

XY 数学 X_1 XY 数学 X_2 XY 数学 X_3

用于自然语言处理情感分类的量子图变换器

Andreas Bartschi
CCS-3 Information Sciences
Los Alamos National Laboratory
Los Alamos, NM, USA
baertschi@lanl.gov

Abdel-Hameed A. Badawy
Klipsch School of Electrical and
Computer Engineering
New Mexico State University
Las Cruces, NM, USA
badawy@nmsu.edu

Stephan Eidenbenz
CCS-3 Information Sciences
Los Alamos National Laboratory
Los Alamos, NM, USA
eidenben@lanl.gov

Abstract

量子机器学习是在构建更高效和更具表达力的模型方面具有潜力的方向，特别是在理解复杂且结构化数据至关重要的领域。我们介绍了量子图转化器 (QGT)，这是一个混合图结构的架构，将量子自注意机制整合到结构化语言建模的消息传递框架中。注意机制通过参数化量子电路 (PQC) 实现，使模型能够捕捉丰富的上下文关系，同时显著减少可训练参数的数量，与经典注意机制相比更为高效。我们在五个情感分类基准上评估了 QGT。实验结果表明，QGT 在准确率上始终高于或相当于现有的量子自然语言处理 (QNLP) 模型，包括基于注意力和非注意力的方法。与相应的经典图转化器相比，QGT 在实际数据集上的平均准确率提高了 5.42%，在合成数据集上提高了 4.76%。此外，QGT 显示出改进的样本效率，在 Yelp 数据集上需要近 50% 更少的标记样本即可达到相当的性能。这些结果强调了基于图的 QNLP 技术在推进高效和可扩展语言理解方面的潜力。

CCS Concepts: • Computer systems organization → Quantum computing; • Computing methodologies → Natural language processing; • Hardware → Quantum technologies.

Keywords: Quantum machine learning, Graph Transformer, Self-attention mechanism, Text classification

1 引言

人工智能 (AI) 已成为现代技术进步中不可或缺的一部分，对医疗、金融、科学发现和通信等领域产生影响 [23]。自然语言处理 (NLP) 通过使机器能够理解、解释和生成人类语言，在 AI 中发挥基础性作用 [21]。这种能力使得 NLP 可以为各种应用提供支持，从情感分析到机器翻译、对话代理以及内容审核 [17, 29]。最近 NLP 模型的进步，如 GPT-4，显著提升了机器生成连贯且具有上下文意识文本的能力 [19]。尽管取得了成功，当前的 NLP 模型仍然面临许多重要限制。训练和推理最先进的模型需要大量标注数据、访问高性能计算 (HPC) 基础设施，并对数以亿计的参数进行优化 [9]。

量子机器学习 (QML) 是一个有吸引力的框架，可以克服与经典机器学习 (ML) 模型相关的一些低效之处 [2, 3]。QML 采用混合量子-经典策略，其中数据被编码到高维 Hilbert 空间中，并使用参数化量子电路 (PQC) 进行学习，这些电路充当可训练层，类似于经典架构中的神经网络层。这些电路利用叠加和纠缠等量子特性更有效地探索复杂函数 [11, 25]。在自然语言处理 (NLP)

的背景下，理解标记之间的上下文关系至关重要，QML 为开发更高效和更具表现力的语言模型提供了一个有前途的方向。量子增强的 NLP (QNLP) 模型有潜力在样本效率、模型准确性、收敛速度和概括能力上实现量子优势。

在量子自然语言处理 (QNLP) 方面，最近的努力探索了将量子原理融入语言建模的各种方法。早期的模型包括量子词袋方法、量子支持向量机 (QSVM) 和采用振幅编码的变分量子分类器 [1, 16]。最近的进展集中于混合架构，例如量子增强的 LSTM [28]、基于自注意力的模型 [4, 13, 34]，以及像 LEXIQL 这样的基于集成的方法 [26]。这些方法显示了可喜的结果，表明 QNLP 作为传统 NLP 模型的可行且可扩展的替代方案具有潜力。

在经典机器学习中，基于图的模型在捕捉复杂数据结构方面表现出强大的性能，包括知识图谱、引用网络和自然语言处理中依赖树 [14, 32]。图神经网络 (GNNs) 将输入表示为图，其中节点表示实体（如自然语言处理中的标记），边捕捉语法或语义关系 [24]。这种基于图的表示使模型能够有效地捕捉数据中的层次和上下文关系。在这项工作中，我们提出了量子图变换器 (QGT)，这是一种用于情感分类的新颖混合量子-经典模型。QGT 将每个句子表示为标记的完全连接图，并应用修改后的量子变压器卷积 (QTransformerConv) 层来在节点之间传播信息。QTransformerConv 层中的自注意机制利用参数化量子电路 (PQCs) 生成查询和键向量，这对于计算相邻节点之间的注意力分数是必要的。具体来说，每个标记节点被编码为量子态，然后应用由可学习权重参数化的 PQCs，以提取查询和键表示。由此产生的注意力分数指导信息在图中的流动，使模型能够在量子增强特征空间中捕捉标记之间的上下文依赖性。我们在五个基准情感分类数据集上评估了 QGT，并证明了 QGT 能有效地学习标记嵌入的上下文表示。我们还将模型在这些基准数据集上的性能与其他 QNLP 模型和一个经典图变压器模型进行比较。

我们的主要贡献如下：

1. 我们引入了一种混合基于图的架构，将量子自注意机制集成到消息传递框架中，用于结构化语言建模。我们的设计利用 PQC 来实现自注意力，与传统的经典注意力机制相比，所需的可训练参数显著减少，同时保持了高表达能力。
2. 我们在五个情感分类基准上评估了所提出的 QGT 模型，包括三个真实世界的数据集 *i.e.* Yelp、IMDB 和 Amazon [12] 以及两个合成数据集 *i.e.* MC 和 RP [16]。QGT 模型始终如一地学习到了有意义

的上下文表示，并在所有数据集中准确预测情感标签。它的表现超出了或者达到了现有基于注意力的 QNLP 模型和其他 QNLP 基线的准确性。

3. 为了进一步评估模型性能，我们将 QGT 与等效的经典图转换器模型进行比较。我们的结果显示，QGT 在三个真实世界数据集上的准确率显著提高，平均提高了 5.42%，在两个合成数据集上提高了 4.76%。
4. 此外，QGT 展现出了更高的样本效率，在 Yelp 数据集上达到类似准确度时所需的训练样本数量几乎减少了 50%。

2 背景 & 动机

2.1 自然语言处理任务

自然语言处理 (NLP) 是人工智能中的一个专门领域，旨在通过使计算机系统能处理和生成自然语言来弥合人类交流与机器理解之间的差距。核心的 NLP 任务包括文本分类、命名实体识别、问答系统和机器翻译。这些任务的统一之处在于需要对文本序列内和序列之间的复杂句法和语义关系进行建模。在情感分类任务中，目标是预测给定句子或文档中所表达的潜在情感或意见，通常为一个类别标签，如积极、消极或中性。由于情感分类在客户反馈分析、用户体验评估和自动评论聚合等领域具有实际意义，因此它是一个广泛研究的问题。在自然语言处理的早期阶段，模型主要依赖于简单的表示方法，如词袋和 N 元特征，这些方法将文本视为无序的词语计数集合，而不考虑语法或语义。随后，更多语言学知情的方法出现，利用诸如依存树和词性标注等语法感知结构，以捕捉句子中词与词之间的句法关系。虽然这些结构化模型提供了改进的语言学洞察，但它们常常难以在多样的语言环境中有效泛化。一个重大进步是词嵌入的引入，如 Word2Vec 和 GloVe，它们将词表示为密集嵌入空间中的连续向量。这些嵌入根据词的共现模式捕捉语义相似性，使模型能更好地在相关的语言结构间泛化。循环神经网络 (RNNs)、长短时记忆网络 (LSTMs) 和门控循环单元 (GRUs) 等模型利用这些嵌入来处理序列数据并学习文本的上下文表示。然而，这些模型在捕获远程依赖性时遇到了困难，例如梯度消失和内存容量有限的问题。

自然语言处理领域的另一个突破性进展是注意力机制的引入，它使得模型能够动态地关注输入序列中相关的部分。这个概念是 Transformer 架构 [30] 开发的核心，它完全用自注意力层取代了循环结构。诸如 BERT [8]、GPT [10] 和 RoBERTa [15] 等模型在许多自然语言处理任务中都取得了最先进的结果。这些模型通常需要大量的标注数据和大量的计算资源，并且在推广到未见过的语言模式时仍然面临挑战。

2.2 现有的量子自然语言处理模型

量子机器学习 (QML) 近年来由于其潜力能够超越经典机器学习的能力并展示量子优势而被广泛研究。QML 通常采用量子-经典混合方法，其中 PQC 被用作可训练的量子层。PQC 表示为作用于输入量子态 $|\psi_{in}\rangle$ 上的幺正

变换 $U(\lambda)$ ，其中 $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_p\}$ 是可调的实值参数。结果输出态由 $|\psi_{out}\rangle = U(\lambda)|\psi_{in}\rangle$ 给出。在 QML 框架中，经典输入数据首先被编码为量子态，随后 PQC 被训练来学习最优参数，以最小化目标学习任务的损失函数。

最近，人们对 QNLP 模型越来越感兴趣，它作为经典 NLP 模型的一个有前途的替代方案。混合量子增强 NLP 模型旨在将自然语言编码为量子态，并优化参数以学习单词之间的语义关系。由于量子态存在于高维希尔伯特空间中，即使采用相对较少的参数，也能够提供更强的表现力。类别组合分布 (DisCoCat) 模型结合了分布语义和语法结构，并自然映射到量子张量乘积空间 [6]。然而，从句子输入中准确提取句法结构的挑战限制了 DisCoCat 模型的适应性，尤其是在处理自然语言的变化和歧义时。其他方法包括量子增强的词袋 (BoW) 和 N-gram 模型，这些模型将词出现或共现模式编码为量子态 [1]。这些模型在合成数据集上表现出了良好的前景，但其实际实施和性能通常受到量子比特需求的限制，限制了其可扩展性和实际应用。此外，量子 LSTM [28] 和量子 SVM [1] 模型已经被提出以解决与经典 NLP 模型类似的情感分类任务。最近的努力集中在开发量子增强的自注意力机制。这些模型在真实世界基准数据集上表现出更好的学习能力 [4, 13, 34]。此外，Silver 的一项近期工作 *et al.* 引入了一种基于集成的方法，采用增量数据注入方法以提高情感分类任务中量子 NLP 模型的泛化能力 [26]。

2.3 经典图变压器

经典图变压器是一类图神经网络模型，它将原始变压器架构中的自注意力机制整合到图结构数据中 [30, 33]。具体而言，图变压器用学习的基于注意力的消息传递替代固定的图卷积。这允许每个节点根据学习的注意力分数关注一部分或所有其他节点。因此，图变压器能够有效地捕捉底层数据中的局部和长距离依赖性。给定一个无向、无权图 $G = (V, E)$ ，节点特征为 $\{x_i \in \mathbb{R}^d\}_{i \in V}$ ，通过三个学习的线性投影，每个节点获得查询、键和值的表示：

$$Q_i = W_Q \cdot x_i, K_j = W_K \cdot x_j, V_j = W_V \cdot x_j, \quad (1)$$

其中 $W_Q, W_K, W_V \in \mathbb{R}^d$ 是共享的可训练参数。注意力分数 $\alpha_{i,j}$ 是通过目标节点 i 与其邻居 j 之间的查询与键向量的缩放点积计算得到的：

$$e_{i,j} = \frac{Q_i \cdot K_j^T}{\sqrt{d}} \quad (2)$$

为确保这些分数在邻居中具有可比性，它们使用 softmax 函数进行了归一化：

$$\alpha_{i,j} = \frac{\exp(e_{i,j})}{\sum_{m \in \mathcal{N}(i)} \exp(e_{i,m})} \quad (3)$$

其中 $\mathcal{N}(i)$ 表示节点 i 的邻居集合。得到的注意力系数 $\alpha_{i,j}$ 反映了在更新节点 i 的特征向量时，每个邻居 j 的相对重要性或相关性。节点 i 的最终输出特征是其邻居

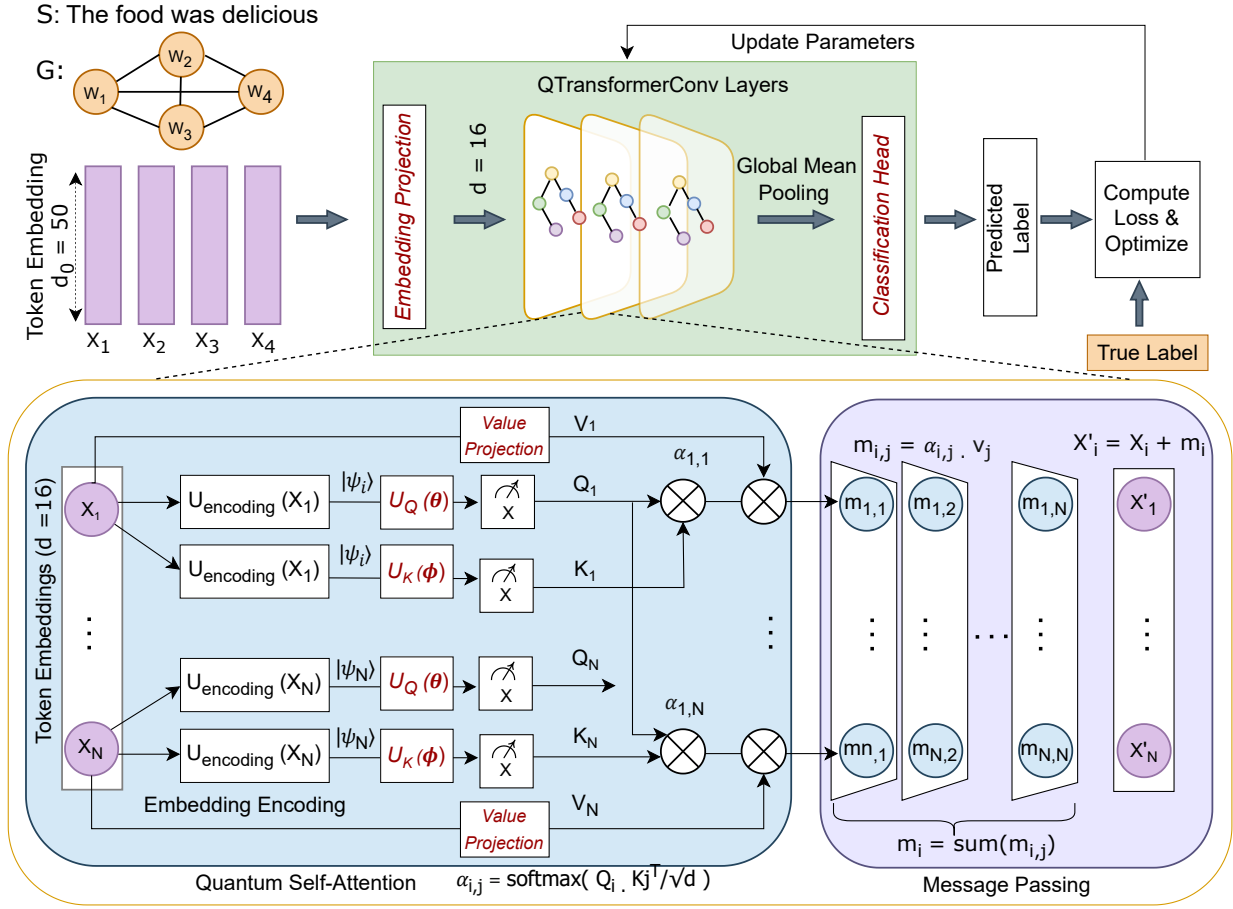


Figure 1. 量子图变压器 (QGT) 架构概述。每个输入句子 S 被标记化并表示为一个完全连接的图 G 。标记使用 GloVe 嵌入 ($d_0 = 50$) 初始化, 并通过 $d = 16$ 投射到 $x_i \in \mathbb{R}^d$, 然后通过堆叠的 $QTransformerConv$ 层处理, 并通过全局均值池化聚合。在每个 $QTransformerConv$ 层中, 节点特征 x_i 被编码为量子状态 $|\psi_i\rangle$, 并使用 $U_Q(\theta)$ 和 $U_K(\phi)$ 处理以提取查询和键向量。这些向量用于量子自注意力计算注意力分数 $\alpha_{i,j}$, 以指导消息传递更新节点特征。聚合特征通过分类头进行标签预测, 模型参数 (用红色标记) 通过反向传播更新以将损失最小化。

值向量的加权和:

$$x'_i = x_i + \sum_{j \in \mathcal{N}(i)} \alpha_{i,j} \cdot V_j \quad (4)$$

此种形式使每个节点能够选择性地关注其邻域的不同部分, 使模型能够根据相关性分数动态地编码结构信息。可以叠加多个基于变压器的卷积层, 以逐步优化节点表示并捕捉图中较高阶的依赖关系。

在当今由人工智能驱动的世界中, 提升自然语言理解变得愈加重要, 量子增强的自注意力模型在推动自然语言处理任务方面显示出了显著的潜力。受此启发, 我们提出了一种用于情感分类的量子图变压器 (QGT) 模型, 通过量子增强的表示捕捉更深层次的上下文。

3 量子图转换器模型

图 1 展示了所提出的 QGT 模型的整体架构。这种混合 QGT 模型将量子自注意力整合到基于图的神经架构中, 用于情感分类。模型的架构由以下关键组件组成: 每个输入句子 $S \in \mathcal{D}$, 其中 $\mathcal{D}$ 表示数据集, S 表示一个句子, 首先被标记化为一个标记序列。这个序列用 $S = \{w_1, w_2, \dots, w_N\}$ 表示, 其中每个 w_i 表示一个标记, N 表示句子在标记化后的总标记数。基于这个标记化序列, 我们构建一个图 $G = (V, E)$, 其中 V 表示被视为节点的标记, E 对应于标记之间的边连接。在 QGT 模型中, 我们考虑一个完整的图结构, 其中每个节点与每个其他节点相连, 即对所有 $i \neq j$ 都有 $(i, j) \in E$ 。密集的连通性允许每个标记在信息传递期间与所有其他标记直接交换信息。这对于捕获情感分析任务中所需的长距离依赖和全

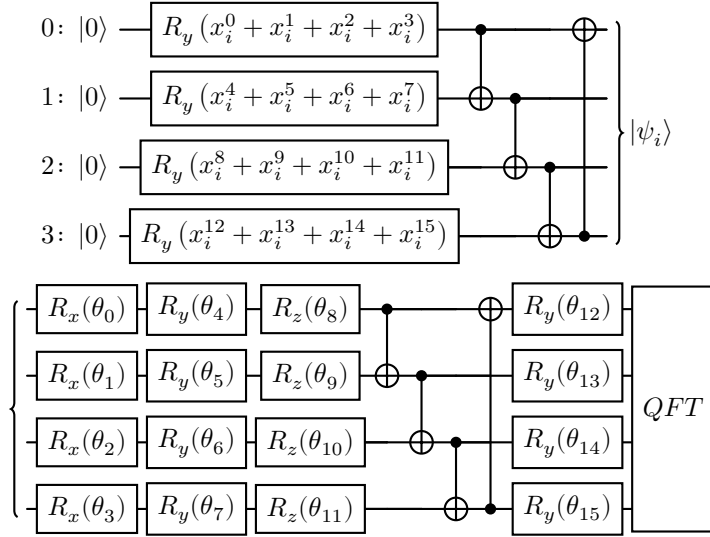


Figure 2. (左) 量子嵌入电路 $U_{\text{encoding}}(x_i)$ 使用分组 R_y 旋转将令牌嵌入向量 $x_i \in \mathbb{R}^{d=16}$ 映射到 4 量子比特的量子态 $|\psi_i\rangle$ ，其中每个旋转角度由 $\sum_{j=4k}^{4k+3} x_i^j$ 提供给量子比特 $k = 0, 1, 2, 3$ 。然后是一个环状 CNOT 纠缠层。(右) 用于查询向量生成的参数化量子电路 $U_Q(\theta)$ 包含每个量子比特的 R_x 、 R_y 和 R_z 层，一个环状 CNOT 纠缠层，第二个 R_y 层，以及最后的量子傅里叶变换 (QFT) [31]。综合电路 $U_Q(\theta) \cdot U_{\text{encoding}}(x_i)$ 被应用，测量得到查询向量 Q_i 。类似地，使用参数 ϕ 生成键向量 K_i 。

局上下文特别重要。或者，为了提高效率和局部上下文建模，可以使用 k -最近邻 (k-NN) 图，其中每个节点仅与其在标记序列中的 k 个直接邻居相连。这种稀疏性可以显著减少计算开销，特别是对于较短的句子（例如 e.g. 推文、聊天消息或搜索查询），其中可能不需要完全连接。

3.1 QTransformerConv: 嵌入编码

每个节点 $i \in V$ （对应于标记 w_i ）最初被映射到一个预训练的 GloVe 嵌入向量 [20] $em_i \in \mathbb{R}^{50}$ ，捕捉标记的语义和语法信息。首先，我们通过线性层将 GloVe 嵌入投射到较低维的表示 $x_i \in \mathbb{R}^d$ 。这里， $d = 16$ 是每个节点特征向量的降维。维度缩减后的输入嵌入向量可以有效地编码到量子电路中，从而需要较少的量子比特。然后投射特征 x_i 使用量子编码电路 $U_{\text{encoding}}(x_i)$ 编码到量子态 $|\psi_i\rangle$ 中，如图 2（左）所示。电路利用 $n = \sqrt{d}$ 个量子比特，其中每个量子比特是通过依次对应于 x_i 组件的 n 个 R_y 门来初始化的。在这些 R_y 旋转之后，通过一系列环状排列的 CNOT 门，将每个量子比特连接到其相邻量子比特，来引入纠缠。这种编码电路可以表示为

$$U_{\text{encoding}}(x_i) = \left(\prod_{k=0}^{n-1} \text{CNOT}_{k, k+1} \right) \cdot \left(\bigotimes_{k=0}^{n-1} \left(\prod_{j=nk}^{nk+n-1} R_y(x_i^j) \right) \right) \quad (5)$$

由于每个量子比特经历 n 顺序 R_y 旋转，这些旋转可以合并为一个具有累计旋转角度的 R_y 门。这种简化表示

为：

$$\prod_{j=nk}^{nk+n-1} R_y(x_i^j) \equiv R_y \left(\sum_{j=nk}^{nk+n-1} x_i^j \right) \quad (6)$$

将此编码电路应用于初始状态 $|0\rangle^{\otimes n}$ 可得到编码的量子态：

$$U_{\text{encoding}}(x_i) = \left(\prod_{k=0}^{n-1} \text{CNOT}_{k, k+1} \right) \cdot \left(\bigotimes_{k=0}^{n-1} R_y \left(\sum_{j=nk}^{nk+n-1} x_i^j \right) \right) \quad (7)$$

$$|\psi_i\rangle = U_{\text{encoding}}(x_i) |0\rangle^{\otimes n} \quad (8)$$

在编码之后，每个节点特征被表示为一个量子态 $|\psi_i\rangle$ 。然后我们应用一个 QTransformerConv 层，该层基于量子增强的自注意力机制进行信息传递。为了提取自注意力的查询向量，我们对每个编码状态 $|\psi_i\rangle$ 应用 $U_Q(\theta)$ 与 n 量子比特。如图 2（右）所示，查询电路 $U_Q(\theta)$ 由三个单量子比特旋转层—— R_x 、 R_y 和 R_z ——组成，应用于每个量子比特，然后是一个环形 CNOT 纠缠层。紧接着是一个附加的 R_y 旋转层和对所有量子比特应用的量子傅里叶变换 (QFT)。一个相似的电路结构 $U_K(\phi)$ 被用于生成键向量，仅在其参数集 ϕ 上有所不同。 $U_Q(\theta)$ 和 $U_K(\phi)$ 均是完全可微分的，它们的参数 θ 和 ϕ 在训练期间被学习和优化。然后通过反复从电路中采样以测量一

些观测量的期望值，从而获得查询 Q_i 和键 K_i 向量：

$$Q_i = (\langle \psi_i | U_Q^\dagger(\theta) | \mathcal{O}_k | U_Q(\theta) | \psi_i \rangle)_{k=0, \dots, n-1} \quad (9)$$

$$K_i = (\langle \psi_i | U_K^\dagger(\phi) | \mathcal{O}_k | U_K(\phi) | \psi_i \rangle)_{k=0, \dots, n-1} \quad (10)$$

观测值 $\mathcal{O}_k$ 可以是泡利矩阵 X 、泡利矩阵 Y 和泡利矩阵 Z 测量的任意组合。在我们提出的模型中，我们对每个量子比特应用泡利矩阵 X 测量，以获得查询 Q_i 和键 K_i 向量。对于具有 N 个节点的图，此过程对每个节点独立重复，以获取其对应的查询和键表示。一旦计算出所有节点的查询 Q_i 和键 K_i 向量，就会计算注意力分数，以确定每个邻近节点 j 与给定节点 i 的相关性。类似于经典的自注意力机制在方程 (2) & (3) 中的表述，节点 i 与节点 j 之间的注意力分数通过一个缩放的点积计算，然后通过 softmax 函数进行归一化：

$$\alpha_{i,j} = \frac{\exp\left(\frac{Q_i \cdot K_j^T}{\sqrt{d}}\right)}{\sum_{m \in \mathcal{N}(i)} \exp\left(\frac{Q_i \cdot K_m^T}{\sqrt{d}}\right)} \quad (11)$$

。这个归一化的分数确定了在消息传递阶段，令牌 j 对令牌 i 施加的影响程度。更高的注意力分数对应于来自源节点的更强影响。

3.2 QTransformerConv: 信息传递

一旦计算出在标记节点之间的量子注意力分数，模型便通过图上的消息传递来传播上下文信息。来自相邻标记的量子增强注意力值会被聚合以更新每个节点的嵌入向量。具体而言，每个标记节点将消息发送到其直接连接的邻居（基于边集 E ），每条消息由量子注意力分数 $\alpha_{i,j}$ 加权，如方程 11 中定义。对于每个节点 i ，从相邻节点 $j \in \mathcal{N}(i)$ 接收到的消息计算为 $m_{i,j} = \alpha_{i,j} \cdot V_j$ ，其中 V_j 是与节点 j 关联的投影特征向量（值向量）。在节点 i 从所有邻居聚合的总消息为 $m_i = \sum_{j \in \mathcal{N}(i)} m_{i,j}$ 。然后通过结合原始节点表示 x_i 与聚合的消息 $x'_i = x_i + m_i$ 来更新节点的特征向量。该更新方案在丰富节点表示的同时保留了原始节点的身份，通过从其邻域派生的上下文感知特征来实现。

在 QTransformerConv 层使用量子注意力引导的信息传递优化了 token（节点）表示后，模型聚合节点级信息以进行句子级别的分类。全局平均池化操作应用于所有节点嵌入 x'_i 。池化后的句子表示 x'_{graph} 计算如下：

$$x'_{\text{graph}} = \frac{1}{N} \sum_{i=1}^N x'_i \quad (12)$$

其中 N 是节点的数量。然后将此池化后的表示输入全连接层以生成原始类别对数 z ，其表示为：

$$z = W^T x'_{\text{graph}} + b \quad (13)$$

其中 W 是权重矩阵， b 是偏置。对数 z 用于交叉熵损失 [7] 函数，该函数内部应用 softmax 函数以计算类别的

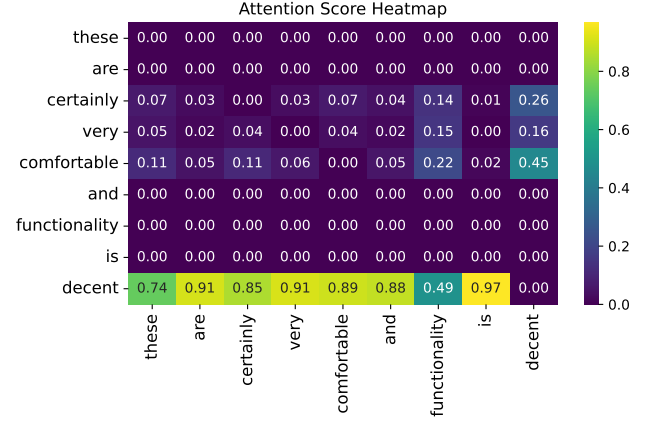


Figure 3. 来自亚马逊数据集的测试样本的注意力分数热图。颜色越浅表示来自其他标记给对应标记的注意力越高。

概率分布。损失定义为：

$$\mathcal{L} = - \sum_{i=1}^c y_i \log \left(\frac{\exp(z_i)}{\sum_{j=1}^c \exp(z_j)} \right) \quad (14)$$

其中 c 是类别数， y_i 是真实标签， z_i 是类别 i 的对数。在反向传播期间，模型的可训练参数被更新以最小化该损失。模型参数来自两个主要组件：(1) 量子自注意力机制，包括 $U_Q(\theta)$ 和 $U_K(\phi)$ ，每个都有 $4 \times 4 = 16$ 个可训练参数；以及 (2) 经典层，包括嵌入投影、值投影和分类头。图 1 用红色显示了可训练参数。详细的训练设置和训练参数在第 ?? 节中讨论。

我们在五个常用的 QNLP 模型基准数据集上评估了 QGT 模型。三个数据集来自 Yelp、IMDB 和 Amazon，分别包含 1,000 个样本。Yelp 数据集包含带有正面或负面情感标签的餐馆评论，IMDB 数据集包含带有正面或负面情感标签的电影评论，而 Amazon 数据集包括 1 到 5 的用户产品评价。三个数据集的最大句子长度分别为 34、45 和 30。其他两个数据集，MC（意义分类）和 RP（相对代词），是合成生成的，如 [16] 所述。它们的词汇量相对较小，每个样本长度为 4 个标记。MC 数据集包含 130 个样本，而 RP 数据集有 105 个样本。这五个数据集集中的每一个都被分为 70%、10% 和 20%，分别作为训练、验证和测试集，并且为每个数据集训练一个独立的模型。训练使用 PennyLane Lightning 设备进行，模型在搭载 Apple M3 Max 芯片和 64 GB 内存的 macOS 上运行。每个数据集都使用第 3 节中描述的 QGT 模型进行训练和训练。在我们的实验中，我们使用 4 个量子比特和单层的 PQC，其中 $U_Q(\theta)$ 和 $U_K(\phi)$ 分别用于计算查询和键向量。参数 θ 和 ϕ 从均值为 0 且标准差为 0.01 的正态分布中初始化。我们选择了 Adam 优化器，学习率为 0.01，并采用交叉熵损失 [7]，这非常适合于包含分类情感标签的分类任务。我们使用步长为 5 和衰减因子 (gamma) 为 0.7 的 StepLR 调度器，批量大小为 32，并基于验证损失应用提前停止，训练最多进行 25 个轮次。为了鼓励对 $U_Q(\theta)$ 和 $U_K(\phi)$ 参数空间的更好探索，我们应用了

一种基于强化学习的正则化策略。奖励定义为负训练损失，并用于直接更新 PQC 的参数。我们还尝试使用多层的 $U_Q(\theta)$ 和 $U_K(\phi)$ 以及多头注意力。然而，由于数据集较小，并未观察到显著的改进。我们首先分析了所提出的 QGT 模型的学习动态以评估其性能。图 ?? 展示了 Yelp、IMDB 和 Amazon 数据集的训练损失曲线。图表显示了训练损失在各个时代中一致减少，表明在所有三个数据集上学习效果良好。在每个案例中，损失在初始时期急剧下降，然后逐渐收敛。图 ?? 展示了相同数据集的训练和验证准确性曲线。验证准确性持续跟随训练准确性，这表明模型在减少损失的同时具有良好的泛化能力。图 3 展示了 Amazon 数据集中一个测试样本的注意力得分热图，其中较浅的颜色表示较高的注意力权重。从图中可以观察到，像“decent”（体面）、“comfortable”（舒适）和“certainly”（当然）这样的标记从其他标记中获得了更高的注意力得分，因为它们对正面情感标签贡献更多。

我们将所提出的 QGT 模型的性能与现有用于情感分类任务的 QNLP 模型进行比较。图 ??（左）展示了 QGT 模型及其他基于注意力的量子 NLP 模型（QSANN [13]、QSAN [34]、QMSAN [4]）在 Yelp、IMDB、Amazon、MC 和 RP 这五个数据集上的测试准确率。该图还包括经典图转换器模型的结果，提供相对于非量子架构的比较视角。QGT 模型在所有五个数据集中都持续取得更高的准确率，表明其在通过量子图形注意力机制捕获语义结构方面的有效性。图 ??（右）展示了 QGT 模型与几个基准模型的性能比较，其中包括语法感知的 DiscoCat [5, 16]、基于集成的 LEXIQL 模型 [26] 和 AmpMean [1]。该图表明，除在 Amazon 数据集上 LEXIQL 以外，QGT 模型优于其他基准量子 NLP 技术。总体而言，QGT 模型在所有基准量子 NLP 模型和经典图转换器模型中表现出强大的泛化性能。在我们提出的 QGT 模型中，相比经典注意力机制，自注意力机制利用 PQC 来计算注意力，显著减少了参数。虽然经典注意力层需要使用大权重矩阵进行高维投影，但我们的量子模型仅使用 32 个可训练参数为每个 $U_Q(\theta)$ 和 $U_K(\phi)$ 层生成查询和键向量。尽管参数数量减少了，图 ?? 中的比较显示，QGT 模型有效地学习了注意力，并优于经典图变换器模型。这个效率的关键在于 PQC 的表达能力，使得模型能够在希尔伯特空间中捕捉丰富的量子表现。

当前经典自然语言处理模型的主要限制之一是它们对于大型数据语料库的依赖以及计算密集型的架构。任何可行的替代模型都应致力于克服这些限制。为了检查样本效率，我们在评估固定测试集的测试准确度时将训练数据从 Yelp 数据集的 10 % 到 100 % 进行变化。图 ?? 展示了量子图和经典图转换器模型在不同训练样本大小下的测试准确度表现。图表显示，量子图转换器（QGT）模型需要较少的训练样本便可达到相似的表现。具体而言，QGT 模型在只使用 Yelp 数据集 30 % 的训练样本时就能达到 82 % 的准确度，而经典图转换器则需要 60 % 的训练样本来达到类似的准确度。这突出了量子自然语言处理模型在自然语言处理任务中数据高效学习的潜力。

4 局限性及未来工作

QGT 模型在小规模数据集上的情感分类表现出色，优于现有的 QNLP 技术和经典的图转换器。然而，其评估仅限于小型、合成数据集。未来的工作应该探讨 QGT 在更大数据集上的性能，例如 SST-5 [27]、Twitter Sentiment [22] 以及 Kaggle IMDB 电影评论数据集 [18]。为了适应这些数据集，我们可能需要结合多层的查询和键 PQC，并使用多头量子自注意力来捕获更丰富的语义模式。多头注意力中的每个头可以通过独立的量子电路并行计算，尽管量子比特的数量会随着头的数量线性增长。此外，我们观察到在小数据集上的参数减少和样本效率的提升。未来的工作还可以研究其他量子优势，如泛化能力和资源需求。此外，还可以在噪声影响下评估 QGT 模型的表现。此外，图结构可以通过结合语义或语法感知的连接进行优化，而不是完全连接的词元图。

在这项工作中，我们为情感分类任务提出了量子图转换器（QGT）模型。该模型引入了一种新颖的 QTransformerConv 层，该层利用量子自注意力在节点之间进行消息传递。该架构能够有效捕捉句子标记和其相应情感标签之间的复杂关系。在基准数据集上的实验结果展示了基于图的量子 NLP 模型在学习和执行 NLP 任务中的潜力。

5 致谢

本文所述的研究得到了国家核安全管理局（NNSA）在洛斯阿拉莫斯国家实验室的摩尔定律之外的高级模拟和计算项目以及洛斯阿拉莫斯国家实验室的实验室定向研究与开发项目（项目编号 20230049DR）的支持。此项工作被指定为 LANL 技术报告编号 LA-UR-25-24685。

References

- [1] Aaranya Alexander and Dominic Widdows. 2022. Quantum Text Encoding for Classification Tasks. In *2022 IEEE/ACM 7th Symposium on Edge Computing (SEC)*. IEEE, 355–361. doi:10.1109/sec54971.2022.00052 arXiv:2301.03715
- [2] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. 2017. Quantum machine learning. *Nature* 549 (Sept. 2017), 195–202. doi:10.1038/nature23474 arXiv:1611.09347
- [3] Marco Cerezo, Guillaume Verdon, Hsin-Yuan Huang, Lukasz Cincio, and Patrick J. Coles. 2022. Challenges and opportunities in quantum machine learning. *Nature Computational Science* 2 (Sept. 2022), 567–576. doi:10.1038/s43588-022-00311-3 arXiv:2303.09491
- [4] Fu Chen, Qinglin Zhao, Li Feng, Chuangtao Chen, Yangbin Lin, and Jianhong Lin. 2025. Quantum mixed-state self-attention network. *Neural Networks* 185 (May 2025), 107123. doi:10.1016/j.neunet.2025.107123 arXiv:2403.02871
- [5] Kenneth Ward Church. 2017. Word2Vec. *Natural Language Engineering* 23, 1 (Jan. 2017), 155–162. doi:10.1017/S1351324916000334
- [6] Bob Coecke, Mehrnoosh Sadrzadeh, and Stephen Clark. 2010. Mathematical Foundations for a Compositional Distributional Model of Meaning. *Linguistic Analysis* 36, 1–4 (March 2010), 345–384. arXiv:1003.4394 Festschrift for Joachim Lambek.
- [7] PyTorch Contributors. 2025. CrossEntropyLoss. <https://pytorch.org/docs/stable/generated/torch.nn.CrossEntropyLoss.html> Accessed: 2025-05-07.

- [8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, 4171–4186. doi:10.18653/v1/N19-1423 arXiv:1810.04805
- [9] Xianzhong Ding, Le Chen, Murali Emani, Chunhua Liao, Pei-Hung Lin, Tristan Vanderbruggen, Zhen Xie, Alberto Cerpa, and Wan Du. 2023. HPC-GPT: Integrating Large Language Model for High-Performance Computing. In *SC-W'23 Workshops of The International Conference on High Performance Computing, Network, Storage, and Analysis*. Association for Computing Machinery, 951–960. doi:10.1145/3624062.3624172 arXiv:2311.12833
- [10] Luciano Floridi and Massimo Chiriatti. 2020. GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines* 30 (Nov. 2020), 681–694. doi:10.1007/s11023-020-09548-1
- [11] Vojtěch Havlíček, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta. 2019. Supervised learning with quantum-enhanced feature spaces. *Nature* 567 (March 2019), 209–212. doi:10.1038/s41586-019-0980-2 arXiv:1804.11326
- [12] Dimitrios Kotzias, Misha Denil, Nando De Freitas, and Padhraic Smyth. 2015. From Group to Individual Labels Using Deep Features. In *21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, 597–606. doi:10.1145/2783258.2783380
- [13] Guangxi Li, Xuanqiang Zhao, and Xin Wang. 2024. Quantum self-attention neural networks for text classification. *Science China Information Sciences* 67 (March 2024), 142501. doi:10.1007/s11432-023-3879-7 arXiv:2205.05625
- [14] Bang Liu and Lingfei Wu. 2022. *Graph Neural Networks: Foundations, Frontiers, and Applications*. Springer, Chapter Graph Neural Networks in Natural Language Processing, 463–481. doi:10.1007/978-981-16-6054-2_21
- [15] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint* (July 2019), 1–13. arXiv:1907.11692
- [16] Robin Lorenz, Anna Pearson, Konstantinos Meichanetzidis, Dimitri Kartsaklis, and Bob Coecke. 2023. QNLP in Practice: Running Compositional Models of Meaning on a Quantum Computer. *Journal of Artificial Intelligence Research* 76 (April 2023), 1305–1342. doi:10.1613/jair.1.14329 arXiv:2102.12846
- [17] Yasmin Moslem, Rejwanul Haque, John D. Kelleher, and Andy Way. 2023. Adaptive Machine Translation with Large Language Models. In *24th Annual Conference of the European Association for Machine Translation*. European Association for Machine Translation, 227–237. arXiv:2301.13294 <https://aclanthology.org/2023.eamt-1.22/>
- [18] Lakshmi N. Narayan. 2018. IMDb Dataset of 50K Movie Reviews. <https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews> Accessed: 2025-05-14.
- [19] OpenAI, Josh Achiam, et al. 2024. GPT-4 Technical Report. *arXiv preprint* (March 2024), 1–100. arXiv:2303.08774
- [20] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global Vectors for Word Representation. In *2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 1532–1543. doi:10.3115/v1/d14-1162
- [21] Sushree Bibhuprada B. Priyadarshini, Amiya Bhusan Bagjadab, and Brojo Kishore Mishra. 2020. *Natural Language Processing in Artificial Intelligence*. Apple Academic Press, Chapter A Brief Overview of Natural Language Processing and Artificial Intelligence, 211–224. doi:10.1201/9780367808495-8
- [22] Sara Rosenthal, Noura Farra, and Preslav Nakov. 2017. SemEval-2017 Task 4: Sentiment Analysis in Twitter. In *11th International Workshop on Semantic Evaluation (SemEval-2017)*. Association for Computational Linguistics, 502–518. doi:10.18653/v1/s17-2088 arXiv:1912.00741
- [23] Stuart J. Russell and Peter Norvig. 2016. *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson. <http://aima.cs.berkeley.edu/>
- [24] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. 2009. The Graph Neural Network Model. *IEEE Transactions on Neural Networks* 20, 1 (Jan. 2009), 61–80. doi:10.1109/TNN.2008.2005605
- [25] Maria Schuld and Nathan Killoran. 2019. Quantum Machine Learning in Feature Hilbert Spaces. *Physical Review Letters* 122, 4 (Feb. 2019), 040504. doi:10.1103/physrevlett.122.040504 arXiv:1803.07128
- [26] Daniel Silver, Aditya Ranjan, Rakesh Achutha, Tirthak Patel, and Devshv Tiwari. 2024. LEXIQL: Quantum Natural Language Processing on NISQ-era Machines. In *SC'24: International Conference for High Performance Computing, Networking, Storage and Analysis*. IEEE, 1–15. doi:10.1109/sc41406.2024.00073
- [27] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. 2013. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In *2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 1631–1642. doi:10.18653/v1/d13-1170
- [28] Jonas Stein, Ivo Christ, Nicolas Kraus, Maximilian Balthasar Mansky, Robert Müller, and Claudia Linnhoff-Popien. 2023. Applying QNLP to Sentiment Analysis in Finance. In *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, 20–25. doi:10.1109/qce57702.2023.10178 arXiv:2307.11788
- [29] Kian Long Tan, Chin Poo Lee, and Kian Ming Lim. 2023. A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research. *Applied Sciences* 13, 7 (April 2023), 4550. doi:10.3390/app13074550
- [30] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems NIPS 2017*. Curran Associates, Inc., 5998–6008. arXiv:1706.03762 <http://papers.nips.cc/paper/7181-attention-is-all-you-need>
- [31] Y. S. Weinstein, M. A. Pravia, E. M. Fortunato, S. Lloyd, and D. G. Cory. 2001. Implementation of the Quantum Fourier Transform. *Physical Review Letters* 86, 9 (Feb. 2001), 1889. doi:10.1103/physrevlett.86.1889 arXiv:quant-ph/9906059
- [32] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S. Yu. 2021. A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems* 32, 1 (Jan. 2021), 4–24. doi:10.1109/tnnls.2020.2978386 arXiv:1901.00596
- [33] Seongjun Yun, Minbyul Jeong, Raehyun Kim, Jaewoo Kang, and Hyunwoo J. Kim. 2019. Graph Transformer Networks. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems NeurIPS 2019*. Curran Associates, Inc., 11960–11970. arXiv:1911.06455 <http://papers.nips.cc/paper/9367-graph-transformer-networks>
- [34] Hui Zhang, Qinglin Zhao, and Chuangtao Chen. 2024. A light-weight quantum self-attention model for classical data classification. *Applied Intelligence* 54, 4 (Feb. 2024), 3077–3091. doi:10.1007/s10489-024-05337-w