基于增强数值表示的数学应用题求解模型研究

胡静丹，李波，杨静
（华中师范大学计算机学院，湖北 武汉 430079）

摘　要：数学应用题（MWP）的计算机自动求解是当前学术界的研究热点。尽管现有研究已取得显著进展，但多数研究将 MWP 中的数值替换为占位符并作为普通文本进行处理，忽略了数值语义对于 MWP 求解的重要性。因此，基于“编码器-解码器”的通用架构提出了一种增强数值表示的 MWP 求解模型。该模型引入了两项新颖的设计来增强数值编码能力并提升 MWP 的求解性能，利用图卷积神经网络（GCNN）来显示建模数值之间以及数值与上下文文本之间的语义关系；引入辅助学习任务来指导模型充分捕捉与任务相关的数值语义，显著增强编码器的数值建模能力。在常用的 MWP 数据集 Math23K 和 MAWPS 上的实验结果表明，所提出的模型充分利用到了数值语义，模型整体性能优于现有的主流 MWP 求解模型。

关键词：数学应用题；图卷积神经网络；辅助学习任务；图编码层；图变换层
中图分类号：TP391　　**文献标志码：**A　　**DOI：**10.19678/j.issn.1000-3428.0070131

Research on Math Word Problem Solving Model Based on Enhanced Numerical Representation

HU Jingdan, LI Bo, YANG Jing
(School of Computer Science, Central China Normal University, Wuhan 430079, Hubei, China)

【Abstract】The automatic resolution of Math Word Problems (MWP) is a current research hotspot in the academic community. Despite the significant progress made in the existing research, most studies treat numerical values in MWPs as placeholders and simply process them as ordinary text, overlooking the importance of numerical semantics in solving MWPs. To address this issue, this study proposes an enhanced numerical representation model for MWP solving based on the generic "encoder-decoder" architecture. The model achieves this by utilizing Graph Convolutional Neural Networks (GCNN) to explicitly model the semantic relationships between numerical values and between numerical values and context text and introducing auxiliary learning tasks to guide the model to fully capture task-related numerical semantics. This significantly enhances the numerical modeling capability of the encoder. Empirical evidence from the commonly used MWP datasets, Math23K and MAWPS, shows that the proposed model can fully consider numerical semantics and outperform mainstream classical models in solving large-scale Chinese application problem sets.

【Key words】Math Word Problems (MWP); Graph Convolutional Neural Networks (GCNN); auxiliary learning tasks; graph encoding layer; graph transformation layer

0　引言

数学应用题（MWP）是以自然语言形式呈现的一类数学问题。准确求解这些问题需要良好的自然语言理解能力和逻辑推理能力。由于数学应用题的计算机自动求解在教育、金融等领域具有重要的应用价值，因而引起了学术界和工业界的广泛关注。该领域内的早期研究主要通过基于规则或传统机器学习的方法对 MWP 进行建模和求解，受限于规则制定或特征工程的高昂代价以及较差的灵活性，性能不够理想[1]。

近年来，深度学习方法通过自动学习文本和数值表征为 MWP 的自动求解提供了有力的支撑。WANG 等[2]提出了基于 Seq2Seq 框架的方法，模型采用循环神经网络学习数学问题文本特征，并以端到端的方式生成模板求解公式。HUANG 等[3]将 Seq2Seq 框架和强化学习方法相结合以优化求解准确性，克服了训练和测试过程中最大似然估计的差异性问题。WANG 等[4]基于递归神经网络和模板方法，解决了生成的目标表达式范围过大时会出现性能下降的问题。LI 等[5]提出了组合注意力方法，以增强 Seq2Seq 框架模型汇总各类

基金项目：教育部人文社会科学研究青年基金（19YJC870012）。
作者简介：胡静丹，女，硕士，主研方向为自然语言处理；李波（通信作者），副教授、博士；杨静，硕士。
收稿日期：2024-07-16　　**修回日期：**2024-10-15　**E-mail：**liboaoe@foxmail.com

计算机工程 www.ecice06.com

数学应用题的能力。上述这些基于 Seq2Seq 框架的方法可以自主学习数学问题的上下文语义并顺序生成正确的目标结果，但它们难以充分理解数学问题的推理过程和逻辑。为了解决这一问题，XIE 等[6]引入目标驱动机制，提出了基于 Seq2Tree 框架的方法，通过使用树状结构解码器以自顶向下或自底向上的方式生成表达式，进而提升模型对数学推理的感知能力。QIN 等[7]基于 Seq2Tree 框架考虑到问题和表达式中的数值特征，增强了模型对数值的理解能力。WU 等[8]基于 Seq2Tree 框架提取问题中的数值进行单独编码并融合的方式来增强数值表征。后续的一些工作还结合外部知识，如句法依赖性[9]和常识性知识[10]，对 Seq2Tree 框架模型进一步改进。此外，不同于上述两类架构，ZHANG 等[11]提出了基于 Graph2Tree 框架的模型，将数学问题建模为图结构，充分利用了数学问题本身的结构化信息，捕捉问题中的数值文本关联性特征和数值比较特征来改进问题表示，LIANG 等[12]利用问题中的语义和数量的组合模态编码文本，提高了模型对问题的理解能力。近年来，借助大语言模型[13]求解数学应用题的研究受到广泛关注[14]，其强大的上下文学习能力使得推理时仅提供少量数学应用题任务示例作为条件，就能够执行目标任务，而无需更新模型参数。但有研究表明[15]，LLM 在测试时间示例之前给出输入输出对的上下文示例，不足以证明在数学推理等具有挑战性的任务上实现高性能。另一方面，使用思维链提示[16]信息指导 LLM 进行数学推理，这种类型的少量学习在不同的上下文示例选择中可能是高度不稳定的[17-18]，即无法确定哪些语境推理实例是最有效的提示语。

尽管上述这些求解数学应用题的模型具有良好的自然语言理解能力和离散数值推理能力。然而，在考虑数值属性特征和上下文提示信息时大多只关注数值本身，忽略了数值之间、数值和上下文文本之间以及上下文提示信息之间包含的多重关联信息，导致模型无法有效利用问题中的关键提示信息进行良好的数值推理。其次，这些方法均采用了数值映射技术，将数值替换为占位符当作普通文本进行处理，侧重于获得有效的语义表征、对齐语义空间（文本描述）和符号空间（符号解决方案），而对 MWP 中的数值属性特征挖掘不充分。例如，若数学应用题“甲、乙两列火车从相距 1 050 km 的两地同时相对开出，甲车每小时行驶 80 km，经过 2.8 h 后，两车相距全程的 60%。求乙车每小时行驶多少千米”中的数值 60%变化为 630 km 时，最终的表达式将由“$x=1\ 050\times(1-60\%)/2.8-80$”变为“$x=(1\ 050-630-2.8\times80)/2.8$”。这种由数值属性特征的改变所导致的表达式结果变化，可能会增加模型对题意理解和推理的难度。

为了解决以上两个问题，本文基于“编码器-解码器”架构并利用图卷积神经网络和辅助学习任务来增强数值表征。本文的主要贡献如下：

1）在 graph-to-tree 架构基础上引入多模态关系的图结构表示。这些增强的图结构通过充分利用问题本身的信息，学习数值之间、数值与上下文文本之间以及上下文提示信息之间包含的多重信息，获取优良的特征。

2）将 graph-to-tree 架构与 4 个数值相关的监督任务相结合。利用辅助任务提供的额外训练信号引导模型捕捉数值语义，从而改善数值表征。

3）在常用的 MWP 数据集 Math23K 和 MAWPS 上的实验结果证明，本文所提出的模型能充分考虑到数学应用题中的数值文本语义，模型组件具有泛化性，能有效提升各种基线模型求解数学应用题的准确率。

1 问题描述

数学应用题是由一组包含词和数值的序列构成的，记为 $X=\{x_1,x_2,\cdots,x_m\}$，其中，x_i 表示问题中的一个词或数值，且 $1\leqslant i\leqslant m$，m 为问题序列的长度。为便于描述，设数学应用题中的词序列 W_x 可表示为$\{w_1,w_2,\cdots,w_q\}$，其中 $1\leqslant q\leqslant m$，并且数值序列 N_x 表示为$\{n_1,n_2,\cdots,n_s\}$，其中，$1\leqslant s\leqslant m$，n_l 表示与词汇 w_j 相关的数值。在此情境中，有 $m=q+s$，其中，q 为问题中词的数量，s 为数值的数量。此外，本文将求解数学应用题需要的常量集合记为 $C=\{c_1,c_2,\cdots,c_d\}$，运算符集合定义为 $O=\{+,-,*,/,\hat{}\}$，其中，c_k 表示一个常量符号，如圆周率 π，且 $1\leqslant k\leqslant d$，d 表示常量总数。

求解 MWP 的最终目标是将 X 映射到一个有效的前缀表达式 Y 并计算结果。Y 中通常包含数值 $n_l\in N_x$、常量 $c_k\in C$ 以及运算符 $o\in O$。

2 模型

本文的模型框架如图 1 所示。模型由 3 个部分构成：图编码器，辅助学习任务和树结构解码器。在图编码器部分，首先在嵌入向量层使用 BiGRU 对 MWP 文本进行编码获得节点向量表示，然后连同图构造层中的 4 种图一起输入到图变换层中学习得

计算机工程 www.ecice06.com

到全局图表示。在树结构解码器部分输入全局图表示并生成最终的前缀表达式。辅助学习任务包含 4 个数值相关的监督任务，这些任务最终与主框架一起训练，引导模型捕捉数值语义。

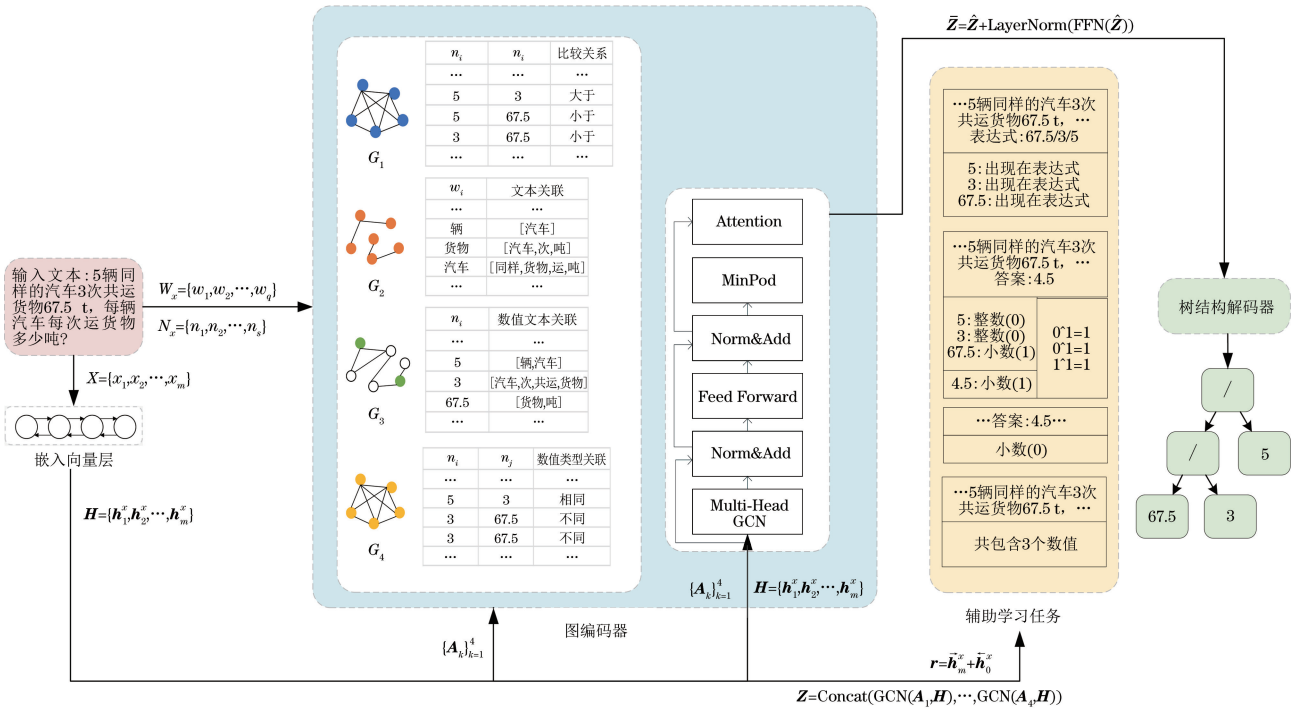


图 1 模型结构

Fig. 1 Model structure

2.1 向量层嵌入

首先使用两层的双向门控循环单元(GRU)[19]将数学应用题 X 编码成一个隐藏状态表示 $\boldsymbol{H}=(\boldsymbol{h}_1^x,\boldsymbol{h}_2^x,\cdots,\boldsymbol{h}_m^x)$，如式(1)所示：

$$\boldsymbol{h}_i^x=\mathrm{BiGRU}(\boldsymbol{E}(\boldsymbol{x}_i),\boldsymbol{h}_{i-1}^x) \tag{1}$$

式中：词嵌入向量 $\boldsymbol{E}(\boldsymbol{x}_i)$ 由词嵌入层 $E(\cdot)$ 得到，词嵌入层 $E(\cdot)$ 为外部预训练词向量[20]初始化权重参数；$\boldsymbol{h}_i^x$ 表示第 i 个时间时刻的隐藏状态。

2.2 图编码层

基于 Graph Transformer 的图编码层主要由图构造层和图变换层两部分构成。

2.2.1 图构造层

在构造图结构前，首先抽取 MWP 中包含数值的实体，并根据这些实体所在的句子，结合词性标注方法，找到对应关联的名词、形容词和动词，最后将所有词语作为属性词列表$\{w_1,w_2,\cdots,w_q\}$。通过这些数值、属性词，模型可以显示地将数值与其相关联的属性词语以及不同数值之间连接起来的方式，间接地在与数值相关的属性词语之间建立联系。以此为基础，构建的 4 种不同模态关系的图结构如下：

1)数值比较图 G_1：如果颠倒减法和除法操作符前后的操作数，运算表示的含义则完全相反。因此，显性地构造比较数值大小的图结构对模型学习操作符前后操作数的位置顺序是必要的。对数学应用题 X 中的任意两个数量 n_i 和 n_j，若 $n_i>n_j$，则将有向边 e_{ij} 添加到图 G_1 中。

2)文本关联图 G_2：数值比较图可以让模型学习操作数之间的位置顺序，但是无法关注文本中有用的描述性词语信息，如在图 1 中，类似“吨、货物”这些属性词能够提供单位、对象信息，这些信息的缺失可能会使模型生成与实际含义完全不同的错误表达式。因此，显性地构造属性词关联图结构有利于模型关注上下文，对于数学应用题 X 中的任意两个属性词 w_i 和 w_j，将 w_i 和 w_j 之间的无向边 w_{ij} 添加到图 G_2 中。

3)数值文本关联图 G_3：数值比较图和文本关联图可以让模型分别关注数值信息和描述性词语信息，但是无法在数值及其属性词之间建立联系。因此，显性地构造数值文本关联图结构对模型联合学习数值和词语两种模态的信息是有必要的。将数学应用题 X 中的每一个数值 n_i 和属性词 w_j 以及任意两个数值 n_i 和 n_j 对应位置处的无向边 f_{ij} 均添加到图 G_3 中。

4)数值类型图 G_4：由于数学运算需要计算出准确的数值结果，而不同的数据类型通常能够表示的数值范围、精度和运算规则存在差异，因此显性地构

造比较数值的数据类型图结构存在必要。对数学应用题 X 中的任意两个数量 n_i 和 n_j，若 n_i 和 n_j 的数据类型相同，则将无向边 g_{ij} 添加到图 G_4 中。

对于每一种图，首先初始化邻接矩阵 $\boldsymbol{A}\in\mathbb{R}^{m\times m}$。如果在第 i 个节点和第 j 个节点之间存在一条边，则需要为这条边的邻接矩阵$(i,j,\boldsymbol{A}_{ij})$对应的位置赋值 1；否则，赋值 0。因此，上述 4 种计算图 $\{G_k\}_{k=1}^4$ 的邻接矩阵为$\{\boldsymbol{A}_k\}_{k=1}^4$。

2.2.2 图变换层

图变换层的输入是由隐藏状态 $\boldsymbol{H}=(\boldsymbol{h}_1^x,\boldsymbol{h}_2^x,\cdots,\boldsymbol{h}_m^x)$和多个图的邻接矩阵$\{\boldsymbol{A}_k\}_{k=1}^4$ 组成的。具体地，首先利用图卷积神经网络(GCN)[21]学习图节点特征，对于输入的多张图采用带有 K 头的图卷积设置，GCN 的特征学习表示为：

$$G_k^{\mathrm{GCN}}(\boldsymbol{A}_k,\boldsymbol{H})=\mathrm{Softmax}(\boldsymbol{A}_k\mathrm{ReLU}(\boldsymbol{A}_k\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}^{(l-1)})\boldsymbol{W}^{(l)}) \tag{2}$$

上述的 GCN 模块包含两个参数不共享的图卷积网络，其中，$k=1,2,3,4$，$\boldsymbol{W}=\mathbb{R}^{d\times d_k}$ 是可学习的权重参数，对于每一个图$\{\boldsymbol{A}_k\}_{k=1}^4$，并行执行 GCN 学习，产生 d_k 维的输出值，并将这些输出值拼接成最终输出结果：

$$\boldsymbol{Z}=\mathrm{Concat}(G_1^{\mathrm{GCN}}(\boldsymbol{A}_1,\boldsymbol{H}),\cdots,G_4^{\mathrm{GCN}}(\boldsymbol{A}_4,\boldsymbol{H})) \tag{3}$$

式中：Concat(·)表示4个 GCN 头输出的拼接操作。最后，图变换层通过前向反馈神经网络、层归一化和残差连接的方式作进一步学习，得到增强的图卷积向量表示：

$$\hat{\boldsymbol{Z}}=\boldsymbol{Z}+\mathrm{LayerNorm}(\boldsymbol{Z}) \tag{4}$$

$$\bar{\boldsymbol{Z}}=\hat{\boldsymbol{Z}}+\mathrm{LayerNorm}(F_{\mathrm{FFN}}(\hat{\boldsymbol{Z}})) \tag{5}$$

式中：$F_{\mathrm{FFN}}(\cdot)$表示前馈神经网络。

2.3 辅助学习任务

在数学应用题求解中，常常涉及到对数值信息的理解和处理。因此，引入 4 个针对文本上下文、表达式和答案的数值监督任务，通过自己额外的监督信号，尝试挖掘数值属性特征，增强模型对数值语义的感知和理解。

2.3.1 数值个数预测任务

数值个数预测任务要求模型能理解 MWP 文本中数值的数量，这要求模型能正确区分数值和非数值文本。因此，此处引入一个预测数值个数的回归目标，其损失函数的公式如下：

$$L_{\mathrm{num_count}}=\mathrm{MSE}(M_1^{\mathrm{MLP}}(\boldsymbol{r}^x),s) \tag{6}$$

$$\boldsymbol{r}^x=\overrightarrow{\boldsymbol{h}}_m^x+\overleftarrow{\boldsymbol{h}}_0^x \tag{7}$$

式中：MSE(·)是均方损失函数；$M_1^{\mathrm{MLP}}(\cdot)$是多层感知机；$\boldsymbol{r}^x$ 是第 2.1 节中 GRU 正向输出序列中最后一个位置的隐藏状态 $\overrightarrow{\boldsymbol{h}}_m^x$ 和反向输出序列中最后一个位置的隐藏状态$\overleftarrow{\boldsymbol{h}}_0^x$ 相加的结果；s 是问题描述 X 中出现的数值个数。

2.3.2 答案的数值类型预测任务

MWP 求解模型对待不同的数值类型可能会产生不同的计算步骤和处理方式，例如，如果一个问题的答案是一个整数，模型可能会选择使用整数运算来避免舍入误差；如果答案是分数，可能会使用分数运算来保留精确度。模型能够关注 MWP 中答案的数值类型信息，利用目标值的属性特征指导模型的求解过程。因此，此处引入一个预测答案数据类型的多分类目标，其损失函数的公式如下：

$$L_{\mathrm{ans_type}}=\mathrm{CrossEntropy}(M_2^{\mathrm{MLP}}(\boldsymbol{r}^x),t) \tag{8}$$

式中：CrossEntropy(·)是交叉熵损失函数；$M_2^{\mathrm{MLP}}(\cdot)$是多层感知机；$t\in\{0,1,2,3\}$是答案的数值类型标签，各个数值分别表示整数标签、小数标签、分数标签和百分数标签。

2.3.3 数值与答案表达式关系任务

对于给定的 MWP，希望模型具备判断问题中的数值是否出现在答案表达式中的能力，以此来提高模型对 MWP 任务相关的数值语义的理解能力。因此，此处引入了一个判断数值是否出现在表达式的二分类目标，其损失函数的公式如下：

$$L_{\mathrm{gate_num}}=\mathrm{BCE}(M_3^{\mathrm{MLP}}(\boldsymbol{Z}),E) \tag{9}$$

式中：BCE(·)是二分类交叉熵损失函数；$M_3^{\mathrm{MLP}}(\cdot)$是多层感知机；$\boldsymbol{Z}$ 是第 2.2.2 节中经过图变换层得到的输出结果；$E=\{e_1,e_2,\cdots,e_q\}$为标签值，q 表示一个小批量样本中的最大文本序列长度，且 $e_i=\begin{cases}1, & x_i\in Y\\ 0, & x_i\notin Y\end{cases}$，$Y$ 是最终的目标表达式，x_i 表示数学应用题中的一个单词或数值。

2.3.4 数值与答案类型匹配任务

通过比较问题中各个数值的数据类型与答案的数据类型之间的差异，模型可以有效地利用数值的不同格式提供的启发式信息。这种信息有助于促进模型对数值特征的进一步学习，并增强 MWP 文本中数值的表示能力。因此，引入一个比较数值与答案类型是否相同的多分类目标，其损失函数的公式如下：

$$L_{\mathrm{type_xor}}=\mathrm{CrossEntropy}(M_4^{\mathrm{MLP}}(\boldsymbol{Z}),u_i\oplus t) \tag{10}$$

式中：CrossEntropy(·)是交叉熵损失函数；$M_4^{\mathrm{MLP}}(\cdot)$是多层感知机；$\boldsymbol{Z}$ 是第 2.2.2 节中经过图变换层得

计算机工程 www.ecice06.com

到的输出结果；$t\in\{0,1,2,3\}$是答案的数据类型标签；$u_i\in\{0,1,2,3\}$是数值 n_i 的数据类型标签，分别表示整数、小数、分数和百分数；⊕表示标签之间的异或运算符。

2.4 树结构解码器

受 Seq2Tree 框架[6,22-24]的启发，本文建立一个基于二叉树结构的树型解码器。其中，将数量节点设置为叶子节点，操作符节点设置为非叶子结点，且每个操作符节点有两个子节点，由于最终的目标表达式是树解码器按照前序遍历顺序生成的序列。作为树构建过程的一部分，首先生成最中心的操作符，然后生成左子节点。重复上述过程，一直到生成叶子节点为止，然后递归地生成其他子节点。

2.5 训练

对于给定的训练样本对$\{(X^i,Y^i)\}_1^N$，其中，N 表示训练集中样本个数，Y^i 表示表达式的前缀遍历序列。通过最小化负对数似然估计损失 L_1 来训练模型，最终的损失函数 L 是 L_1 和辅助学习任务的损失函数之和：

$$P(Y^i\mid X^i)=\prod_{j=1}^{m}\text{prob}(y_j\mid X) \tag{11}$$

$$L_1=-\sum_{i=1}^{N}\log_a P(Y^i\mid X^i) \tag{12}$$

$$L=L_1+L_{\text{num_count}}+L_{\text{ans_type}}+L_{\text{gate_num}}+L_{\text{type_xor}} \tag{13}$$

3 实验

将模型与 MWP 中 6 个具有代表性的基线模型在两个公开数据集上进行对比。此外，还进行了消融实验、模型变体比较实验和案例分析，以验证模型中各个模块的有效性。

3.1 数据集

本文使用在 MWP 研究中广泛使用的两个数据集：中文数学应用题数据集 Math23K[2]和英文数学应用题数据集 MAWPS[25]。其中，Math23K 由 23 162 条数学问答样本组成，带有问题、算术表达式和解决方案的答案。MAWPS 由 2 373 条英文数学应用题问答样本组成，数据格式与 Math23K 一致。

3.2 实验设置与评价指标

实验选择数据集中出现频率最高的 4 000 个单词作为词汇表，并用一个特殊的标记 UNK 替换剩余的单词，问题截断最大序列长度为 120，表达式截断最大序列长度为 45。词嵌入层的维度设置为 300，隐藏层大小设置为 512，批大小设置为 64，使用 Adam 优化器来优化模型，学习率设置为 0.001，且每 20 个 Epoch 减少 1/2，Dropout 设置为 0.5。此外，模型的训练时间为 80 个 Epoch，在测试过程中，将 Batch Size 设置为 5。

对于 Math23K 数据集，采用 MWP 研究工作的一般处理方式，划分验证集 1 000 个样本，测试集 1 000 个样本，其余样本作为训练集。对于 MAWPS 数据集，使用五折交叉验证的方式进行评估。此外，本文使用在 MWP 研究中广泛采用的评价指标：表达式准确率和答案准确率。表达式准确率用于度量模型生成的表达式是否与目标表达式相同，答案准确率则用于度量模型生成的表达式的计算结果是否与目标值相等。由于不同的表达式可能产生相同的计算结果，因此两项评价指标具有一定的互补性。

3.3 实验结果与对比分析

3.3.1 基线模型

为了验证模型的有效性，本文选取 MWP 中 6 个具有代表性的基线模型：

1）DNS[2]是一种通用数学求解器。

2）T-RNN[4]是在预测的树状结构模板上应用递归神经网络。

3）GTS[6]是一种以目标驱动的方式生成表达树的树状结构神经网络。

4）NS-Solver[7]是一个考虑数学符号约束的神经符号求解器。

5）Num_S2T[8]是一个考虑外部知识库和数值属性的数学应用题求解器。

6）Graph2Tree[11]是通过构建数量单元图和数量比较图来提取 MWP 中的有用信息。

将上述 6 个基线模型在 2 个数据集上进行实验，模型的准确率结果如表 1 和表 2 所示，其中“—”表示由于代码私有等原因无法获取对应结果，“*”表示基于开源代码重新复现的结果。

表 1 不同模型在 Math23K 数据集上的准确率

Table 1 Accuracy of different models on the Math23K dataset %

模型	表达式准确率	答案准确率
DNS	—	58.10
T-RNN	—	66.90
GTS	62.10*	75.60
NS-Solver	64.90*	75.67*
Num_S2T	66.47*	77.26*
Graph2Tree	65.10*	76.60*
Our model	67.40	78.50

www.ecice06.com 计算机工程

表 2 不同模型在 MAWPS 数据集上的准确率

Table 2 Accuracy of different models on the MAWPS dataset %

模型	表达式准确率	答案准确率
DNS	—	59.50
T-RNN	—	66.80
GTS	81.45*	82.60
Graph2Tree	80.99*	83.70
Our model	82.40	83.45

由表 1、表 2 可知，本文模型在 Math23K 和 MAWPS 数据集上的答案准确率分别为 78.50%和 83.45%。与较好的 Graph2Tree 模型相比，在 Math23K 数据集上的答案准确率提升了 1.9 百分点，在 MAWPS 数据集上的表达式准确率提升了 1.41 百分点。证明了本文模型不仅能有效处理中文数学应用题，在英文数学应用题的求解中仍然存在有效性。对比模型在两种不同语言数据集上的性能表现，发现模型对大规模中文数学应用题的求解准确率更高，主要原因在于：1）模型在更大规模更多样化的中文数学应用题训练集上进行训练；2）中文和英文在语法、句法结构、词汇用法等方面存在显著差异，这些语言特性可能影响模型对不同语种题意的理解和表达式的生成能力。

3.3.2 消融实验

为了分析模型中不同组件对性能的影响，将 4 个数值相关的辅助学习任务和 4 种图结构一共 8 个影响因子在 Math23K 数据集上进行消融研究，对应的答案准确率结果如表 3 所示，其中“↓”表示与完整模型相比答案准确率的减少量。

表 3 不同模型配置在 Math23K 数据集上的答案准确率

Table 3 Answer accuracy of different model configurations on Math23K dataset %

类别	影响因子	答案准确率	变化量
辅助学习任务组	去除数值个数预测任务	78.14	0.36↓
	去除答案的数值类型预测任务	78.19	0.31↓
	去除数值与答案表达式关系任务	78.34	0.16↓
	去除数值与答案类型匹配任务	76.98	1.52↓
不同图结构组	仅使用数值比较图	77.90	0.60↓
	仅使用文本关联图	76.50	2.00↓
	仅使用数值文本关联图	76.30	2.20↓
	仅使用数值类型图	77.10	1.40↓
本文模型		78.50	0.00

通过表 3 可以发现，移除 4 个数值相关的辅助学习任务后，模型性能均有所下降，可见 4 个数值监督任务均能使模型学习到更丰富的数值属性特征。其中，数值与答案类型匹配任务对模型的性能增益明显优于其他 3 个任务，推测其主要原因是该任务综合考虑了问题中的不同数值与答案的数据类型信息，使得模型可以有效地学习从输入值到目标值的直接映射知识，同时，将数值类型信息细分为小数、百分数、分数、整数 4 类，有别于仅区分整型、浮点型两类浅表启发式信息，能进一步促进模型学习数值的不同数据格式特征进行数学运算。此外，在表 3 的不同图结构组中，使用单一的图结构与同时使用 4 种不同的图结构相比仍然有较大差距，可见显式建模数值之间、数值与上下文文本之间以及上下文提示信息之间的多重信息，能够进一步提高模型对问题中的关键提示和约束信息的理解。

3.3.3 模型变体比较实验

为了评估模型中不同模块的泛化能力，将图结构和辅助学习任务两个模块分别应用在 GTS 和 Graph2Tree 基线模型，计算它们在 Math23K 数据集上的表达式准确率，如表 4 所示。

表 4 不同模型变体在 Math23K 数据集上的表达式准确率

Table 4 Equation accuracy of different model variants on Math23K dataset %

类别	模型变体	表达式准确率
GTS 及其变体	GTS	62.1
	GTS+辅助学习任务	64.4
	GTS+4 种图结构	64.7
Graph2Tree 及其变体	Graph2Tree	65.1
	Graph2Tree+辅助学习任务	66.2
	Graph2Tree 替换 4 种图结构	66.1
本文模型		67.4

从表 4 的实验结果可以发现，对比单任务训练的 GTS 和 Graph2Tree，引入辅助学习策略后，表达式准确率分别提升了 2.3 和 1.1 百分点，这些额外的训练信号充分引导模型挖掘数值属性特征，改善数值表征。4 种图结构对 GTS 和 Graph2Tree 的性能增益分别达到 2.6 和 1.0 百分点。不同模态关系的图结构充分利用了问题本身的信息，学习数值之间、上下文文本之间以及数值与上下文文本之间的关系。

对比上述不同模型的原始版本（GTS、Graph2Tree）和相应变体结果，增加了 4 种图结构和辅助学习任务后，模型的表达式准确率均有所提升，表明这两种组件均能有效提升模型性能，具有泛化能力。

3.3.4 案例研究

在所有的基线模型中，Graph2Tree 性能最优，因此从 Math23K 数据集中选择了两个代表性的示

计算机工程 www.ecice06.com

例问题，将 Graph2Tree 和本文模型生成的解决方案进行案例研究，相关结果如下：

案例 1 我国第一长河长江全长 6 380 km，黄河比长江的(9/10)要短 278 km。黄河全长多少千米？

Graph2Tree 模型：6 380×(9/10)+278(错误)

本文模型：6 380×(9/10)−278(正确)

案例 2 食堂大米比面粉多(5/6)，正好多 300 kg，食堂面粉有多少千克？

Graph2Tree 模型：300/(1+(5/6))(错误)

本文模型：300/(5/6)(正确)

分析案例结果可以发现：在案例 1 中，“长江”与数值“6 380”、“(9/10)”有关，数值“278”与“短”有关，“短”与“黄河”有关。在这种情况下，本文通过引入 4 种不同的图结构，可以有效地对数值之间、描述性词语之间以及数值与上下文文本之间的关系进行建模，增强模型捕获异构信息的能力，从而能提升对问题的理解并生成正确的表达式和结果。在案例 2 中，Graph2Tree 虽然学习了数量和描述性词语之间的表示，但是忽略了数值属性特征，导致产生了问题中不存在的数值“1”。本文模型通过引入 4 个数值相关的辅助任务来引导模型建模数值语义，提高模型对数值语义的感知和理解，从而能产生正确的表达式和结果。

4 结束语

本文提出一种新颖的 MWP 求解器，通过使用图编码器挖掘文本中的数值和实体之间的隐含关系，利用辅助学习任务挖掘数值属性特征以改善数值表征。在广泛的公开测试集数据实验中使用基线评估提出 MWP 求解模型，实验结果表明，本文提出的模型在 MWP 任务上的表现优于主流的经典基线模型。本文模型目前仅能解决答案唯一型数学应用题，针对多语言的数学应用题语料库，模型的求解能力尚存在不足。未来将探索借助大语言模型解决多元方程的数学应用题问题，使其能够更好地适用于数学文本领域。

参考文献

[1] 王明书. 机器解答小学数学应用题研究[D]. 武汉：华中师范大学，2020.
WANG M S. Research on machine solving elementary school mathematics application problems [D]. Wuhan: Central China Normal University, 2020. (in Chinese)

[2] WANG Y, LIU X J, SHI S M. Deep neural solver for math word problems[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2017: 845-854.

[3] HUANG D Q, LIU J, LIN C Y, et al. Neural math word problem solver with reinforcement learning[C]//Proceedings of the 27th International Conference on Computational Linguistics. Washington D. C., USA: IEEE Press, 2018: 213-223.

[4] WANG L, ZHANG D X, ZHANG J P, et al. Template-based math word problem solvers with recursive neural networks[J]. Artificial Intelligence, 2019, 33(1): 7144-7151.

[5] LI J R, WANG L, ZHANG J P, et al. Modeling intra-relation in math word problems with different functional multi-head attentions[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA: Association for Computational Linguistics, 2019: 6162-6167.

[6] XIE Z P, SUN S C. A goal-driven tree-structured neural model for math word problems[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence. Washington D. C., USA: IEEE Press, 2019: 5299-5305.

[7] QIN J H, LIANG X D, HONG Y N, et al. Neural-symbolic solver for math word problems with auxiliary tasks[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Washington D. C., USA: IEEE Press, 2020: 3780-3789.

[8] WU Q Z, ZHANG Q, WEI Z Y, et al. Math word problem solving with explicit numerical values[C]// Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. Bangkok, Thailand: [s. n.], 2021:5859-5869.

[9] LIN X, HUANG Z Y, ZHAO H K, et al. HMS: a hierarchical solver with dependency-enhanced understanding for math word problem[J]. Artificial Intelligence, 2021, 35(5): 4232-4240.

[10] WU Q Z, ZHANG Q, FU J L, et al. A knowledge-aware sequence-to-tree network for math word problem solving[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2020: 7137-7146.

[11] ZHANG J P, WANG L, LEE R K, et al. Graph-to-tree learning for solving math word problems[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA: Association for Computational Linguistics, 2020: 3928-3937.

[12] LIANG Z W, ZHANG J P, GUO K H, et al. Compositional mathematical encoding for math word problems [C] // Proceedings of the Findings of the Association for Computational Linguistics. Stroudsburg, USA: Association for Computational Linguistics, 2023: 10008-10017.

[13] BROWN T B, MANN B, RYDER N, et al. Language models are few-shot learners [C] // Proceedings of the Advances in Neural Information Processing Systems. Cambridge, USA: MIT Press, 2020:1877-1901.

[14] MIN S, LÜ X X, HOLTZMAN A, et al. Rethinking the role of demonstrations: what makes in-context learning work? [C]//Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2022: 11048-11064.

[15] RAE J W, BORGEAUD S, CAI T, et al. Scaling language models: methods, analysis & insights from training gopher [EB/OL]. [2024-06-10]. https: // arxiv. org/abs/2112.11446.

[16] WEI J, WANG X Z, SCHUURMANS D, et al. Chain of thought prompting elicits reasoning in large language models[C]// Proceedings of the Advances in Neural Information Processing

计算机工程 www.ecice.com

Systems. Cambridge, USA: MIT Press, 2024: 24824-24837.
[17] RUBIN O, HERZIG J, BERANT J. Learning to retrieve prompts for in-context learning[C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, USA: Association for Computational Linguistics, 2022: 2655-2671.
[18] LIU J C, SHEN D H, ZHANG Y Z, et al. What makes good in-context examples for GPT-3? [C]// Proceedings of DeeLIO'22. Washington D. C., USA: IEEE Press, 2022: 100-114.
[19] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2014: 1724-1734.
[20] LI S, ZHAO Z, HU R F, et al. Analogical reasoning on Chinese morphological and semantic relations [C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA: Association for Computational Linguistics, 2018: 138-143.
[21] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks [C]// Proceedings of International Conference on Learning Representations. Washington D. C., USA: IEEE Press, 2017: 443-455.
[22] 肖菁, 何岱俊, 曹阳. 一种自动求解数学应用题的双路文本编码器[J]. 华南师范大学学报(自然科学版), 2023, 55(1): 36-44.
XIAO J, HE D J, CAO Y. A dual channel text encoder for solving math word problems [J]. Journal of South China Normal University(Natural Science Edition), 2023, 55(1): 36-44. (in Chinese)
[23] 林加艺, 夏鸿斌, 刘渊. 基于类比学习的数学应用题求解模型[J]. 计算机工程, 2024, 50(7): 63-70.
LIN J Y, XIA H B, LIU Y. Math word problem solving model based on analogical learning [J]. Computer Engineering, 2024, 50(7): 63-70. (in Chinese)
[24] LIU Q Y, GUAN W, LI S J, et al. Tree-structured decoding for solving math word problems[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2019: 2370-2379.
[25] KONCEL-KEDZIORSKI R, ROY S, AMINI A, et al. MAWPS: a math word problem repository[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, USA: Association for Computational Linguistics, 2016: 1152-1157.

文字编辑 索书志
栏目编辑 赖玉玲

计算机工程 www.ecice06.com