

# 基于 Capsule Network 的数学简答题自动反馈

姚淑佳<sup>1</sup> 谭红叶<sup>1</sup> 李茹<sup>1</sup> 卢宇<sup>2,3</sup> 段庆龙<sup>2</sup>

<sup>1</sup>(山西大学计算机与信息技术学院 山西 太原 030006)

<sup>2</sup>(北京师范大学未来教育高精尖创新中心 北京 100875)

<sup>3</sup>(北京师范大学教育技术学院 北京 100875)

**摘要** 自动反馈是指通过参考答案、教师意见等自动给出学生答案的修改意见。数学简答题答案具有类型有限和易混淆的特点,对此,提出一种自动反馈的方法。使用基于静态路由的 Capsule network 预测答案类型,根据答案类型通过相似度计算确定反馈内容。构建一个数学简答题自动反馈数据集并据此进行实验。实验结果表明,基于静态路由的 Capsule network 在确定答案类型任务上表现更好,相对于 CNN 有 3 百分点至 8.3 百分点的提升,相对于基于动态路由的 Capsule network 有 0.3 百分点至 2.5 百分点的提升,总体上能达到 85.3% 以上的准确率。

**关键词** 自动批阅 自动反馈 文本分类 胶囊网络

中图分类号 TP3 文献标志码 A DOI: 10.3969/j.issn.1000-386x.2021.09.005

## AUTOMATIC FEEDBACK OF MATHEMATICAL SHORT-ANSWER QUESTIONS BASED ON CAPSULE NETWORK

Yao Shujia<sup>1</sup> Tan Hongye<sup>1</sup> Li Ru<sup>1</sup> Lu Yu<sup>2,3</sup> Duan Qinglong<sup>2</sup>

<sup>1</sup>(School of Computer & Information Technology, Shanxi University, Taiyuan 030006, Shanxi, China)

<sup>2</sup>(Advanced Innovation Center for Future Education, Beijing Normal University, Beijing 100875, China)

<sup>3</sup>(School of Education Technology, Beijing Normal University, Beijing 100875, China)

**Abstract** Automatic feedback refers to the modification of students' answers through reference answers, teachers' opinions, etc. The students' answers to the short answer questions of mathematics have the characteristics of limited types and easy to be confused. In view of the above characteristics, a method to determine the feedback content is proposed. The Capsule network based on static routing was used to predict the answer type, and then the feedback content was determined through similarity calculation according to the answer type. An automatic feedback data set of mathematical short answer questions was constructed and the experiment was carried out accordingly. The experimental results show that the Capsule network based on static routing performs better in determining the answer type tasks, with an improvement of 3 percentage points to 8.3 percentage points compared with CNN, and 0.3 percentage points to 2.5 percentage points compared with the Capsule network based on dynamic routing, with an overall accuracy of more than 83.3%.

**Keywords** Automatic marking Automatic feedback Text classification Capsule network

## 0 引言

自动批阅 (Automatic Marking) 是指根据参考答

案、题目考察点等对学生答案进行自动评分和自动反馈<sup>[1]</sup>,其中自动评分 (Automatic Grading)<sup>[2]</sup>是指根据专家提供的参考答案预测学生答案的分数,目前已经有了较为成熟的研究成果并投入使用。但是在应用中

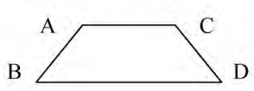
发现,同一个得分错误类型可能不同。例如:同样都是0分,可能是与题目完全无关的答案,也可能是看出了考察点但并没有回答正确。对于这种情况,需要给出更为具体的信息来辅助学习,因此需要提供反馈。

反馈本身是一件很耗费时间的工作,教师在进行反馈的时候需要先判断学生的错误属于什么类型,再根据不同的类型给出不同的意见和建议。对于老师而言,通常需要对几百甚至上千份学生答案进行反馈,耗时耗力,而且存在学生答案错误类别相似,但仍然需要重复给出相似反馈的情况。对于学生而言,由于老师反馈所需要的时间较长,因此学生并不能收到即时反馈<sup>[3]</sup>。自动反馈<sup>[4]</sup>(Automatic Feedback)为了有效减少老师的工作量,提供即时反馈,提高学生的学习效率。但是不同科目、不同题型的反馈重点不同,例如:英文一般关注时态、短语搭配等;语文一般更注重语义一致、语句通顺等问题。

本文主要研究数学简答题自动反馈,数学简答题是指有阅读材料描述,答案为数学概念的题目,图1所示为一个题目和答案举例。

题目

在下面横线上填上推理依据:



如图 $\angle A + \angle B = 180^\circ$ , 求证 $\angle C + \angle D = 180^\circ$

证明:  $\because \angle A + \angle B = 180^\circ$

$\therefore AC \parallel BD$  ( ) (1)

$\therefore \angle C + \angle D = 180^\circ$  ( ) (2)

答案

(1) 同旁内角互补, 两直线平行

(2) 两直线平行, 同旁内角互补

图1 数学简答题题目及答案举例

本文构建的数据集具有以下特点:(1)数据量较小。本文所使用的数据来源于真实考试中的学生答案,因此数据量有限,每道题大概有4000条左右的数据并且有一定的重复率。(2)答案的类型有限。通过对学生答案的分析将其分为完全正确、概念描述错误、概念不适合这道题目和答案完全错误四类。(3)答案类型易混淆。例如“同旁内角互余,两直线平行”和“同旁内角互补,两直线平行”只差一个字,但分别属于上述最后两类答案,传统的分类模型很难正确地将其进行分类。此外,在进行自动反馈之前,本文还为每个类型的学生答案定义了一个反馈模板,但是为了使模板具有针对性,模板中的一部分内容是需要通过相似度计算进行填充的,具体如表1所示。

表1 答案类型及反馈模板

分数	类型	示例	反馈模板
0分	完全无关	360度	请认真读题,回答相关内容。本题的考察点是“同旁内角性质”
	概念错误	同旁内角相等两直线平行	你对于“同旁内角性质”知识点掌握不足,请重新学习“同旁内角性质”相关知识。并且思考本题的考察点
	概念不适于本题	垂直于同一直线的两直线平行	你回答的是“垂线段的公理”,本题要考察的是“同旁内角的性质”,你没有理解本文的考察点
1分	回答正确	同旁内角互补两直线平行	你的答案完全正确,再接再厉,继续加油

注:引号内填充的为概念属性

本文提出一种分两步进行的数学简答题自动反馈方法,第一步确定学生答案类型,第二步通过相似度计算填充对应的反馈模板。本文使用基于静态路由的胶囊网络用于确定答案类型。Capsule network是2017年Hunter提出的一种方法,用神经元向量代替传统神经网络的单个神经元节点,以动态路由的方式去训练这种全新的神经网络。与传统的卷积神经网络相比,其具有以下特点:(1)传统分类模型输出类别,但是Capsule network输出的是向量;(2)CNN要得到比较好的结果需要大量的训练数据,而Capsule network因为在胶囊内部进行特征学习,因此较少的数据就能得到更好的结果;(3)CNN池化会丢失一些信息,而每个capsule的都会携带大量的信息;(4)capsule间的传递和运算更符合人脑的思考形式。本文将胶囊层间的连接方式改为静态路由,与动态路由相比,降低了计算的复杂度,提高结果精度,更适合文本分类。

确定答案类型之后,每个类型对应一个反馈模板,但是反馈模板中的一部分内容是需要填充的。因此通过相似度计算对模板进行填充。通过上述方法在已构建数据集上可以达到83.3%以上的正确率。

## 1 相关工作

Page<sup>[5]</sup>早在1966年就已经开始进行自动批阅相关研究,但早期的自动批阅主要是使用规则的方法为学生提供一个合理的分数。如Bachman等<sup>[6]</sup>提出WebLAS方法,根据答案自动生成正则表达式,每个正则表达式与一个分数相关联,因此当学生答案与表达式相匹配就会给出该答案的分数。但是基于规则的方法由于规则表达的有限性,泛化能力较差,随着机器学

习的发展,人们开始用文本分类或回归模型来预测分数。Sultan等<sup>[7]</sup>使用文本相似度、词项权重等特征构建了随机森林分类器,在SemEval-2013评测数据集SCIENTSBANK上进行测试,F1值达到55%。近年来,由于深度学习方法在许多任务上已经达到了最先进的水平,研究人员开始使用词向量和神经网络进行评分研究。如Riordan等<sup>[8]</sup>使用CNN-LSTM构成的神经网络进行自动评分,获得的效果比非神经网络方法好。随着评分技术的日渐成熟,人们发现如果只是给出分数并不能让学生很清楚地认识到自己的不足之处,无法达到智能批阅的真正目的,因此人们开始尝试研究自动反馈。

在自动反馈研究初期,Marvaniya等<sup>[9]</sup>提出了一种针对于英语填空的自动反馈方法。对于每个空,由老师提前预测每个空的所有可能错误答案,例如动词时态变化、名词单复数形式的变化、短语的固定搭配等。对于预测的每一个答案教师都给出了一个反馈,通过计算学生答案和所有预测答案的相似度,找出和学生答案相似度最高的预测答案,将相对应的反馈提供给学生。通过该方法有效地减少了老师对相似答案重复提供类似反馈的工作量。但是并不是所有题目都可以预测到所有错误类型,而且这项工作需要老师有丰富的教学经验。因此,Brooks等<sup>[10]</sup>对于从慕课网上收集的学生答案进行度量聚类,老师对于每一类聚类结果给出反馈,使得老师的工作量由几百甚至几千降到几十。但是聚类方法的准确性不高。Zhang等<sup>[11]</sup>提出了一种基于深度学习+聚类+学生建模的方法,追踪学生的学习轨迹为学生提供反馈。这个方法需要大量的学生学习数据,并不是很容易实现。同时,Dzikovska等<sup>[12]</sup>提出将答案按照错误类型分类而不是分数,将答

案分为正确、部分正确、非领域、不相关、矛盾五类,为本文的反馈提供了思路。

文献[13]发表后,关于胶囊网络的研究受到了许多关注。Afshar等<sup>[14]</sup>将胶囊网络用于肿瘤图片的分类,通过改变capsule的特征层数来提高精度。Du等<sup>[15]</sup>提出一种新的基于胶囊的混合神经网络用于情感分类。阿里研究发现胶囊网络应用到投诉文本模型上的整体表现优于原先TextCNN<sup>[16]</sup>。Zhao等<sup>[17]</sup>将基于动态路由的胶囊网络用于文本分类任务,在六个经典文本分类数据集上进行实验,其中四个都取得了当前最好的结果。

## 2 方法设计

本文提出一种分两步进行的数学简答题自动反馈方法。第一步是对学生答案类型进行建模,给定学生答案,通过基于静态路由的胶囊网络判断学生答案类型。该模型的最大特点是以胶囊的形式进行特征传递,能更好地保留文本内部信息,在小数据集上得到精度更高的分类结果。第二步是通过计算学生答案和正确概念的相似度确定概念属性,将概念属性填充到反馈模版中。这一步的优点是可以较为准确地预测答案类型,同时提供的反馈与学生答案紧密相关。

### 2.1 基于静态路由的胶囊网络确定答案类型

本文采用一种基于静态路由的胶囊网络确定答案类型,模型架构示意图如图2所示,由输入层、N-gram卷积层、初级胶囊层、卷积胶囊层、全连接胶囊层、输出层六层构成。胶囊能够表示局部实体的属性,并用向量而不是标量来表达语义。

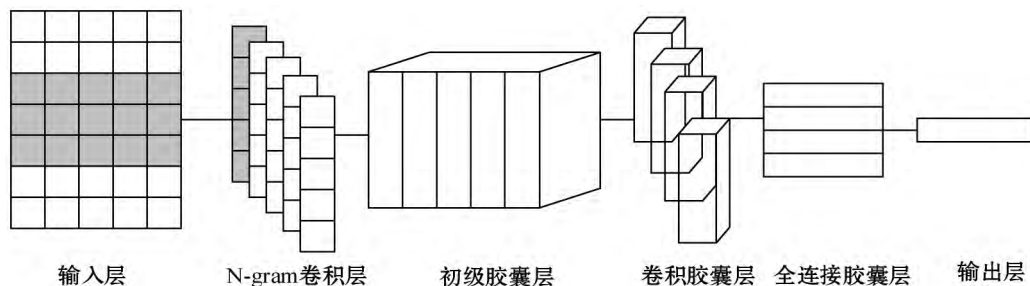


图2 胶囊网络模型架构示意图

模型的输入层是学生答案  $S \in \mathbf{R}^{L \times V}$ ,  $L$  代表句子长度,  $V$  代表词向量维度。

第二层是 N-gram 卷积层。在这一层用  $B$  个大小相同、权重不同的滤波器  $W^a \in \mathbf{R}^{K \times V}$  抽取低级特征,每个滤波器都会生成一个神经元  $m^a \in \mathbf{R}^{L-K+1}$ ,这个神经元中的每一个元素按式(1)进行计算。

$$m_i^a = f\left(\sum W^a * X_{i:i+K-1} + b\right) \quad (1)$$

式中:  $X_{i:i+K-1}$  代表第  $i$  到  $i+K-1$  个词的词向量;  $W^a$  为宽度  $K$ 、长度为  $V$  的滤波器,目的是在不同位置抽取特征;  $*$  代表单元乘法;  $f$  为激活函数(本文采用 sigmoid 函数);  $b$  代表偏置。对于  $a=1, 2, \dots, B$ , 一共有  $B$  个拥有相同的 N-gram 尺寸的滤波器,可以生成  $B$  个特征

映射,如下所示:

$$M = [m_1 \ m_2 \ \dots \ m_B] \in \mathbf{R}^{(L-K+1) \times B} \quad (2)$$

下一层是初级胶囊层  $p_i \in \mathbf{R}^d$  代表一个胶囊  $d$  代表胶囊的大小。由  $m_i \in \mathbf{R}^B$  作为输入,  $W_b \in \mathbf{R}^{B \times d}$  是不同窗口共享的滤波器。窗口的滑动步幅为 1,则会产生  $C$  个  $(L-K+1) \times d$  的胶囊,每一个胶囊  $p_i$  的计算为:

$$p_i = g(W^b M_i + b_1) \quad (3)$$

式中:  $g$  是 Squashing 激活函数;  $b_1$  是胶囊偏置项。对于所有的  $C$  个滤波器,生成的胶囊特征映射可以重新排列:

$$P = [p_1 \ p_2 \ \dots \ p_B] \quad (4)$$

激活函数具体如下:

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \cdot \frac{s_j}{\|s_j\|} \quad (5)$$

激活函数  $g$  既保留了向量的大小,也将向量的模长控制在  $[0, 1)$  之间。

第四层卷积胶囊层,这一层会自动学习“部分-整体”关系,也就是由低级特征(子胶囊)  $u_i$  预测高级特征  $u_{j|i}$ (父胶囊),如式(6)所示。

$$u_{j|i} = W_j^c \cdot u_i + b_{j|i} \quad (6)$$

式中:  $b_{j|i}$  是胶囊偏置项;  $u_i$  是低级特征;  $W_j^c \in \mathbf{R}^{D \times d \times d}$  是一个共享权重矩阵。然后,本文使用不同的路由方式连接胶囊层。

下一层是全连接胶囊层,卷积胶囊层的胶囊被放入了一个胶囊列表中,并输入到全连接的胶囊层中,胶囊与胶囊乘以变换矩阵  $W^d \in \mathbf{R}^{E \times d \times d}$ ,通过路由协议产生最终的胶囊  $v_j \in \mathbf{R}^d$  及其概率  $a_j \in \mathbf{R}$ 。

接下来通过比较胶囊层间的不同的连接方式选出适合文本分类的连接方式。

### 2.1.1 动态路由

Sabour 等<sup>[13]</sup>指出,胶囊网络通过动态路由过程更新耦合系数  $c_{ij}$  的权重,并确定低级特征和高级特征之间的匹配程度。耦合系数由正确高级特征和预测的高级特征的相似程度决定,计算式如下:

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \quad (7)$$

式中:  $i \in [1, a]$ ;  $j \in [1, k]$ ;  $k$  代表总的类别数;  $b_{ij}$  是由子胶囊和父胶囊的相似程度决定的;  $c_{ij}$  在每次迭代过程都会根据  $b_{j|i}$  进行更新。具体如算法 1 所示。

#### 算法 1 动态路由算法

procedure ROUTING( $u_{j|i}$ ,  $a_{j|i}$ ,  $r$ ,  $l$ )

初始化耦合系数的 logits

$b_{j|i} = 0$

for  $r$  iterations do

for all capsule  $i$  in layer  $l$  and capsule  $j$  in layer  $l+1$

$$c_{j|i} = a_{j|i} \cdot \text{leaky-softmax}(b_{j|i})$$

for all capsule  $j$  in layer  $l+1$ :

$$v_j = g(\sum_i c_{j|i} u_{j|i}) \quad a_j = |v_j|$$

for all capsule  $i$  in layer  $l$  and capsule  $j$  in

$$\text{layer } l+1: b_{j|i} = b_{j|i} + u_{j|i} \cdot v_j$$

return  $v_j$ ,  $a_j$

### 2.1.2 静态路由

与基于动态路由的胶囊网络相比,静态路由主要有以下两个优势:(1)降低了计算复杂度;(2)提高计算精度。Zhao 等<sup>[17]</sup>首次将胶囊网络用于文本分类时采用的是动态路由的连接方式。这是直接借鉴 Hunter 在图像上的两个胶囊层之间的连接方式通过静态路由决定,在图像上考虑到实体间的空间结构是合理的,但是在语言中,文字的表达方式有很大不同,因此,本文提出静态路由的方式将子胶囊和父胶囊进行连接:

$$s_j = \sum_i w_{ij} h_i \quad (8)$$

### 2.2 自动反馈

确定答案类型后不同的类型有不同的模板,本文通过相似度计算的方法进行反馈模板填充。在进行反馈模板填充之前,本文人工构建了初中数学概念库,包含了初中阶段所有用到的数学概念 184 条。概念库中的概念以表 2 所示的形式存在。概念库构建完成之后,进行模板填充。本文的答案类型包括完全无关、概念错误、概念不适合这道题目、回答正确四类。通过字重叠确定相似度最高的概念,将对应的概念属性填充进模板,完成自动反馈。

表 2 概念库举例

概念属性	概念内容
矩形的性质	矩形的对角线平分且相等
矩形的性质	矩形的四个角都是直角
矩形判定定理	有一个角是直角的平行四边形叫做矩形
矩形判定定理	对角线相等的平行四边形是矩形
矩形判定定理	有三个角是直角的四边形是矩形
菱形的定义	邻边相等的平行四边形
菱形的性质	菱形的四条边都相等

## 3 实验

### 3.1 数据集及评价指标

本文所使用的数据集来源于某中学八年级数学试

卷的三道数学简答题,每个题目有两个空,每个空的答案数量分别为 3 632、3 632、4 290、4 290、4 913、4 913。由于学生答案有一定的重复和空白,因此对数据进行去重之后进行统计,数据集的具体信息如表 3 所示,其中 Math2 重复率最高,Math1 次之,Math3 最低。

表 3 数据集统计

题目	总数	去重	重复率/%
Math1_A	3 632	1 928	46.92
Math1_B	3 632	2 162	40.48
Math2_A	4 290	237	94.48
Math2_B	4 290	259	93.96
Math3_A	4 913	3 532	28.11
Math3_B	4 913	3 416	30.47

本文采用准确率  $P$  作为实验评价指标。假定  $a$  表示被正确分到某类的文档数,  $b$  表示错误分到该类的文档数。则准确率计算如下:

$$P = \frac{a}{a+b} \quad (9)$$

### 3.2 实验参数设置

在参数设置上,本文使用 Skip-gram 训练维度为 60 维的字向量和词向量,通过对词级别和字级别的句子长度进行统计发现(包括学生答案+参考案),词级别的最大长度为 15,字级别为 40。因此本文输入为  $15 \times 60$  和  $40 \times 60$  的矩阵,在 N-gram 层,采用 64 个 3-gram 的卷积层抽取低级特征。胶囊的维度  $d=6$ ,答案类别设为四类,共享滤波器的窗口维度为 2。

### 3.3 实验结果及分析

#### 3.3.1 实验结果

为了说明本文方法的有效性,本文分别采用传统机器学习模型 SVM 和深度学习模型 CNN 作为基线模型,同时和最近在各个任务上都表现优异的 Bert 进行了对比实验。

SVM<sup>[15]</sup> 是一个进行文本分类的经典模型,使用 N-gram 来提取输入特征,使用 SVM 作为分类器,得到答案类型的分类结果。

CNN<sup>[16]</sup> 是目前应用范围最广的神经网络模型之一,许多学者利用其解决多种 NLP 任务并获得很好效果。本文分别使用字级别和词级别的特征作为输入,得到答案类型的分类结果。

Bert<sup>[17]</sup> 是 2018 年 Google 提出的通用预训练语模型,目前在许多文本分类任务上都取得了比传统方法更好的结果。

表 4 列出了基于动态路由的 Capsule network(CPN

\_d) 和基于静态路由的 Capsule network(CPN\_s) 和上述模型的对比结果。

表 4 各模型准确率对比(%)

题目	SVM	CNN	Bert	CPN_d	CPN_s
Math1-A	80.3	83.1	89.4	90.2	91.4
Math1-B	79.5	85.1	90.1	90.5	91.8
Math2-A	89.3	93.5	96.2	95.1	96.5
Math2-B	86.2	90.1	94.5	94.9	95.2
Math3-A	73.1	78.6	82.1	82.8	85.3
Math3-B	74.2	79.1	83.7	84.6	86.5

可以看出,CPN\_s 在三道题上表现均为最优,Bert 和 CPN\_d 的实验结果在三道题上的结果都明显高于 SVM 和 CNN,但稍低于 CPN\_s。在 Math1 上两种胶囊网络的结果均优于 Bert,准确率达到了 91.4% 和 91.8%。在 Math2 上胶囊网络的准确率为 96.5% 和 95.2%,是三道题里最高的。可能是由于题目难度较低,学生答案重复率较高。在 Math3 上 Bert 的表现稍优于 CPN\_d。实验结果表明在文本分类任务上胶囊形式优于单个的神经元,且静态路由优于动态路由。

#### 3.3.2 数据去重

由表 3 对数据的统计发现数据的重复率较高,通过对题目的分析发现和题目难度呈正相关。本文为了证明模型的有效性在去重数据上重新进行了实验,实验参数与之前一致。实验结果如表 5 所示。

表 5 数据去重后各模型准确率(%)

题目	SVM	CNN	Bert	CPN_d	CPN_s
Math1-A	69.1	70.0	73.4	75.6	77.5
Math1-B	69.8	71.3	74.5	76.9	78.1
Math2-A	30.3	30.4	35.8	45.1	47.3
Math2-B	30.8	32.5	36.7	47.2	50.6
Math3-A	75.1	78.4	80.1	81.2	82.4
Math3-B	74.5	77.7	79.4	80.2	82.1

通过对去重后的实验结果进行分析发现,去重之后三道题的准确率均有大幅下降,在 Math2 上的准确率下降幅度最大,Math1 次之,Math3 最小,准确率与数据重复率成正相关。

#### 3.3.3 不同级别特征

为了检验不同级别的特征对实验结果的影响,本文使用字向量在 CNN 和基于静态路由的 Capsule network 上进行对比实验,具体如图 3 所示,其中:  $t$  (token) 代表字级别的特征,  $w$  代表词级别的特征。

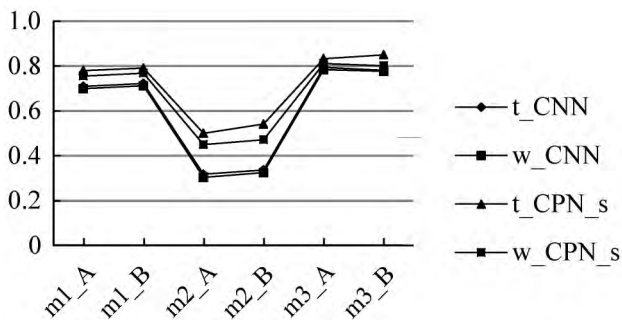


图3 不同级别特征在 CNN 和 CPN\_s 上的结果

通过对图3观察发现,不论是 CNN 还是 CPN\_s, 字级别的实验结果都优于词级别的实验结果。

## 4 结 语

本文首先构建了数学简答题数据库,然后提出将数学简答题的自动反馈分为两步进行,通过基于静态路由的 Capsule network 确定答案类别。为了证明该方法的有效性,与 SVM、CNN、Bert 进行了对比实验。确定答案类型后,不同的类型有不同的模板,本文通过相似度计算的方法进行反馈模板填充。但是该方法的反馈模板需要手动定义,答案类别也并没有细分,因此只能给出较为概括性的反馈。下一步计划通过构建数学概念知识图谱为学生提供更细致的反馈。

## 参 考 文 献

- [1] 郑庆华,董博,钱步月,等.智慧教育研究现状与发展趋势[J].计算机研究与发展,2019,56(1):209-224.
- [2] Madnani N, Loukina A, Cahill A. A large scale quantitative exploration of modeling strategies for content scoring [C]// Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications. ACL, 2017: 457-467.
- [3] 余胜泉.人工智能教师的未来角色[J].开放教育研究,2018,24(1):11-12.
- [4] Ashford S J, Cummings L L. Feedback as an individual resource: Personal strategies of creating information [J]. Organizational Behavior and Human Performance, 1983, 32(3): 370-398.
- [5] Page E B. The imminence of grading essays by computer [J]. Phi Delta Kappan, 1966, 47(5), 238-243.
- [6] Bachman L F, Carr N, Kamei G, et al. A reliable approach to automatic assessment of short answer free responses [C]// 19th International Conference on Computational Linguistics. DBLP, 2002.
- [7] Sultan M A, Salazar C, Sumner T. Fast and easy short answer grading with high accuracy [C]//The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. ACL, 2016: 1070-1075.
- [8] Riordan B, Horbach A, Cahill A, et al. Investigating neural architectures for short answer scoring [C]//Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications. ACL, 2017: 159-168.
- [9] Marvaniya S, Saha S, Dhamecha T I, et al. Creating scoring rubric from representative student answers for improved short answer grading [C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. ACM, 2018: 993-1002.
- [10] Brooks M, Basu S, Jacobs C, et al. Divide and correct: Using clusters to grade short answers at scale [C]//2014 1st ACM Conference on Learning @ Scale. ACM, 2014: 89-98.
- [11] Zhang Y, Shah R, Chi M. Deep learning + Student modeling + clustering: A recipe for effective automatic short answer grading [C]//Proceedings of the 9th International Conference on Educational Data Mining. IEDMS, 2016: 562-567.
- [12] Dzikovska M O, Nielsen R D, Brew C. Towards effective tutorial feedback for explanation questions: A dataset and baselines [C]//Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. ACL, 2012: 200-210.
- [13] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules [C]//Annual Conference on Neural Information Processing Systems, 2017: 3856-3866.
- [14] Afshar P, Mohammadi A, Plataniotis K N, et al. Brain tumor type classification via capsule networks [C]//2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018: 3129-3133.
- [15] Du Y P, Zhao X Z, He M, et al. A novel capsule based hybrid neural network for sentiment classification [J]. IEEE Access, 2019, 7: 39321-39328.
- [16] Kim Y. Convolutional neural networks for sentence classification [EB]. arXiv: 1408.5882, 2014.
- [17] Zhao W, Ye J B, Yang M, et al. Investigating capsule networks with dynamic routing for text classification [EB]. arXiv: 1804.00538, 2018.
- [18] 薛亮.基于 SVM 的中文文本分类系统的设计与实现 [D].重庆:重庆大学.
- [19] Dumoulin V, Visin F. A guide to convolution arithmetic for deep learning [EB]. arXiv: 1603.07285, 2016.
- [20] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [EB]. arXiv: 1810.04805v2, 2018.