Computer Science and Application 计算机科学与应用, 2022, 12(2), 356-365
Published Online February 2022 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2022.122036

Hans汉斯

基于图卷积网络的分子气味印象预测

邱晓芳，骆德汉，魏若冰

广东工业大学信息工程学院，广东 广州

收稿日期：2022年1月17日；录用日期：2022年2月14日；发布日期：2022年2月22日

摘 要

自20世纪90年代以来，嗅觉技术越来越受欢迎，并以各种方式进入了商业用途。从化妆品到洗发水，以及带有香味的博物馆和主题公园，嗅觉消费产品已经陡然流行起来，消费者不仅虚心接受，甚至积极寻求嗅觉产品。然而，目前关于嗅觉的研究大多来自于气味分子的电子鼻数据和质谱数据的角度，而这些数据的获取需要耗费大量的人力和时间。因此，我们从一个新的角度出发，将气味分子的结构视为一个由节点和边组成的图，并引入图卷积网络作用于这个图结构来预测气味分子的气味印象。我们在公开的气味数据集上进行了模型训练，预测了气味分子的气味愉悦度、强度和熟悉度得分，均取得了较好的结果，其中气味愉悦度得分预测的平均绝对误差MAE = 8.532，皮尔逊相关系数为$r = 0.520$ ($p < 0.0000001$)，证实了将气味分子的结构视为图结构而获得的分子信息能够预测气味分子的气味印象。

关键词

气味分子，图形卷积网络，气味印象

Molecular Odor Impression Prediction Based on Graph Convolutional Networks

Xiaofang Qiu, Dehan Luo, Ruobing Wei

School of Information Engineering, Guangdong University of Technology, Guangzhou Guangdong

Received: Jan. 17th, 2022; accepted: Feb. 14th, 2022; published: Feb. 22nd, 2022

Abstract

Since the 1990s, olfactory technology has grown in popularity and entered commercial use in a variety of ways. From cosmetics to shampoos, to scented museums and theme parks, olfactory consumer products have exploded in popularity, with consumers not only humbly accepting but actively seeking them. However, most of the current research on olfaction comes from the elec-

文章引用: 邱晓芳, 骆德汉, 魏若冰. 基于图卷积网络的分子气味印象预测[J]. 计算机科学与应用, 2022, 12(2): 356-365. DOI: 10.12677/csa.2022.122036

tronic nose data and mass spectrometry data of odor molecules, and the acquisition of these data requires a lot of manpower and time. Therefore, from a new perspective, we treat the structure of odor molecules as a graph consisting of nodes and edges, and introduce a graph convolutional network to act on this graph structure to predict the odor impression of odor molecules. We trained the model on the public odor data set, and predicted the odor pleasantness, intensity and familiarity scores of odor molecules, and achieved good results. The mean absolute error of odor pleasantness score prediction was MAE = 8.532, and Pearson's correlation coefficient was r = 0.520 (p < 0.0000001), confirming that the Molecular information obtained from the structure of odor molecules as a graph structure can predict the odor impression of odor molecules.

Keywords

Odor Molecules, Graph Convolutional Networks, Odor Impression

1. 引言

气味在我们的日常生活中无处不在。随着合成化学的出现，香味产品(香水、化妆品、洗涤剂等)已经以各种方式进入商业应用。然而，气味的合成通常需要熟悉芳香成分及其气味印象[1]。如果我们能够获得每个香味分子的气味印象，将大大简化香味产品的生产过程，并减少从天然作物中提取香味的生态影响[2]。

气味印象是人们对闻到的物质的一种感官印象。在日常生活中，当人们找不到合适的例子来描述某种化学物质的气味印象时，就可以用日常口述的形容词来形容，如“酸”、“草”等。在最简单的情况下，人们使用“愉快”和“熟悉”等口头描述来描述他们闻到的物质的气味。近年来，气味印象预测(如愉悦度、强度、熟悉度等)已成为嗅觉研究的热门话题。研究方法可分为三类：基于物理化学特性的、基于电子鼻(e-nose)的和基于质谱的。

基于物理化学特征的方法是通过分子计算软件(DRAGON)计算出每个分子的描述符信息，并将其作为每个分子的特征信息，从而通过建模得到分子气味印象的预测。2007 年，Khan R.M.等人从语言描述符构建了一个定性空间，并从化学描述符构建了一个类似的物理化学空间[3]。他们发现感知的主轴是气味的愉悦度，并使用主成分分析(PCA)来预测分子的愉悦度。2017 年，Shang L.等人使用分子计算软件(DRAGON)获取气味分子的物理化学参数，并通过 BR-C SVM 模型预测气味感知[4]。同年，Keller A.等人在 Dream Olfaction Prediction Challenge 中，利用正则化线性模型和随机森林，根据分子的化学信息特征预测分子的感官特性，从而成功预测气味的强度和愉悦度[5]。

基于电子鼻的方法利用电子鼻传感器测量分子气味，获取高维气味数据信息作为分子特征信息，利用机器学习或深度学习算法进行气味预测。电子鼻技术应用在生活的很多方面，如食物异味检测、工业气体检测、疾病检测等[6] [7] [8] [9]。关于分子气味印象预测的研究很多。2010 年，Haddad 等人使用手动特征提取方法从传感器中提取 120 个特征，并基于对电子鼻气味宜人性的评估，使用 MATLAB 构建三层前馈反向传播神经网络[10]。2019 年， Wu D.L.等人设计了一个 POP-CNN 模型，根据从电子鼻获得的气味信息的特征来预测精油的愉悦感。该模型还可用于气味分类和检测[11]。

基于质谱的方法是利用质谱仪通过实验获得各气味分子的质谱数据，然后以原始质谱数据为输入，

感官数据为输出，建立质谱数据到感官数据的映射模型。2013 年，Nakamoto T.等人使用精油和食用香料的质谱数据库，通过非负矩阵分解(NMF)和非负最小二乘法研究了一些气味成分的选择[12]。2016 年，Nozaki Y.等人设计了一个九层前馈神经网络预测模型，从气味的质谱数据中预测气味印象。该模型以化学分子的质谱数据为输入，嗅觉感官描述符的有无为输出，建立了从质谱特征空间到感官数据特征空间的映射函数[13]。基于九层前馈神经网络预测模型，2018 年，Nozaki Y.等人通过在建模中引入自然语言处理，成功地利用单分子气味物质的质谱预测气味特征。在他们的方法中，首先对数据中的描述符进行聚类，然后设计一个六层神经网络预测模型[14]。

所有这些关于气味印象预测的研究都取得了可以接受的结果；然而，电子鼻数据和质谱数据的获取往往需要大量的人力和时间进行广泛的实验[15] [16]。DRAGON 计算的分子描述符信息非常多样化，往往包含大量冗余信息，甚至不利于分子气味印象的预测。同时，复杂的分子描述符信息不适合小数据样本的实验，可能导致过拟合[17]。众所周知，分子由原子和连接它们的化学键组成。数据的结构不同于非结构化数据，例如电子鼻数据和质谱数据。这是一个图形结构，重点显示分子内部结构之间的关系。由于分子是一种图结构数据，所以它是一种网格数据。以前常用的机器学习和深度学习方法如随机森林(RF)、支持向量机(SVM)、卷积神经网络(CNN)和自动编码器(AE)都不适用于处理这种数据结构，因此我们提出了一种基于图神经网络(GNN) [18] [19]的方法。GNN 是一种基于深度学习处理非欧几里得空间数据的方法。主要思想是通过总结顶点本身的特征和顶点邻居的特征来生成顶点表示。早期的图神经网络使用循环神经网络，但随着其在图像处理和文本中的大规模普及，研究人员开始尝试将卷积扩展到图结构，利用顶点特征和拓扑结构信息进行预测，然后将图卷积网络产生。

由于构成每个气味分子的原子都有自己的特征状态，因此原子之间存在不同的联系。正是这些不同的状态信息和不同的联系，使每个气味分子都有自己独特的气味印象[20]。为了更好地利用分子的内部结构信息，我们设计了三个图卷积层。在每个图卷积层中，顶点(原子)可以向与其相连的顶点发送顶点状态消息。一个顶点将自己的状态信息与其相连的顶点结合起来作为新的状态信息，从而建立分子之间的内部连接。经过 3 次迭代更新顶点信息，我们得到一个代表分子信息的特征矩阵。之后我们设计了一个均值聚合层，将这个特征矩阵聚合成一个向量，实现分子信息的向量化。最后，我们通过全连接层的输出预测了分子的气味印象。首先，我们使用 RDKit 将分子的 SMILE 字符串转化为图，提取图信息，包括特征矩阵和邻接矩阵。然后将处理后的邻接矩阵和特征矩阵输入到三层图卷积层中。经过 3 次顶点信息更新迭代，通过平均聚合层和全连接层获得气味分子印象。

与其他数据相比，分子结构包含了气味分子的所有嗅觉信息。这使我们能够获得更全面的分子嗅觉信息来预测分子气味印象。同时我们发现通过 RDKit 计算分子图信息不仅快捷方便，而且得到的顶点特征信息可以解释，更通用。同时，与分子的物理特性相比，我们可以使用较少的顶点特征信息来表示分子结构信息。最后，我们证明了将气味分子的结构处理为图结构得到的分子图信息，作为图卷积网络的输入，可以预测气味分子的气味印象。

2. 数据集与方法

2.1. 数据集

为了预测气味分子的气味印象，我们使用了 Dream Olfaction Prediction Challenge 发布的最大的气味数据集。气味数据集提供了 476 种单分子化学物质，我们将其用作实验对象来预测它们的气味印象。同时，气味数据集包含 55 名志愿者，他们通过实验对 476 种单分子化学物质的气味印象进行评分，包括不常见、有气味甚至无气味的分子。在对单分子化学品的气味印象评分的过程中，志愿者首先判断该单分子化学品是否有气味，如果有，则根据气味的强度和质量等几个类别进行评分，评分范围为 0~100。这

个分数最终作为我们预测分子气味印象的标准，来判断预测结果的好坏。

由于 55 名志愿者中有 6 名没有完成所有单分子化学物质的气味测试，我们在实验过程中去掉了这 6 名志愿者的其他气味测试数据，只使用了其余 49 名志愿者对愉快度、熟悉度、和强度作为我们实验的数据。此外，由于该数据集给出了每种单分子化学品的 CAS 编号，这与我们实验所需的分子表示不兼容。因此，我们从 PubChem (有机小分子生物活性数据)中批量导出每个分子的 SMILE 字符串，其中 SMILE 字符串是标准编码结构。

众所周知，不同的人在闻到相同的气味时可能会有不同的气味印象。在我们使用的气味数据集中，志愿者对相同单分子化学物质的愉悦度、强度和熟悉度的评分差异很大。一些志愿者对愉快的评价较低，而另一些则相反。为了获得对气味分子气味印象的更统一评分，我们采用高斯拟合方法拟合了 49 名志愿者对同一分子的气味印象评分，例如愉悦度评分，并使用评分的平均值作为单分子化学物质的愉悦。对强度和熟悉度的评分也进行了类似的操作，得到了 476 种单分子化学物质中每种气味的三种气味印象的评分，表示为 476 × 3 的评分标签矩阵。

2.2. 分子的图表示

图是一种通常包含节点和边的数据结构。在现实生活中，许多重要的数据集以图的形式存储，如社会网络信息、知识图、蛋白质网络、分子结构等。这些图网络不是结构化信息图像，而是非结构化信息。

通常，分子表示为 SMILE 字符串。为了将其转化为更直观的图结构，我们使用 RDKit，将原子作为顶点，将原子间的化学键作为连接顶点的边，得到了分子结构的图表示，如图 1 所示。图一般包含两部分信息，一是邻接矩阵，即关于顶点之间连接关系的信息；二是特征矩阵，即关于顶点状态的信息。类似地，我们可以将分子中原子之间的连通性关系作为邻接矩阵，将原子的特征信息作为特征矩阵，得到分子结构图的信息，这些信息可以输入图学习网络中进行预测。

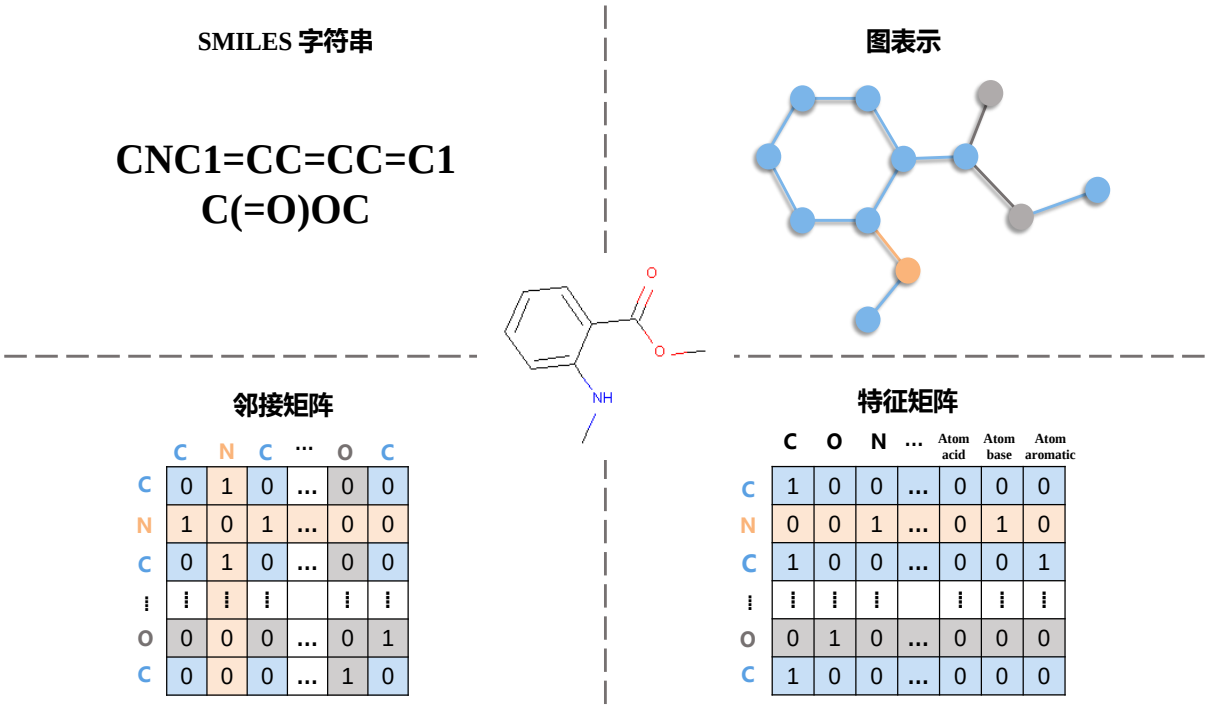

Figure 1. Molecular graph represents structural diagram
图 1. 分子图表示结构图

邻接矩阵 A 是一个 $n \times n$ 的对称矩阵，n 是分子中的原子数。邻接矩阵 A 中的元素代表了分子中任何两个原子之间的连接关系，具体来说，如果第 i 个原子与第 j 个原子相连，那么 $A_{ij}=1$，否则 $A_{ij}=0$，以此类推。

特征矩阵 F 是一个 $n \times m$ 的矩阵，n 是分子中的原子数，m 是为每个原子描述符数量。每个原子的不同类型的原子描述符的信息从 RDKit 中计算出来，并表示为一个 $1 \times m$ 的二进制向量，该向量被表示为原子特征向量。然后将原子特征向量按照原子的顺序拼接成一个 $n \times m$ 的原子特征矩阵。

2.3. 计算并预处理原子特征

RDKit 是化学信息的开源软件，它提供了许多基于分子描述符和原子描述符的计算方法。我们从 RDKit 中计算出 11 种不同类型的原子描述符信息，并将描述符信息以单次编码的形式表示出来，从而使每个原子的特征信息形成一个 1×60 的二进制向量。

由于从 RDKit 获得的原子特征矩阵是以独特的热编码形式表示的，它是一个非常稀疏的二进制矩阵，其中有一些冗余的信息，对分子嗅觉的预测没有贡献，但会增加预测时间。所以我们对 11 种不同类型的原子描述符的信息进行过滤。我们分别使用了三种方法来处理原子特征。主成分分析(PCA)，自动编码(AE)和基于相关性的特征选择(PCC)。其中，用 CFS 获得的信息来预测分子嗅觉效果最好。

PCA 是一种传统的降维算法，它利用正交变换将原始数据投影到多个高方差方向(维数)上，将原来众多具有特定相关性的指标重新组合成一组新的相互不相关的复合指标。

AE 是一种无监督的人工神经网络。它通过消除重要特征上的噪声和冗余来寻找低维数据的表示。编码器将高维数据编码成低维数据，解码器接收到低维数据后，尝试重构原始的高维数据。虽然可能存在过拟合，但可以通过正则化等方法来解决。

PCC 是一种基于特征间相关性来选择特征的算法。这种相关性也可以看作协方差，它反映了两个变量是在的同一方向还是相反的方向上变化。取区间[−1, 1]内的值，越靠近两端越表示两个变量之间存在明显的线性关系，只需要保留其中一个。

2.4. 图卷积网络

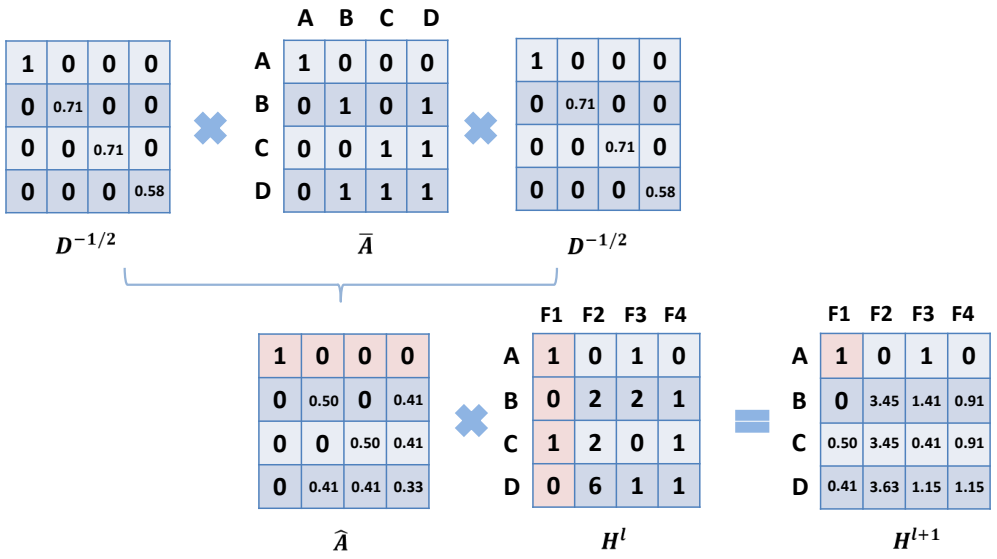

Figure 2. Computation of feature data update for Graph Convolutional Networks
图 2. 图卷积网络的特征数据更新的计算

具体来说，图卷积网络中的每个图卷积层都有两个输入，一个是正则化的邻接矩阵 $\hat{A}$ 和上一层的原子特征信息 H^l (最初始层 H^0 是特征矩阵 F)。如图 2 所示，我们在图卷积网络的每一层之后都加入了非线性激活函数 ReLU，这样，图卷积网络每一层的具体操作如下：

$$H^{l+1} = \mathrm{ReLU}\left(\hat{A}H^lW^l\right) \tag{1}$$

$\hat{A}$ 是通过对直接从图中得到的邻接矩阵 A 进行正则化处理，原因有二：一是解决原子信息自转移问题；二是对邻接矩阵 A 进行正则化。解决自传递问题主要是在原图的基础上给每个原子增加一个自环，具体做法是 $\tilde{A} = A + I$ 。邻接矩阵 A 的归一化主要是通过引入邻接矩阵 A 的度矩阵 D(D 是一个对角线矩阵，对角线上的元素是对应顶点的度)，具体操作如下：

$$\hat{A} = D^{-1/2}\tilde{A}D^{-1/2} \tag{2}$$

X 是最终获得的代表整个气味分子的特征信息。

2.5. Odor-GCN 分子气味印象预测模型

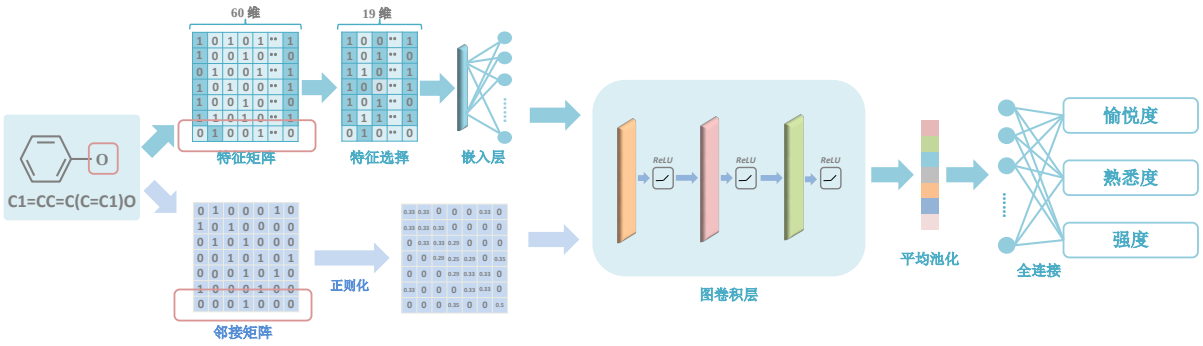

Figure 3. Odor-GCN model for odor impression prediction
图 3. Odor-GCN 分子气味印象预测模型

为了预测气味分子的气味印象，我们建立了一个 Odor-GCN 分子气味印象预测模型，该模型包括输入层、嵌入层、三个图卷积层、平均聚集层、全连接层和输出层，如图 3 所示。我们使用 RDKit 将气味分子 SMILE 串转换为图表示，并获得分子图信息，包括邻接矩阵和特征矩阵。然后，对邻接矩阵和特征矩阵分别进行正则化和选择，得到图卷积网络的输入。

众所周知，one-hot 编码是一个非常稀疏的向量，由于我们用它来表示顶点特征信息，这使得最终得到的特征矩阵非常稀疏。然而，深度学习的结构特点是不利于稀疏向量的处理，所以我们首先设计一个嵌入层将每一行的特征信息的顶点从低维特征矩阵稀疏线性变换到高维密度状态。这不仅使相互独立的顶点特征信息更加内在地联系在一起，而且使一般的顶点特征信息相互分离。

在嵌入层之后，为了更好地利用分子内部的结构信息，我们设计了一个三层的图卷积层，利用顶点之间的连接关系，在三次迭代中更新每个顶点的状态信息。最后，提取整个分子的气味信息。然后，通过平均聚集层，我们将分子的气味信息表示为一个气味向量。平均聚合过程的具体计算过程如(3)所示。将经过三层图卷积后得到的顶点特征矩阵在行上求平均，得到表示分子信息的向量。

$$X = \mathrm{Mean}\left(\mathrm{ReLU}\left(\hat{A}H^3W^3\right)\right) \tag{3}$$

其中 X 是最终获得的代表整个气味分子的特征信息。

3. 结果与分析

3.1. 对顶点特征降维结果的分析

我们主要使用 PCC 对从 RDKit 中提取的 60 维顶点特征进行过滤。首先，通过计算 60 个顶点特征之间的 PCC，得到相关系数矩阵；但是，我们发现相关矩阵中存在间隙，因为在我们的数据集中，构成分子的所有顶点都不包含与间隙对应的特征，所以我们首先消除这些顶点特征。接下来，我们从相关系数矩阵中选取两个高度相关的顶点特征(Pearson 系数绝对值接近 1)，剔除其中一个，只保留另一个作为顶点特征信息，预测气味分子的气味印象。通过一系列的选择，我们总共筛选出了 41 个顶点特征，只使用剩下的 19 个顶点特征对气味分子的气味印象进行预测。表 1 列出了 19 个顶点的特征，包括原子的类型、原子连接的化学键数、原子的电荷量、原子是否在苯环中。

Table 1. Molecular characteristic type
表 1. 分子特征类型

顶点特征	类型	数量
原子类型	C、O、N、S、Cl、P、I、Na	9
氢键数量	0、1、2、3、4	5
原子价态	0、1、2、3、4	4
是否为芳香烃	是、否	1
	总共	19

虽然顶点特征的维度减小，但最终预测的分子气味印象与实际分子气味印象(愉悦度得分)之间的皮尔逊相关系数(r)继续增加，如表 2 所示。初始顶点特征为 60 维，$r = 0.33$，最终顶点特征为 19 维，$r = 0.52$。在此过程中，由于剔除了一些计算时间较长且与其他特征高度相关的顶点特征，如氢键、环大小等，因此预测分子气味印象的速度从最初的预测开始不断加快。最初，当使用 60 维顶点特征时，预测分子气味印象需要 594.45 秒，但使用 19 维顶点特征后，预测分子气味印象只需 48.45 秒。当然，我们也继续对剩下的 19 个顶点特征进行降维处理，然而继续剔除顶点特征使得气味印象预测的相关系数不断下降，但模型操作的时机变化不大。

Table 2. Prediction results after dimension reduction of atomic features
表 2. 原子特征降维后的预测结果

特征维度	时间(s)	r
60 维	594.45	0.33
40 维	572.63	0.38
28 维	394.14	0.48
19 维	48.45	0.52

3.2. 分子气味印象预测结果的分析

我们将气味数据集中的 476 个气味分子随机分成训练集和测试集，比例为 4:1，然后用构建的 Odor-GCN 分子气味印象预测模型分别预测了气味分子的愉悦度得分、强度得分和熟悉度得分，其中模型对气味愉悦度得分的预测更为准确，多次实验预测得到的气味愉悦度得分的平均绝对误差 $\mathrm{MAE} = 8.532$，皮尔逊相关系数为 $r = 0.520\ (p < 0.0000001)$。对气味强度和气味分子熟悉程度的多次预

测的 MAE 和相关性分别为 $\mathrm{MAE}=11.561$，$r=0.422\left(p<0.0000001\right)$ 和 $\mathrm{MAE}=6.30317$，$r=0.451\left(p<0.0000001\right)$。

同时，从分子结构的角度，我们利用不同的分子结构信息来表示具有代表性的分子，并使用常用的机器学习方法来预测分子的气味印象。这包括从分子的原子信息开始，从分子的摩根指纹信息开始。我们采用随机森林(RF)、部分最小二乘回归(PLS)、高斯过程回归(GPR)和卷积网络(CNN)，利用分子摩根指纹、分子的原子特征拼接和利用图形卷积网络提取的特征信息作为分子信息，对气味分子的愉悦度分值、强度分值和熟悉度分值进行预测。使用这些方法，预测得到的平均绝对误差几乎都高于使用 Odor-GCN 分子气味印象预测模型的预测结果，而且相关系数都远远低于 Odor-GCN 分子气味印象预测模型的预测结果，有些预测结果甚至出现了负相关系数。

Table 3. Prediction results after dimension reduction of atomic features
表 3. 原子特征降维后的预测结果

		愉悦度		熟悉度		强度	
		MAE	*r*	MAE	*r*	MAE	*r*
Odor-GCN		8.53	0.52	6.30	0.45	11.56	0.42
摩根指纹信息	RF	10.78	0.43	6.43	0.34	11.69	0.24
	GP	11.71	0.12	6.43	0.24	11.83	0.10
	PLS	11.01	0.41	7.98	0.18	14.30	0.15
	CNN	10.53	0.29	7.76	0.18	15.58	0.19
特征信息拼接	RF	9.22	0.39	6.56	0.19	11.62	0.22
	GP	9.62	0.04	6.69	0.01	12.11	-0.06
	PLS	9.37	0.35	8.43	0.07	12.56	0.13
	CNN	10.54	0.28	8.95	0.22	16.00	0.14
图结构信息	RF	8.94	0.43	6.57	0.33	11.57	0.30
	GP	8.60	0.45	6.58	0.32	11.98	0.01
	PLS	10.29	0.27	7.50	0.15	13.12	0.21
	CNN	12.24	0.23	8.47	0.24	14.04	0.28

从表 3 可以得出结论，对于不同的分子特征信息，使用普通机器学习算法预测分子气味印象的结果比使用 Odor-GCN 预测的结果要差。其中，使用 Odor-GCN 模型预测分子气味印象的气味愉悦度得分与真实的气味愉悦度得分最接近，预测得分与真实得分的 Pearson 相关系数为 $r=0.52$。

此外，利用分子摩根指纹、分子原子特征拼接和图形特征提取获得的特征信息，利用 RF、PLS、GPR 和 CNN 等算法预测分子气味印象(愉悦度、熟悉度和强度)。从表 3 可以看出，虽然 Morgan 指纹也从分子结构的角度描述分子，但将其作为分子信息预测分子气味印象不如将分子结构作为图形表示从而作为分子信息进行预测。在 Morgan 指纹比对检验中，随机森林算法的预测效果最好，但其分子气味印象预测的 Pearson 相关系数比 Odor-GCN 模型的预测效果低约 0.1。同样，忽略原子与分子气味印象预测之间的联系关系，仅仅将原子特征拼接成分子信息，甚至表明预测结果与实际结果之间存在负相关关系。例如，使用 GP 算法对分子气味印象的强度进行预测，其皮尔逊相关系数为 $r=-0.06$。最后，虽然我们使用常用的机器学习方法在经过三层图卷积层(即原子信息经过三次迭代后的特征信息被更新为分子信息)后预测分子气味印象，但最终结果不如直接使用 Odor-GCN 模型预测的结果。由此可见，利用我们搭建的

Odor-GCN 模型，并以气味分子的结构信息作为输入，以气味分子的气味印象评分作为输出，可以较为精准地预测分子的气味印象。

4. 结论

气味分子的结构信息受到构成气味分子的原子类型信息和原子相互连接方式信息的强烈影响；然而，分子的结构信息在气味研究中并没有引起足够的重视。以往的研究大多是从气味分子的电子鼻数据、质谱数据和分子描述符数据进行的，这些数据大多需要大量的人力和时间来做实验来获得。本文将气味分子的结构视为图结构，直接从 RDKit 中获取气味分子的图结构信息，从而大大节省了进行实验所需的人力和时间。

因此，本文从气味分子的分子结构信息角度出发，搭建了用来预测分子气味印象的 Odor-GCN 模型，并较为精准地预测了分子的气味印象，包括分子气味的愉悦度、熟悉度和强度。其中分子气味愉悦度的预测效果最好，其预测得到的平均绝对误差 $\mathrm{MAE}=8.532$，皮尔逊相关系数为 $r=0.520\ (p<0.0000001)$。相较于传统的机器学习模型预测的结果，其平均绝对误差 MAE 最高降低了 4.44。该结果证明了利用图卷积网络从气味分子的结构信息来预测气味分子的气味印象的可行性。这项工作为分子气味的研究提供了一个新的方向，同时也为气味产品的研发带来便利。接下来，本文研究可以进一步从分子的三维结构出发，将分子的空间结构信息加入到二维结构中，使预测结果更加准确。

基金项目

国家自然科学基金(61571140)。

参考文献

[1] Takamichi, N. (2016) Essentials of Machine Olfaction and Taste. Vol. 1, Wiley, Hoboken. https://doi.org/10.1002/9781118768495.ch1

[2] Castro, J.B. and Seeley, W.P. (2014) Olfaction, Valuation, and Action: Reorienting Perception. *Frontiers in Psychology*, **5**, Article No. 299. https://doi.org/10.3389/fpsyg.2014.00299

[3] Khan, R.M., Luk, C.H., Flinker, A., *et al*. (2007) Predicting Odor Pleasantness from Odorant Structure: Pleasantness as a Reflection of the Physical World. *Journal of Neuroscience*, **27**, 10015-10023. https://doi.org/10.1523/JNEUROSCI.1158-07.2007

[4] Shang, L., Liu, C., Tomiura, *et al*. (2017) Machine-Learning-Based Olfactometer: Prediction of Odor Perception from Physicochemical Features of Odorant Molecules. *Analytical Chemistry*, **89**, 11999-12005. https://doi.org/10.1021/acs.analchem.7b02389

[5] Keller, A. and Vosshall, L.B. (2016) Olfactory Perception of Chemically Diverse Molecules. *BMC Neuroscience*, **17**, Article No. 55. https://doi.org/10.1186/s12868-016-0287-2

[6] Cheng, Y., Wong, K., Hung, K., Li, W., Li, Z. and Zhang, J. (2019) Deep Nearest Class Mean Model for Incremental Odor Classification. *IEEE Transactions on Instrumentation and Measurement*, **68**, 952-962. https://doi.org/10.1109/TIM.2018.2863438

[7] Zhang, S., Cheng, Y., Luo, D., He, J., Wong, A.K.Y. and Hung, K. (2021) Channel Attention Convolutional Neural Network for Chinese Baijiu Detection with E-Nose. *IEEE Sensors Journal*, **21**, 16170-16182. https://doi.org/10.1109/JSEN.2021.3075703

[8] Guo, J., Cheng, Y., Luo, D., Wong, K.-Y., Hung, K. and Li, X. (2021) ODRP: A Deep Learning Framework for Odor Descriptor Rating Prediction Using Electronic Nose. *IEEE Sensors Journal*, **21**, 15012-15021. https://doi.org/10.1109/JSEN.2021.3074173

[9] Bruhn, C. (2013) Electronic Noses: How to "Smell" Diseases. *Deutsche Medizinische Wochenschrift*, **138**, 1040-1041. (In German) https://doi.org/10.1055/s-0032-1330190

[10] Haddad, R., Medhanie, A., Roth, Y., *et al*. (2010) Predicting Odor Pleasantness with an Electronic Nose. *PLoS Computational Biology*, **6**, e1000740. https://doi.org/10.1371/journal.pcbi.1000740

[11] Wu, D.L., Luo, D.H., Wong, K.-Y. and Hung, K. (2019) POP-CNN: Predicting Odor Pleasantness with Convolutional

Neural Network. *IEEE Sensors Journal*, **19**, 11337-11345. https://doi.org/10.1109/JSEN.2019.2933692

[12] Nakamoto, T. and Nihei, Y. (2013) Improvement of Odor Approximation Using Mass Spectrometry. *IEEE Sensors Journal*, **13**, 4305-4311. https://doi.org/10.1109/JSEN.2013.2267728

[13] Nozaki, Y. and Nakamoto, T. (2016) Odor Impression Prediction from Mass Spectra. *PLoS ONE*, **11**, e0157030. https://doi.org/10.1371/journal.pone.0157030

[14] Nozaki, Y. and Nakamoto, T. (2018) Predictive Modeling for Odor Character of a Chemical Using Machine Learning Combined with Natural Language Processing. *PLoS ONE*, **13**, e0198475. Erratum in: *PLoS ONE*, **13**, e0208962. https://doi.org/10.1371/journal.pone.0208962

[15] Karakaya, D., Ulucan, O. and Türkan, M. (2020) Electronic Nose and Its Applications: A Survey. *International Journal of Automation and Computing*, **17**, 179-209. https://doi.org/10.1007/s11633-019-1212-9

[16] McLafferty, F. (2011) A Century of Progress in Molecular Mass Spectrometry. *Annual Review of Analytical Chemistry*, **4**, 1-22. https://doi.org/10.1146/annurev-anchem-061010-114018

[17] Mauri, A., Consonni, V., Pavan, M., Todeschini, R. and Chemometrics, M. (2006) Dragon Software: An Easy Approach to Molecular Descriptor Calculations.

[18] West, D.B. (2001) Introduction to Graph Theory. Vol. 2, Prentice Hall, Upper Saddle River.

[19] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C. and Yu, P.S. (2021) A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, **32**, 4-24. https://doi.org/10.1109/TNNLS.2020.2978386

[20] Buck, L. and Axel, R. (1991) A Novel Multigene Family May Encode Odorant Receptors: A Molecular Basis for Odor Recognition. *Cell*, **65**, 175-187. https://doi.org/10.1016/0092-8674(91)90418-X