** for drinking it.*
*I feel **bad** for doing it.*

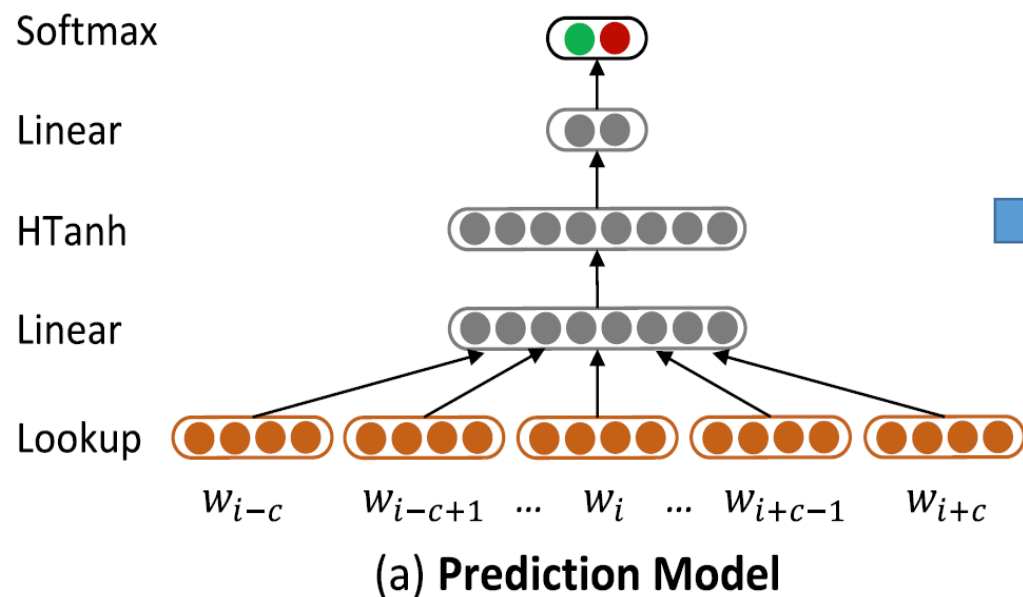




情感词向量学习模型

AI DISCOVERY

情感词向量模型是在词向量模型的基础上扩展得到的，不同之处在于引入了句子的情感信息作为监督指导



给定一个句子，设置一个固定大小的窗口在这个句子上滑动，然后根据窗口内词的词向量来预测这个窗口的情感极性。

- 模型使用正确的情感标签与预测的情感标签之间的交叉熵作为 *Softmax* 层的损失函数
- 神经网络里经过线性和非线性变换后得到的特征通常更加抽象更具有判别性
- 第二个 Linear 层的目的是将之前抽取出的特征映射到情感类别

模型假设： 同一个句子中（窗口内）的单词的情感极性趋于一致

当窗口足够大时这种假设非常适合 Twitter 语料，tweet 的篇幅通常很短，且情感极性比较集中，很少出现一条 tweet 表达两种或以上情感的情况，但这种做法并不适合篇章级的语料。

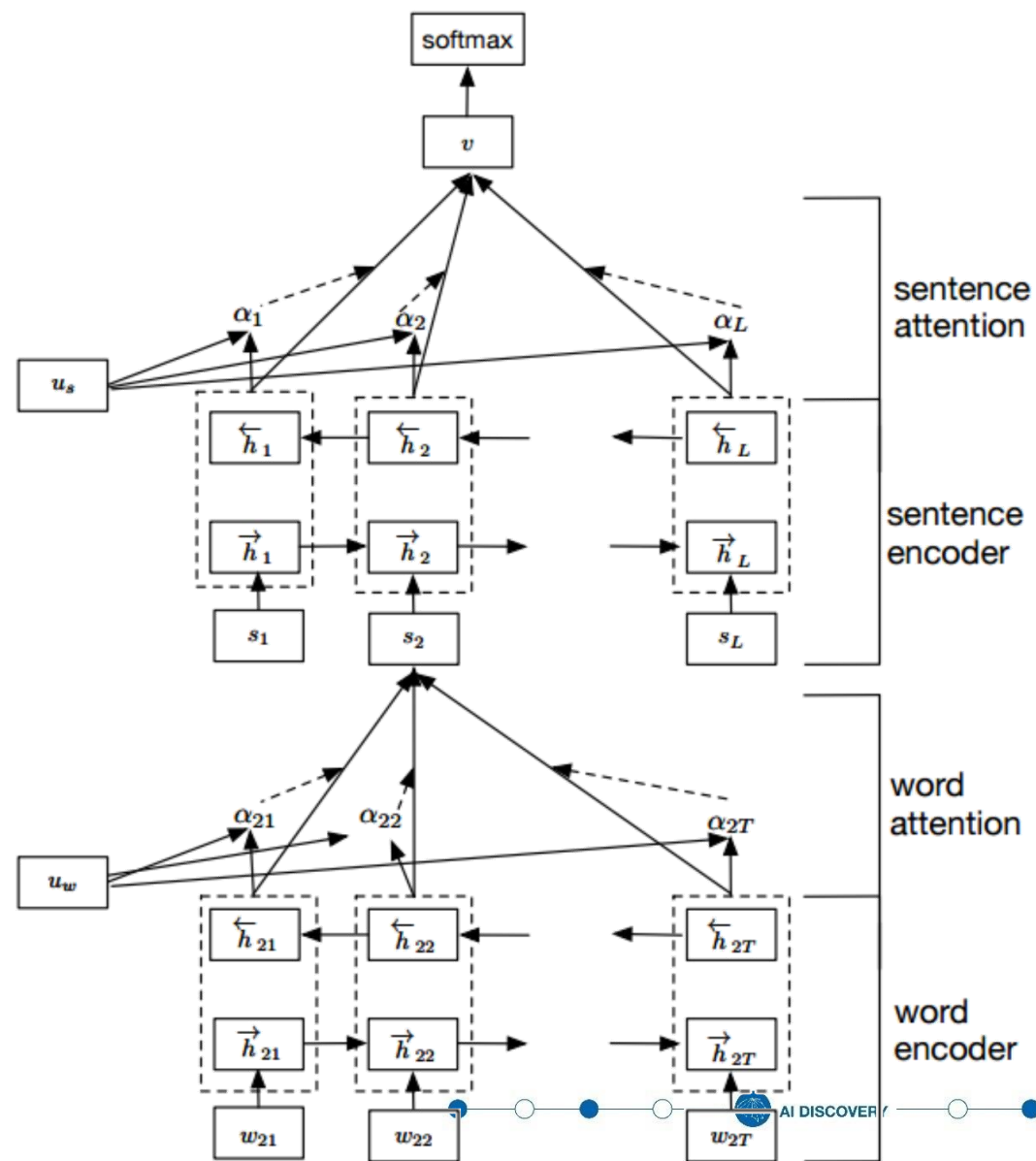


篇章级情感分类模型



AI DISCOVERY

- ◆ 篇章级情感分析是指对**整篇文档**的**全局情感**极性进行分析判断
- ◆ 基本思想是**对词->句子->篇章**逐层进行语义编码表示，得到篇章的向量表示
- ◆ 好处是让机器像人一样对篇章进行阅读：从简单的字词组成句子，句子组成篇章，最后形成思想，这就是自然语言处理中的**层级 (Hierarchical) 概念**





句子级情感分类模型



- ◆ 句子级情感分析是指对**单句**的情感极性进行分析判断的任务
- ◆ 基于**深度神经网络**的方法主要是对句子所包含的语义信息进行表示, 进而对其情感极性进行判别
- ◆ 常用的方法有基于**CNN**、**RNN**、**Recursive-NN**的方法以及最新的**BERT**等

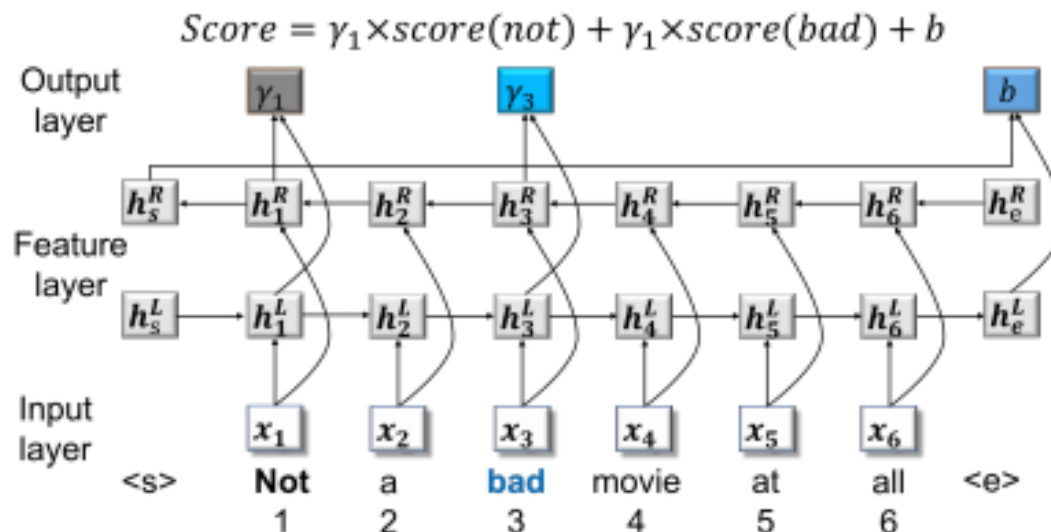
The performances were all **really** It's an **insignificant** **[criticism]**_{-1→-0.5}.

Nobody gives a **[good]**_{+3→-1} performance in this movie

She's **not** **[terrific]**_{+5→+1} but **not** **[terrible]**_{-5→-1} either.

It's **not** a **very** **[good]**_{+3→-0.25} movie song!

It **removes** my **[doubts]**_{-3→+1}.





属性级情感分类模型



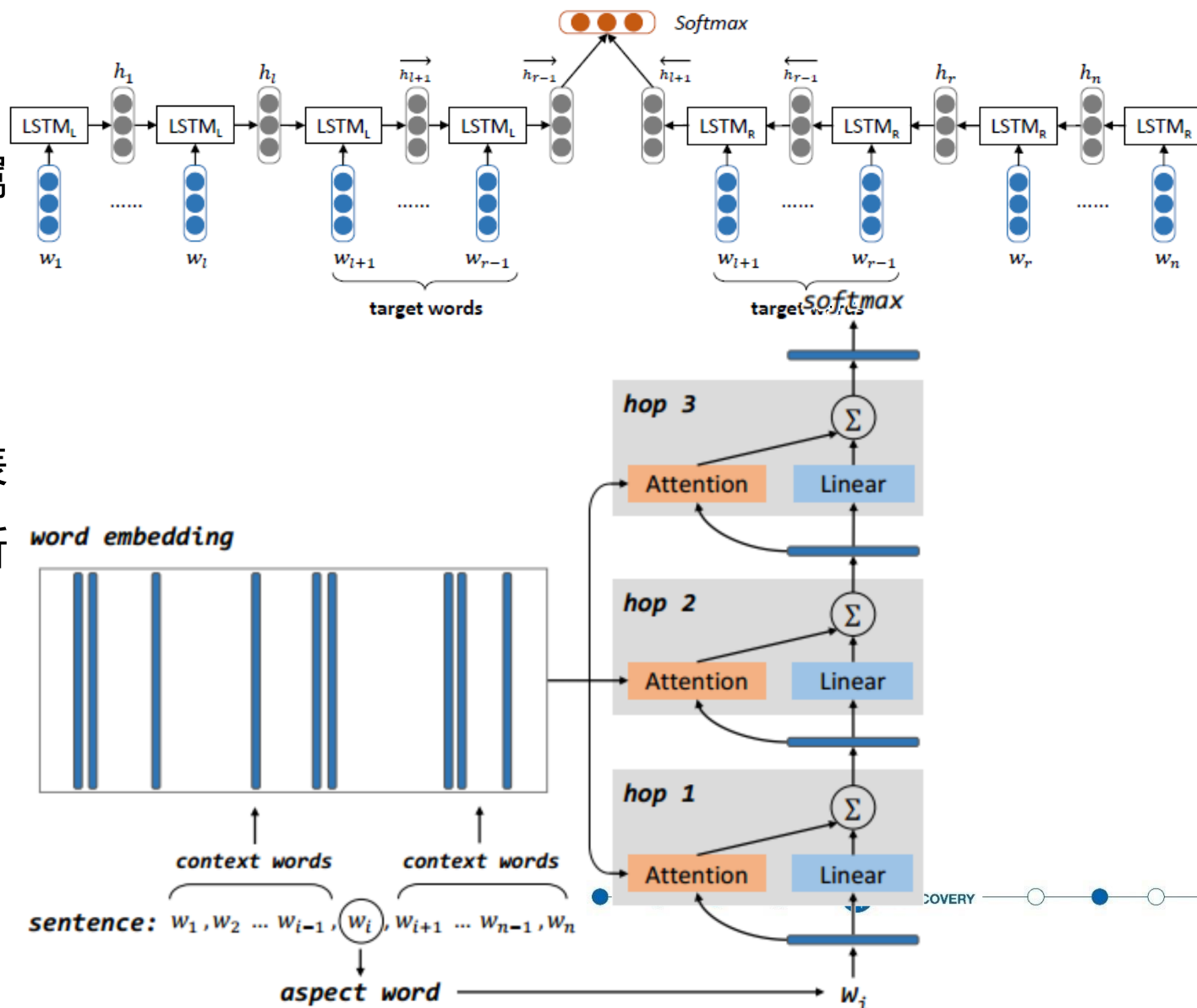
- ◆ 属性级情感分析 (Aspect-Level) 是 **细粒度情感分析**，指对所描述事物的属性情感极性进行判断

➢ 我非常**喜欢**这款手机的**摄像头**

- ◆ 基本思想是对**属性词和其上下文**进行表示，并建立它们之间的关系，进而判断情感极性

- ◆ 有两类方法

- **分段式表示** (RNN、CNN)
- **整体表示** (记忆网络MemNN)

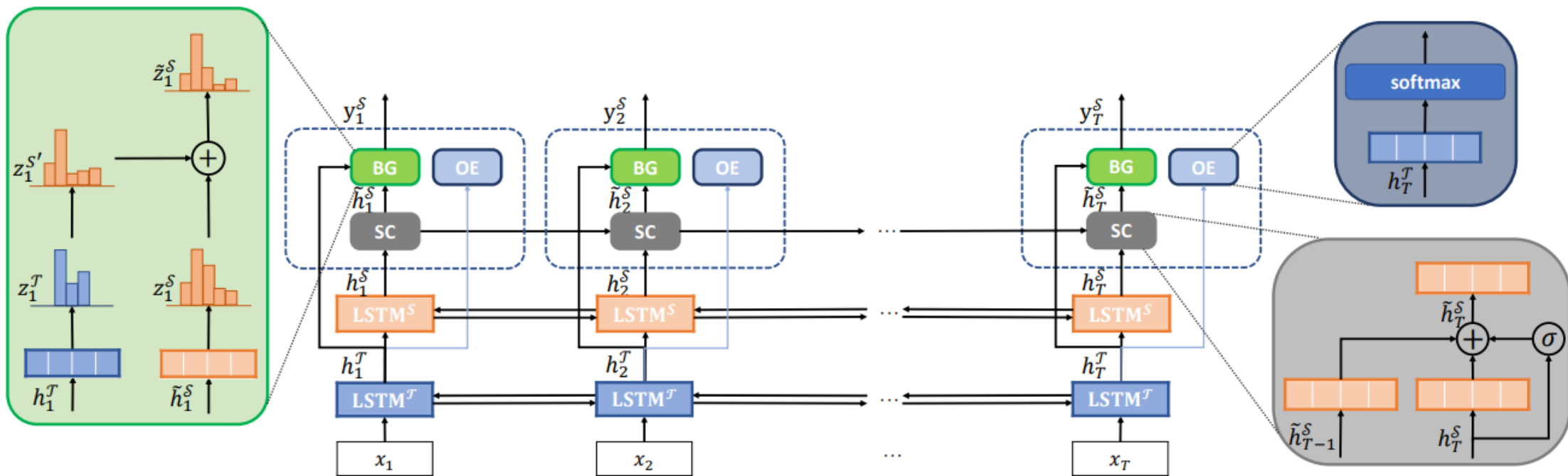




意见目标抽取和目标情感预测

◆ A Unified Model for Opinion Target Extraction and Target Sentiment Prediction, AAI2019

- 通过改进基本的序列标注模型来提升端到端抽取用户评论中的意见目标和相应的情感倾向的性能
- 提出了一个新的统一模型并在模型中采用了一种联合的标注方案，模型使用两层循环神经网络：
 - 上层的RNN用来预测联合的标签，这些标签会作为意见目标的情感分析的输出
 - 下层的RNN用于意见目标实体的边界检测。
- 设计一个组件对同一个意见目标内的情感一致性进行维持，基于门控制机制来对上一个时刻的特征和当前时刻的特征之间的关系进行建模

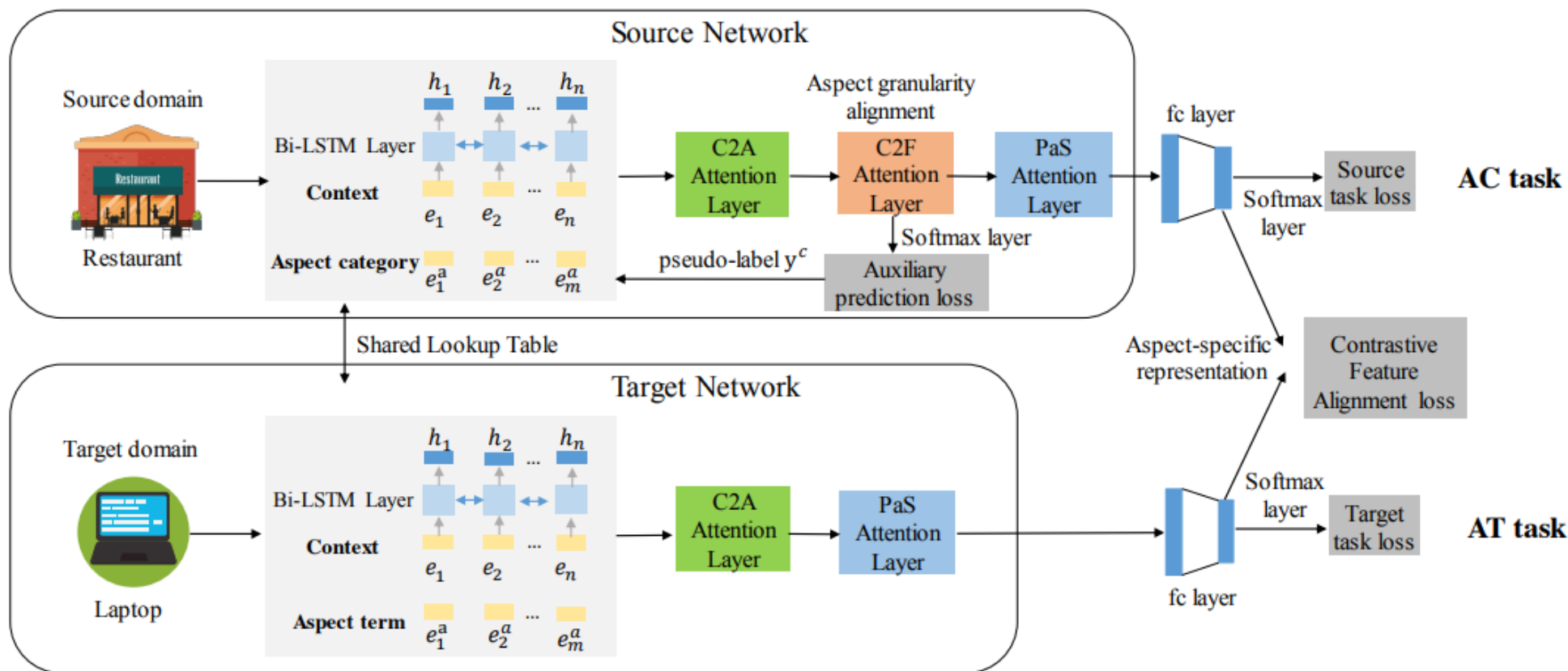




从粗粒度到细粒度的迁移学习

◆ Exploiting Coarse-to-Fine Task Transfer for Aspect-level Sentiment Classification, AAAI2019

- 当前AT-level的公开数据集都相对较小,受限于这些小规模的数据集,当前大多数方法依赖于复杂的结构,从而限制了神经网络模型的有效性。
- 该论文从数据相对充足的粗粒度任务 (AC-level) 到数据稀疏的细粒度任务 (AT-level) 进行迁移学习, 提出一个多粒度对齐网络 (MGAN)





自然语言处理



AI DISCOVERY

情感分析

机器阅读

自动问答

文本生成



AI DISCOVERY

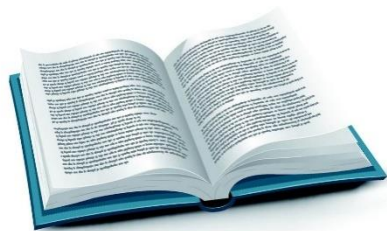


什么是机器阅读



AI DISCOVERY

- ◆ **机器阅读**旨在让AI代替人类，自动阅读新闻报道等信息并根据问题给出答案，可应用于舆情分析和情报挖掘等多个场景。
- ◆ **机器阅读**是自然语言处理领域“皇冠上的明珠”，涉及到**语义理解**、**知识推理**等多项复杂技术，极具挑战性。



情报信息



AI阅读者



问题：谁是美国网络部队司令官？

答案：亚历山大



AI DISCOVERY



AI DISCOVERY



机器阅读的难点挑战



AI DISCOVERY

语言博大精深

“作为失败的典型，你
真的太成功了”

语义
推理难

语义
关联难

语义表示难

字同意不同

“谁是特朗普的儿子”
“谁的儿子是特朗普”

分词引歧义、一词有多义

“昨天/下午” & “昨/天下/午”
“我爱吃苹果” & “苹果发布会”



AI DISCOVERY



机器阅读的主要步骤

将文本表示成机器能理解的符号

寻找问题和原文句子的语义关联

对语义关联进行加工和推理

对候选答案进行排序和输出

文本表示

语义匹配

理解推理

结果推荐

Word Vector
Character CNN
ELMO

Sentence Vector
Bi-Attention
Self-Attention

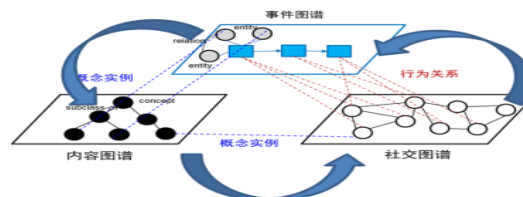
Memory Network
Knowledge Graph

Pointer Network
Learning2Rank

深度学习模型



神经网络推理模型





机器阅读数据集——MC Test



◆ 问题描述

P: **James the Turtle was always getting in trouble.** Sometimes he'd reach into the freezer and empty out all the food. Other times he'd sled on the deck and get a splinter. His aunt Jane tried as hard as she could to keep him out of trouble, but he was sneaky and got into lots of trouble behind her back.

Q: What is the name of the trouble making turtle?
(A) Fries (B) Pudding (C) James (D) Jane

A: James

➤ 给定一篇文章、一个问题以及四个选项，预测哪个选项是问题的答案





机器阅读数据集——CNN/Daily Mail



◆ 问题描述

P: the ent381 producer allegedly struck by ent212 will not press charges against the “ ent153 ” host , his lawyer said friday . ent212 , who hosted one of the most - watched television shows in the world , was dropped by the ent381 wednesday after an internal investigation by the ent180 broadcaster found he had subjected producer ent193 “ to an unprovoked physical and verbal attack . ”

Q: producer X will not press charges against ent212 , his lawyer says .

A: ent193

➤ 给定一篇文章和一个问题，预测问题中被X替换的实体是文章中的哪个实体





机器阅读数据集——SQuAD



◆ 问题描述

P: In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under gravity. The main forms of precipitation include drizzle, rain, sleet, snow, graupel and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals within a cloud. Short, intense periods of rain in scattered locations are called “showers”.

Q: What causes precipitation to fall?

A: gravity

➤ 给定一篇文章、一个问题，预测文章中的一个区间（一个或几个词）作为答案



机器阅读数据集——Quasar-T



◆问题描述

P1: In July 2007 , 7-Eleven redesigned some stores to look like Kwik-E-Marts in select cities to promote **The Simpsons Movie**.

P2: Tie-in promotions were made with several companies, including 7-Eleven, which transformed selected stores into Kwik-E-Marts.

P3: “ 7-Eleven Becomes Kwik-E-Mart for ‘ **Simpsons Movie** ’ Promotion ” .

Q: 7-Eleven stores were temporarily converted into Kwik E-marts to promote the release of what movie?

A: the simpsons movie

➤ 给定几篇文章、一个问题，预测文章中的一个区间（一个或几个词）作为答案



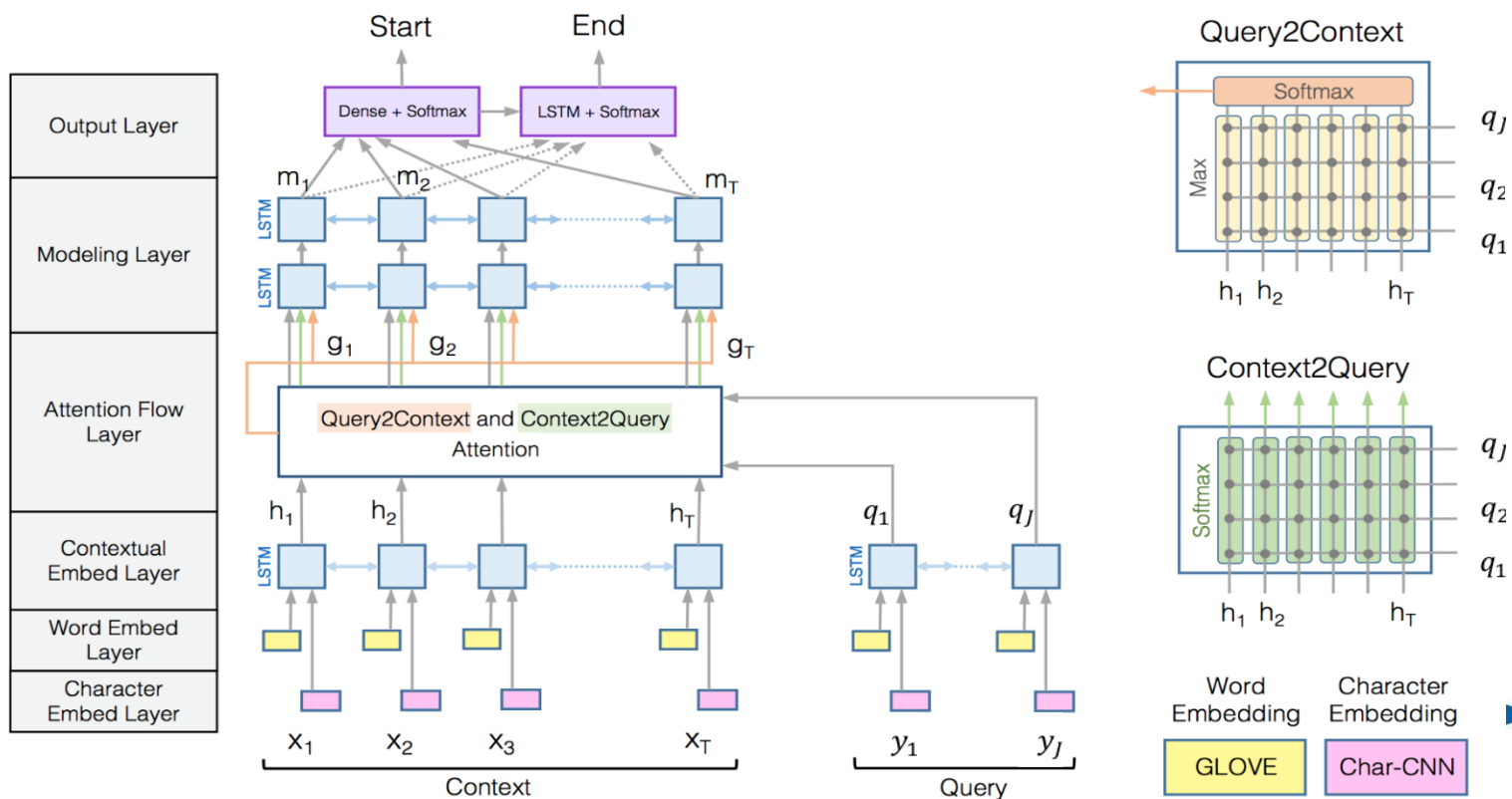
机器阅读模型 (BiDAF)



AI DISCOVERY

◆ Bidirectional Attention Flow For Machine Comprehension, ICLR 2017

- 输入：一篇文章 X 和一个问题 Y
- 输出：文章中每一个词作为答案开始的概率和答案结束的概率



AI DISCOVERY



机器阅读模型 (BiDAF)

AI DISCOVERY

◆ 输入部分

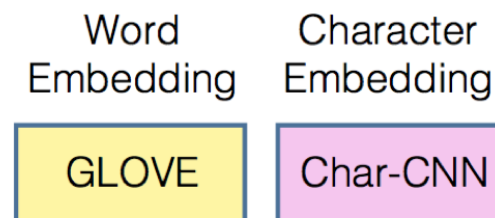
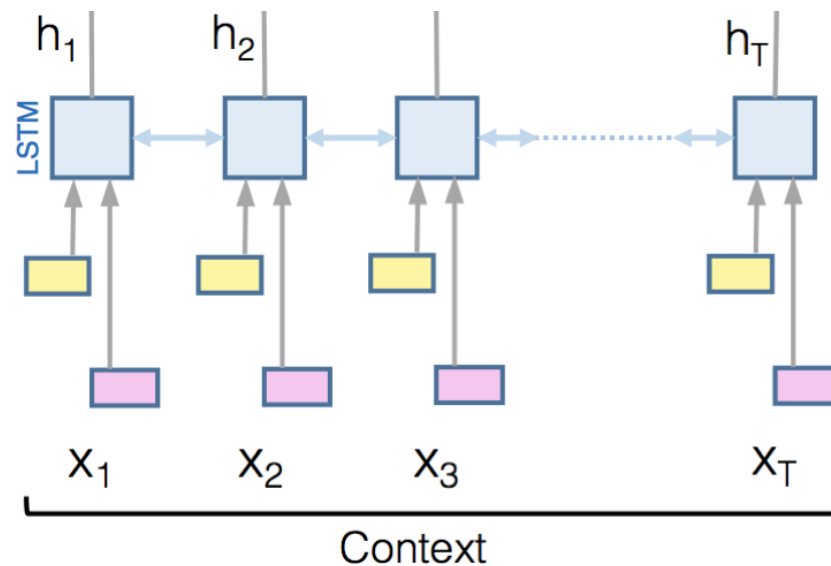
Context: $[x_1, x_2, \dots, x_T]$

Query: $[y_1, y_2, \dots, y_I]$

Context: $[h_1, h_2, \dots, h_T]$

Query: $[q_1, q_2, \dots, q_I]$

- 拼接每一个词对应的**GLOVE词向量**和**Char-CNN词向量**
- 经过一层BiLSTM层得到**Context**和**Query**的向量表示。





机器阅读模型 (BiDAF)

AI DISCOVERY

◆ 双向注意力

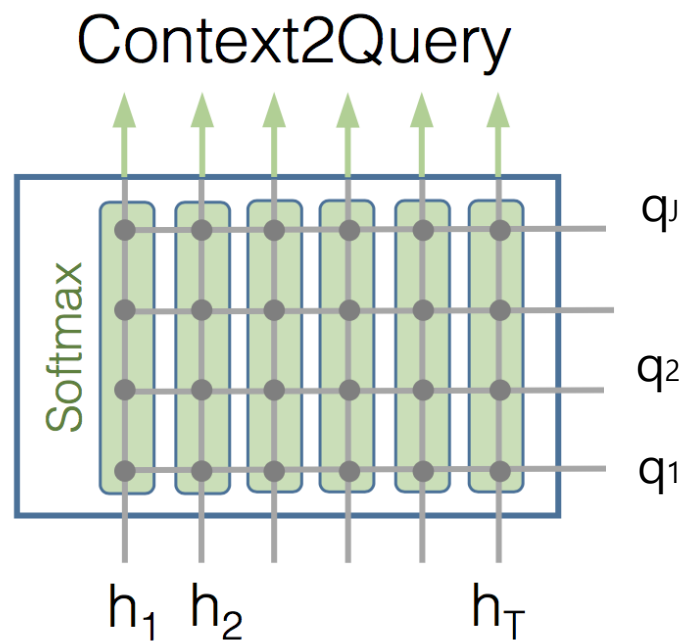
Context: $[h_1, h_2, \dots, h_T]$

Query: $[q_1, q_2, \dots, q_J]$

$$a_{ij} = \mathbf{w}_1 \cdot \mathbf{h}_i + \mathbf{w}_2 \cdot \mathbf{q}_j + \mathbf{w}_3 \cdot (\mathbf{h}_i \odot \mathbf{q}_j)$$

$$p_{ij} = \frac{e^{a_{ij}}}{\sum_{j=1}^{n_q} e^{a_{ij}}}$$

$$\mathbf{c}_i = \sum_{j=1}^{n_q} \mathbf{q}_j p_{ij}$$





机器阅读模型 (BiDAF)

AI DISCOVERY

◆ 双向注意力

Context: $[h_1, h_2, \dots, h_T]$

Query: $[q_1, q_2, \dots, q_J]$

$$a_{ij} = \mathbf{w}_1 \cdot \mathbf{h}_i + \mathbf{w}_2 \cdot \mathbf{q}_j + \mathbf{w}_3 \cdot (\mathbf{h}_i \odot \mathbf{q}_j)$$

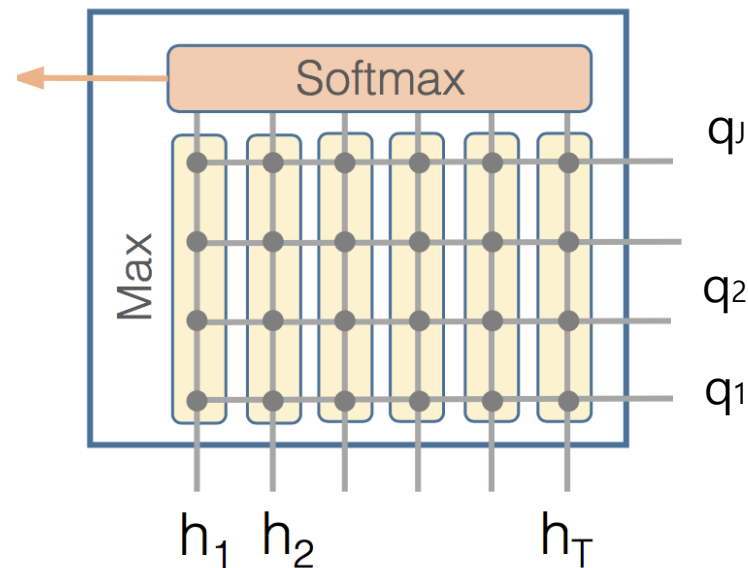
$$m_i = \max_{1 \leq j \leq n_q} a_{ij}$$

$$p_i = \frac{e^{m_i}}{\sum_{i=1}^{n_c} e^{m_i}}$$

$$\mathbf{q}_c = \sum_{i=1}^{n_c} \mathbf{h}_i p_i$$

$\mathbf{G} : [g_1, g_2, \dots, g_T]$, $g_i = [h_i; c_i; h_i \odot c_i; h_i \odot q_c]$

Query2Context





机器阅读模型 (BiDAF)



◆ 输出部分

Context: $[h_1, h_2, \dots, h_T]$

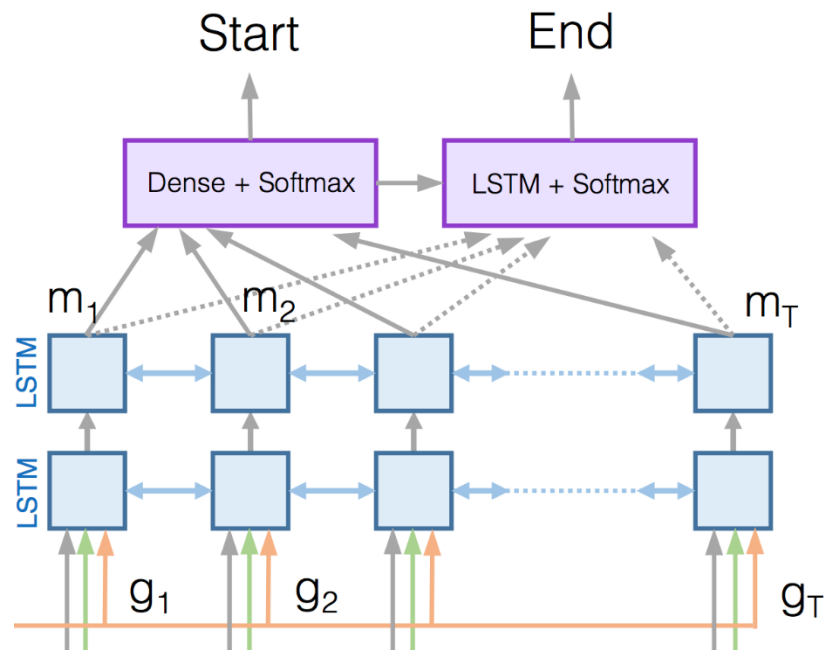
Query: $[q_1, q_2, \dots, q_J]$

$$M = BiLSTM(G)$$

$$M^2 = BiLSTM(M)$$

$$p^1 = \text{softmax}(w_{(p^1)}^T [G; M]),$$

$$p^2 = \text{softmax}(w_{(p^2)}^T [G; M^2])$$





机器阅读最新进展

AI DISCOVERY

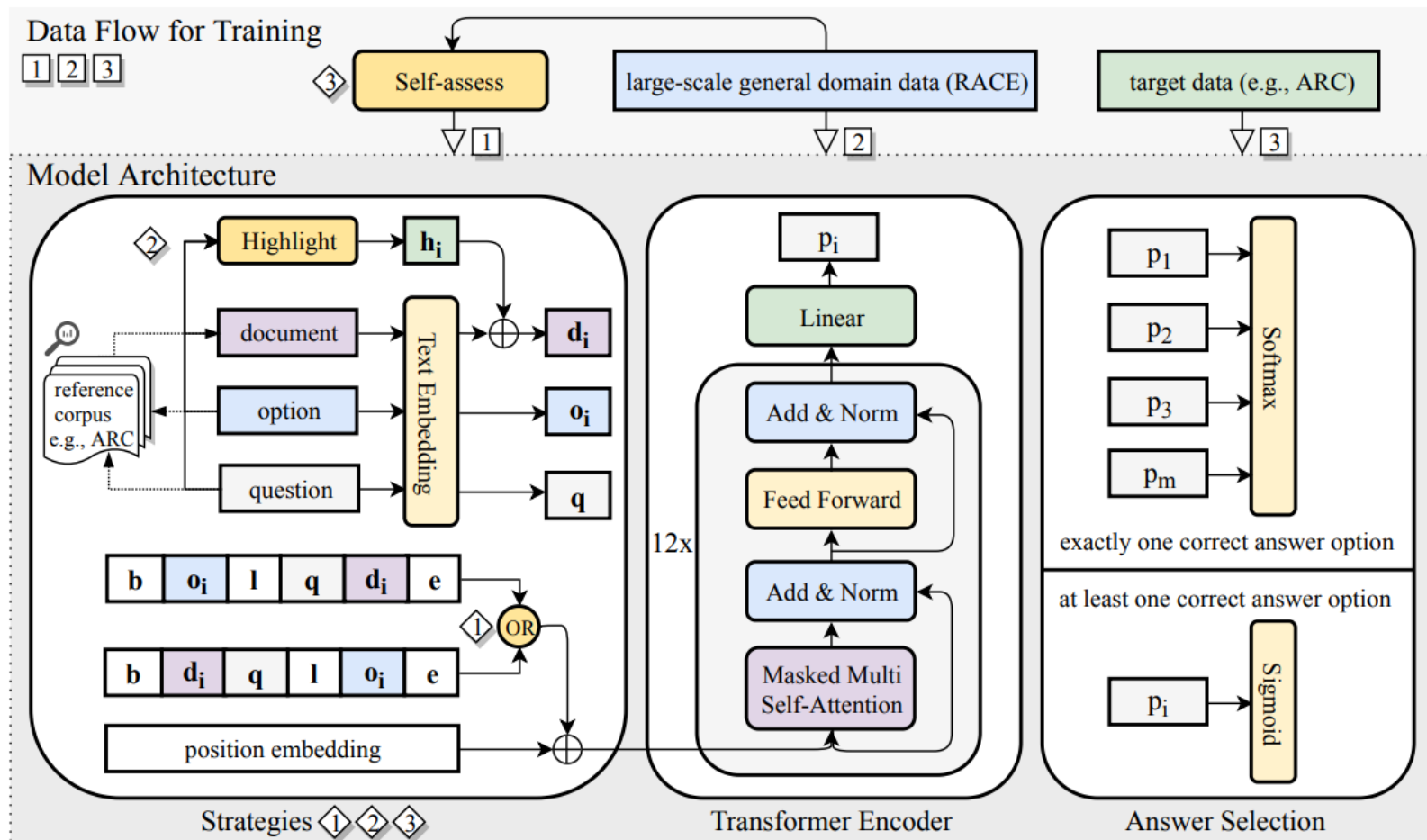
◆ Improving Machine Reading Comprehension with General Reading Strategies, NAACL2019

➤ 阅读策略被证明可以提高阅读理解的水平，本文受此启发提出三种阅读策略

- Back and Forth Reading
- Highlighting
- Self-Assessment

➤ 该工作的创新之处是提出三种阅读策略，因此其阅读模型是基于已有模型的改进。

➤ 该模型用到了GPT (Generative Pre-trained Transformer)





机器阅读新问题—对抗阅读



◆ Adversarial Examples for Evaluating Reading Comprehension Systems, EMNLP2017

- 文本针对斯坦福问答数据集 (SQuAD) 提出一个对抗评估方案
- 自动生成并在段落中插入对抗语句，然后测试系统是否能围绕这一段回答
- 这些对抗语句的目的在于干扰计算机回答系统，但并不会改变问题的正确答案，也不会对人类造成干扰
- 在这种对抗方案中，16 个已发表模型的精确度从 75% (F1 分数) 降到了 36%。当对抗系统被允许加入不符合语法的短语串，有四个模型的平均精度下降到了 7%

Article: Super Bowl 50

Paragraph: “*Peyton Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver’s Executive Vice President of Football Operations and General Manager. Quarterback Jeff Dean had jersey number 37 in Champ Bowl XXXIV.*”

Question: “*What is the name of the quarterback who was 38 in Super Bowl XXXIII?*”

Original Prediction: John Elway

Prediction under adversary: Jeff Dean





自然语言处理



AI DISCOVERY

情感分析

机器阅读

自动问答

文本生成



AI DISCOVERY





什么是问答系统

- ◆ 问答系统被认为是图灵测试的原始形式
- ◆ 问答系统是下一代搜索引擎的基本形态



Search needs a shake-up

On the twentieth anniversary of the World Wide Web's public release, **Oren Etzioni** calls on researchers to think outside the keyword box and improve Internet trawling.

Two decades after Internet pioneer Tim Berners-Lee introduced his World Wide Web project to the world using the *alt.hypertext* newsgroup, web search is on the cusp of a profound change — from simple document retrieval to question answering. Instead of poring over long lists of documents that contain requested keywords, users need direct

for bygone technologies such as electric typewriters and vinyl records. But this transformation could be unreasonably delayed. As a community, computer scientists have underinvested in tools that can synthesize sophisticated answers to questions, and have instead focused on incremental progress in lowest-common-denominator search. The clas-

search. They must invest much more in bold strategies that can achieve natural-language searching and answering, rather than providing the electronic equivalent of the index at the back of a reference book. Today, the "book" is distributed over billions of web pages of uneven quality, and much effort has been directed at ranking the most useful results. Such engines readily



Prof. Oren Etzioni

Turing Center
University of Washington

以直接而准确的方式回答用户自然语言提问的自动问答系统将构成下一代搜索引擎的基本形态

— 《Nature》2011.8

□输入：自然语言的问句，而非关键词的组合

谁获得1987年的诺贝尔文学奖？

□输出：直接答案，而非文档集合

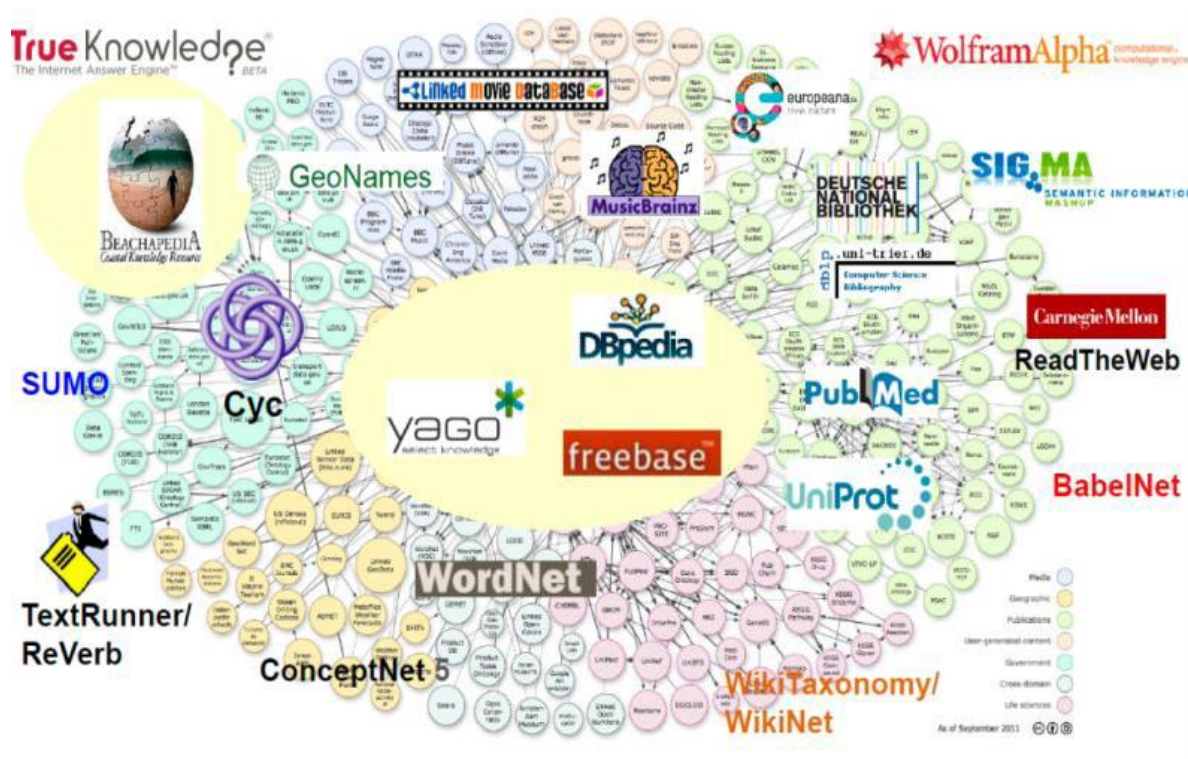
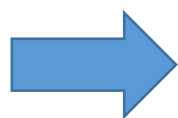
约瑟夫·布罗茨基



基于知识图谱的问答系统



我国有两个省接壤的省份最多，与它们接壤的省分别有哪些？



- 内蒙古：黑龙江、吉林、辽宁、河北、山西、陕西、宁夏、甘肃
- 陕西：内蒙古、宁夏、甘肃、四川、重庆、湖北、河南、山西



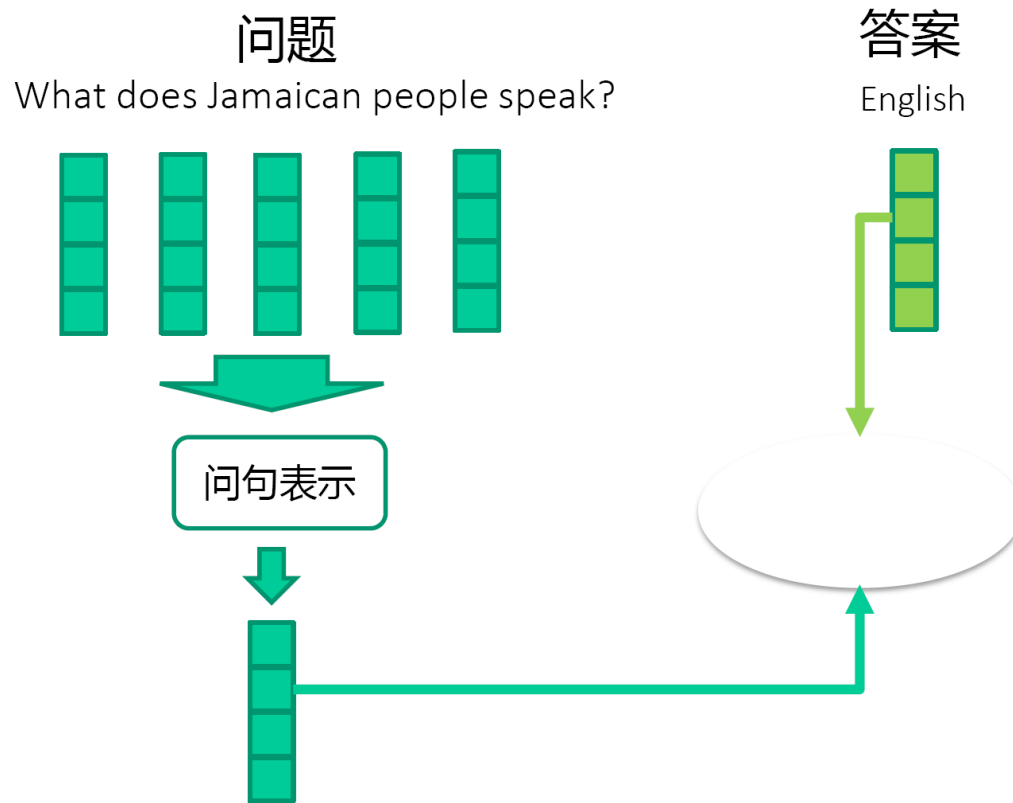


基于知识图谱的问答——深度学习学习方法



◆ 深度学习方法的关键问题

- 对问题的表示
- 对答案的表示
- 建立问题和答案之间的语义关联





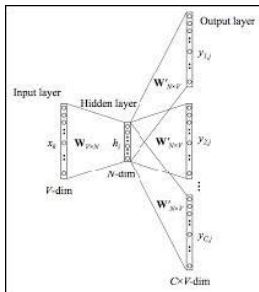
文本、知识的深度表示



AI DISCOVERY

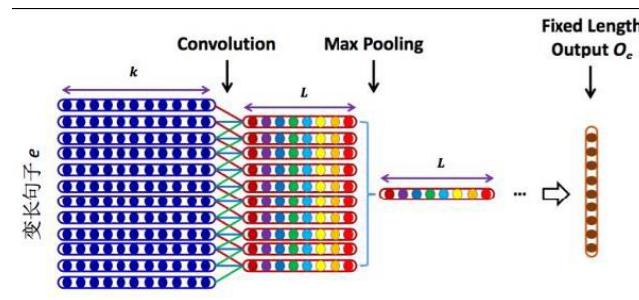
◆词的向量化

- NNLN, C&W, Skip-gram



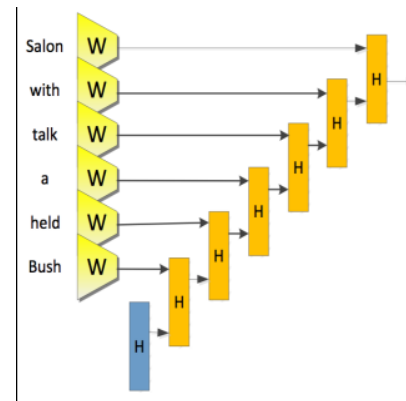
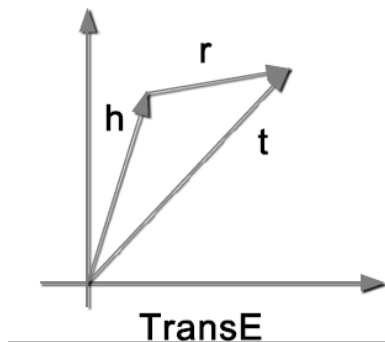
◆句子（文本）的向量化

- 直接累加
- 卷积神经网络 (Convolutional Neural Network)
- 递归神经网络 (Recursive neural network)
- 循环神经网络 (Recurrent neural network)



◆知识（事实、命题）的向量化

- REACAL
- TransE
- NTN



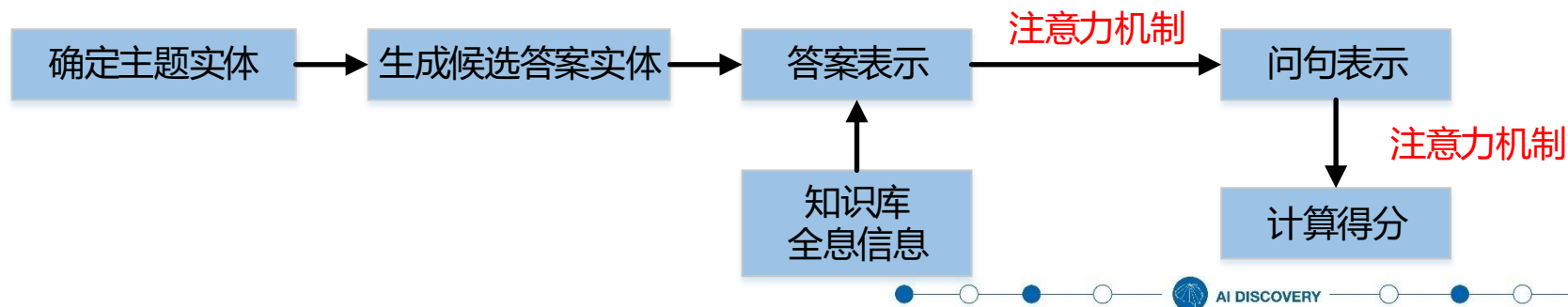
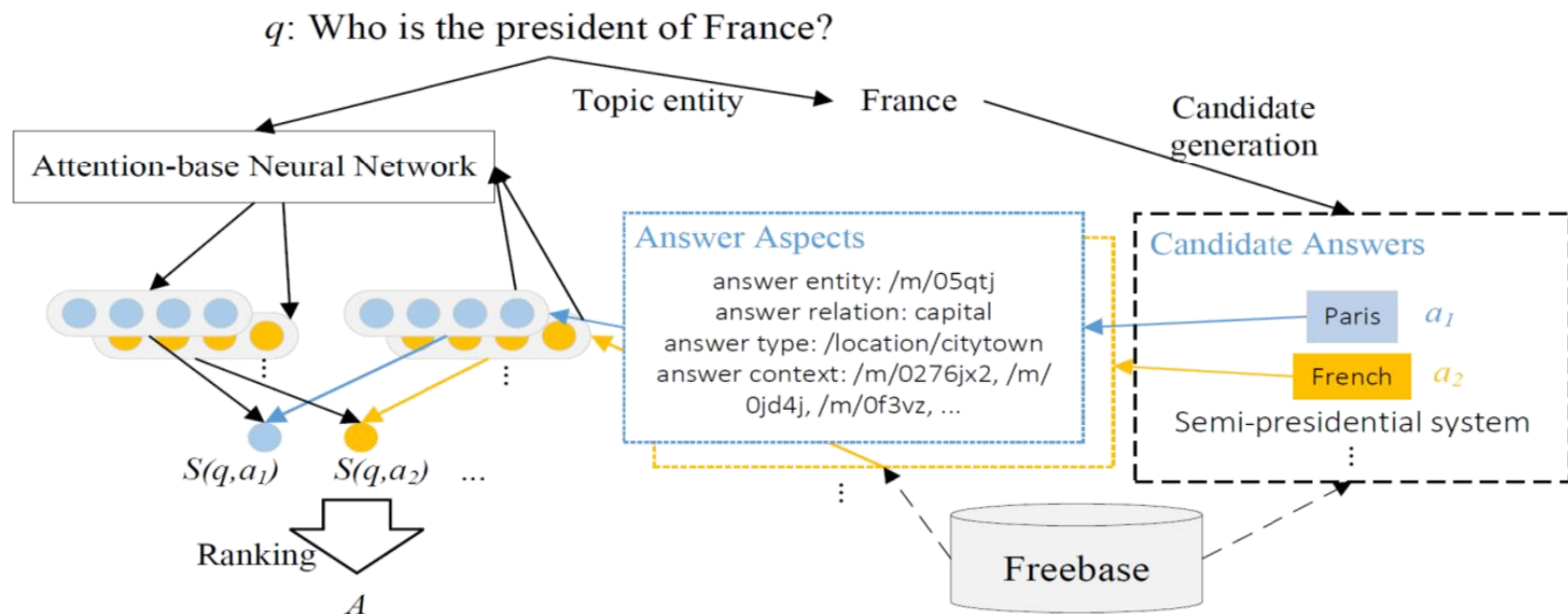
AI DISCOVERY



基于知识图谱的问答模型

✓ **Motivation:** 一个答案是否正确，应该从这四个方面去考虑：候选实体本身、它和主题实体的关系、它的类别以及知识库当中的关联实体。

✓ **方法:** 对于一个特定的候选答案，基于它的四个方面利用 cross-attention 的机制分别表示问句，进而计算候选答案和问题的相似度





基于推理的问答系统



AI DISCOVERY

推理问答主要考验机器的**智能理解能力**，它可以通过对已知知识的推理来得到**未知的知识**。推理问答的输入不仅有**问题**，还有**上下文**，它能够在阅读理解上下文之后，对知识进行推理，然后得到问题的**正确答案**。

John is in the playground.
Bob is in the office.
Where is John? **A:playground**

John is in the playground.
Bob is in the office.
John picked up the football.
Bob went to the kitchen.
Where is the football? **A:playground**

bAbI数据集

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

Squad数据集



Memory Network

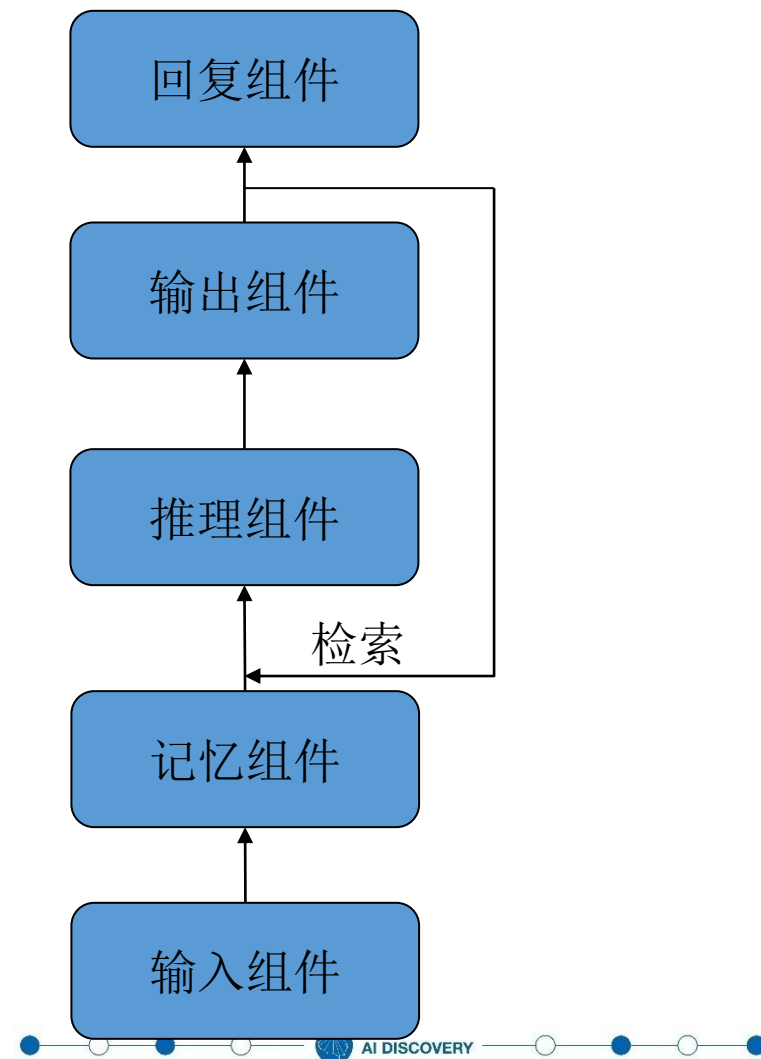


AI DISCOVERY

◆使用额外的Memory保存上下文，使用Attention进行检索

Memory，使用多步（multi hop）操作进行推理。

- 输入组件将上下文encode 成多个Memory slot，存储在记忆组件中。
- 推理组件会维护一个推理状态 s_t ，首先将问题或上一次推理状态 s_{t-1} 与记忆组件中的Memory slot进行Attention，检索出最适合下一次推理的事实 m_i 。
- 用检索到的 m_i 更新上一次推理结果 s_{t-1} 得到 s_t ，即一次推理。
- 输出组件输出结果可以用来进行下一次检索，或者用于生成答案。
- 最终回复组件将输出组件的结果转换成问题相关的答案。



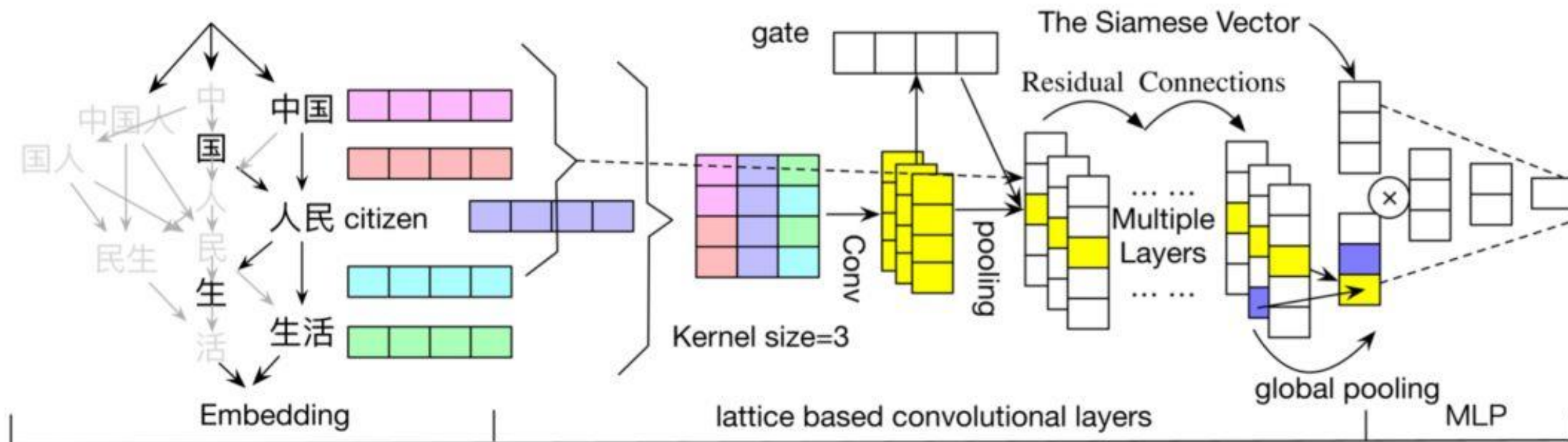


基于Lattice CNN的中文问答匹配

AI DISCOVERY

◆ Lattice CNNs for Matching Based Chinese Question Answering, AAI2019

- 问答系统是普通用户使用知识库最直接的渠道。匹配用户问题这种短文本，通常面临相同语义的单词和表达方式不唯一的挑战。中文这种还需要额外分词的语言中，这种现象尤为严重
- 研究者提出一个基于Lattice CNN的模型，利用在单词语义中多粒度的信息来更好地处理中文问答的匹配





机器超越人类水平



- ✓ 在斯坦福大学举办的 CoQA (Conversational Question Answering) 挑战赛中, 由微软亚洲研究院 (MSRA) NLP 组和微软雷德蒙德语音对话研究组组成的团队排名第一
- ✓ 此挑战赛衡量了机器理解文本、在对话中回答一系列相互联系的问题的能力
- ✓ CoQA 中的问题非常简短, 以模仿人类对话。除此之外, 第一个问题之后的每一个问题都基于对话历史, 这加大了机器解析那些短问题的难度

Rank	Model	In-domain	Out-of-domain	Overall
	Human Performance <i>Stanford University</i> (Reddy & Chen et al. '18)	89.4	87.4	88.8
1 Mar 29, 2019	Google SQuAD 2.0 + MMFT (ensemble) <i>MSRA + SDRG</i>	89.9	88.0	89.4
2 Mar 29, 2019	ConvBERT (ensemble) <i>Joint Laboratory of HIT and iFLYTEK Research</i>	88.7	85.4	87.8
2 Mar 29, 2019	Google SQuAD 2.0 + MMFT (single model) <i>MSRA + SDRG</i>	88.5	86.0	87.8
3 Mar 28, 2019	ConvBERT (single model) <i>Joint Laboratory of HIT and iFLYTEK Research</i>	87.7	84.6	86.8
3 Jan 25, 2019	BERT + MMFT + ADA (ensemble) <i>Microsoft Research Asia</i>	87.5	85.3	86.8

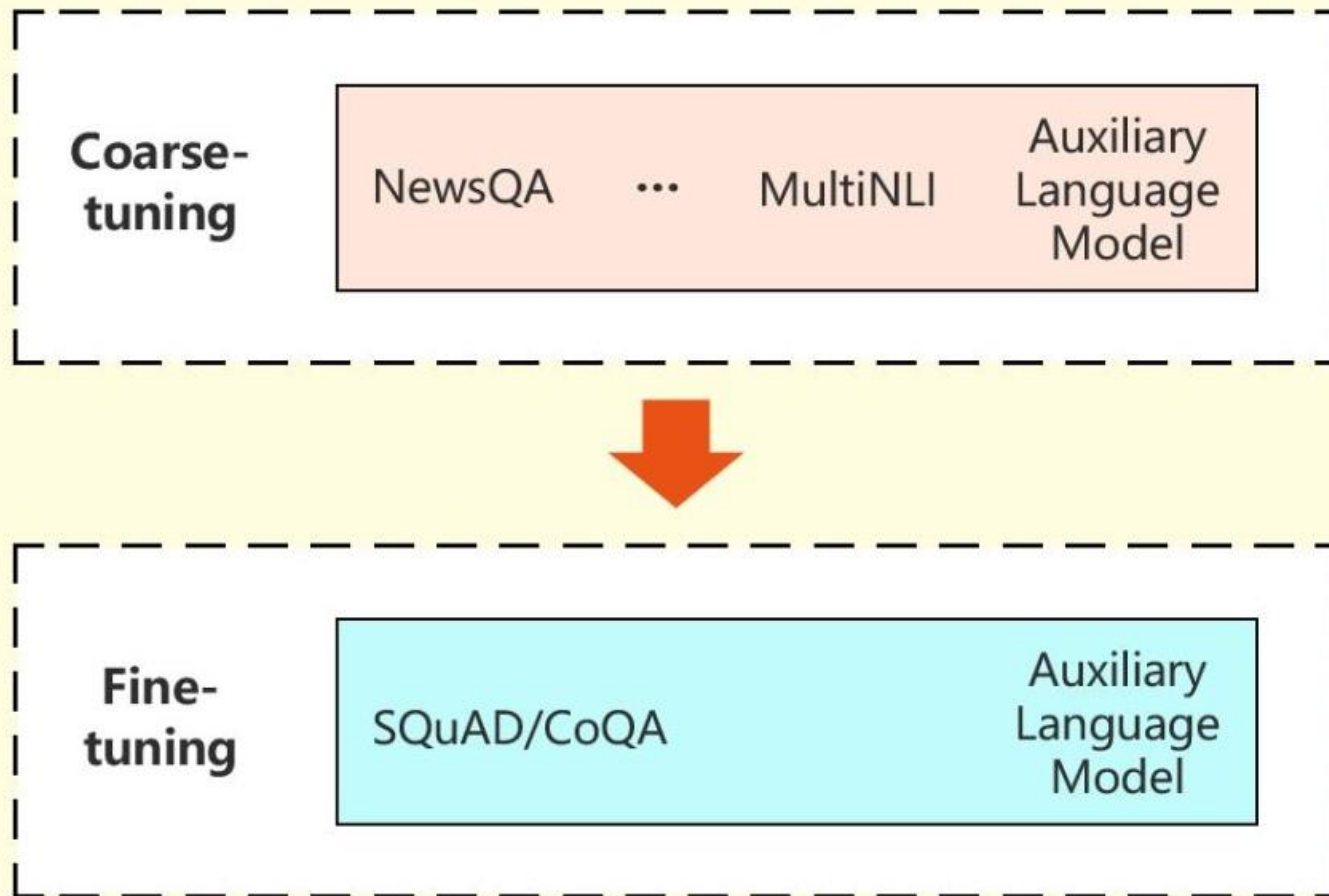


机器超越人类水平



- ✓ 微软研究团队的方法使用多个相关任务学习到的信息用于改进最终的机器阅读理解任务
- ✓ 在此多阶段、多任务精调方法中，研究人员首先在多任务环境下从相关任务中学习 MRC 关联背景信息，然后在目标任务上精调模型
- ✓ 在两个阶段，都使用语言建模作为辅助任务，帮助降低对话问答模型的过拟合
- ✓ 该方法的集成系统分布取得了 89.9 的领域内得分、88.0 的领域外得分、89.4 的整体 F1 得分，而人类的表现分别为 89.4/87.4/88.8

Multistage Multitask Fine-tuning





自然语言处理



AI DISCOVERY

情感分析

机器阅读

自动问答

文本生成



AI DISCOVERY





什么是文本生成



AI DISCOVERY

- ◆ **文本生成**是自然语言处理领域的一个研究热点，是实现机器像人一样使用自然语言与人类进行交流的关键技术
- ◆ **文本生成**是希望计算机能够像人类一样会表达，能够撰写出高质量的自然文本

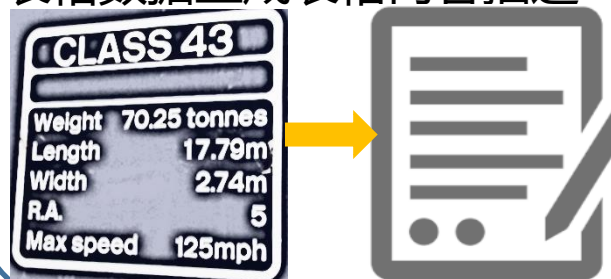
文本-文本

文本到文本的生成主要指对给定文本进行理解、变换和丰富，从而得到新的文本



数据-文本

数据到文本的生成主要指对给定数据生成相关文本的技术，比如表格数据生成表格内容描述



图像-文本

图像到文本的生成是指根据给定的图像生成描述该图像内容的自然语言文本



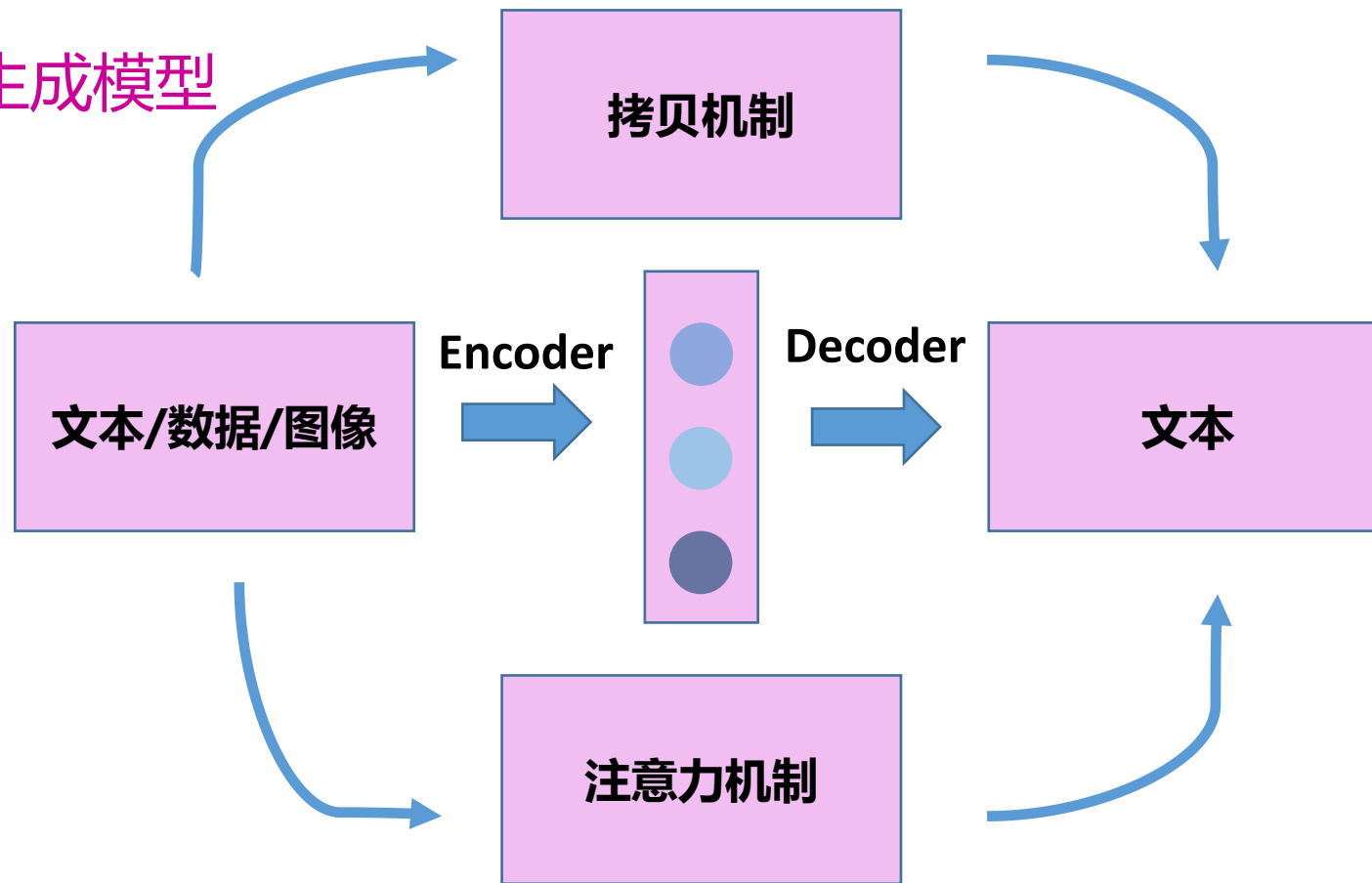
AI DISCOVERY



文本生成模型

◆ 基于神经网络的生成模型

- 编码器-解码器
- 注意力机制
- 拷贝机制



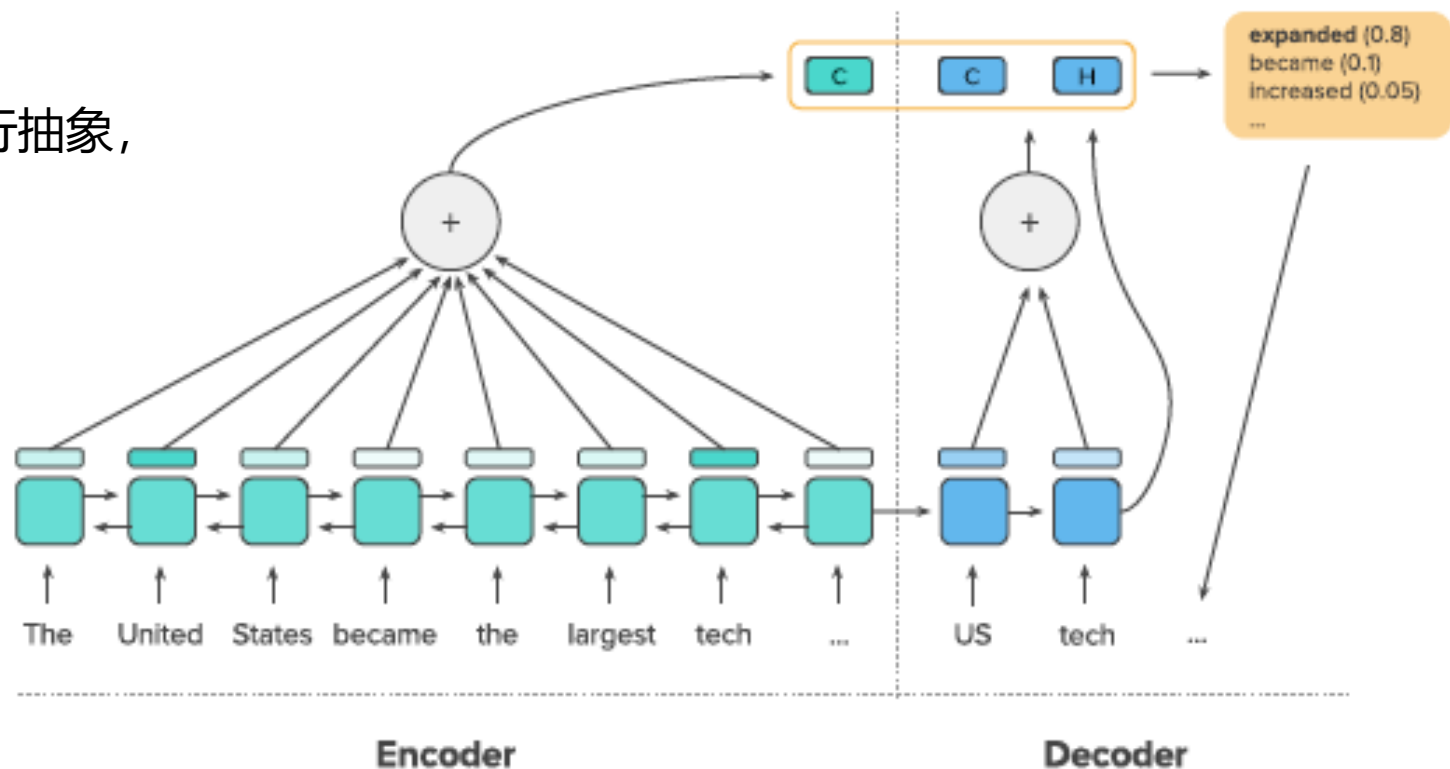


生成式摘要



◆ A Deep Reinforced Model for Abstractive Summarization, ICLR 2018

- 抽取式摘要从原文中找到和主题相关的句子或短语组合成摘要
- 生成式摘要是在理解原文基础上，对原文进行抽象，使用语义相近的词或不同表述产生文本摘要，
- 本模型使用两种attention机制
 - Intra-temporal attention
 - Intra-decoder attention
- 生成摘要的评价指标为ROUGE值，通过比较参考摘要和生成的摘要，给出摘要的评价
- 由于ROUGE不可导，无法直接对ROUGE进行梯度计算。因此，本模型用强化学习将ROUGE指标加入训练目标



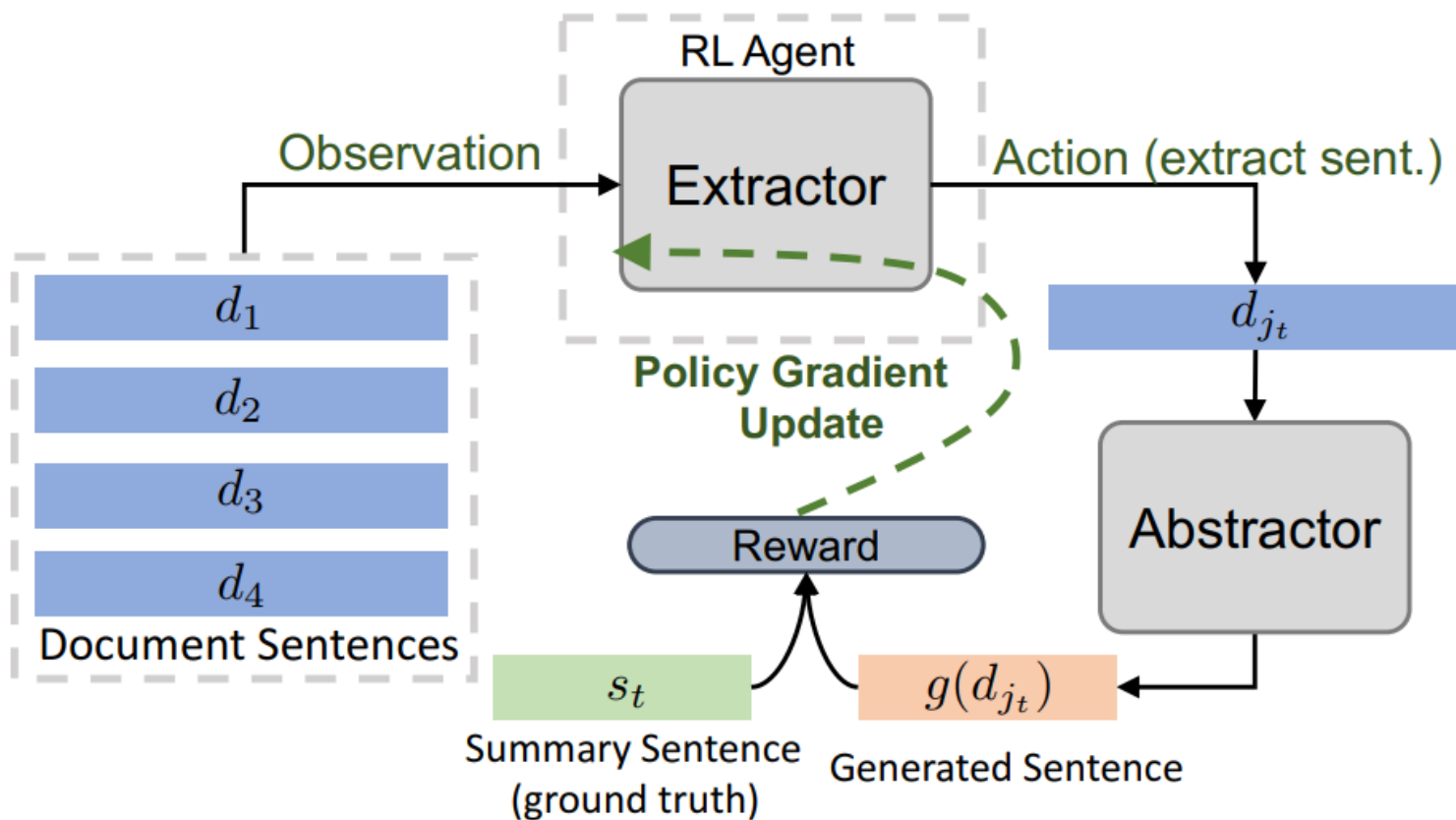


生成+抽取摘要

AI DISCOVERY

◆ Fast Abstractive Summarization with Reinforce-Selected Sentence Rewriting, ACL2018

- 为了解决生成摘要对于较长的文档作摘要时存在着冗余，不准确以及慢的问题，本工作提出一个句子级的增强学习模型，在不需要标注关键句子语料的情况下，实现抽取和生成结合的摘要
- 该模型包含两个模块：
 - Extractor agent
 - Abstractor agent
- 把抽取式模型看成一个分类任务，关键句子即为正类
- 生成模块是端到端的Seq2seq模型





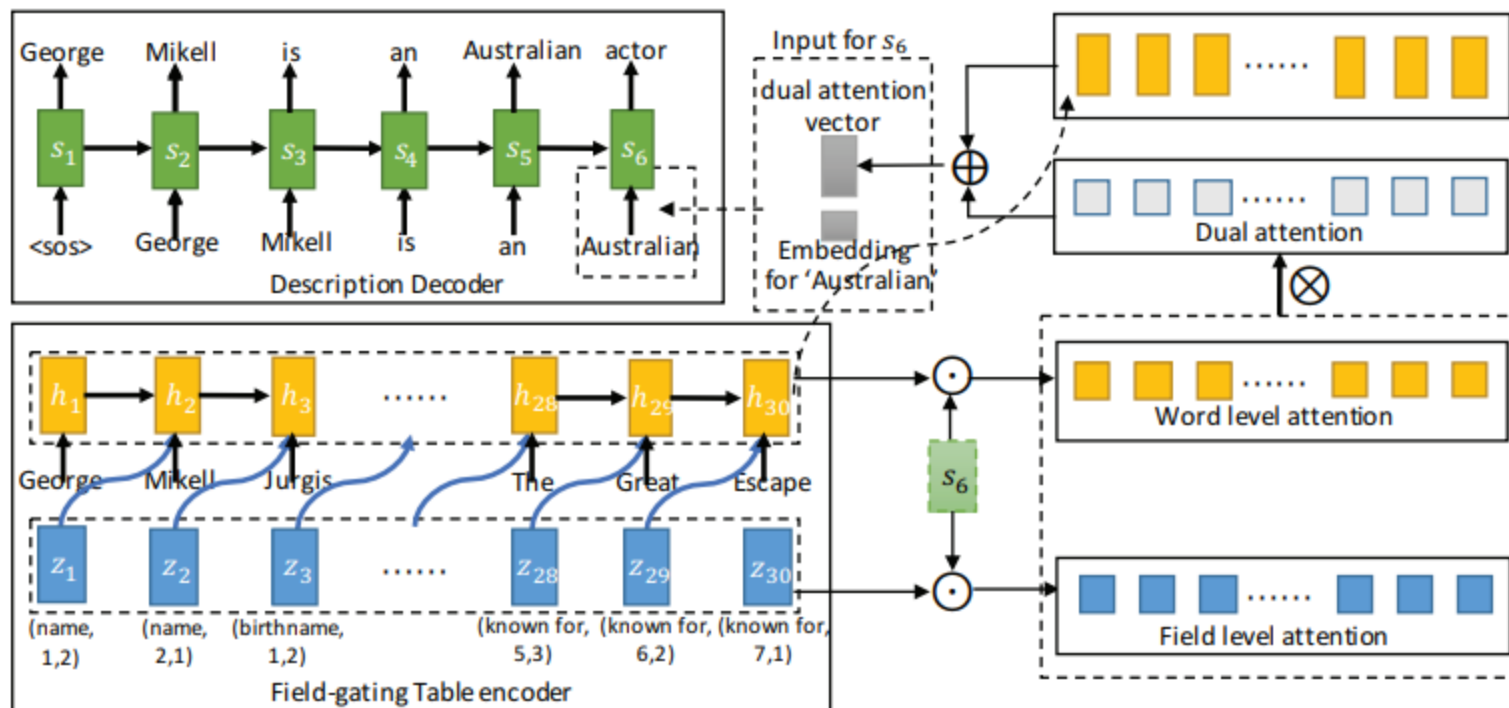
结构化数据生成



AI DISCOVERY

◆ Table-to-Text Generation by Structure-aware Seq2seq Learning, AAAI 2018

- 为结构化表格生成自然语言描述是结构化数据生成文本的一项重要任务
- 已有工作是在有限的预定义模式中对结构化数据进行建模
- 该工作提出一种结构感知的Seq2seq生成框架来对内容以及表格结构进行建模，将字段信息编码为结构化表格的表示
- 提出字段门控编码器和双重注意力机制，用于在结构化表格内容和字段信息之间进行局部 (local) 和全局 (global) 寻址



AI DISCOVERY



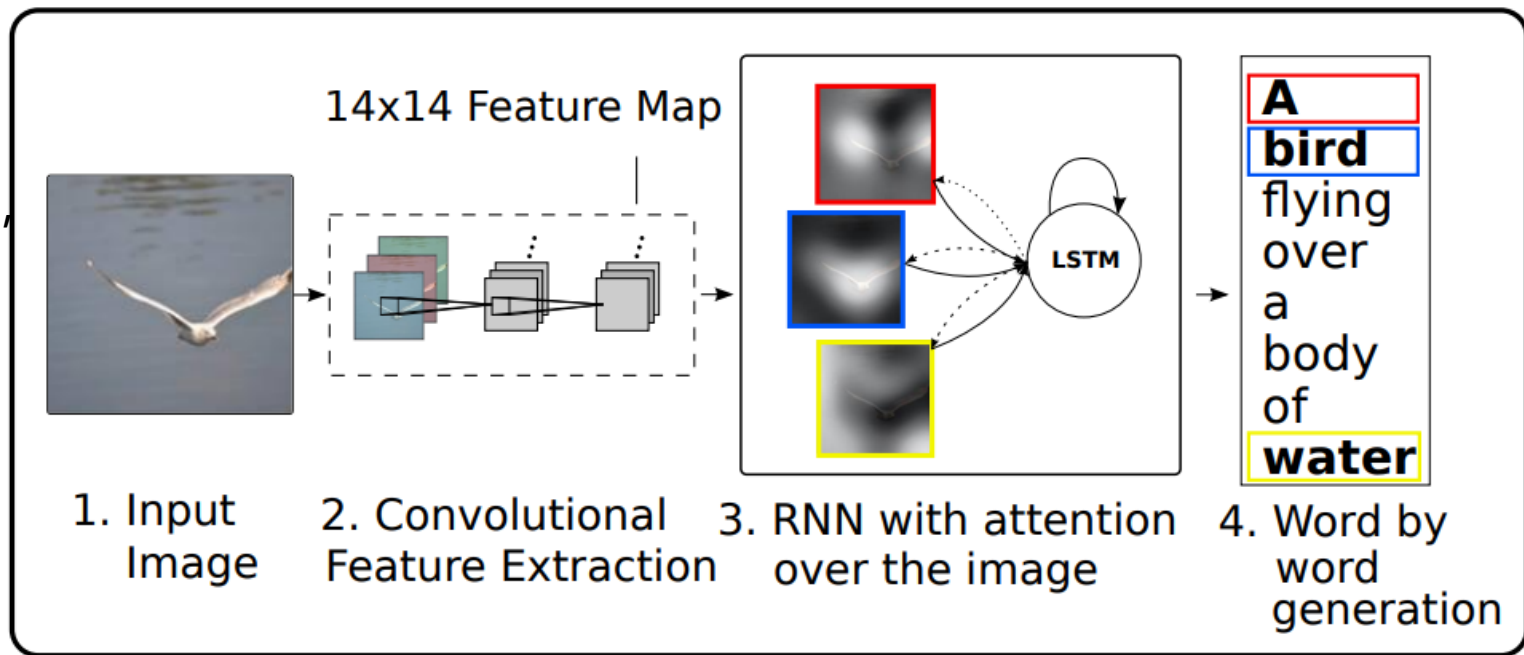
图像描述生成



AI DISCOVERY

◆ Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, ICML 2015

- 该工作是2015年ICML上的一篇文章，把attention机制引入到图像领域的文章
- Image caption是计算机视觉的最初始任务，不仅要获得图片里的物体，还要表达他们之间的关系
- 目前现存的方法大都是Encoder-Decoder架构，利用CNN、RNN、LSTM等神经网络完成caption工作
- 两种基于注意力机制的图像描述生成模型：
 - 反向传播训练的Soft Attention
 - 强化学习训练的Hard Attention



AI DISCOVERY



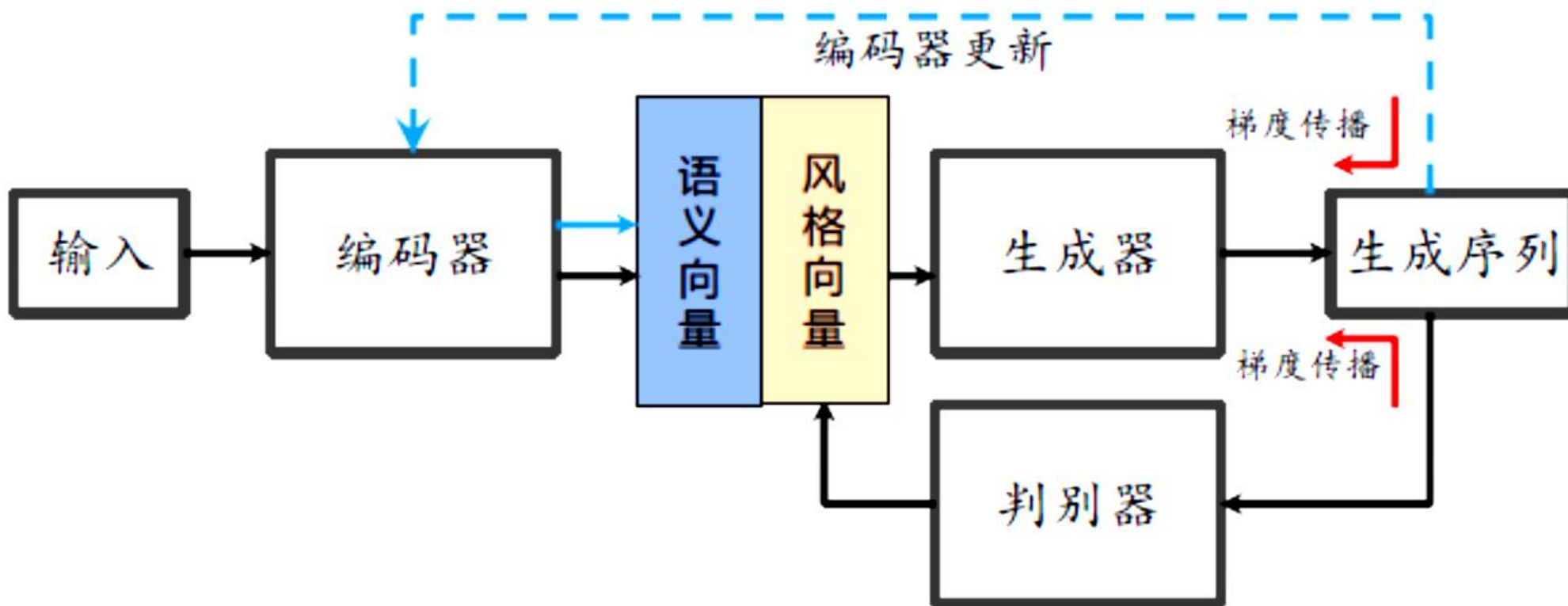


可控文本生成



AI DISCOVERY

◆ 加入可控因子



AI DISCOVERY

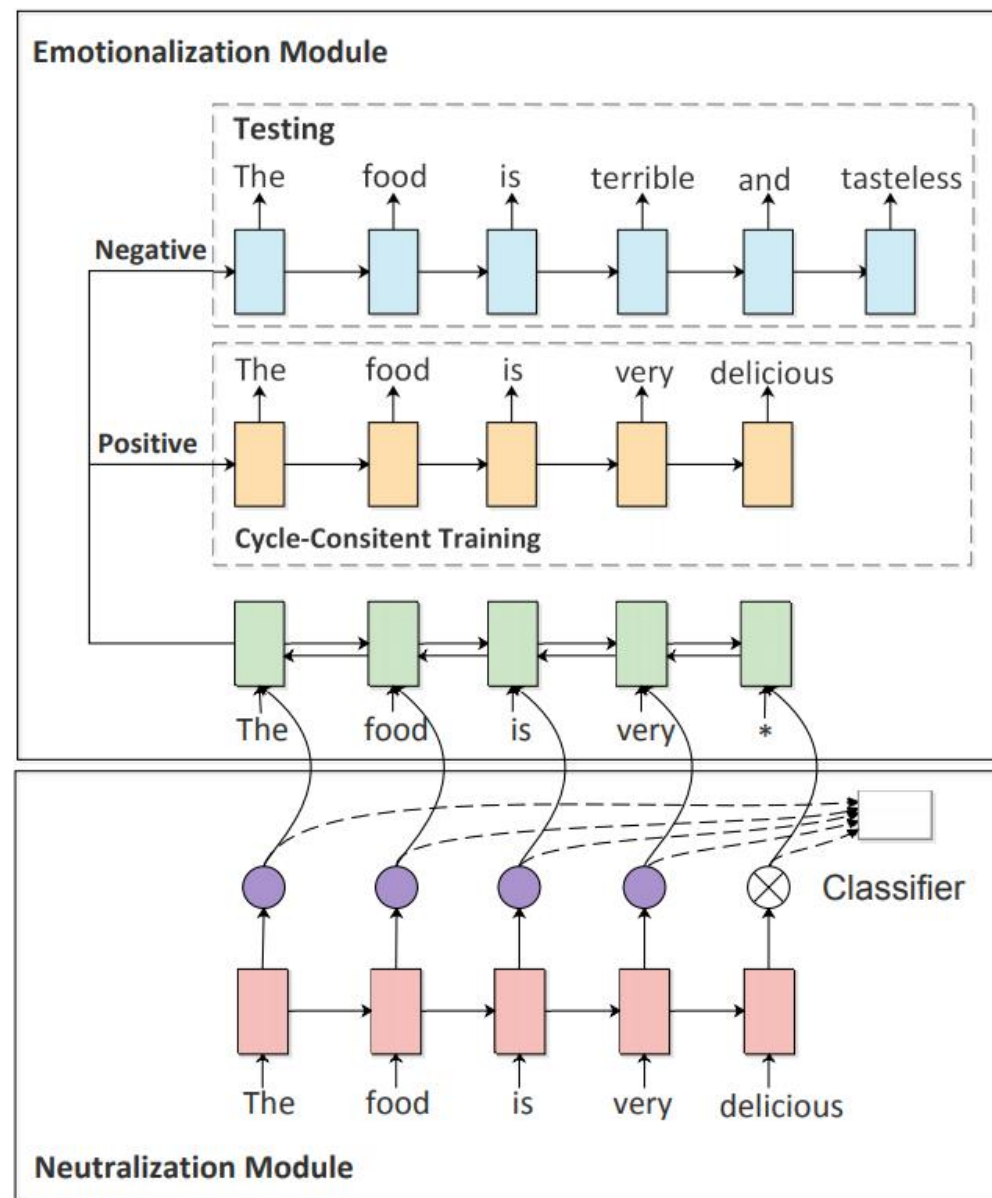


学习不同风格的解码器

AI DISCOVERY

◆ Unpaired sentiment-to-sentiment Translation: A cycled reinforcement learning approach

- 本文提出了一种新的方法用于无平行语料的情感风格迁移
- 现有的情感风格迁移的模型存在的问题是，在情感迁移的同时不能保证内容不变。例如，一句话 “The food is delicious”，生成却是 “What a bad movie”，虽然情感的极性改变了，但主语从food变成movie，内容也发生了变化
- 本文提出了循环强化学习模型，包括中立化模块和情感化模块。中立化模块作用是去除情感词，以提取非情感的语义信息。情感化模块作用是添加情感词，让中性句子情感化
- 其核心思想是：第一步，中立化模块先去除情感，然后情感化模块根据原始情感和语义内容重构原句，让情感化模块在有监督的情况下学习增加情感。第二步，将情感词取反，这样就可以实现添加相反的情感词





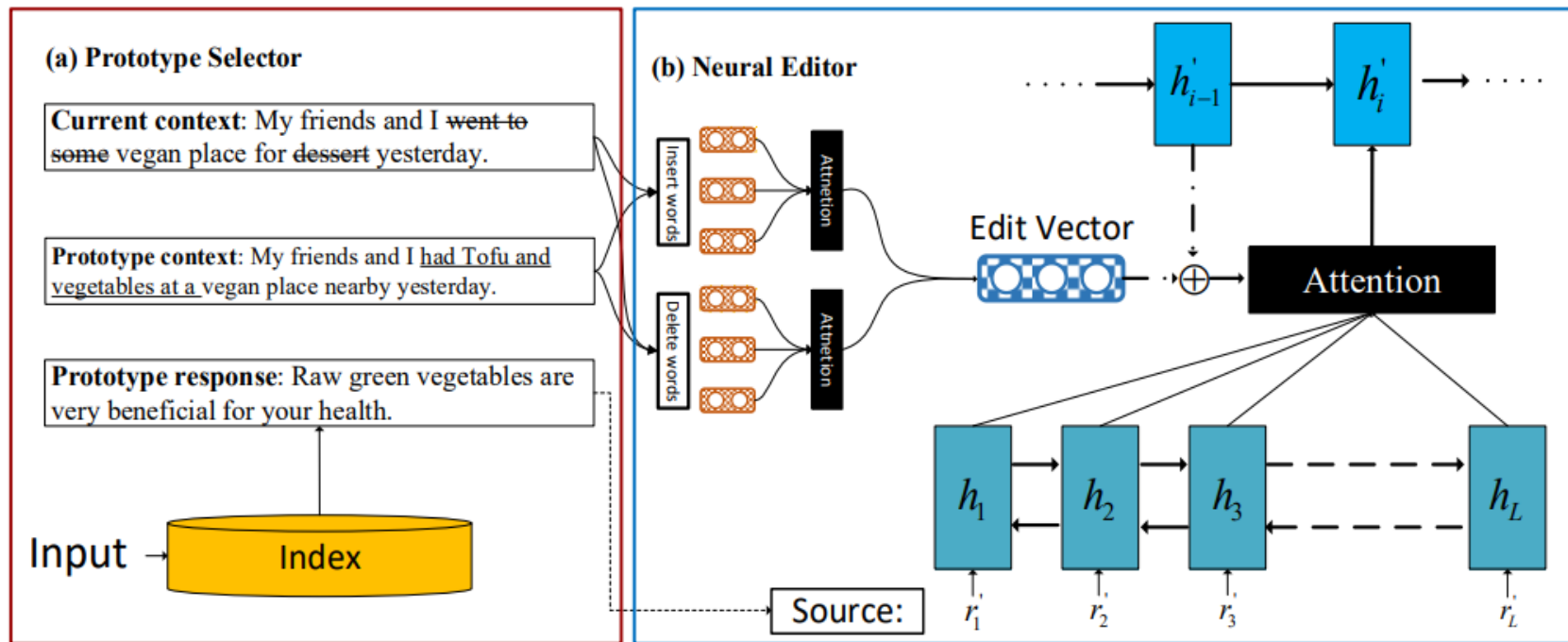
文本改写



AI DISCOVERY

◆ Response Generation by Context-aware Prototype Editing, AACL 2019

- 本文的主要目的是为了解决开放域对话生成中，生成句子较短且无意义的问题
- 主要思想是将检索与生成相结合，首先从预定义的语料中检索到原回复，然后根据原问题和当前问题之间的差异，对检索到的回复进行编辑，从而使得到的回复既保证了语法和信息，也获得了上下文的语意连贯性



AI DISCOVERY



考虑读者评论的生成式文本摘要



AI DISCOVERY

◆ Abstractive Text Summarization by Incorporating Reader Comments, AAI 2019

在基于神经网络的生成式文本摘要研究领域，传统的基于序列到序列的模型对文档主题的建模经常出错。为解决这个问题，研究者提出了读者感知的摘要生成任务，它利用读者的评论来帮助生成符合文档主题的更好的摘要。与传统的生成式摘要任务不同，读者感知型摘要面临两个主要挑战：

- (1) 评论是非正式的，有噪音的；
- (2) 如何联合建模新闻文档和读者评论

document	On August 28, according to a person familiar with the matter, <u>Toyota Motor Corporation will invest 500 million U.S. dollars into the Uber, a taxi service company, with a valuation of up to 72 billion U.S. dollars.</u> The investment will focus on driverless car technology. However, its development path is not smooth. In March of this year, a <u>Uber driverless car hit a woman and caused her death.</u> In last year, Softbank also invested into Uber with a valuation of \$48 billion.
comments	Toyota's investment in Uber is a wise choice.
	\$500 million investment is really a lot of money!
good summary	Toyota invests \$500 million into Uber with a valuation of \$72 billion
bad summary	An Uber driverless car hits a passerby to death



AI DISCOVERY





考虑读者评论的生成式文本摘要

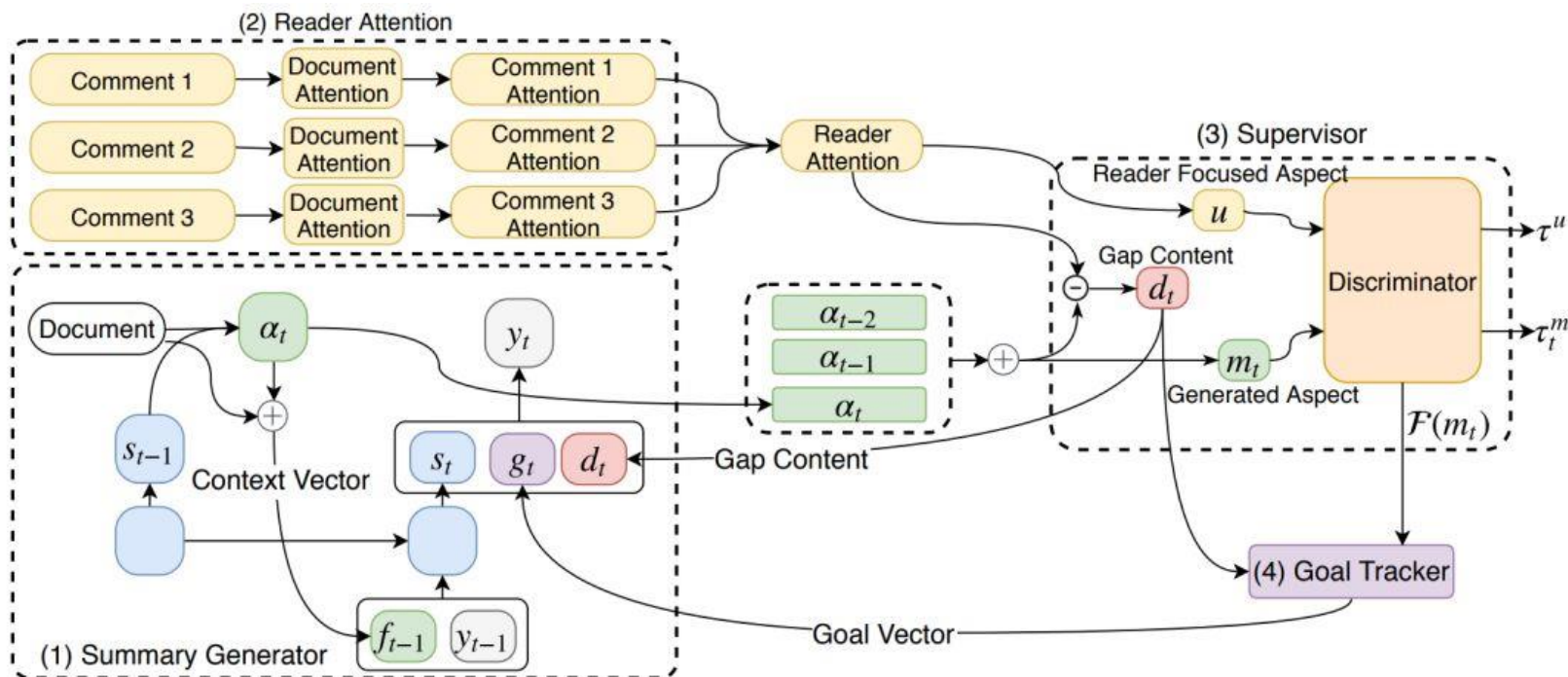


该工作设计了一个基于对抗学习的读者感知型摘要生成模型 (RASG)

模型由四部分组成:

1. 基于序列到序列的摘要生成模块;
2. 读者注意力建模模块来捕捉读者关注的主题;
3. 督导建模模块来描述生成的摘要和读者关注主题之间的语义差距;
4. 目标跟踪模块来产生每个生成步骤的目标。

督导和目标定位模块以对抗学习的方式来指导我们框架的训练。





目录



AI DISCOVERY

1 自然语言处理概述

基本概念、技术概览

2 自然语言处理进阶

情感分析、机器阅读、自动问答、文本生成

3 多模态融合

多模态分类、多模态检索、命名实体识别

4 应用与实践

优化方法、调参方法、可视化方法
实践：机器翻译



AI DISCOVERY



多模态融合



AI DISCOVERY

多模态分类

多模态检索

命名实体识别



AI DISCOVERY





什么是多模态情感分类



AI DISCOVERY

单模态上的信息往往不全面或者带有歧义，多模态数据可以对单模态数据形成多视角补充。



Gorgeous!

(a)



Doing hard labor again!

(b)



"As long as the road is right, not afraid o distant future; if you believe it's worthwhi do not care about vicissitudes."

(c)



Hello there sweetie. :)

(a)



PD Achilles meets a new friend. Special post for one of our followers who I met last night and had a good chat to

(b)



If anyone woke up in edinburgh this morning to discover their car missing i think i know where it is

(c)

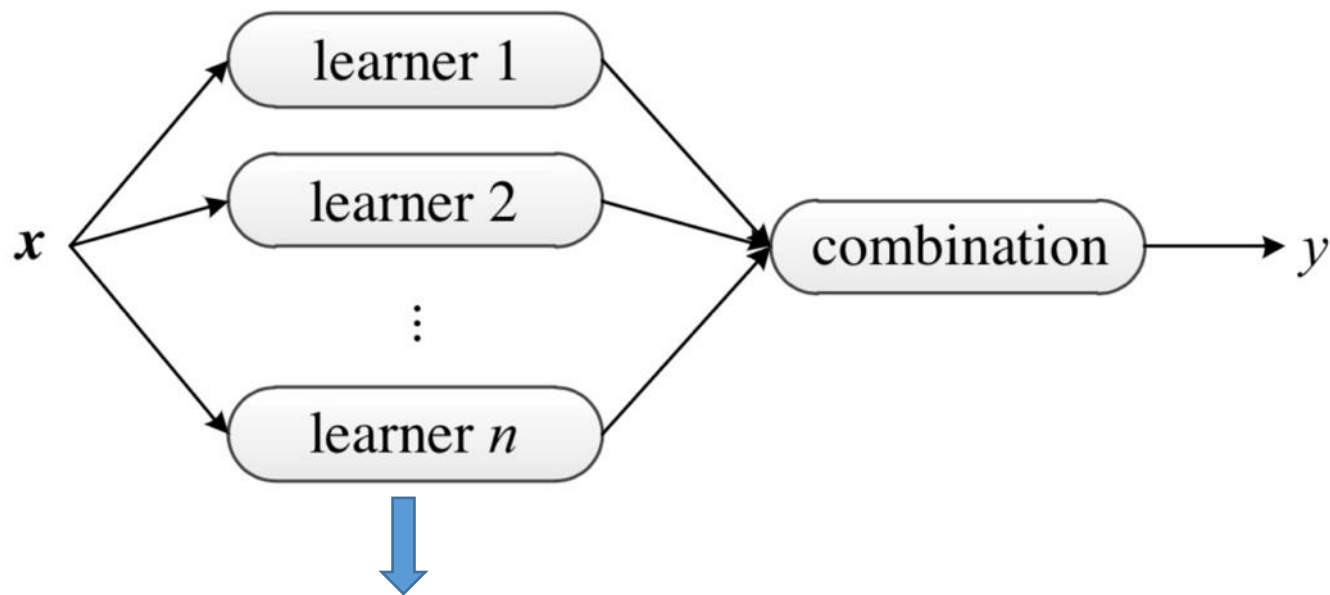


AI DISCOVERY



传统多模态融合方法

通过组合多个学习器来实现多模态融合，每个单独的学习器被称为个体学习器，可以分别设置为文本、图片、语音等单视角分类器。



个体学习器可以是SVM、决策树、神经网络或者其他学习算法。





集成学习什么情况下有效



	测试例1	测试例2	测试例3		测试例1	测试例2	测试例3		测试例1	测试例2	测试例3
h_1	✓	✓	X	h_1	✓	✓	X	h_1	✓	X	X
h_2	X	✓	✓	h_2	✓	✓	X	h_2	X	✓	X
h_3	✓	X	✓	h_3	✓	✓	X	h_3	X	X	✓
集成	✓	✓	✓	集成	✓	✓	X	集成	X	X	X

(a) 集成提升性能

(b) 集成不起作用

(c) 集成起负作用

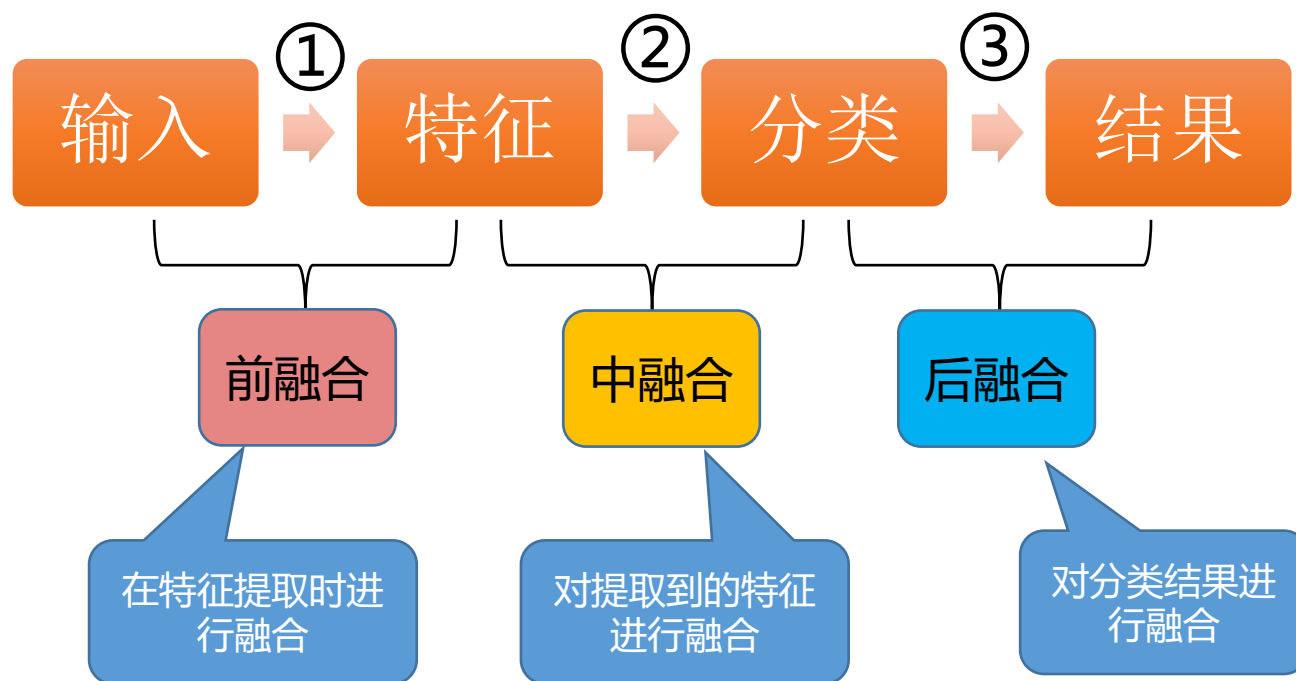
✓ 个体学习器应该 **“好而不同”**，既要有一定的**准确性**，又要有**差异性**





思考题

- ✓ 如果把一个机器学习问题分解为四部分三阶段，那么通过**集成学习**实现的多模态数据融合属于**哪一个阶段的融合**？



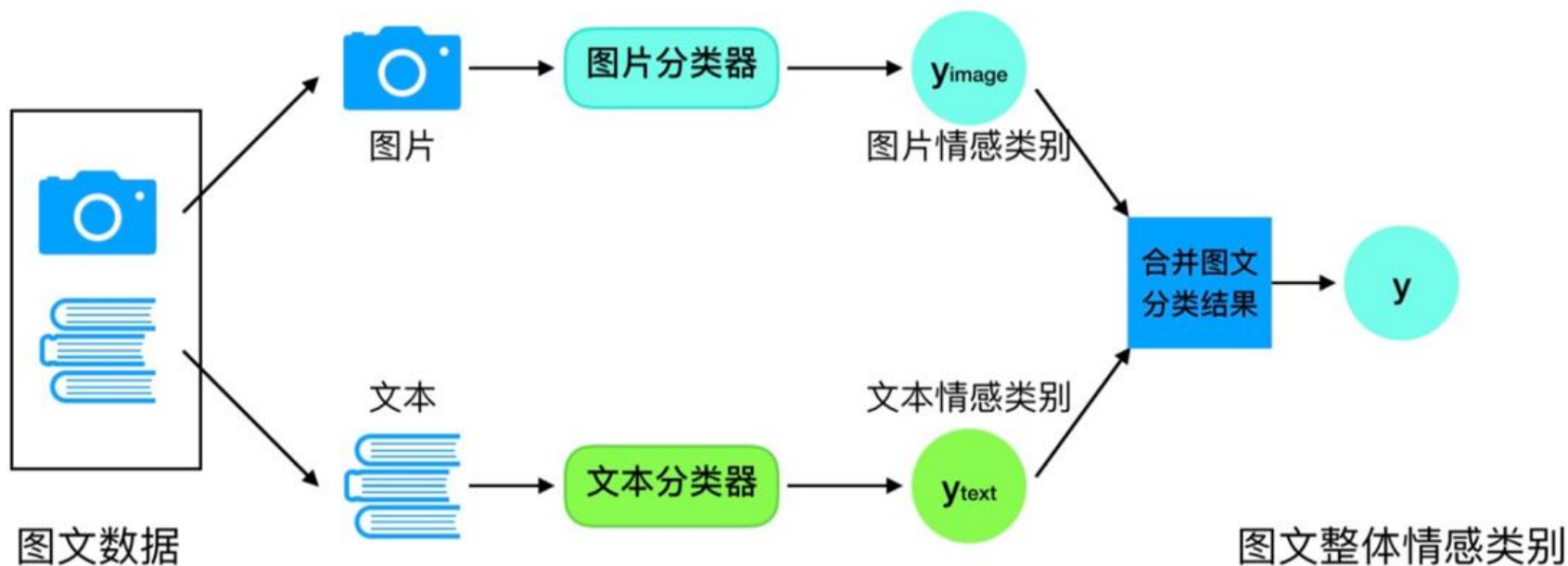


基于深度学习的多模态情感分类 (1)



◆ 基于后融合的多模态情感分类模型的关键点有两个

- 如何更加有效的对单个模态进行情感分类
- 如何有效的组合多个单模态的情感分类结果



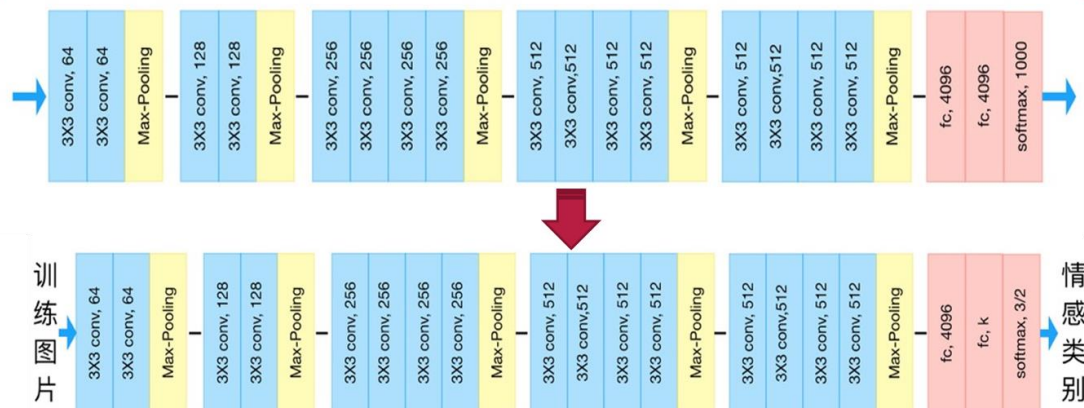


基于深度学习的多模态情感分类 (2)

AI DISCOVERY

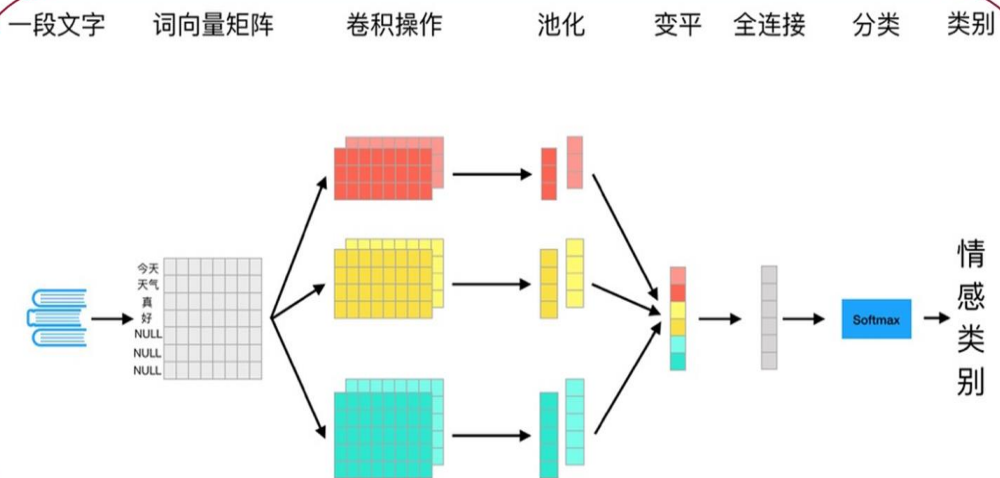
✓ 深度学习模型对训练数据的规模是有要求的，因参数较多，如果训练数据不足，则效果不好。所以在训练图片分类器时采用迁移学习思想。

图片分类器

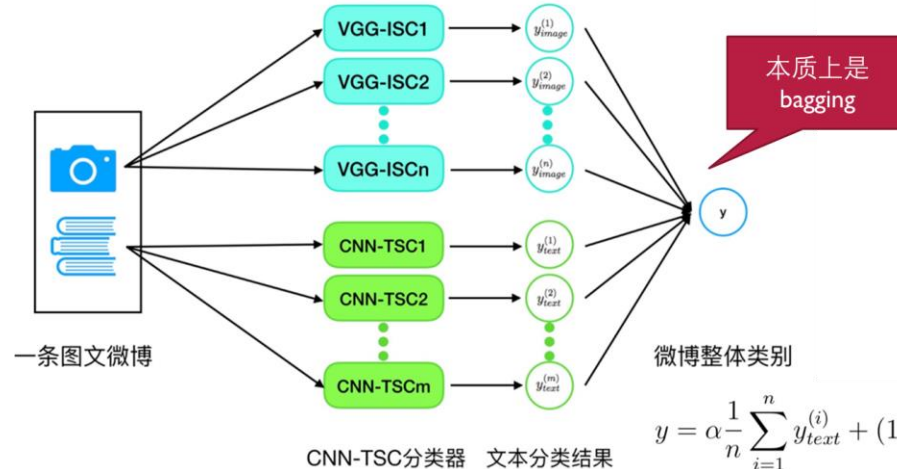


VGG-NET网络示意图

文本分类器



VGG-ISC分类器 图片分类结果

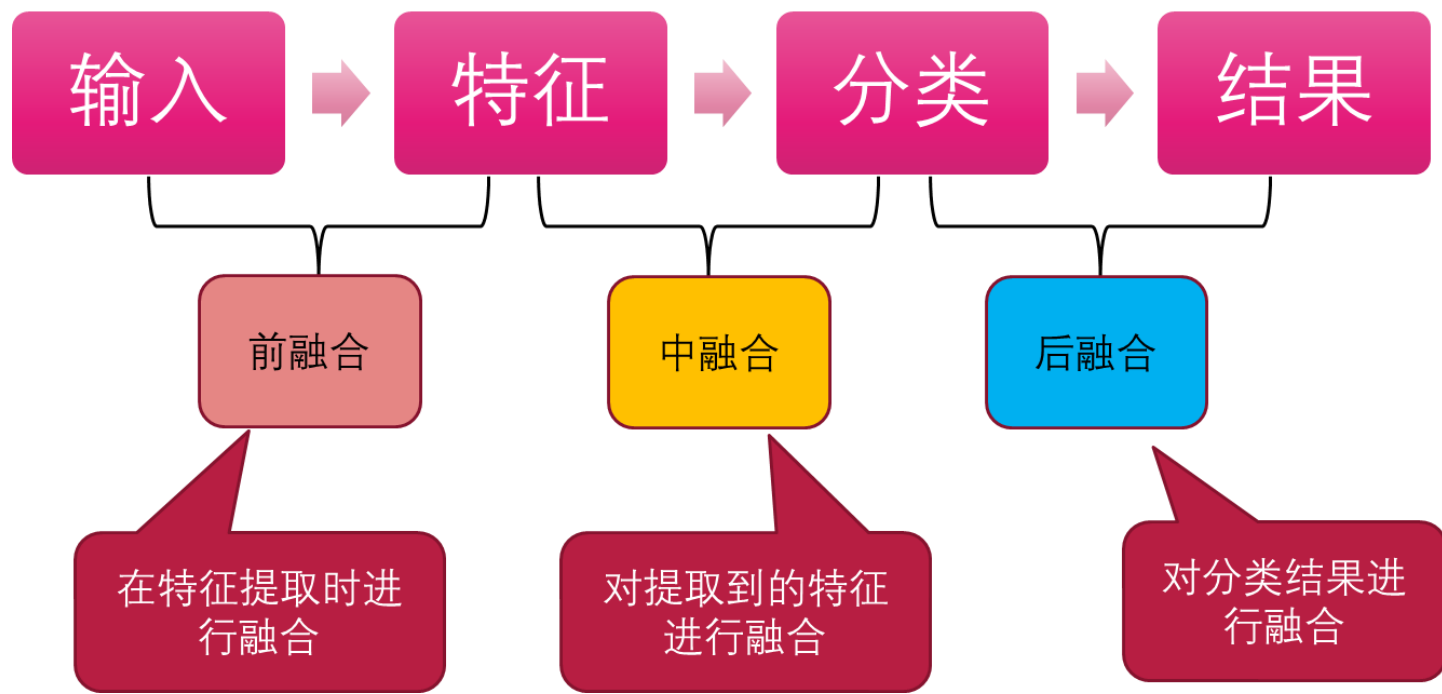


$$y = \alpha \frac{1}{n} \sum_{i=1}^n y_{text}^{(i)} + (1 - \alpha) \frac{1}{m} \sum_{i=1}^m y_{image}^{(i)}$$



如何进行前融合

- ◆ **前融合**是指通过学习不同模态数据之间的语义关联，进行**联合特征抽取**，使得抽取出的特征不仅能表征单模态数据特点，还能表征多模态数据之间的**关联信息**。
- ◆ **后融合和中融合**都无法做到这一点，因为无论后融合还是中融合，特征提取都是一个**独立的过程**。



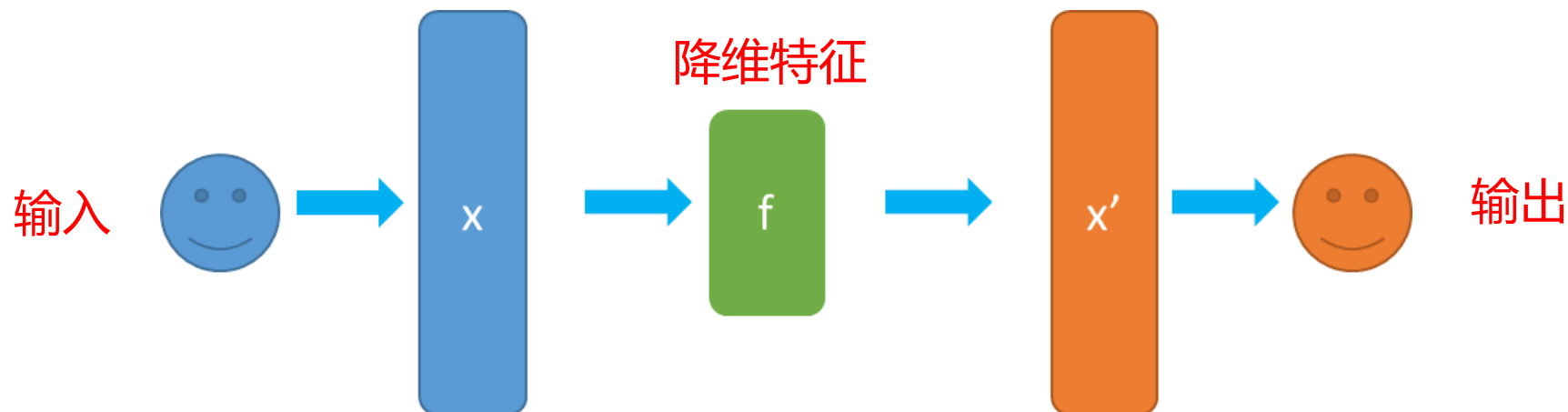


自编码器是什么



AI DISCOVERY

自编码器(AutoEncoder)是一种前馈神经网络，目标是尽可能的让输出与输入一致。
AutoEncoder 使用反向传播进行训练，是无监督模型，主要是用于数据的降维或者特征的抽取上。



前融合是指在不同模态数据之间进行联合特征抽取，而自编码模型主要用于数据降维或者特征抽取，正是因为自编码模型具有提取特征的能力，所以可用于前融合

AI DISCOVERY



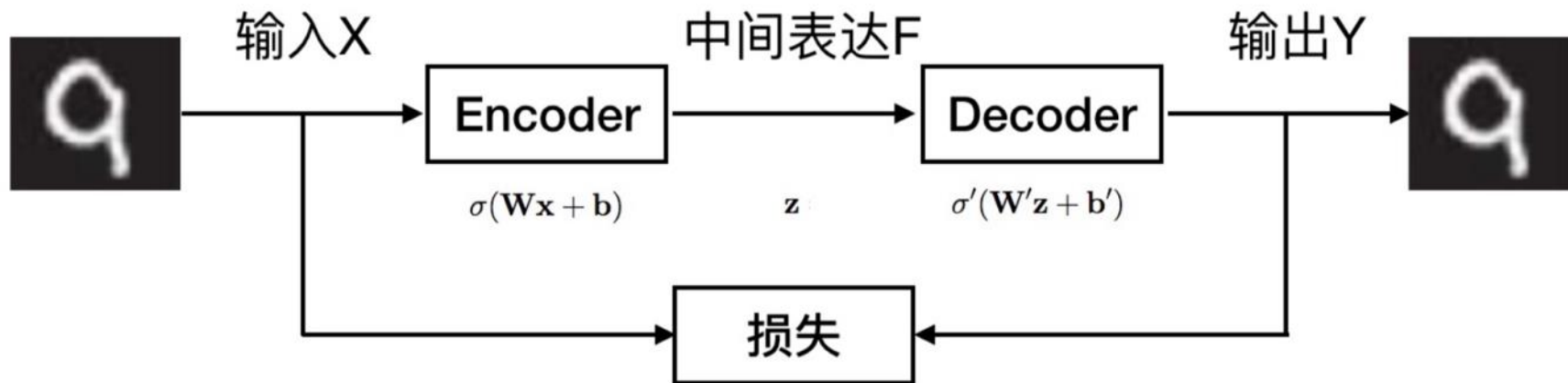
自编码器原理



AI DISCOVERY

◆ AutoEncoder 通常包含两个部分:编码器和解码器。

- 编码器和解码器可以是单层的也可以是多层的, 多层的编码器和解码器往往性能更好。
- 输入X经过编码器得到中间表示层F, 又称为编码(encode), 然后 F 经过解码器得到输出 Y, 两个过程分别称为编码过程和解码过程。
- AutoEncoder 的目标函数可以表示为输出和输入的差值最小。

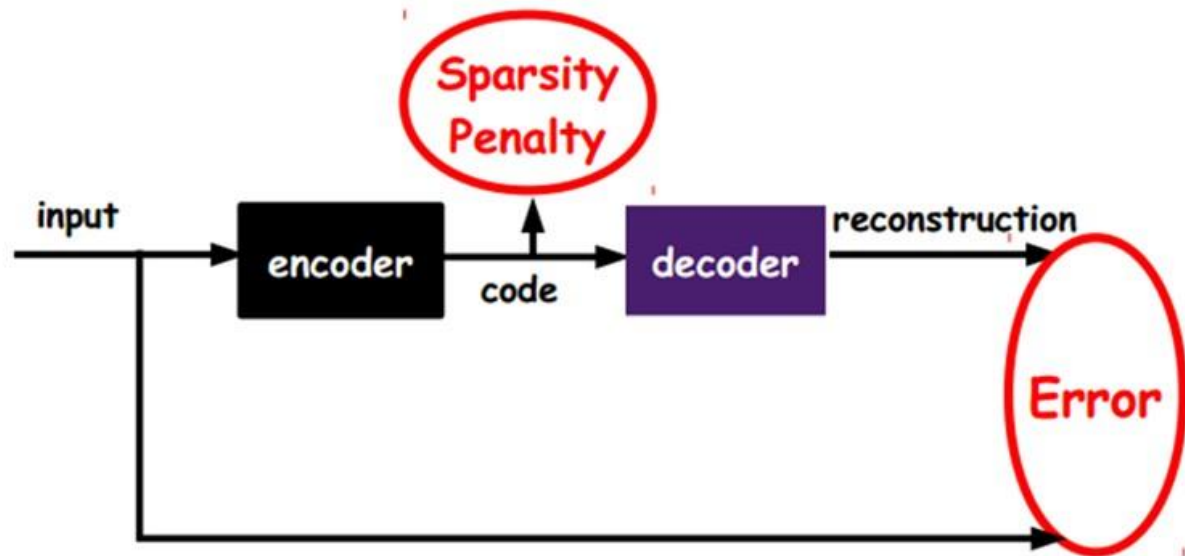


AI DISCOVERY





什么是稀疏自编码



- input: X code: $h = W^T X$

- loss: $L(X; W) = \|W h - X\|^2 + \lambda \sum_j |h_j|$ ← L1正则

- ✓ 稀疏自编码器 (Sparse AutoEncoder) 可以约束中间表达尽可能的稀疏, 能够学习到更加有用的特征。
- ✓ 如果在AutoEncoder的基础上加上L1正则限制 (L1主要是约束每一层中的节点中大部分都要为0, 只有少数不为0, 这就是Sparse名字的来源), 可以得到Sparse AutoEncoder。



多模态融合



AI DISCOVERY

多模态分类

多模态检索

命名实体识别



AI DISCOVERY


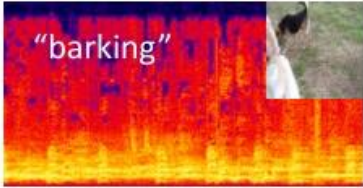
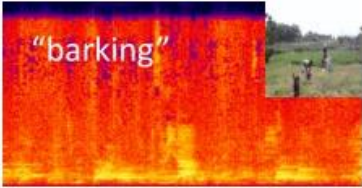

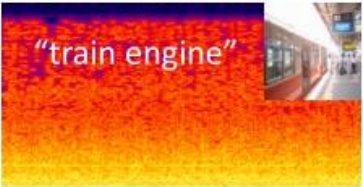










什么是多模态检索



当输入是一张小狗的图片的时候，如果以图搜声，则可得到狗叫声的音频；如果以图搜文，则可得到关于小狗的描述。

Input Query	Sound Retrievals	Text Retrievals	Image Retrievals
	 "barking"	 "barking"	<ul style="list-style-type: none">- A dog lying down on the beach- The dog belongs to the homeowner 
 "train engine"	 "train passing"	 "train passing"	<ul style="list-style-type: none">- Steel tracks under the train.- The train platform 
The choppy water the man is riding	 "water crashing"	 "boat engine"	<ul style="list-style-type: none">- A person stands on water skis in the water- A couple of kayakers paddling through water 

✓ 值得注意的是：不同模态之间的共同部分对跨模态检索任务而言是最重要，不同模态的特有部分对跨模态检索任务并不是必要的，甚至可能会影响跨模态检索的质量。



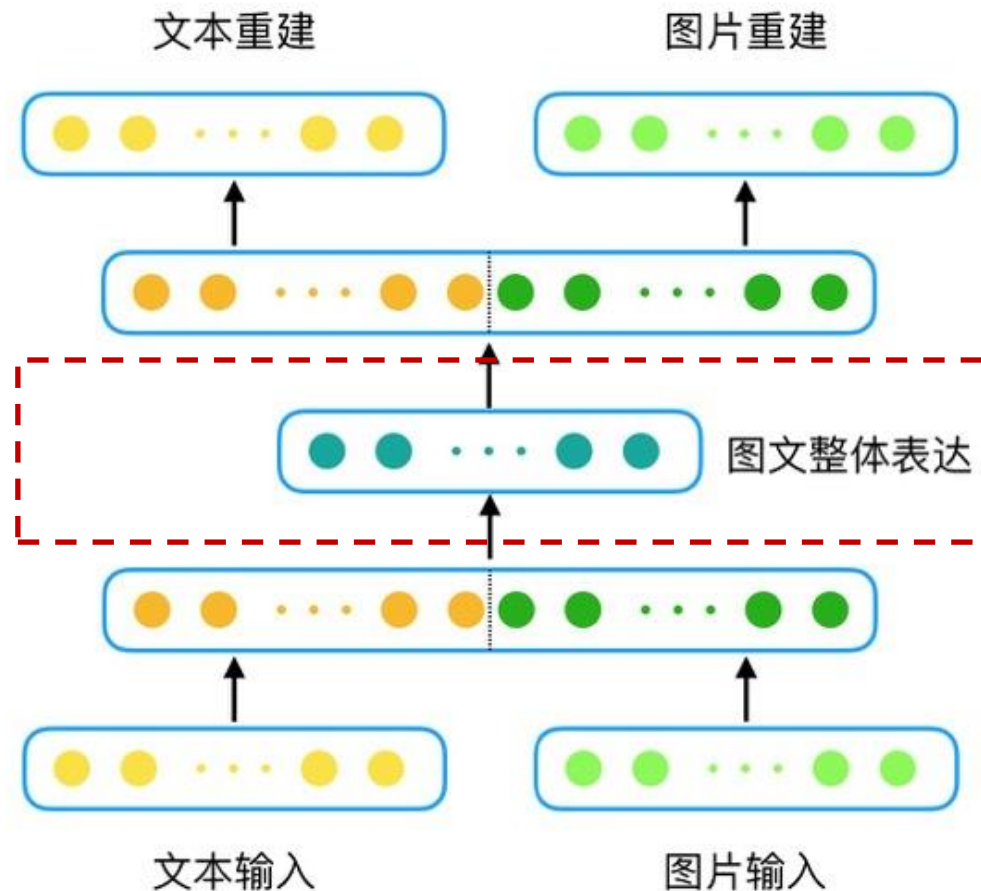


Bimodal DBN



◆ Bimodal DBN模型

- ✓ 该模型通过一个共享层来连接图像和文本模态，也可以看作一个把图像和文本输入拼接起来作为一个输入的受限玻尔兹曼机。
- ✓ 在Bimodal DBN中，图片和文本共享同一个特征表达。

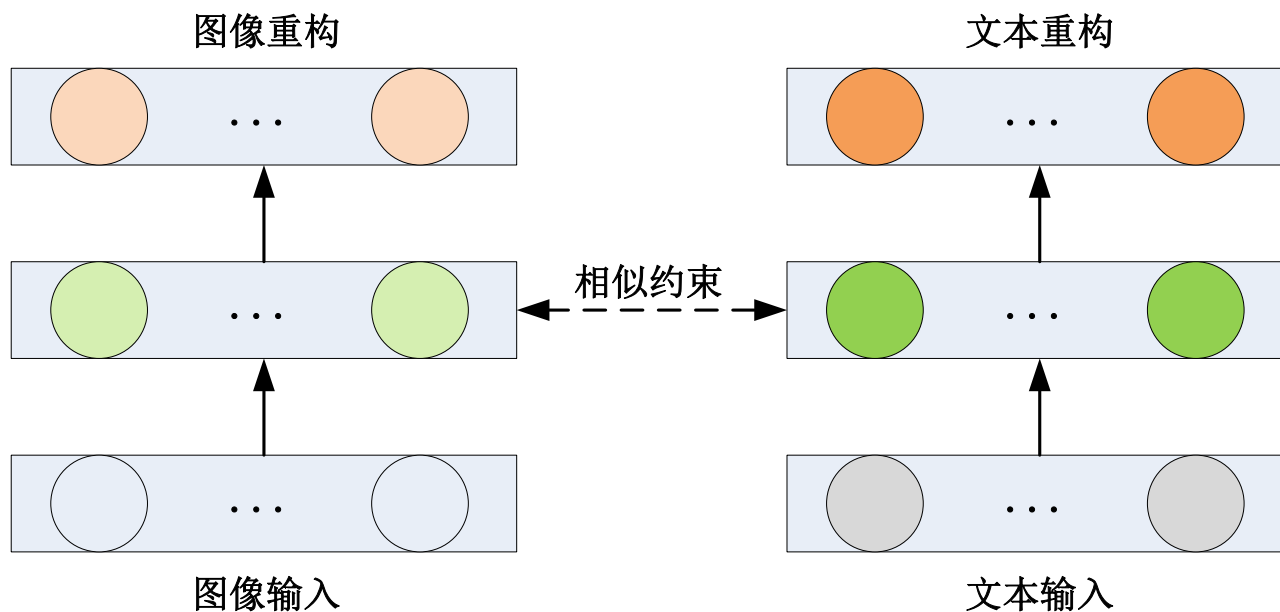




对应自编码器 (Correspondence Autoencoder)



图片、文本共享特征模型不同：对应自编码器是由两个单模态的自编码器组成，每个自编码器负责其相应模态的表示学习。两个自编码器通过在表示层引入一个**相似性约束**组成了一个整体模型。



相似性约束和两个单模态自编码器本身的重构约束使得这个模型能够具备两方面的能力：

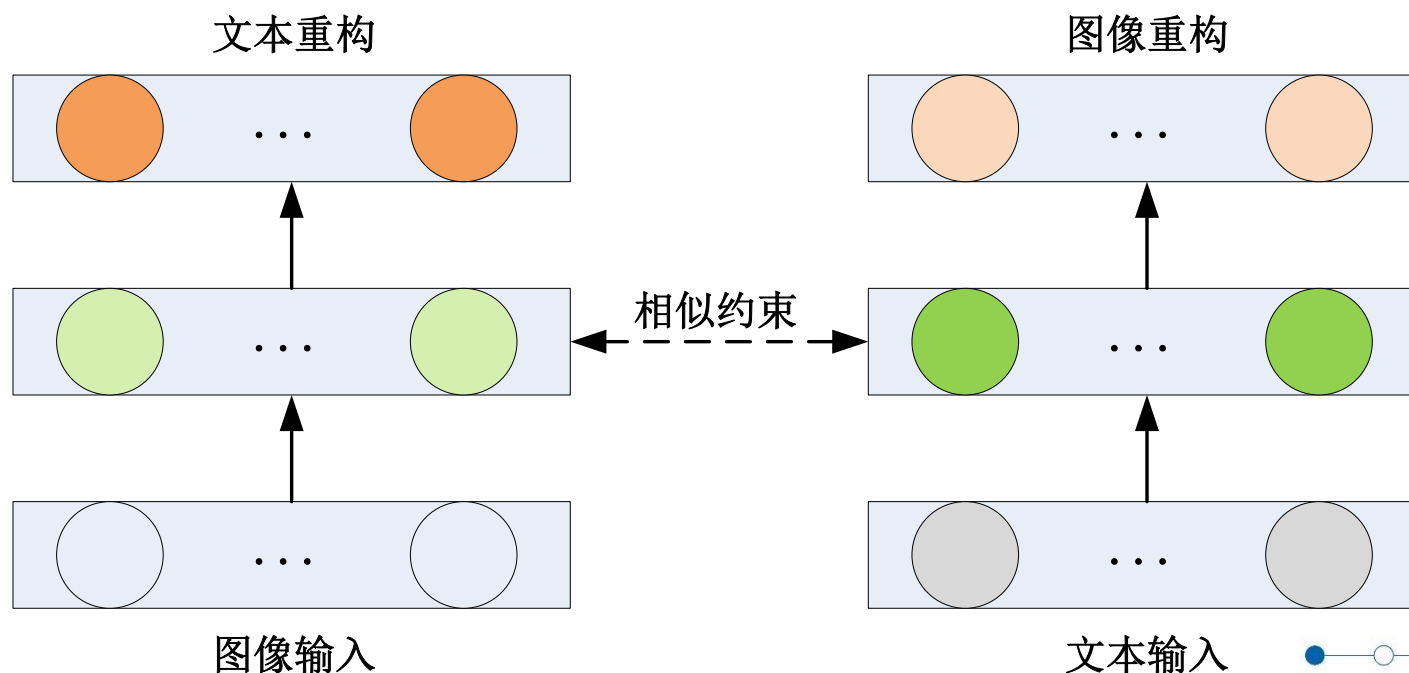
- ✓ 对齐的图像和文本输入在表示空间中距离相近
- ✓ 相似的单模态输入在表示空间中距离相近





对应跨模态自编码器 (Correspondence Cross-Modal Autoencoder)

对应自编码器只重构自己模态的输入数据，而在**对应跨模态自编码器**中，左右两部分都是跨模态自编码器，图像模态的表示学习考虑了文本模态的信息，同样文本模态的表示学习也考虑了图像模态的信息。这使得模型在表示学习模块也**学习到了不同模态数据之间的关联性信息**。

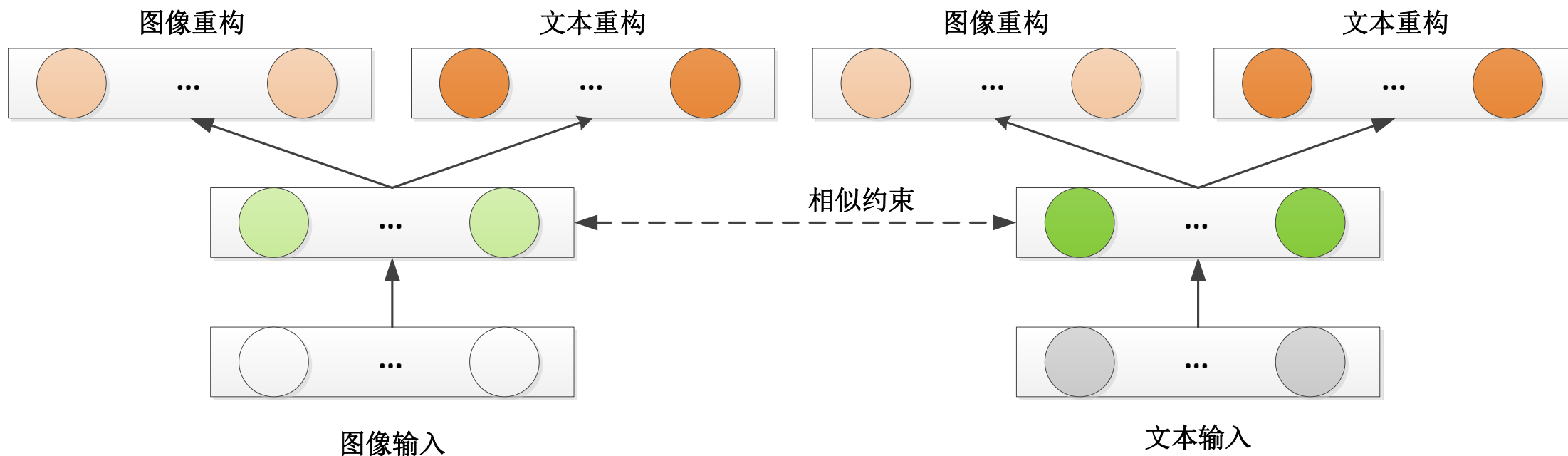




对应全模态自编码器 (Correspondence Full-Modal Autoencoder)

AI DISCOVERY

对应全模态自编码器，左右两侧分别以单模态为输入，在输出端不仅重构图像，而且重构文本，实际上是**综合**了前面介绍过的对应自编码器和对应跨模态自编码器。全模态自编码器最早被提出用来建模**语音和视频**数据，当然也可以用于建模**图像和文本**。



AI DISCOVERY



哪种多模态神经网络好



	MAP			Top 20%		
	以图检文	以文检图	平均	以图检文	以文检图	平均
Bimodal DBN	0.189	0.222	0.206	26.19	31.6	28.90
对应自编码	0.326	0.361	0.344	56.06	55.19	55.63
对应跨模态自编码	0.336	0.341	0.338	55.41	58.66	57.04
对应全模态自编码	0.335	0.368	0.352	57.36	57.79	57.58

- ✓ 相比 Bimodal（共享层），三类对应自编码模型在**两种性能评价标准上**都有**明显的优势**。
- ✓ Bimodal DBN学到的隐层表示包含了不同模态间的公共部分和各模态自己独特的部分。因此，它们更加关注多模态数据表示的**补全**，而没有把重点放在学习不同模态数据的关联上。
- ✓ 三类对应自编码模型性能都差不多。这类模型在表示层引入关联约束，将不同模态的输入映射到同一个表示空间中。这类模型还可以让单模态表示学习和多模态关联学习放在一个阶段中整体优化。



多模态融合



AI DISCOVERY

多模态分类

多模态检索

命名实体识别



AI DISCOVERY

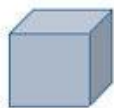




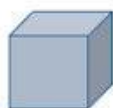
图文混合命名实体识别 (1)



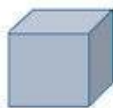
- ◆ 在 **tweet 命名实体识别** 中，传统的方法往往只使用了文本内容。但是许多推文还包含图像等内容。这种视觉信息在实体识别任务是极有价值的。



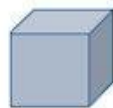
Person name



Location name



Organization name



Misc



- ✓ 在只使用文本信息时可能会把「Rocky」识别为人名；但是结合 tweet 图片，我们知道「Rocky」其实是一条狗。

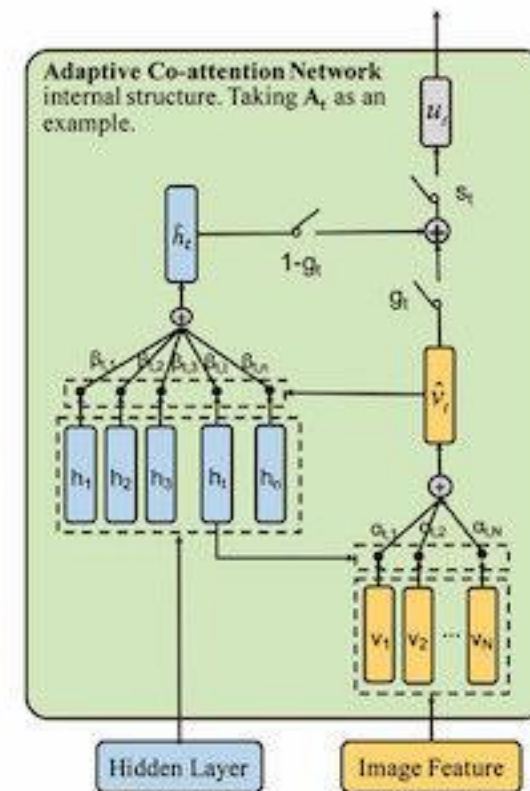
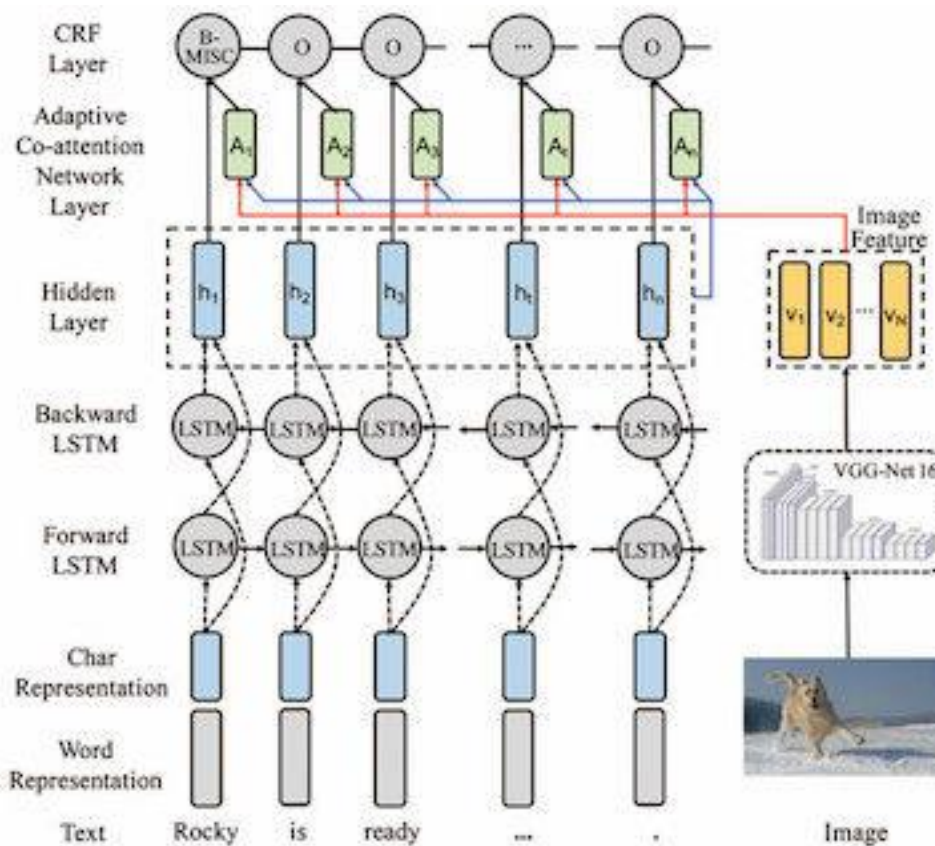




图文混合命名实体识别 (2)

Adaptive Co-attention Network

- 1) 命名实体识别的经典神经网络框架：双向LSTM+CRF
- 2) 在特征层引入图像特征 (VGGNet)
- 3) 通过Co-attention机制来增强文本和图片之间的关联，从而将相关文本和图片内容进行对齐



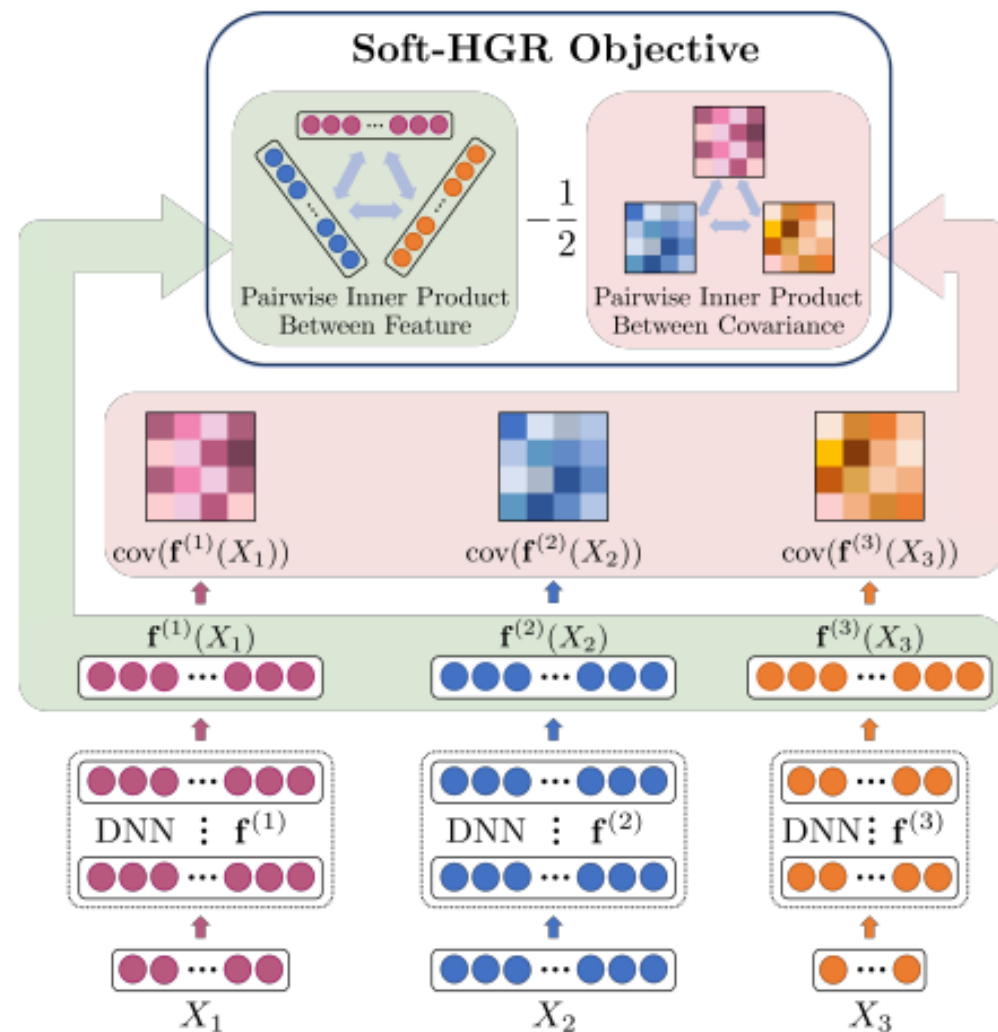


多模态新进展

AI DISCOVERY

◆ An Efficient Approach to Informative Feature Extraction from Multimodal Data, AAAI2019

- HGR最大相关度因较好的理论性质，经常被作为优化的目标函数。然而，HGR最大相关度中对于白化的严格约束，部分限制了其应用。
- 为解决这一问题，本文提出了Soft-HGR的新框架，以解决从多个数据模态中提取有效特征的问题。
- Soft-HGR的优化目标仅包含两个内积项，可以保证优化过程的求解效率和稳定性。研究者进一步将该方法泛化，用于解决超过两个数据模态以及部分模态缺失的问题。对于数据中仅有部分标注信息的情况，研究者可以通过半监督适应的方法，使得所提取的特征更具有判别力。





目录



1 自然语言处理概述

基本概念、技术概览

2 自然语言处理进阶

情感分析、机器阅读、自动问答、文本生成

3 多模态融合

多模态分类、多模态检索、命名实体识别

4 应用与实践

优化方法、调参方法、可视化方法
实践：机器翻译



应用与实践



AI DISCOVERY

优化方法

调参方法

可视化方法



AI DISCOVERY





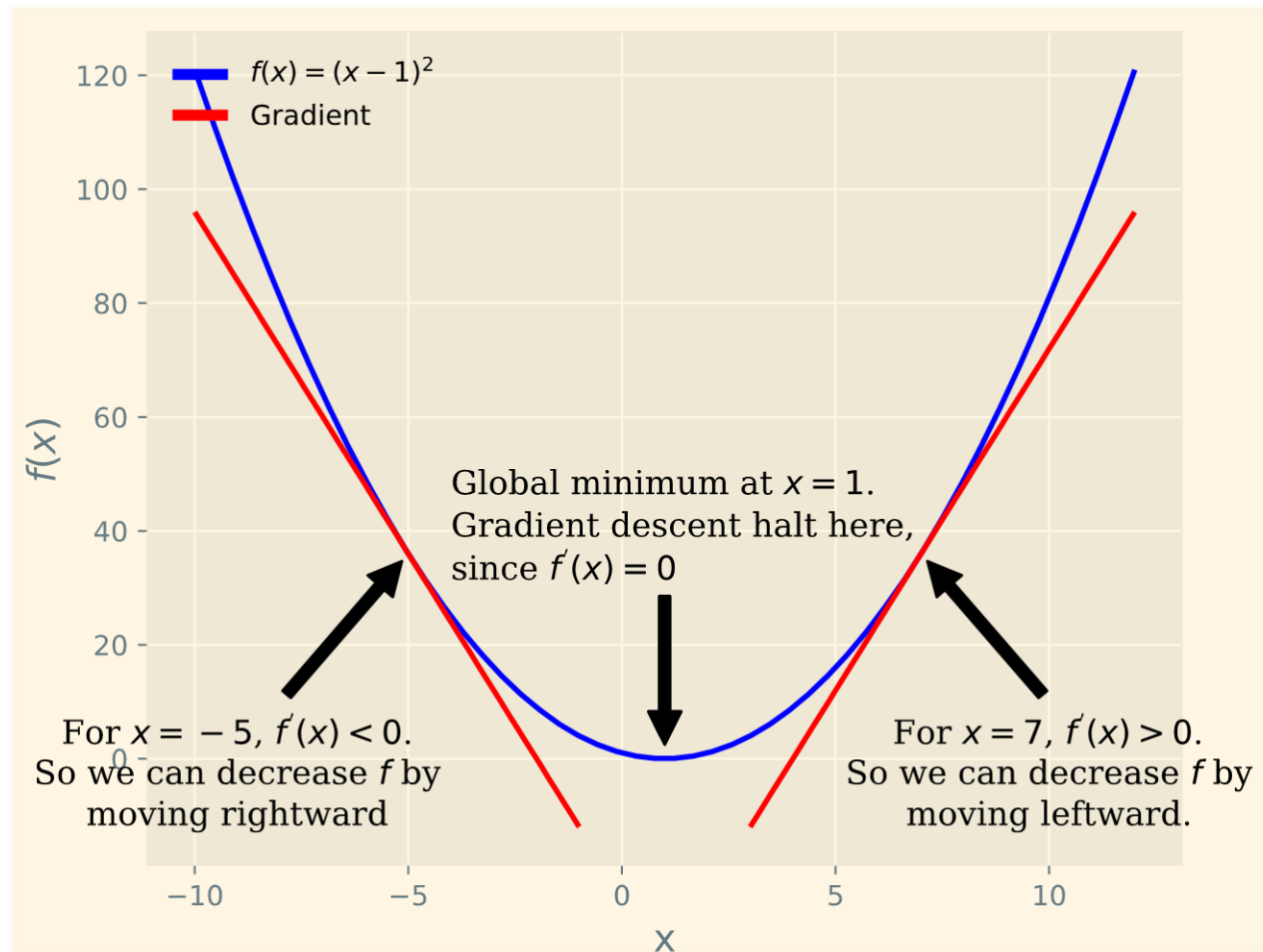
什么是优化 (optimization)



◆ $x = \arg \min f(x)$

➤ 优化是指在给定自变量 x ，及关于 x 的函数 $f(x)$ ，求使得函数 $f(x)$ 取得最大或最小值的自变量 x 。我们把 $f(x)$ 叫做目标函数 (Objective Function)，在某些情况下也叫做损失函数 (Loss Function or Cost Function)。

➤ 如右图为例，使用梯度下降方法求 $\arg \min(x - 1)^2$ 。





优化在深度学习中的应用



- ✓ 在深度学习中，我们通常使用多层神经网络求解问题。我们希望找到神经网络合适的参数 θ ，使得损失函数取得最小值。因此， θ 是优化中的自变量，而损失函数是优化目标，即， $\theta = \arg \min f_{\theta}$
- ✓ 而在神经网络中，参数往往具有较为复杂的运算过程。若要找到损失对于每个参数的梯度，我们需要使用链式法则。
- ✓ 先对参数进行随机初始，然后使用链式法则求得损失对所有参数 θ 的梯度 ∇_{θ} ，我们使用学习率 ϵ 对参数进行更新。即， $\theta = \theta - \epsilon * \nabla_{\theta}$ 。重复上述过程至收敛，我们便完成了参数求解。





举例



AI DISCOVERY

✓ $f(x) = \sigma(g(x)) = \sigma(Wx + b)$ 。 $W = [w_1, w_2]$, $x = [x_1, x_2]$, $\sigma(x) = \text{sigmoid}(x)$, $g(x) = Wx + b$ 。

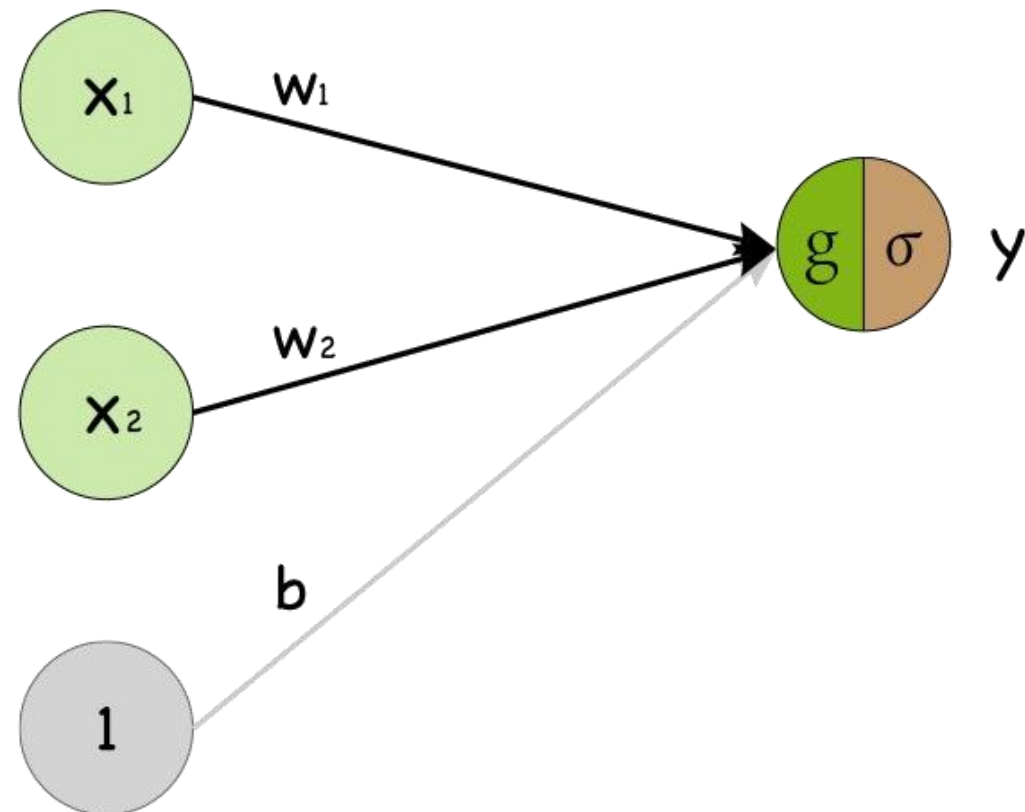
✓ 假设 $W = [0.1, 0.3]$, $b = 0.1$, $x = [2, 3]$, $y = 0.4$, $\epsilon = 0.1$ 。

✓ $f(x) = \sigma(Wx + b) = \sigma\left((0.1 \quad 0.3) \begin{pmatrix} 2 \\ 3 \end{pmatrix} + 0.1\right) = \sigma(1.2) \approx 0.769$

✓ $\text{Loss} = L = \frac{1}{2}(f(x) - y)^2 = (0.769 - 0.4)^2 = 0.068$

✓ $\nabla_{w_1} = \frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial f(\cdot)} \frac{\partial f(\cdot)}{\partial g(\cdot)} \frac{\partial g(\cdot)}{\partial w_1} = (f(x) - y) * g(x)(1 - g(x)) * x_1 =$
 $(0.769 - 0.4) * 1.2 * (1 - 1.2) * 2 = -0.17712$

✓ $w_1 = w_1 - \epsilon \nabla_{w_1} = 0.1 - 0.1 * (-0.1772) = 0.11772$



AI DISCOVERY





存在问题及解决方法

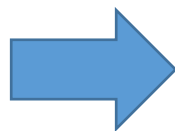
◆ $\epsilon = 0.1$ 的情况下对参数 w_1 进行了优化。

◆ 优化中有如下三大问题：

➤ 如何计算参数的梯度使参数收敛更快，效果更好？

➤ 学习率 ϵ 的取多大比较合适？

➤ 对于数据集中所有数据，如何进行优化？



✓ 问题一是优化算法的主要研究问题，主要有二阶优化方法，模拟退火，使用动量等等。

✓ 学习率的选择问题促使自适应学习率（Ada-）的优化方法诞生。

✓ Batch优化方法同时求得若干个数据的梯度的均值，是介于单条数据和全部数据的折中策略，Batch size也变成了一个最常见的超参（hyper parameter）。



各类优化方法简介



AI DISCOVERY

- ✓ **SGD**: 针对单个样本进行更新权重。
- ✓ **Mini-Batch GD**: 针对一小批量的样本进行更新权重。该方法加上精细的学习率调参有取得最好结果的可能性。
- ✓ **Momentum**: 使用最近两次的导数计算梯度，加速SGD算法在“沟壑”情况下的收敛速度。
- ✓ **Nesterov Accelerated Gradient (NAG)**: 对Momentum的方法进行估计然后修正，避免走的太快。
- ✓ **Adaptive Gradient Algorithm (Adagrad)**: 对低频参数进行较大更新，对高频参数做较小更新。学习率自适应，但是会急剧下降。
- ✓ **Adadelta**: 对Adagrad进行改进，使用均方差进行频度估计，解决学习率急剧下降问题。
- ✓ **RMSprop**: 也是为了解决学习率急速下降问题。使用指数加权平均估计梯度。
- ✓ **Adam**: RMSprop + Momentum, 实际表现中效果较好，也是应用最广泛的一类算法。



AI DISCOVERY

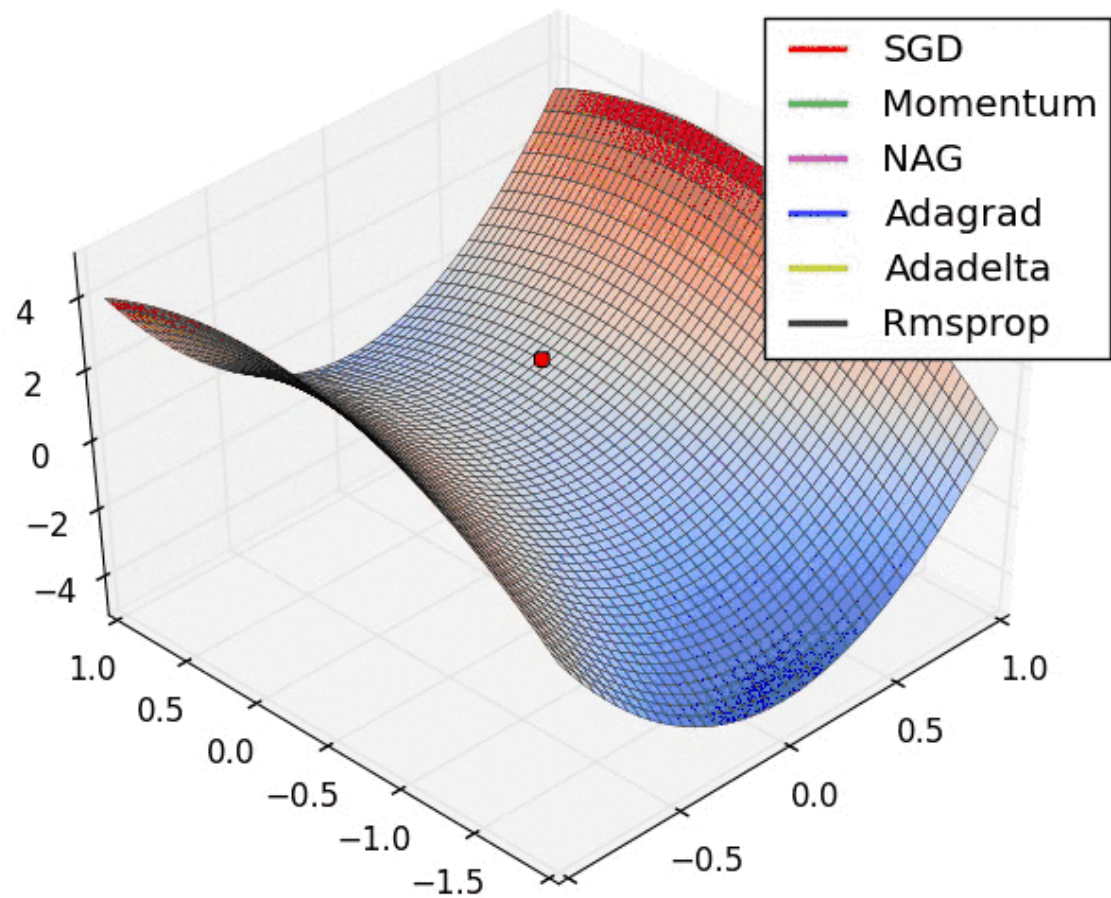


应用比较

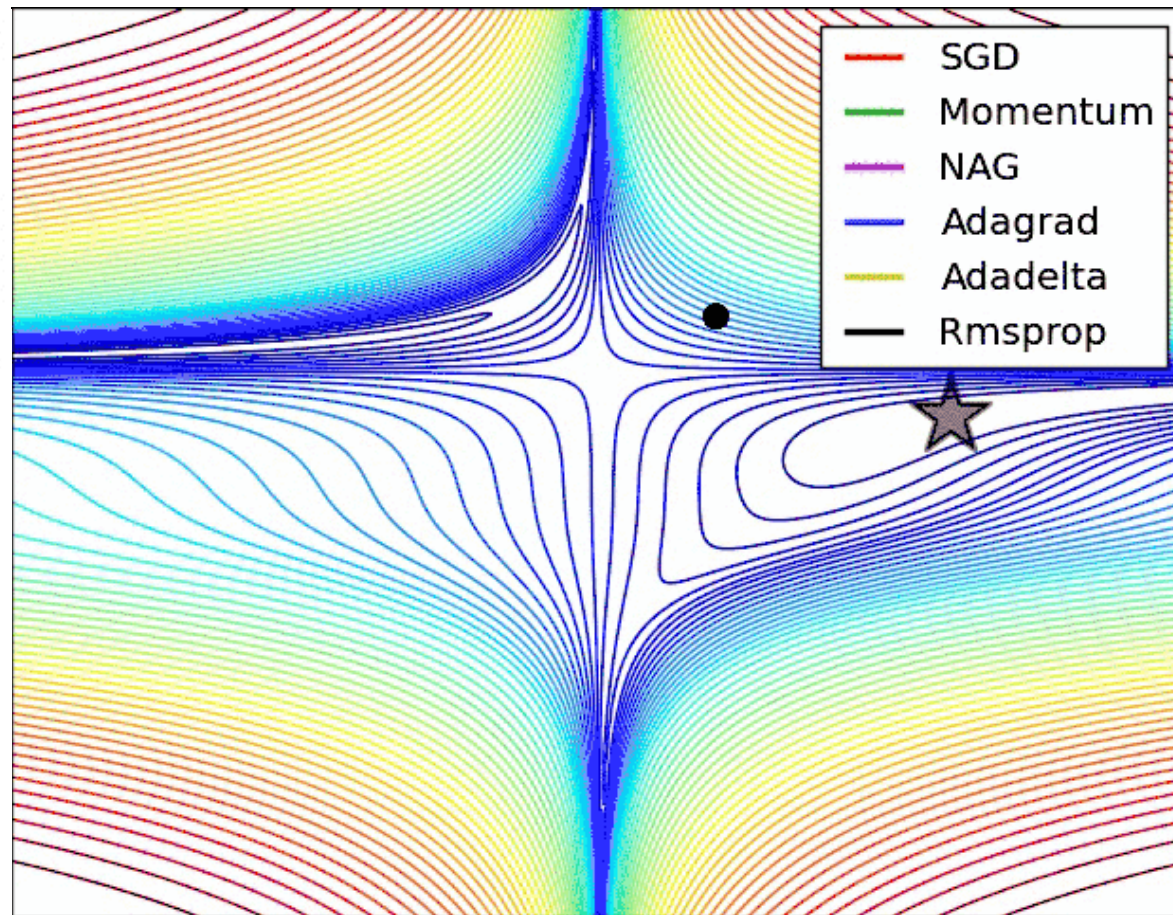


AI DISCOVERY

过鞍点能力



收敛速度



AI DISCOVERY



应用与实践



AI DISCOVERY

优化方法

调参方法

可视化方法



AI DISCOVERY





调参技巧

最优解与过（欠）拟合。



AI DISCOVERY

- ✓ 调参(tune)的目标是调试超参数的取值，以期望模型在测试集达到最优的泛化能力。
- ✓ 最初数据集分为**训练集(training set)**和**测试集(test set)**。研究人员经常会根据测试集的结果调整超参数，而这样做存在一个问题，即我们得到的模型是**最适应测试集**的模型，而无法评价模型在测试集的**泛化能力**。
- ✓ 因此，我们从训练集分离出来一小部分数据集作为**开发集(development set)**，这样我们在开发集上进行调参，便可以评价模型在测试集的泛化能力。
- ✓ 调参往往希望达到**两个目标**，一个是尽可能拟合训练集的数据分布（**较好的局部最优解**），另一个是尽可能在测试集上达到最优效果（**较强的泛化能力**）。



AI DISCOVERY





格搜索(Grid Search)



AI DISCOVERY

- ◆虽然模型参数可以进行自学习，但是不可避免的有一些超参 (hyper parameter) 需要我们去手动调试。
- ◆最常用的方法是格搜索 (Grid Search) ，即对每个超参进行离散地枚举。选择结果最好的一组取值。

Algorithm: GridSearch

FOR bs IN (4, 8, 16, 32, 48, 64):

FOR ϵ IN (0.0001, 0.0005, 0.001, 0.005):

Optimize(θ)

✓其中bs是Batch size大小， ϵ 是学习率。



AI DISCOVERY





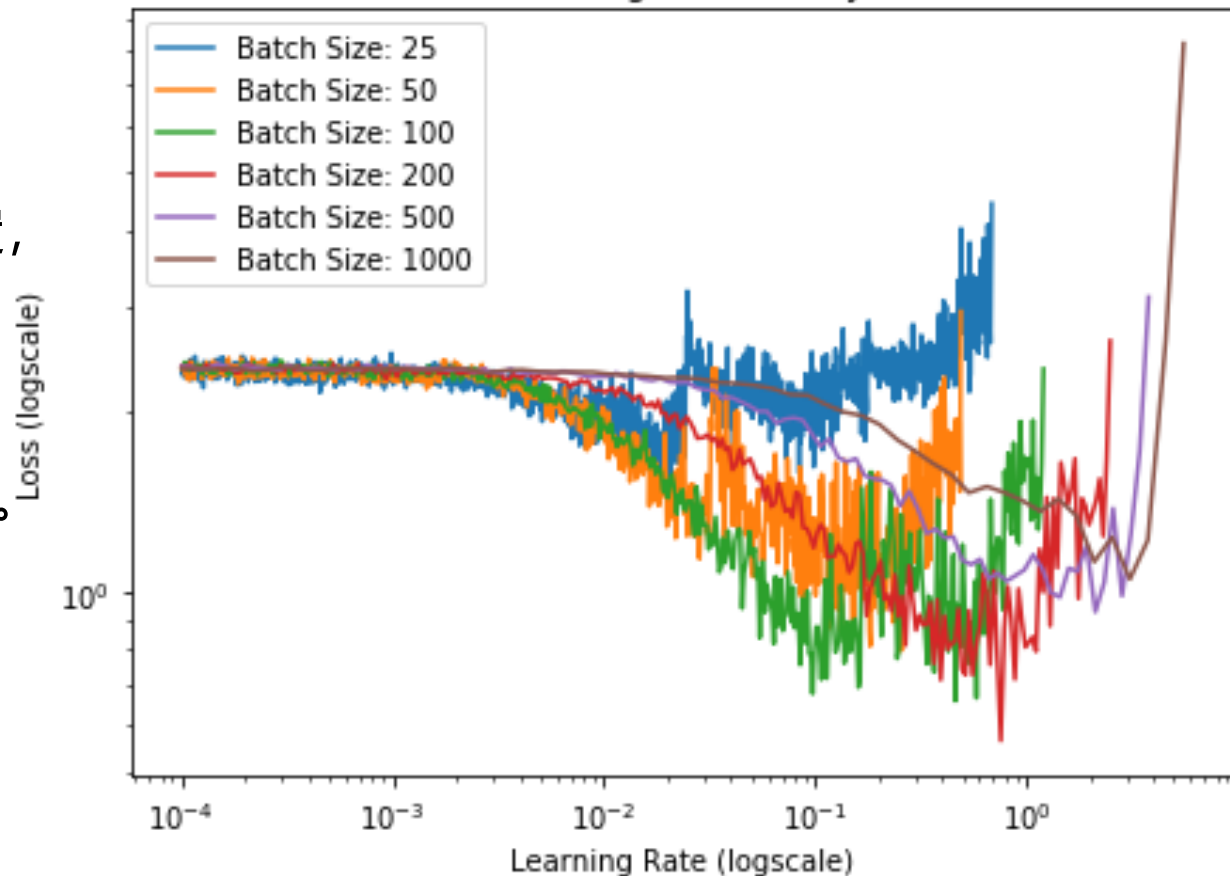
最优解



AI DISCOVERY

- ✓ 学习率较大会在最优解附近产生震荡，导致误差上升，如右图右侧呈上升趋势的曲线。
- ✓ 右图展示了学习率，Batch size 对loss的影响。学习率过大过小，loss都较高。过小的时候稳定，但是过大的时候震荡上升。
- ✓ 对于相同数据数量不同batch size的结果对比如右图。size较小的时候波动剧烈，size较大稳定。在100的时候取到最优值。
- ✓ 最新的研究发现，增大batch size是减小学习率另外一种替代策略。

Learning Rate Survey



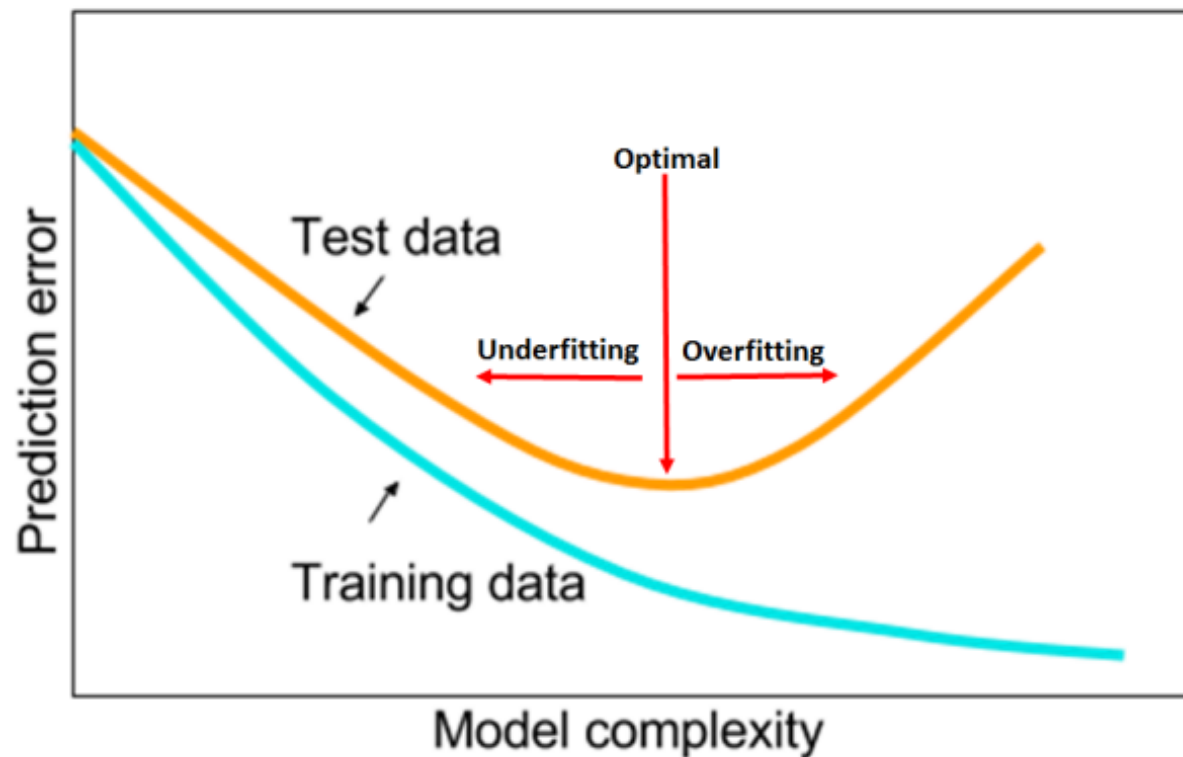
AI DISCOVERY



欠拟合与过拟合



- 如果一个模型在训练集上的误差很大，那么模型可能存在**拟合能力不足**的问题，即模型不能拟合该问题，我们把这种情况叫做**欠拟合**，此时常用做法是**增加神经网络层数**，使神经网络变深，参数变多，神经网络的拟合能力便随着增强。
 - 如果模型训练集误差逐步下降，并在一个比较小的区间继续震荡下降，此时模型的拟合能力足够，而我们要担心模型是否**过拟合**。过拟合的模型会在训练集保持较低的误差，但是泛化能力将会变弱。因此模型在测试集的误差将会先下降后上升，而此时的拐点将是我们想要的**最优值**。
- ✓ 解决上述问题的方法：**正则化**，**dropout**

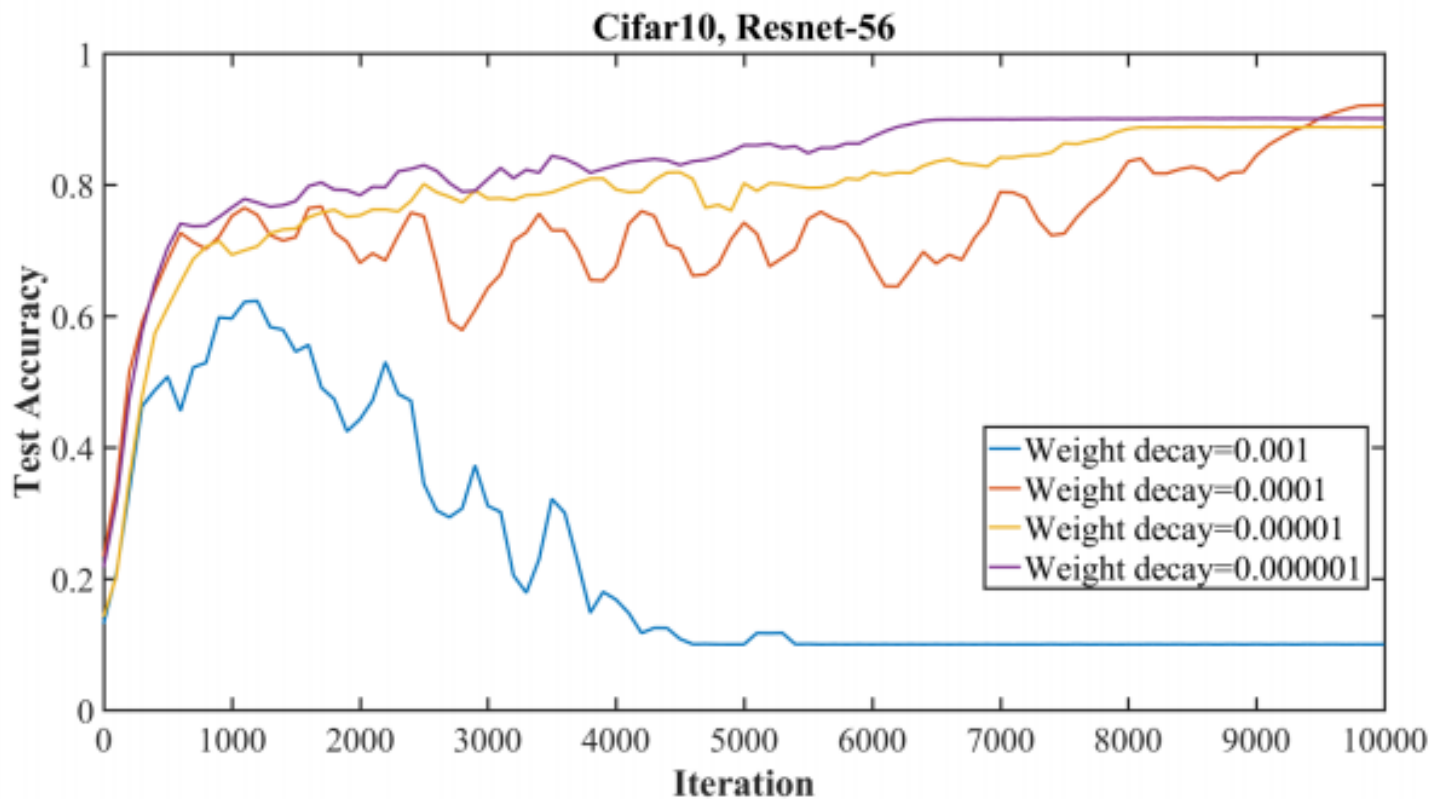




防止过拟合



- ◆ **正则化**: 在目标函数中加入参数的**二范数**，以最小化参数的二范数的方法来约束参数的大小，期望参数取值**尽可能靠近0**。下图展示了正则化项的**权重weight decay**在Cifar数据集对准确率的影响。
- ◆ **dropout**: 模型以一定概率随机丢弃隐含节点的值，置为0。**一般丢弃概率取0.5**。



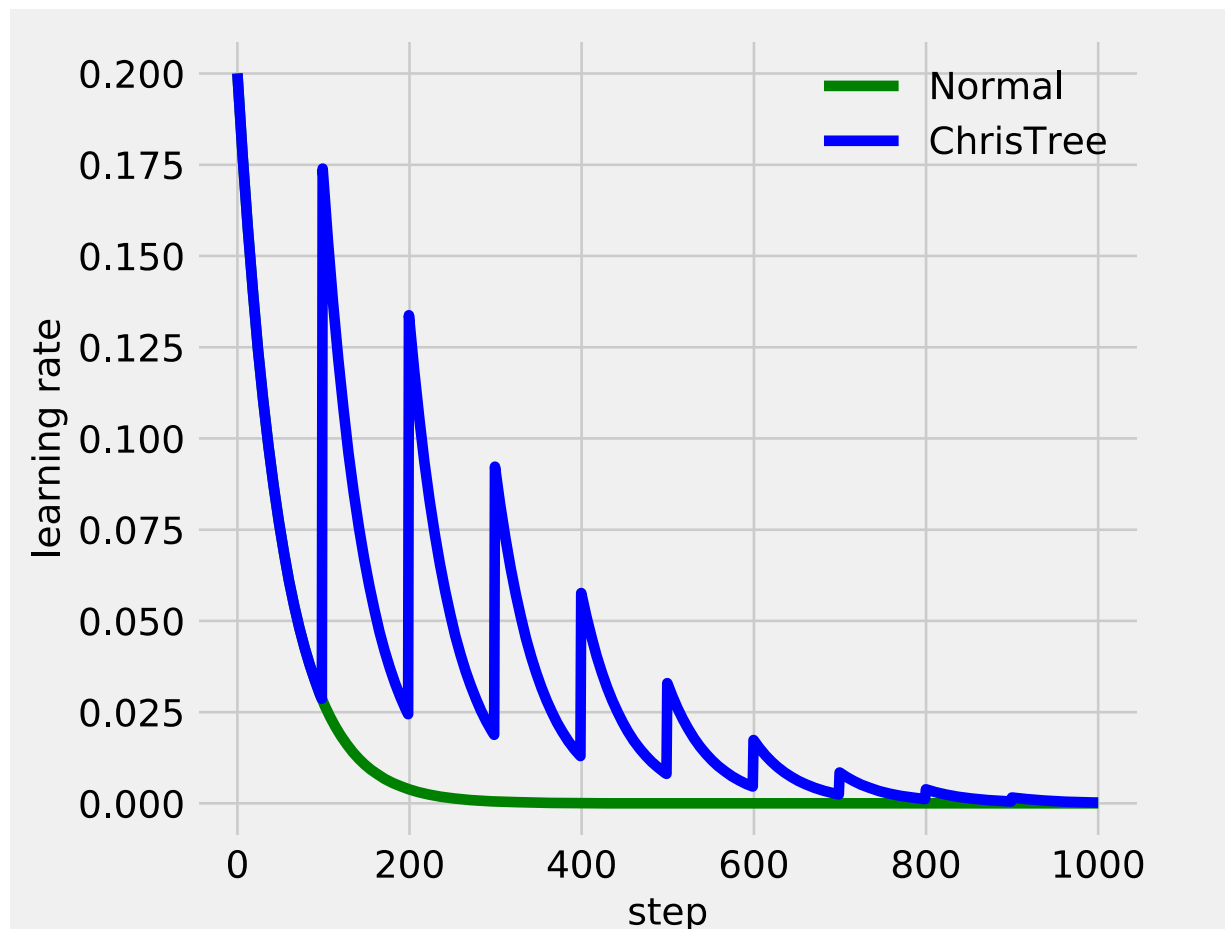


调参进阶



AI DISCOVERY

- ✓ 好的参数初始化策略会让你事半功倍，常用的有高斯初始化，正交初始化，迁移学习等。
- ✓ 由于参数初始化不同，持续减小的学习率可能导致模型收敛到一个较差的局部最优解（Normal），这时候我们需要跳出该局部最优解。我们以某一周期扩大学习率的取值，形成歪倒松树状的学习率取值（ChrisTree），如右图。
- ✓ 随着网络层数的加深，梯度在最底层可能特别小，且小于正则化项提供的梯度，这时候我们可以**对每层参数单独设置其正则化权重weight decay**。



AI DISCOVERY



应用与实践



AI DISCOVERY

优化方法

调参方法

可视化方法



AI DISCOVERY



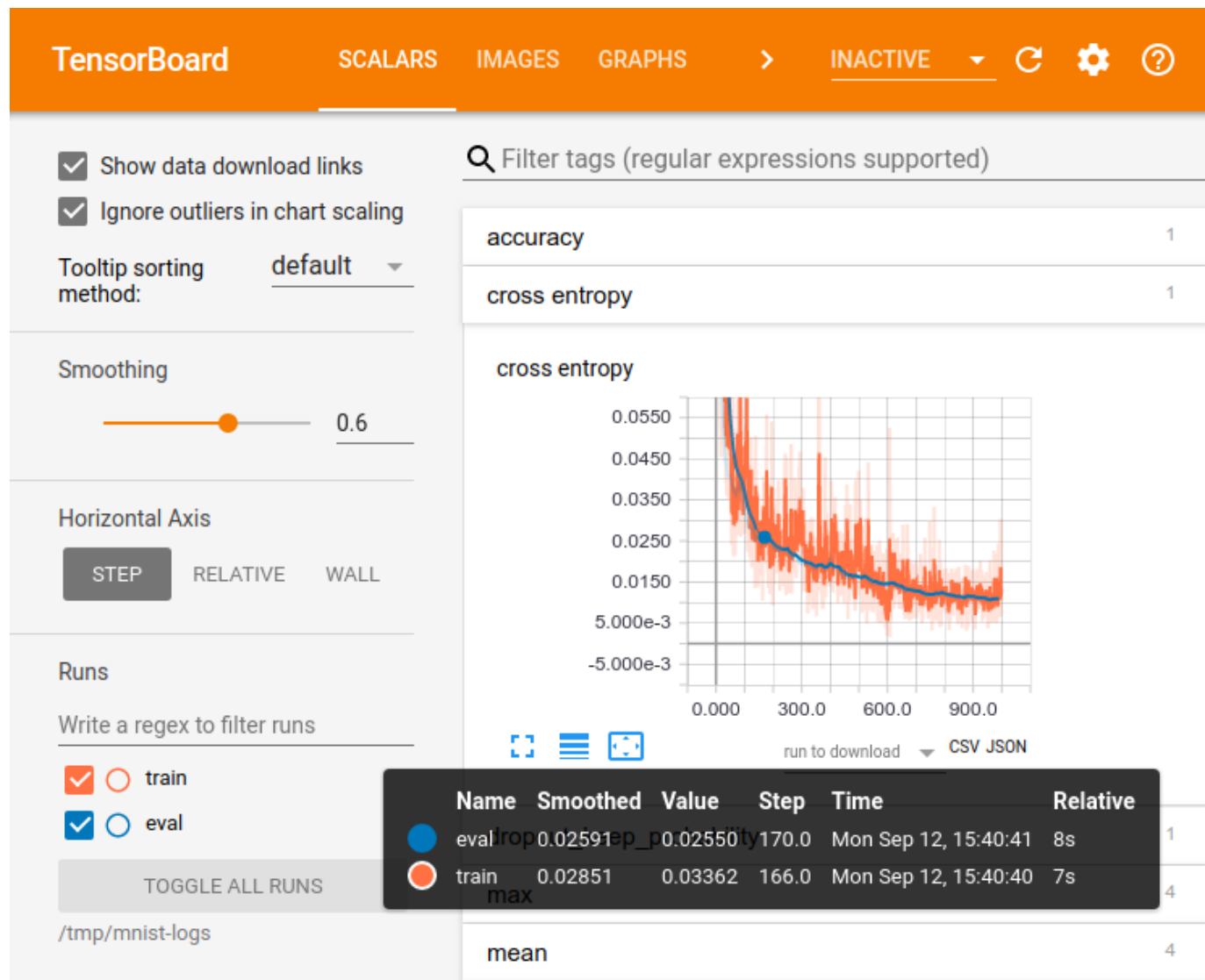


模型可视化



AI DISCOVERY

- ✓ TensorBoard: TensorFlow 集成可视化工具
- ✓ 为了方便 TensorFlow 程序的理解、调试与优化, Google 发布了一套叫做 TensorBoard 的可视化工具。你可以用 TensorBoard 来展现你的 TensorFlow 图像, 绘制图像生成的定量指标图以及附加数据
- ✓ PyTorch 1.1 正式支持 tensorboard 可视化工具



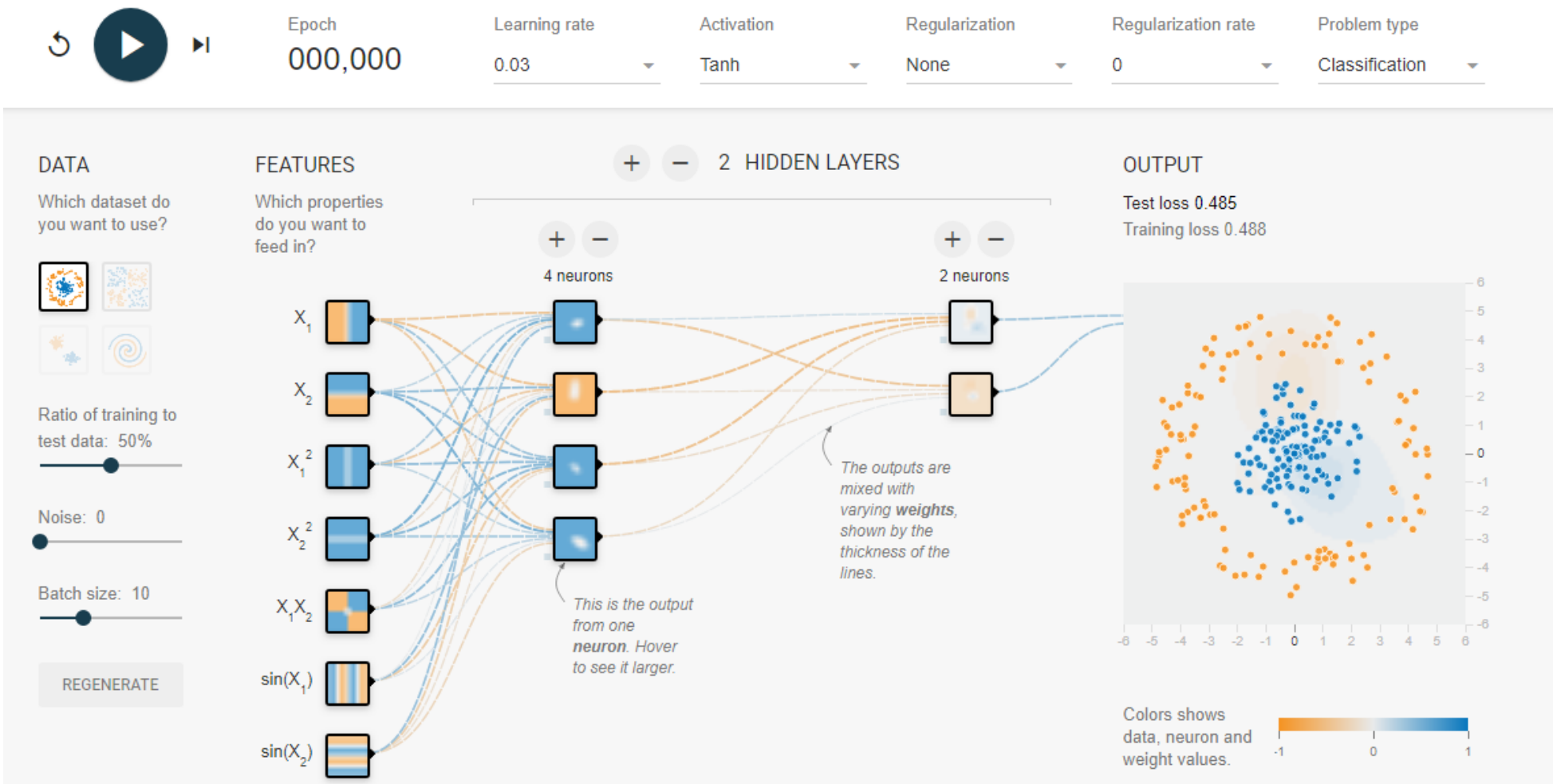


模型可视化



AI DISCOVERY

✓ Playground 是一个图形化用于教学目的简单神经网络在线演示、实验的平台，非常强大地可视化了神经网络的训练过程。<http://playground.tensorflow.org>





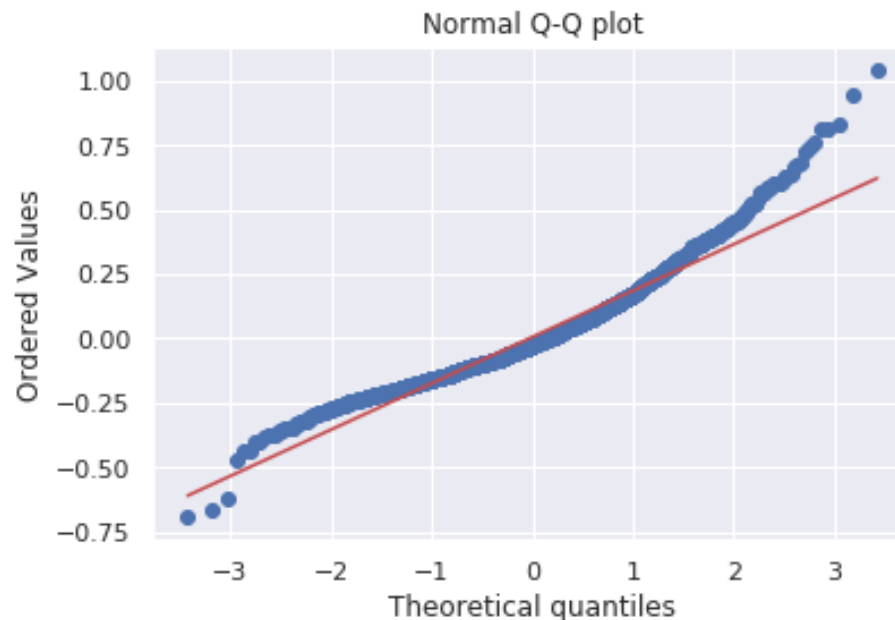
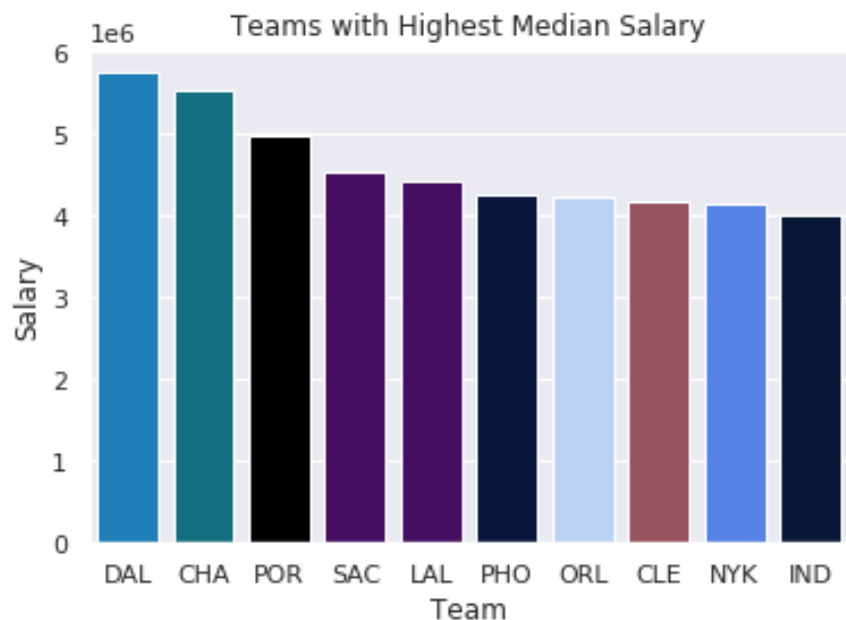
可视化工具



AI DISCOVERY

Matplotlib、Seaborn、Pandas

- Seaborn 和 Pandas 是建立在 Matplotlib 之上的，在用 Seaborn 或 Pandas 中的 `df.plot()` 时，用的其实是 Matplotlib 的代码
- Matplotlib 是比较低级的库，所以它具有很强的自定义程度，可以选择样式 (style selection)，它模拟了像 ggplot2 和 xkcd 等很流行的美化工具



VERY



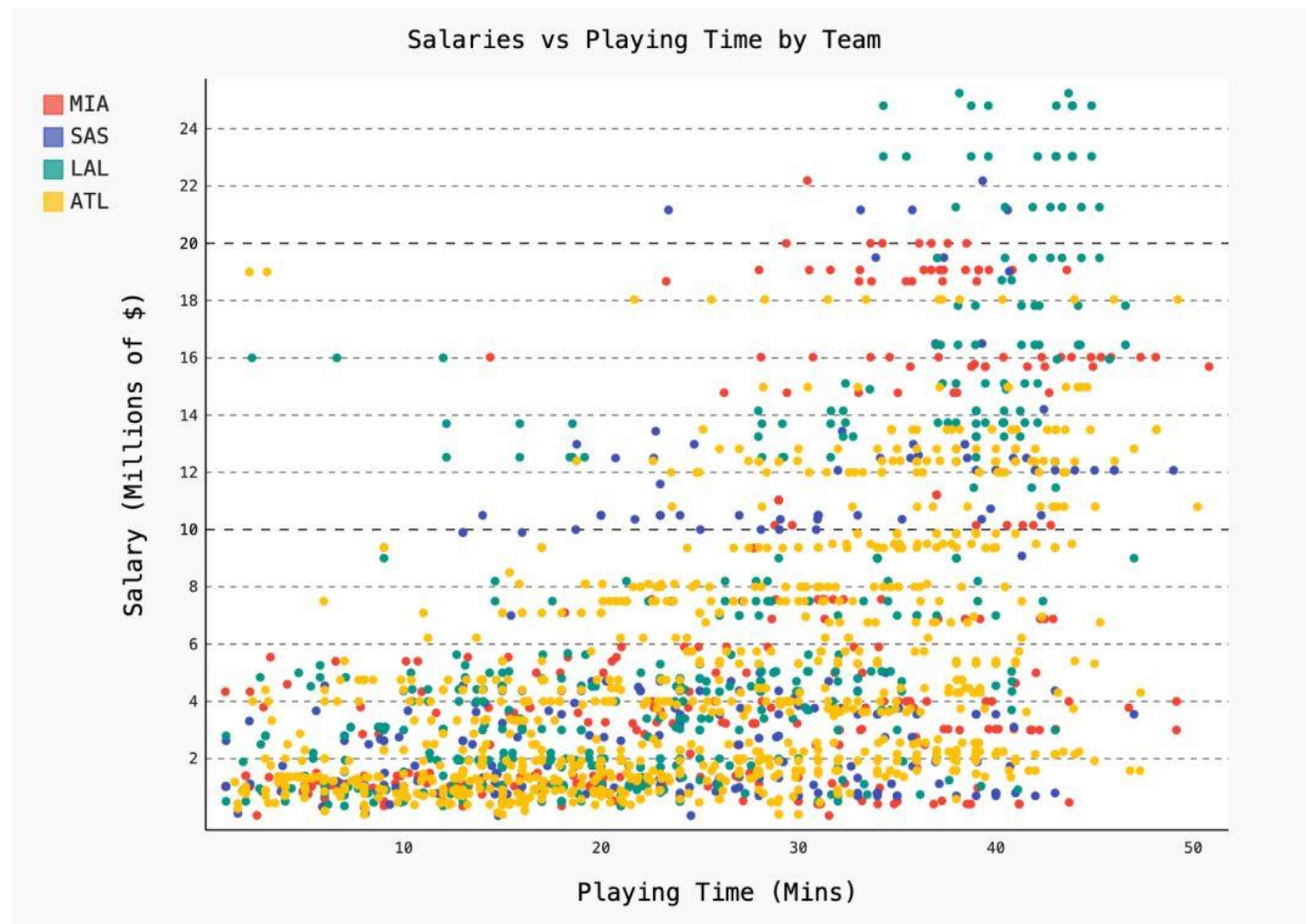


可视化工具

AI DISCOVERY

可以交互的Pygal

- 它是用图形框架语法来构建图像的,由于绘图目标比较简单,因此这是一个相对简单的绘图包
- 使用 Pygal 非常简单
 - 实例化图片
 - 用图片目标属性格式化
 - 用figure.add()将数据添加到图片中
- 因为图片是交互式的,有令人满意而且便于自定义的美化功能



AI DISCOVERY



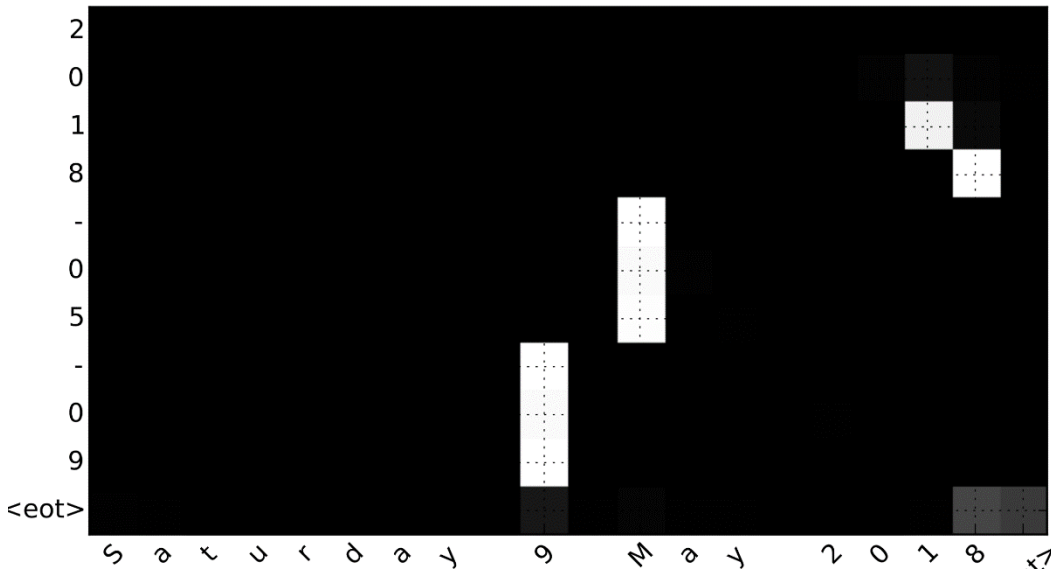
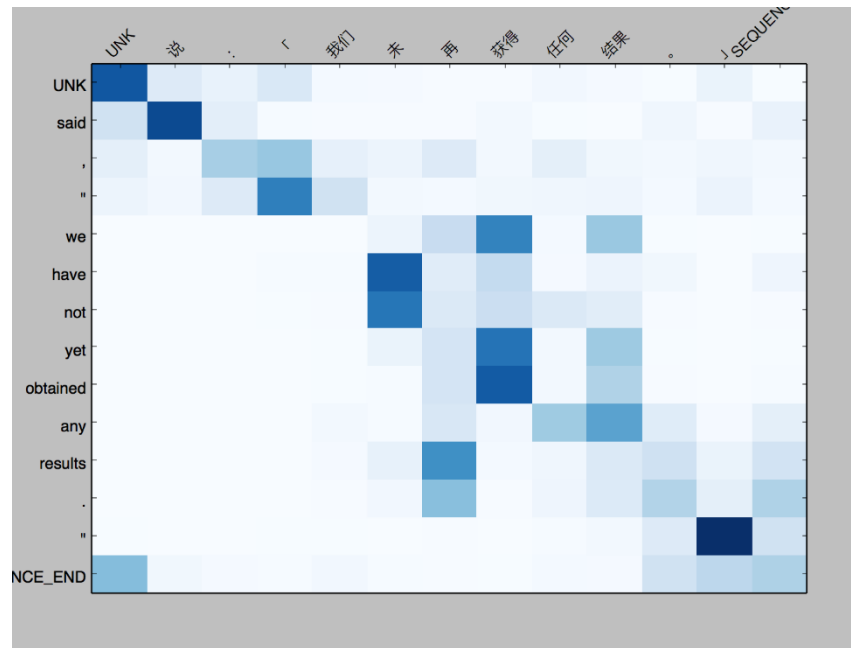
Attention矩阵可视化

AI DISCOVERY

Attention矩阵可视化

- 将Attention参数矩阵从模型中导出
- 利用绘图工具可视化

- <https://medium.com/datalogue/attention-in-keras-1892773a4f22>
- <https://github.com/zhaocq-nlp/Attention-Visualization>
- <https://github.com/Aurelius84/SPWE>



序号	Word_attention	POS_attention
1	都不错 反应也很快 像素也还行吧	都不错 反应也很快 像素也还行吧
2	超大屏 速度不错 这个价位值的	超大屏 速度不错 这个价位值的
3	手机开机 比较慢 其他还好	手机开机 比较慢 其他还好
4	手机很轻薄 外观很漂亮 她喜欢相信京东	手机很轻薄 外观很漂亮 她喜欢相信京东
5	一点都不好用 信号一点都不好	一点都不好用 信号一点都不好
6	爸爸很喜欢 屏幕清 字体大	爸爸很喜欢 屏幕清 字体大
7	总体还不错 吧就是太耗电 续航能力差	总体还不错吧 就是太耗电 续航能力差
8	一次不错的 的选择很满意	一次不错的 的选择很满意



课程实践



AI DISCOVERY

实践：机器翻译



AI DISCOVERY

