R 网︰ M创作R峰顶的小道C与 OMPREHENSION

S小精灵-M正在N网络\_

自然语言计算组，微软研究亚洲y

A断章取义

在本文中，我们介绍了 R-网络，为端到端的神经网络模型

阅读理解样式问题的回答，其目的是要回答的问题

从给定的通道。我们第一次匹配问题和与通道门控

基于注意递归网络获得的问题意识到通道表示形式。

然后我们提出自匹配的注意机制，以完善

通过匹配通过反对本身，它有效地将编码表示形式

整篇文章的信息。我们最后雇用到指针网络

找到答案从段落的位置。我们进行广泛的实验

对球队和 MS-马尔科数据集和我们的模型达到最佳

结果在两个数据集之间所有发表的结果。

1 我公司概况

在本文中，我们专注于阅读理解样式问题回答，旨在回答

给一段或文档的问题。我们主要专注于斯坦福大学问答

数据集 （队） （Rajpurkar et al.，2016年） 和微软机阅读理解

（MS-马可） 数据集，为阅读理解和问答两个大型数据集

两者都通过众包手动创建的。队需要回答的问题

给出了一篇短文。它约束的所有可能的范围内参考通道，空间的答案

这是不同的从完形填空式阅读理解数据集 （Hermann et al.，2015;小山 et al.，

2016 年） 在哪个答案是单个词或实体。此外，球队需要不同形式的

逻辑的推理来推断出答案 （Rajpurkar et al.，2016年）。另一种真实数据集，MS-马可 ·

提供几个相关文件收集从 Bing 索引的一个问题。答案

在 MS 马可的问题由人类活动产生和答案的话不只能来自

给定的文本。

自球队数据集的发布以来取得了飞速进展。王 & 江 (2016b)

构建与比赛-LSTM (王 & 江，2016a)，问题意识到通道表示和预测

回答中与指针网络 （Vinyals et al.，2015年） 通道的边界。Seo 等人 （2016 年）

引入多层次模型问题通过对双向注意流网络

粒度。熊等人 （2016 年） 提出了动态的共同关注网络参加问题

并通过同时和迭代完善答案预测。李等人 （2016 年） 和俞 et al。

（2016 年） 预测答案由排名连续文本范围内的段落。

由王 & 江 (2016b) 的启发，我们引入 R 网，端到端图 1 所示

阅读理解和回答问题的神经网络模型。我们的模型由组成

四个部分︰ 1) 构建问题和段落表示递归网络编码器

另外，2) 封闭的匹配层，以匹配问题和通道，3) 自匹配层

从整篇文章和 4 的聚合信息） 指针网络基础答案边界

预测层。这项工作的主要贡献有三个。

\_ 这是我们的系统和算法，即为 R 网的这台机器的工作在进步技术报告

阅读理解任务。时有显著的改善，我们会更新本技术报告

R-网球队排行榜。早期版本的此技术的报告，即"门控自匹配

网络阅读理解和回答问题。文惠，南阳，腐乳王维宝宝

张和明周"，已被接受并将 ACL 2017 年提出。

y请联系腐乳魏晋明周机中的阅读理解研究微软

亚洲研究中心。

1

通道︰ 特斯拉后来走近摩根寻求更多的资金来建立一个更强大的发射机。

当被问到哪里去了所有的钱，特斯拉回应说，他受

1901 年恐慌，他 （摩根） 造成了。摩根感到震惊的提醒他的部分

在股票市场崩溃和由特斯拉的违反合同要求更多的资金。特斯拉写道︰

另一个呼吁，摩根，但它也是徒劳的。摩根仍然欠原特斯拉钱

协议和泰斯拉过去面对丧失抵押品赎回权甚至塔施工开始之前。

问题︰ 什么做特斯拉归咎于初始金钱的损失？

答案︰ 恐慌 1901 年

表 1: 来自球队数据集的示例。

首先，我们建议封闭的注意基于递归网络，这增加了额外的大门

基于注意递归网络 （Bahdanau et al.，2014;Rockt¨aschel et al.，2015;王 & 江，

2016a)，占的单词的不同的重要性，要回答的事实

对于阅读理解和回答问题的特定问题。在王 & 江 (2016a)

与他们相应的加权注意问题上下文的一段话中被编码在一起

以生成问题意识到通道表示形式。通过引入一个浇注的机制，我们

门控的注意基于递归网络为通道部分取决于分配不同级别的重要性

对其相关性问题，掩蔽了不相干的过流部件和强调

是重要的。

第二，我们引进一个自匹配的机制，它可以有效地聚合从证据

整篇文章来推断出答案。通过门控匹配层，由此产生的问题意识

通过表示有效地将编码通道的每个字的问题信息。然而，

递归网络仅能记住有限的语篇语境中实践尽管其理论的能力。

一个答案候选人往往没有意识到段落的其他部分中的线索。到地址

这一问题，我们建议自匹配层动态完善通过交涉

整篇文章的信息。基于问题意识到通道表示，我们聘请

门控注意基于递归网络上通过对经文本身，聚合有关的证据

对当前通道单词的每一个字在这篇文章。A 门控基于注意复发

网络层和自匹配层动态丰富信息与每个通道表示

从问题和通道，使后续的网络，以更好地预测聚合

答案.

最后，提出的方法产生反对强基线的先进的结果。我们单模型

达到 72.3%精确匹配精度上隐藏的阵容测试集，同时集成模型

进一步推高至 76.9%，结果目前1在球队排行榜第一的位置。

此外，我们的模型也达到 MS 马可数据集上最好的公布的结果 (阮 et al.，

2016 年）。

2 T问D描述

阅读理解样式问题回答，一篇文章P和问题Q给出，我们的任务

是预测答案A问题Q基于中找到的信息P。进一步的阵容数据集

约束的答案A要通过连续分跨度P。答案 A 通常包括非实体

并且可以更长的时间短语。此安装程序挑战我们去理解和关于两个原因

问题和通道以推断出答案。表 1 显示了一个简单的例子，从球队

数据集。对于 MS 马可数据集，几个相关的段落 P 从 Bing 索引为提供

一个问题Q。此外，答案 A 在 MS MARCO 由人类活动产生的不能

这篇文章的连续子范围。

3 R-网 S结构

图 1 给出了 R 网概述。首先，通过双向处理问题和通道

递归网络 （Mikolov et al.，2010年） 分开。我们然后匹配问题和

通过与门控注意基于递归网络，获取问题意识表示

1在 2017 年 5 月 6 日

2

问题通道

问题与通道编码

问题通过匹配

通道自匹配

单词

字符

答案预测

𝑢𝑄 𝑢𝑃

𝑣𝑃

ℎ𝑃

𝑟𝑄

池

qp

注意

PP

注意

P(Begin) P(End)

图 1: R 网结构概述。

时光流逝。最重要的是，我们应用自匹配注意从整体聚合的证据

通道和细化的通道表示形式，然后送入输出层，预测

答案范围的边界。

3.1 Q令和P青杨ENCODER

考虑问题 Q =fwQ

tgmt

1 =和一篇短文 P =fwP

tgnt

1 =。我们首先将转换到单词

他们各自的字级嵌入 (feQ

tgm t1 =和fePt

gn t1 =) 和字符级嵌入

(fcQ

tgmt

1 =和fcPt

gnt

1 =)。字符级嵌入生成的最终隐藏

各国的双向递归神经网络 (RNN) 应用于嵌入的字符在

令牌。这种字符级嵌入已经被证明是有助于处理出的词汇

(OOV) 令牌。然后，我们使用双向 RNN 来产生新的代表 uQ

1; : : : ; uQ

m和

uP1

; : : : ; uPn

所有的词语中的问题和通过分别︰

uQ

tBiRNN =Q(uQ

t􀀀1; [eQ

tc;Q

t]) (1)

uPt

BiRNN =P(uP t􀀀1; [ePt

c;Pt

]) (2)

我们选择使用门控经常性单位 (GRU) （Cho et al.，2014年） 在我们的实验因为它执行

类似于 LSTM (Hochreiter & 施密，1997年) 但计算更便宜。

3.2 G输精管A顽疾-基于RECURRENTN网络

我们建议封闭的关注基于递归网络问题信息纳入通道

表示形式。这是基于注意的递归网络，用额外的门的变异

要确定一个问题关于文章信息的重要性。给定的问题

和通过表示法fuQ

tgmt

1 =和fuPt

gnt

1 =Rockt¨aschel 等人 （2015 年） 提出了生成

句子对表示fvP

tgnt

1 =通过软对齐的问题以及作为通行字

如下所示︰

vP

t= 递归神经网络 (vP

t􀀀1c;t) (3)

在哪里ct= att(uQ; [uPt

v;P

t􀀀1]) 是整个问题的注意池向量 (uQ):

st

jv =T双曲正切 (WQ

uuQ

j+WP

uuPt

+WP

vvP

t􀀀1)

ti

= exp (sti

)=\_mj

1 =exp (st

j)

ct= \_mi

1 =ti

uQ

我(4)

3

每个通道表示vP

t动态合并聚集的匹配信息，从

整个问题。

王 & 江 (2016a) 介绍比赛-LSTM，以uPt

作为经常性的额外投入

网络︰

vP

t= 递归神经网络 (vP

t􀀀1; [uPt

c;t]) (5)

若要确定过流部件的重要性和参加的有关问题，我们将添加

输入另一个大门 ([uPt

c;t]) 的递归神经网络︰

gt= 乙状结肠 (Wg[uPt

c;t])

[uPt

c;t]\_ = gt \_ [uPt

c;t] (6)

不同于在 LSTM 或 GRU 盖茨，额外的门基于当前通道单词

和其注意力集中的问题，侧重于问题之间的关系向量

和当前通道单词。大门口有效模型现象通过那只是部分

是在阅读理解和问答问题有关。[uPt

c;t]\_ 利用

在以后的计算，而不是[uPt

c;t]。我们称这个封闭的关注基于递归网络。

3.3 S小精灵-M正在A顽疾

通过门控注意基于递归网络，问题意识到通过代表fvP

tgnt1 =

生成以找出课文中的重要部分。这种表示法的一个问题是

它有非常有限背景的知识。一个答案候选人往往是没有注意到重要

在其周围的窗口外的走廊的线索。此外，存在着某种形式的词汇或

之间的问题和在大部分的阵容数据集 (Rajpurkar 通过句法差异

et al.，2016年）。通过上下文有必要推断出答案。为了解决这一问题，我们建议

直接匹配问题意识到通道表示反对本身。它动态地收集

从通道中的词整篇文章的证据和编码有关的证据

当前通道单词和成通道表示其匹配问题信息hPt

:

hPt

= BiRNN (hP t􀀀1; [vP

tc;t]) (7)

在哪里ct= att(vP ; vP

t) 是整篇文章关注池向量 (vP):

st

jv =T双曲正切 (WP

vvP

j+W

~ p

vvP

t)

ti

= exp (sti

)=\_nj

1 =exp (st

j)

ct= \_ni

1 =tivP

我() 8

额外的门和门控注意基于递归网络应用于[vP

tc;t] 向自适应

递归神经网络的输入控制。

自匹配提取证据从整篇文章根据当前通道单词和

问题的信息。

3.4 O输出L艾耶尔

我们遵照王 & 江 (2016b) 和使用指针网络 （Vinyals et al.，2015年） 来预测开始

和答案的结束位置。此外，我们使用注意力集中在问题表示法

要生成的初始的隐藏的矢量的指针网络。给出了通道表示

fhPt

gn t1 =，注意机制利用作为选择的起始位置的指针 (p1) 和结束

位置 (p2) 从这篇文章，这可以清楚地表述如下︰

st

jv =T双曲正切 (WP

hhPj

+W

hh在

􀀀1)

ti

= exp (sti

)=\_nj

1 =exp (st

j)

pt= argmax (t

1; : : : ; at

n) (9)

在这里ht􀀀1表示最后的答案经常网络 （指针） 的隐藏的状态。

输入的答案递归网络是基于电流预测的注意池矢量

概率t:

ct= \_ni

1 =ti

hPi

h在

= 递归神经网络 (h在

􀀀1c;t) (10)

4

当预测的开始位置，ht􀀀1表示答案经常隐藏的初始状态

网络。我们利用问题矢量 rQ 作为答案递归网络的初始状态。

rQ= att(uQ;V Q

r) 基于参数的问题注意池向量 V Q

研发:

sjv =T双曲正切 (WQ

uuQ

j+WQ

vVQ

研发)

我= exp (s我)=\_mj

1 =exp (sj)

rQ= \_mi

1 =我uQ

我() 11

对网络进行训练，我们尽量减少地面真相开始的负对数概率的总和

和结束位置的预测分布。

4 E试验

4.1 我广播式D在表述

我们主要专注于要训练和评估我们的模型，得到了一个巨大的阵容数据集

在过去几个月的注意。球队是组成的 100000 + 人群提出的问题

工人对 536 维基百科的文章。数据集随机分成训练集 （80%)，

发展集 （10%)，和一个测试集 （10%)。每个问题的答案是一段

相应的段落。

我们使用从斯坦福大学 CoreNLP （曼宁 et al.，2014年） 记号来预处理每篇文章

和问题。门控的经常性单位 （Cho et al.，2014年） LSTM 变量，该变量在整个使用我们

模型。对于嵌字，我们使用预先训练区分大小写的手套嵌入2 (边宁顿

et al.，2014年） 额问题和通道，和它固定在训练;我们使用零矢量

表示所有出单字单词。我们利用 1 层的双向 GRU 计算 characterlevel

嵌入和 3 层的双向 GRU 编码问题和通道，门控

基于注意递归网络问题和通道匹配也是编码的双向

在我们的实验。隐藏的矢量长度设置为 75 的所有图层。所用到的隐藏的大小

计算注意分数也是 75。我们也适用辍学 （斯里瓦斯塔瓦等人，2014年） 层间

辍学率为 0.2。与 AdaDelta (Zeiler，2012年) 与最初的优化模型

学习速率的 1。 \_ 和 \_ 在 AdaDelta 中使用是 0.95 和 1e􀀀6 分别。

4.2 M艾因R壳体

两个指标用以评价模型性能的阵容︰ 精确匹配 (EM) 和 F1 的分数。

EM 的措施与地面真相答案之一完全匹配预测的百分比。

F1 措施预测和地面真相答案之间的重叠，这需要最大

在所有地面真相 F1 回答。Dev 套分数计算由官方脚本3.

由于测试集隐藏的我们需要提交到斯坦福 NLP 组获得模型

考试成绩。

表 2 显示精确匹配和 F1 分数上的开发和测试设置的我们的模型和竞争

方法4。合奏模型包括 18 培训运行与相同的体系结构和

超参数。在测试时，我们选择的信心分数总和最高的答案

其中每个问题 18 运行。正如我们所看到的我们的方法明显优于基准

和几个强大的国家---艺术系统单一模型和合奏。R 网 (3 月，

2017年) 条目所指与 ACL 提交后我们改进的结果。后原

自匹配层的通道，我们利用双向 GRU 进行深度整合匹配

之前他们进答案指针层的结果。它有助于进一步传播的信息

通过自匹配的通道进行聚合。

2从下载http://nlp.stanford.edu/data/glove.840B.300d.zip.

3从下载http://stanford-qa.com

4从球队排行榜的提取http://stanford-qa.com在 2017 年 5 月 6 日。

5

Dev 集测试集

单模型EM / F1 EM / F1

LR 基线 （Rajpurkar et al.，2016年） 40.0 / 51.0 40.4 / 51.0

动态块读者 （玉等人，2016年） 62.5 / 71.2 62.5 / 71.0

细心的 CNN 上下文与 LSTM （刘成、 CASIA）-/-63.3 / 73.5

比赛-LSTM 与 Ans Ptr (王 & 江，2016b) 64.1 / 73.9 64.7 / 73.7

动态 Coattention 网络 （熊 et al.，2016年） 65.4 / 75.6 66.2 / 75.9

迭代的 Coattention 网络 （复旦大学）-/-67.5 / 76.8

FastQA （Weissenborn et al.，2017年）-/-68.4 / 77.1

BiDAF （Seo et al.，2016年） 68.0 / 77.3 68.0 / 77.3

T-浇注 （北京大学）-/-68.1 / 77.6

打字 （李 et al.，2016年）-/-69.6 / 77.7

SEDT + BiDAF （刘 et al.，2017年）-/-68.5 / 78.0

多角度匹配 （Wang 等人，2016年）-/-70.4 / 78.8

FastQAExt （Weissenborn et al.，2017年）-/-70.8 / 78.9

助记符读者 （国防科学技术大学和复旦大学）-/-69.9 / 79.2

文档阅读器 （Chen 等人，到 2017年）-/-70.7 / 79.4

ReasoNet （Shen 等人，2016年）-/-70.6 / 79.4

反刍的读者 (龚 & 鲍曼，2017年)-/-70.6 / 79.5

jNet （张 et al.，2017年）-/-70.6 / 79.8

交互式的 AoA 读者 （打联合实验室和研究讯飞）-/-71.2 / 79.9

R-网 （Wang 等人，到 2017年） 71.1/ 79.5 71.3/ 79.7

R-网 (3 月 2017) 72.3/ 80.6 72.3/ 80.7

集成模型

细粒度浇注 （杨 et al.，2016年） 62.4 / 73.4 62.5 / 73.3

比赛-LSTM 与 Ans Ptr (王 & 江，2016b) 67.6 / 76.8 67.9 / 77.0

QFASE (NUS)-/-71.9 / 80.0

动态 Coattention 网络 （熊 et al.，2016年） 70.3 / 79.4 71.6 / 80.4

T-浇注 （北京大学）-/-72.8 / 81.0

多角度匹配 （Wang 等人，2016年）-/-73.8 / 81.3

jNet （张 et al.，2017年）-/-73.0 / 81.5

BiDAF （Seo et al.，2016年）-/-73.7 / 81.5

SEDT + BiDAF （刘 et al.，2017年）-/-73.7 / 81.5

助记符读者 （国防科学技术大学和复旦大学）-/-73.7 / 81.7

ReasoNet （Shen 等人，2016年）-/-75.0 / 82.6

R-网 （Wang 等人，到 2017年） 75.6/ 82.8 75.9/ 82.9

R-网 (3 月 2017) 76.7/ 83.7 76.9/ 为 84.0

人类性能 （Rajpurkar et al.，2016年）-/-82.3 / 91.2

表 2︰ 我们的研发网和相互对立的方法的性能4在球队数据集上。

6

单模型胭脂-L / BLEU1

FastQAExt （Weissenborn et al.，2017年） 33.7 / 33.9

预测 (王 & 江，2016b) 37.3 / 40.7

ReasoNet （Shen 等人，2016年） 38.8 / 39.9

R-网 42.9 / 42.2

表 3︰ 我们的研发网和相互对立的方法的性能5女士马可数据集上。

4.3 MS-马可 ·结果

我们也适用于 MS 马可数据集 （阮 et al.，2016年） 我们的方法。MS-马尔科是另一个

机器理解数据集，与球队的两个关键区别。在 MS-马可，每

问题有几个相应的段落，因此，我们只是连接的一个问题的所有段落

在顺序中，给出了在数据集。其次，在 MS 马可答案不一定是 subspans

段落，在官方工具中的 MS 马可评价指标是蓝带

和胭脂-L，广泛应用于多个领域。在这方面，我们选择与跨度

与引用胭脂 L 得分最高为黄金的跨度，在训练中，回答和预测

最高得分范围作为在预测期间的答案。我们训练我们 MS 马可数据集上的模型

(表 3)，结果表明，我们的方法性能上远远胜过其他竞争的基线5.

4.4 D讨论

在本节中，我们报告，并讨论一些未能在我们的实验带来改善的努力。

随着所有的经验主义的研究结果在阵容上，在这里报道的研究结果只适用于我们确切

设置。结果并不一定表明讨论了方法的有效性时

用于其他数据集或结合基线模型与我们的不同。我们相信这些方向

都是有价值的研究课题和我们正在尝试这些想法不同的模式和

实现。

1。 句子排名中的阵容，通过包括几个句子和答案

跨度总是陷入一个句子。它是自然地考虑是否排名句子

将有助于找到最终的答案。我们已经试过两种整合句子排序方法

信息: (a) 我们训练一个单独的句子排序模型，并结合该模型

与跨度预测模型;（b） 我们对待跨度预测和作为句子预测

两个相关的任务，并培训多任务模型。这两种方法未能改善决赛

结果。分析显示，句模一贯下-执行跨度

即使是在句子预测任务的预测模型。我们最好的句子模型精度达到

86%，而我们的跨度预测模型有超过 92%的精度预测答案

句子。这表明，确切的跨度信息其实是选择的关键

正确的答案的句子。

2. 的语法信息我们有试过三种方法来集成语法信息我们

模型。首先，我们要添加一些语法功能作为输入的编码层。这些

语法功能包括 POS 标签、 押尼珥结果、 线性的文法树标记和依赖项

标签。第二，我们试图整合我们的编码后的树 LSTM 风格模块

图层。我们使用多输入的 LSTM 打造隐的状态后依赖项树路径

在自顶向下和自底向上传递。最后，我们试图使用依赖项解析为

在多任务环境中的其他任务。所有上述的失败，给我们的模式带来任何利益

关于球队的数据集。

3. 多跳推理我们试图添加多跳推理模块在答案

指针层，但是未能得到改善，在当前上下文中的最终结果

R-网结构。原因之一可能是需要这样的问题

推理都过于复杂，根据当前的设置，尤其考虑到学生有效地学习

没有关于明确推理过程在阵容中的批注。

5结果除了我们从 MS 马可排行榜中提取http://www.msmarco.org/

leaders.aspx在 2017 年 5 月 6 日。

7

4. 问题生成数据驱动的方法，标记的数据可能成为瓶颈

为更好的性能。丰富文本时，它不是容易找到问题通过

对符合球队的风格。若要生成更多的数据，我们训练序列 tosequence

使用数据集队 （周 et al.，2017年），和生产问题生成模型

大量的伪问题通过对从 EnglishWikipedia。我们的训练

这个伪语料库与球队训练数据，R 网模型和我们分配

小重量到自动生成的样本所以，伪语料库的总权重和

真实语料是大约相等。到目前为止，这种做法未能在决赛中获得任何收益

结果。分析表明，质量生成问题需要改进。

5 R扬眉吐气WORK

阅读理解和问题回答的数据集基准数据集发挥重要

在阅读理解和问题回答研究的最新进展中的作用。现有的

数据集可以分为两大类是否手动标记的地方。那些

这由标记的人始终是在高质量 （理查森等人，2013;Berant et al.，2014;

杨 et al.，2015年），但太小，培养现代数据密集型模式。那些自动

从自然发生数据生成可以非常大 （山等人，2016;赫尔曼 · et al.，

到 2015 年），使模型更具表现力的训练。然而，他们是在完形填空的风格在

其目标是要预测在一篇文章中所缺的单词 （通常是一个命名的实体）。此外，陈

等人 （2016 年） 表明，CNN / 每日新闻数据集 （Hermann et al.，2015年） 要求较低

推理比先前认为的结束，性能已经接近饱和。

从数据集不同，球队提供大型和高质量的数据集。在答案

球队经常包括非实体和可以是多长的短语，是比更具挑战性

完形填空式数据集。此外，Rajpurkar 等人 （2016 年） 表明，该数据集保留一整套不同的

答案，需要不同形式的逻辑推理，包括多句子推理。女士

马可 （阮 et al.，2016年） 也是一个大型的数据集。在数据集问题是真实的

通过 Bing 或柯塔娜和通道发出匿名的查询是相关的 web 页。为

在数据集中的每个问题，几个相关提供了通道。然而，答案是人类

生成，有别于球队在哪里答案必须通过跨度。

端到端的阅读理解能力的神经网络完形填空式的数据集，以及几个

学习模式 （Hermann et al.，2015; 强大深小山 et al，2016;Chen 等人，2016;卡德尔克

et al.，2016;Sordoni et al.，2016;崔 et al，2016;特里施勒等人，2016;Dhingra et al.，2016;沈

et al.，2016年） 介绍了解决这个问题。赫尔曼等人 （2015 年） 首先介绍关注

到阅读理解的机制。山等人 （2016 年） 提出了一个基于窗口的记忆

CBT 数据集的网络。卡德尔克等人 （2016 年） 引入指针网络与一个注意的一步

要预测出实体冲裁。Sordoni 等人 （2016 年） 提出了迭代的交替注意

机制，较好地模拟问题与段落之间的联系。特里施勒等人 （2016 年） 解决

完形填空式问题应答任务由周到模型与 reranking 模型相结合。Dhingra

等人 （2016 年） 提出了迭代通过乘以浇注选择段落的重要部分

问题有代表性的功能。崔等人 （2016 年） 提出了一种双向的注意机制

要进行编码通道并相互质疑。Shen 等人 （2016 年） 提出了以迭代方式推断

用动态数量的推理步骤回答和与强化学习培训。

神经网络模型在球队数据集上的有效性。王 & 江

(2016b) 结合匹配 LSTM 和指针网络用来产生边界的答案。熊

等人 （2016 年） 和搜索引擎优化等人 （2016 年） 采用备选案文 coattention 机制以匹配问题

和相互的通道。熊等人 （2016 年） 提出了动态指针网络来反复推断

答案。玉等人 （2016 年） 和李等人 （2016 年） 解决球队排名连续文本范围的

内通道。杨等人 （2016 年） 目前动态地组合在细粒度的门机制

字级和字符级表示和模型问题与段落之间的相互作用。

王志伟 （2016 年） 建议匹配的通道的问题从多个方面

观点。

与上述的模式不同，我们在我们的模型介绍自匹配的注意。它动态地

细化的通道表示通过纵观整篇文章和聚合的证据

有关当前通道单词和问题，让我们的模型充分利用通道

信息。几个作品中提出了法参加到词的上下文。岭 et al。

8

（2015 年） 建议考虑基于窗口的上下文词不同词和

它的相对位置。郑等人 （2016 年） 提出了一种新型的 LSTM 网络编码在句子中的单词

其中认为当前正在处理的标记和在其过去的标记之间的关系

内存。帕里克等人 （2016 年） 应用此方法进行编码根据词句子中词语

窗体和它的距离。因为通道信息有关的问题，更有助于推断

答案在阅读理解中，我们应用自匹配基于问题意识表示

和门控的注意基于递归网络。它可以帮助我们主要专注于问题相关的模型

在这篇文章的证据和动态看看整篇文章聚合的证据。

我们的另一个关键组件是模型的基于注意的递归网络，这表明

在范围广泛的任务中的成功。Bahdanau 等人 （2014 年） 首先提出了基于注意复发

网络用来生成目标词时推断词水平对齐方式。赫尔曼 · et al。

（2015 年） 引入词级的重视阅读理解模型之间的相互作用

问题和通道。Rockt¨aschel 等人 （2015 年） 和王 & 江 (2016a) 提出了确定

通过词对词匹配的蕴涵。门控的关注基于递归网络是一个变种

基于注意递归网络用额外的门模型过流部件的事实

阅读理解和问题回答特定问题的不同重要性。

6 C结论

在此技术的报告中，我们给 R 网阅读理解和回答问题。

我们介绍的门控的注意基于递归网络和自配的注意机制

获取表示问题和通道，和然后使用指针网络来查找

回答的界限。我们的模型实现了先进的结果对球队和 MS-马尔科

数据集，胜过几个强大的竞争对手的系统。为今后的工作，我们将尝试使用语法

并在我们系统知识库信息。此外，我们也正在对新设计

网络的结构来处理问题，需要复杂的推论。

ACKNOWLEDGEMENT

我们感谢普拉 Samir Rajpurkar 和珀西梁队提交的帮助。

REFERENCES

迪米特里 · Bahdanau、 Kyunghyun Cho 和 Yoshua Bengio。由神经机器翻译联合

学习对齐和翻译.科尔，2014年。

乔纳森 · Berant、 Vivek 瑞、 裴春陈、 艾比范林登、 布列塔尼 · 哈丁、 布拉德 ·

黄、 彼得克拉克和克里斯托弗 · D.曼宁。阅读的生物过程建模

理解。在中自然语言的实证方法 2014年会议论文集

处理，EMNLP 2014 年，2014 年 10 月 25-29 日，多哈，卡塔尔，SIGDAT，会议

ACL 的特别兴趣组，2014年。

陈丹，杰森博尔顿和克里斯托弗 · D.曼宁。Cnn 彻底检查

阅读理解任务的邮件。在协会第 54 次年度会议纪要

为计算语言学，2016 年。

陈丹、 亚当甘草、 杰森 · 韦斯顿和安东尼 Bordes。读维基百科来回答 opendomain

的问题。在计算语言学 (ACL) 协会、 2017年。

郑建鹏、 李东和 Mirella Lapata。长期和短期记忆网络机

阅读。在中自然语言的实证方法 2016年会议论文集

加工、 EMNLP 到 2016 年，奥斯汀，得克萨斯州，美国，2016 年 11 月 1 4 日，2016 年。

Kyunghyun Cho，巴特 · 范 Merrienboer，C¸ aglar G¨ulc¸ehre，迪米特里 · Bahdanau，Fethi Bougares，Holger

Schwenk 和 Yoshua Bengio。学习短语表示使用递归神经网络编解码器

的统计机器翻译。在实证方法 2014年会议论文集

在自然语言处理，EMNLP 2014 年，2014 年 10 月 25-29 日，卡塔尔，多哈会议

SIGDAT、 ACL 特别兴趣组，页 1724年 — — 1734 年 2014年。

9

一鸣崔志鹏陈寺圩、 石津王、 亭刘、 国平胡。注意 overattention

神经网络的阅读理解。科尔，2016 年。

Bhuwan Dhingra、 憨笑刘、 威廉 · 科恩和鲁斯兰 Salakhutdinov。门控注意

读者对文本的理解。科尔，2016 年。

一宸龚和撒母耳 R 鲍曼。反刍的读者︰ 推理与门控多跳的关注。

arXiv 预印本 arXiv:1704.07415、 2017年。

卡尔莫里茨 · 赫尔曼，Tom´as Kocisk´y，爱德华 Grefenstette，Lasse Espeholt，将凯，穆斯塔法

苏莱曼 ·和菲尔 Blunsom。教学机器阅读和理解。在神经研究进展

信息处理系统 28︰ 年会上神经信息处理

系统 2015 年，页 1693年 — — 1701，2015年。

费利克斯 · 希尔、 安东尼 Bordes、 提交 Chopra 和杰森 · 韦斯顿。金凤花姑娘原则︰ 阅读

儿童书与外显记忆的表示法。在的国际诉讼

学习表示会议，2016 年。

Sepp Hochreiter 和 J¨urgen 施密。长期和短期记忆。神经计算，9(8):

1735 — — 1780 年，1997年。

鲁道夫 · 卡德尔克、 马丁 · 施密德、 Ondrej Bajgar 和 Jan Kleindienst。与语篇理解

注意总和阅读器网络。在协会第 54 次年度会议纪要

计算语言学，2016 年。

肯顿李、 汤姆 Kwiatkowski、 Ankur • 帕里克和 Dipanjan Das。学习经常性跨度申述

为采掘问答。arXiv 预印本 arXiv:1611.01436，2016 年。

王凌，尤利娅 Tsvetkov，西尔维奥 · 埃米尔 · 拉蒙 Fermandez，克里斯戴尔，艾伦 · 黑，伊莎贝尔

萨尔瓦多和褚澄林。并不是所有的上下文中平等的︰ 更好的字眼表示法

与变量的注意。在中自然的实证方法 2015年会议论文集

语言处理，EMNLP 到 2015 年，里斯本，葡萄牙，2015 年 9 月 17-21 日、 2015年。

瑞刘、 胡俊杰、 魏魏、 子阳、 埃里克 · 伯格。结构性嵌入的句法树

机器理解。arXiv 预印本 arXiv:1703.00572、 2017年。

克里斯托弗 · D 曼宁，米哈伊苏尔代亚努，约翰 · 鲍尔，珍妮玫瑰芬，史蒂文 Bethard 和

大卫 McClosky。斯坦福大学 corenlp 自然语言处理工具包。在 ACL (系统

示威)，第 55-60，2014年。

托马斯 Mikolov 马丁 Karafi´at、 卢克斯策划、 Jan Cernock'y、 桑吉 Khudanpur。复发性

基于神经网络语言模型。在 Interspeech，2010年。

三阮，和平号空间站罗森博格，夏松，剑锋高，Saurabh Tiwary，Rangan Majumder 和

李邓小平。MS 马可︰ 阅读理解数据集生成的人机。科尔,

abs/1611.09268，2016年。

Ankur 体育帕里克、 奥斯卡 T¨ackstr¨om、 Dipanjan Das 和 Jakob Uszkoreit。可分解的注意

自然语言推理的模型。在实证方法 2016年会议论文集

在自然语言处理，EMNLP 到 2016 年，美国德克萨斯州奥斯汀，2016 年 11 月 1 4 日,

2016。

杰弗里 · 彭宁顿，理查德 Socher 和克里斯托弗 · D.曼宁。手套︰ 全球矢量字

表示。在中自然语言的实证方法 2014年会议论文集

处理，EMNLP 2014 年，2014 年 10 月 25-29 日，多哈，卡塔尔，SIGDAT，会议

ACL 的特别兴趣组，页 1532年 — — 1543 年，2014年。

普拉 Rajpurkar、 张謇、 康斯坦丁 · Lopyrev 和珀西梁。阵容︰ 100000 + 问题

机器理解的文本。在中的实证方法会议论文集

自然语言处理，2016 年。

马修 • 理查森、 克里斯托弗 · J.C.伯吉斯和艾琳 Renshaw。Mctest︰ 挑战数据集

为开放域机器理解的文本。在 2013年会议论文集

实证方法在自然语言处理，第 193-203，2013年。

10

Tim Rockt¨aschel、 爱德华 Grefenstette、 卡尔 · 莫里茨 · 赫尔曼，Tom´as Kocisk´y 和菲尔 Blunsom。

关于蕴涵与神经注意推理。科尔、 2015年。

Minjoon Seo、 Aniruddha Kembhavi、 阿里迪和 Hannaneh Hajishirzi。双向关注

流的机器理解。arXiv 预印本 arXiv:1611.01603，2016 年。

柏森黄总经理覃业龙沈剑峰高，andWeizhu 陈。Reasonet︰ 学习停止在读

机器理解。在认知计算研讨会论文集︰ 整合

神经和象征性的办法 2016 30 年会上神经与共存

信息处理系统 （NIPS 2016 年），2016 年 12 月 9 日在西班牙巴塞罗那。，2016 年。

亚历山德罗 Sordoni、 菲利普 · 巴克曼和 Yoshua Bengio。迭代交替神经注意事项

机器阅读。科尔，abs/1606.02245，2016年。

尼蒂什斯里瓦斯塔瓦、 杰弗里 E.辛、 亚历克斯 Krizhevsky，伊利亚 · Sutskever 和鲁斯兰 Salakhutdinov。

辍学︰ 简单的方法，以防止神经网络拟合。杂志的机器学习

研究，2014年。

特里施勒亚当、 郑叶、 兴迪元、 Kaheer Suleman。自然语言理解

与 epireader。在中自然语言的实证方法会议论文集

处理，2016 年。

Oriol Vinyals、 老君要梅若福尔和 Navdeep Jaitly。指针网络。在神经信息研究进展

处理系统 28︰ 年会上神经信息处理系统

到 2015 年，2015 年 12 月 7-12，蒙特利尔，魁北克加拿大，页 2692年 — — 2700，2015年。

Shuohang 王和荆江。学习自然语言推理与 LSTM。在 NAACL

HLT 2016，2016年会议协会北美章的计算

语言学︰ 人类语言技术，美国圣地亚哥加利福尼亚 San，6 月 12-17，

2016 年，2016a。

Shuohang 王和荆江。使用匹配 lstm 和答案指针的机器理解。

arXiv 预印本 arXiv:1608.07905，2016b。

WenhuiWang、 FuruWei、 宝宝昌、 南阳市明洲。门控自匹配网络

为阅读理解和回答问题。在计算语言学协会

(ACL)、 2017年。

王治国、 海涛 Mi、 Wael 喉塞音符和拉杜弗洛里安。多视角的上下文匹配

机器理解。arXiv 预印本 arXiv:1612.04211，2016 年。

DirkWeissenborn、 GeorgWiese 和劳拉 Seiffe。Fastqa: 简单高效神经体系结构

为问答。arXiv 预印本 arXiv:1703.04816、 2017年。

菜名熊、 维克托 · 钟和理查德 Socher。问题的动态 coattention 网络

回答。arXiv 预印本 arXiv:1611.01604，2016 年。

义阳、 文头益和克里斯托弗 · 温顺。Wikiqa: 挑战数据集为开域问题的

回答。在的 EMNLP 程序，页 2013年-2018年。Citeseer，2015年。

智林阳，Bhuwan Dhingra，叶元、 胡俊杰、 威廉 · 科恩、 鲁斯兰 Salakhutdinov。

字或字符？ 细粒度浇注的阅读理解。科尔abs/1611.01724，

2016。

于洋、 张伟、 少将哈桑，墨玉，Bing 乡和宝荣周。端到端阅读

理解与动态回答块排名。arXiv 预印本 arXiv:1610.09996，2016 年。

马太福音 》 D.Zeiler。ADADELTA: 自适应学习速率方法。 CoRR，abs/1212.5701，2012年。

Junbei 张小丹朱、 前陈、 立荣戴、 江辉。值得探讨的问题

的理解和适应中基于神经网络的问题的回答。arXiv 预印本

arXiv:1703.04617、 2017年。

庆余周、 南阳、 腐乳圩、 传奇谭、 Hangbo 宝明洲。神经问题

从文本生成︰ 初步的研究。arXiv 预印本 arXiv:1704.01792、 2017年。

11