论文分享 | Capsule Graph Neutral Network

原创： 苏涛 [极验](javascript:void(0);) 今天

图卷积生成的高质量节点 embedding 被广泛应用到各种基于节点的应用和方法。本文的作者受到17年胶囊网络的启发，提出胶囊图神经网络CapsGNN。

通过以胶囊的形式抓取节点的特征，

利用路由机制来捕获图表级别的重要信息。

这样，CapsGNN 模型

生成多个graph级别的 embeddings，他们从不同的方面捕获 graph 的属性。

CapsGNN 模型中的注意力模块用于处理各种尺寸的 graph，使模型能够专注于图的关键部分。

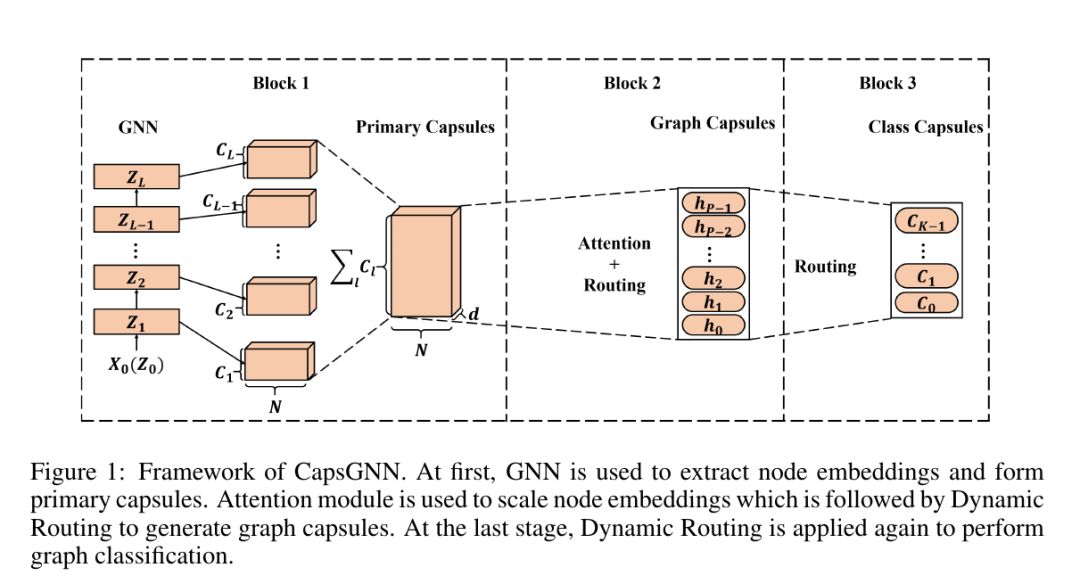
作者在10个图结构数据集上，对模型做了广泛的评估，表明 CapsGNN 拥有捕获图宏观属性的强大机制，在一些图分类任务上，优于其他 state-of-the-art 的方法。

# 胶囊图神经网络

CapsGNN 使用提取自 GNN 的节点特征生成一系列高质量的图 embeddings。每个 graph 都拥有多个 graph 级别的 embedding，每个 embedding 从一个独特的角度反映图的属性。进一步说，通过 GNN 以胶囊的形式提取基础的节点特征，路由机制生成高级的图capsules和类capsules。attention 模块在生成 graph capsules 的过程中解决图尺寸不一致的问题。

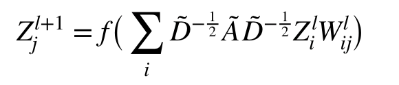
胶囊图神经网络由三个主要的 block 组成：

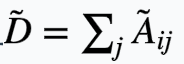
* 基础节点胶囊提取block：GNN提取具有不同感知域的局部顶点特征，产生初级的胶囊。
* 高级图胶囊提取block：注意力模块和动态路由融为一体，生成图的多个胶囊。
* 图分类block：动态路由再次被调用，用于生成图分类的类胶囊。



# 基础节点胶囊

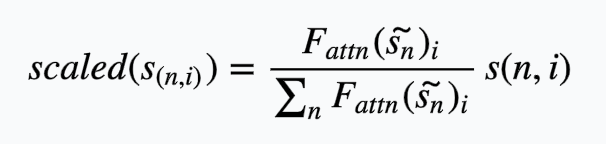
使用 Kipf & Welling (2017) 的 GCN 框架结构作为基础的节点特征抽取器。在没有节点属性的情况下，节点的度可以作为属性。唯一不同的是，从GCN 的每一层抽取 multi-scale 的节点特征，将这些抽取的特征以胶囊的形式表示。



其中https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VqUyIhCcnrEKFgfic4t9hqqvJxticlPXtOHM9PfmyDTvcDkbObSOzQW9A/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1表示参数矩阵， 它是从第 l 层的 i channel 到第l+1层的 j channel 的滤波器。选择激活函数https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VqeycWQM2j02v27BystFJwSj182QCMDUZ1M0IicmdlgWovY4owpZvGSw/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1，https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VD9IWH4PiaZDAsrxluFUX55nzsnWRrZMgHOibmq8Lgq3Uuegnyxv2uoIQ/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1 ，https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VP2l07wjVzBNnp01nQ9ZynfMONv8KuRqeZPWvrxibQS8EYgEl2fIVTFQ/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1并且。为了保留不同尺寸的子组件的特征，作者从 GNN 的所有层来抽取节点特征，来生成高级别的 capsules。

# 高级图胶囊

得到了局部节点的胶囊，通过全局路由机制生成图胶囊。block 的输入是 N 组节点胶囊。每组是https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VkaEJ2eJT81Wq65oiaVnD4iajM7sXnvTc9IianGH4G1JibXiaPUj65YSwfbg/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1，https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VNBypwAqtBUnicBJuF9HPpbjveGX7v2ibuHdBmaRzvUlibjgTlEy91AFJw/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1其中 https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VweYEG2z40WYpKldw24u1zSdk7OribMZ1qRic6yEqWS21icXDjxYIXicc0w/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1 是 GNN 第l层的 channel 数，d 表示每个胶囊的维度。block 的输出是一系列图胶囊https://mmbiz.qpic.cn/mmbiz_jpg/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VOL9HY7VWE1zZIUCbZ3ibzD02cicr0ES4fGqibnHqEjIxCxuoyfXDPNTqg/640?wx_fmt=jpeg&wxfrom=5&wx_lazy=1&wx_co=1。每个图胶囊从不同方面反映了图的属性。胶囊的长度反映了这些属性存在的可能性，角度反映了图形属性的细节。在使用节点胶囊生成图胶囊之前，注意力模块归一化节点胶囊。

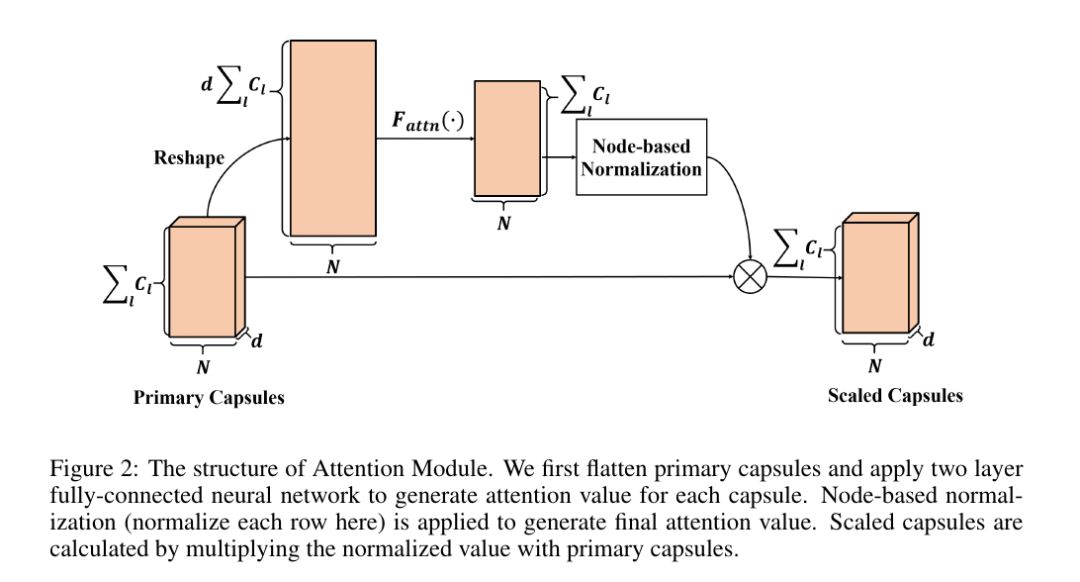


## Attention Module

初级胶囊是基于节点提取的，这就意味着，初级胶囊的个数依赖于输入图表的大小。这样，如果路由机制直接应用在初级胶囊上，生成的高级胶囊的值也严重依赖图表大小，这不是理想的状态。引入 attention 机制来解决这个问题。

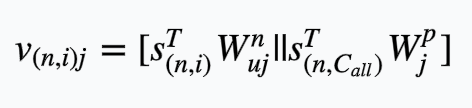
这里选择的 attention 方法是一个两层的全连接网络。输入是的矩阵，其中，输出维度是。应用基于节点的归一化来在每个通道中生成注意值，缩放原始的节点胶囊。

上式中， 是由节点n的所有 capsules 通过 concate 得到。 节点 n 的第 i 个 capsule。 是生成的 attention 值。通过这种方式消除图尺寸的影响，系统结构将侧重于输出 graph 的更重要部分。



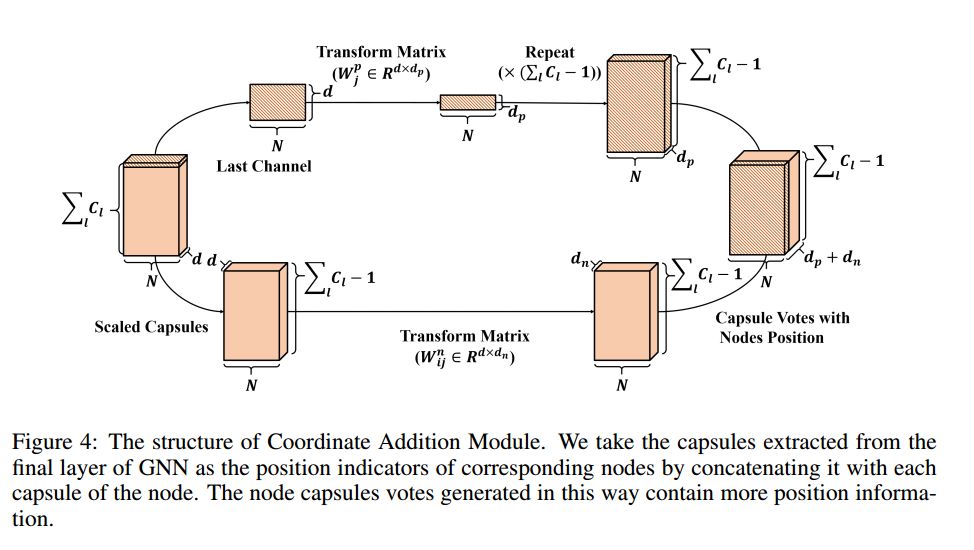
## 选择性优化

接着 Attention 模块之后，是一个坐标加法模块，它可以在生成节点 capsule 投票的过程中，保持节点的坐标信息。作为一个选择性优化，在一些数据集上并不需要使用到它。当 GNN 层数加深时，抽取的节点特征包含更具体的位置信息。利用节点位置指针来计算投票的过程

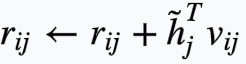


其中，https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VicnEgJFbw5GEhZDSWuVia9x6k9dB3QZW0SZetOq6wg2wSwDfaW13kMGg/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1 代表从第 n 个节点的 channel i 到图capsules 的 channel j 的 capsules 投票。 https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VlYAxV9HJYC8U5GzX3aepGfv1FS40icRAauicBIRPibRrKMz2Kpp4JbZsw/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1 是转化矩阵。https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VMbEtJUGzD6ugictWncJ4F5uBgGx0CicWPZicO2NbZVSibhytRz0MYGiacaw/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1

节点 n 的第 i 个 capsule，https://mmbiz.qpic.cn/mmbiz_jpg/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VOFg9j7JjV4UOIAdAh8uVHYfxJhxQQSONPhJTnXndicNU1SNTxoQxYGg/640?wx_fmt=jpeg&wxfrom=5&wx_lazy=1&wx_co=1是 concatenate 操作。



## 动态路由机制

动态路由机制，输入子 capsules S ,一系列训练转化矩阵 W ,迭代次数 t。具体操作如下，对于所有子 capsules i 有 https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8V4vKC7PRiaSUCdrtlCKKZHvV4sXnoXIKeqRSE9QLc0icMrF4jIloTVUaA/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1 ; 对于子胶囊 i 到父胶囊 j 初始化https://mmbiz.qpic.cn/mmbiz_jpg/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VEgcWI9BmAib0ersSEt8HI0OwrXGqcXZice6OlrysgIWPXMLfIbAO2XNQ/640?wx_fmt=jpeg&wxfrom=5&wx_lazy=1&wx_co=1，进行t次迭代，每次迭代中，对子胶囊 i，更新https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VapFQMrCwIiaHviczdo594uZjCegs0qbnWgv0RnHicwyHxN2FqtSoia1iaFw/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1 ；对于父胶囊 j，更新  https://mmbiz.qpic.cn/mmbiz_jpg/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VQ0EMUgsibVS7O4PtEibvWFDvGPdprcheyuHmEPLgyhJPxSO42JW0174A/640?wx_fmt=jpeg&wxfrom=5&wx_lazy=1&wx_co=1，https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VnOibFbsfgplF9gCicvV0y2Kgfte3gljJo3Oso0bhWPMXeJkb1TiaCvqHA/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1 ；对子胶囊 i 到负胶囊 j，更新 。t 次迭代后得到的https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VKrE8Mj8lbuX7FeCbm90g97kfhic4KPiaG1fcWYInc0xMlShHibRTCMfHQ/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1 就是输出的父胶囊。

因此，生成多个图 capsules 的过程归纳为：

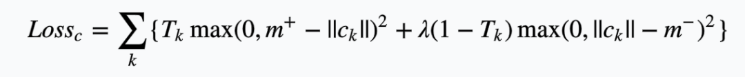
1、缩放初始 capsules，运用 Attention 模块来缩放初始 capsules。它的输出是https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VibS0tlKcDcAeAKSlznHnwhnYW0kicQCYwC9knDSm456iab4zhdicwl2Ejw/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1。

2、计算投票。当计算投票时，来自同一 channel 的不同节点的 capsules共享同一个转化矩阵。这一步的结果是一组投票 https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VvMSO1ve0xNpFgLCopicxpHmz6tK6YI0NAticDBBAOHWR4vWVdFCW0z0A/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1，其中，https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VgsOeCfuLicmunAZaWmskZnJib7iapYkuHKepqMX5lemDeOHyaO6ya3sow/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1表示 channels数，P 表示定义的图 capsules 的个数。

3、动态路由机制。基于前一步产生的投票，计算出高级的图 capsules。

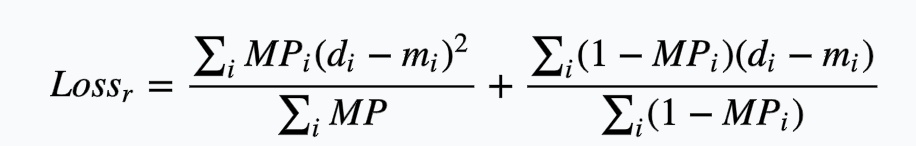
# 图分类

这个block的设计，是用来利用graph capsules做图分类任务。分类Loss，再次利用动态路由，生成分类 capsules https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VPhJycCibhHbdYVJxEicyU0hb7T71dQc5Cybb5GVTdKvsnawjvAHqOlPQ/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1，其中K是图的类别数量。使用差额损失函数，计算如下：



其中，https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VeL33Svic1NrIwfSlSqyBfTW7eIFquKafmiclZfiaO9WzgrglWXsMTR58g/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1 如果输入的graph属于类别 k，那么https://mmbiz.qpic.cn/mmbiz_jpg/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VNd5nuU9F7aVlO0bsUWEhDScNOkgGad7xbUBrKQNWfwaSzqcDn1rl3g/640?wx_fmt=jpeg&wxfrom=5&wx_lazy=1&wx_co=1。使用https://mmbiz.qpic.cn/mmbiz_jpg/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8VYgkFBnWQbwykx9iaUOUHsTgLnyTSlFaqjEYCjLnYXicqiaGJYwibqJM2kQ/640?wx_fmt=jpeg&wxfrom=5&wx_lazy=1&wx_co=1停止初步学习，避免类胶囊的长度，尤其是在K非常大的时候。

使用重建损失作为正则化方法。除了正确的胶囊之外，所有类胶囊都被屏蔽，并且用两个全连接层解码，以重建输入信息。这里重建的是输入节点的直方图 ：



其中 https://mmbiz.qpic.cn/mmbiz_png/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8V5YjtiaLJNKLFXOfc8RiarWq6udvEW2dk2qBibJNqkKfuKDeVbCpg8vA2A/640?wx_fmt=png&wxfrom=5&wx_lazy=1&wx_co=1 代表属性i出现在输入grraph中的节点数，https://mmbiz.qpic.cn/mmbiz_jpg/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8Vp75zZHwfrw7PcmC3FmnqFweCUl1ic7nKDUzZJsAJPvZgCvR3d0byKoQ/640?wx_fmt=jpeg&wxfrom=5&wx_lazy=1&wx_co=1表示对应的解码值。如果输入graph包含属性i的节点，那么https://mmbiz.qpic.cn/mmbiz_jpg/8VOiack4f29ewx0rgn8hQFDAFhAIzGc8V4ka0CbkhURpFXSCvktYPKwE7OeYyBrNiaB4SexF0YWZe6icrhanqGN3Q/640?wx_fmt=jpeg&wxfrom=5&wx_lazy=1&wx_co=1。

**实验**

图分类的目的是通过分析图的结构和节点标签信息，来预测图的类别。作者使用基于核的方法和基于深度学习的方法作为baseline。分别在5组生物图数据集和5组社交网络图数据集上，比较CapsGNN的效果。作者还设置了分析Gapsules效率的对比实验。

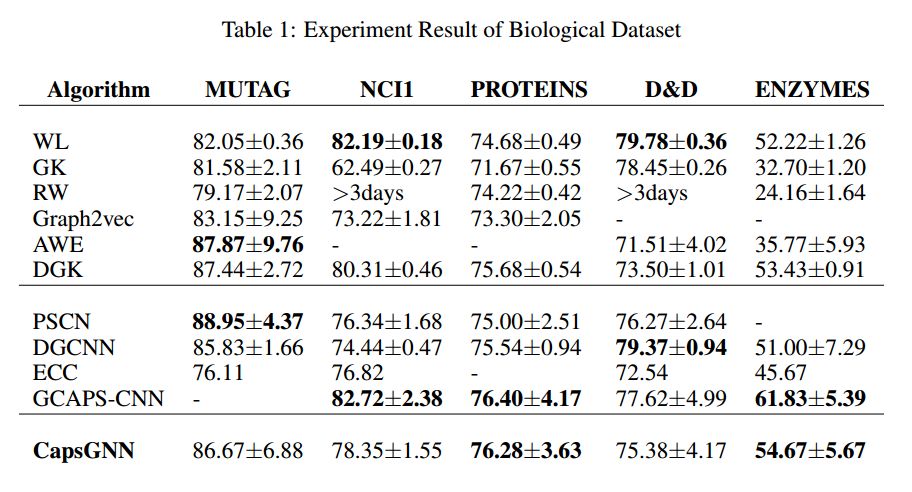
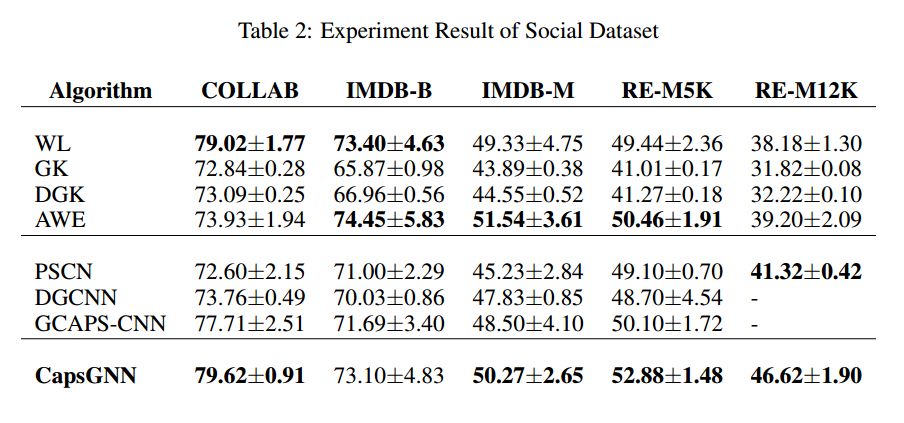
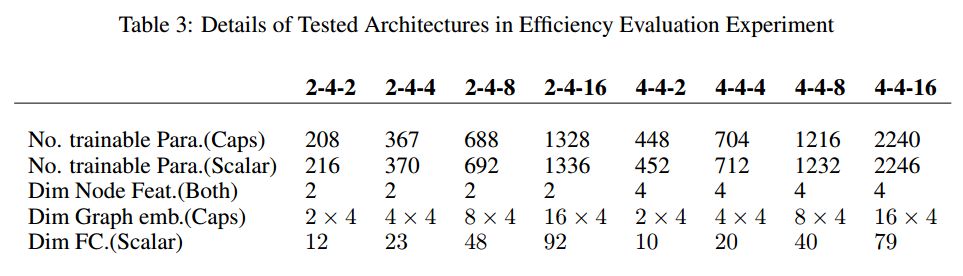


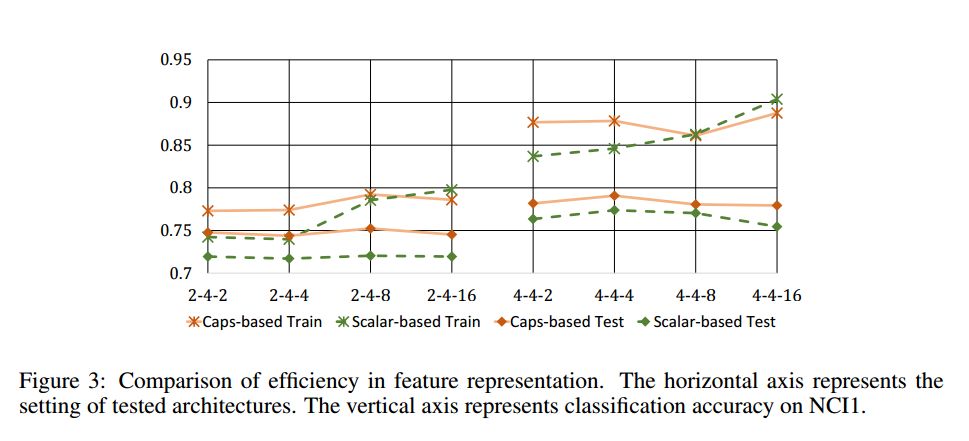
表1列出在生物图数据集上的实验结果，表2展示模型在社交网络数据集上的表现。每个数据集，高亮准确度 top2 的结果。比较其他算法，CapsGNN 在其中6个数据上取得了 top2 的表现。与其他端到端的架构相比，GapsGNN 在社交网络数据上均达到了 top1。



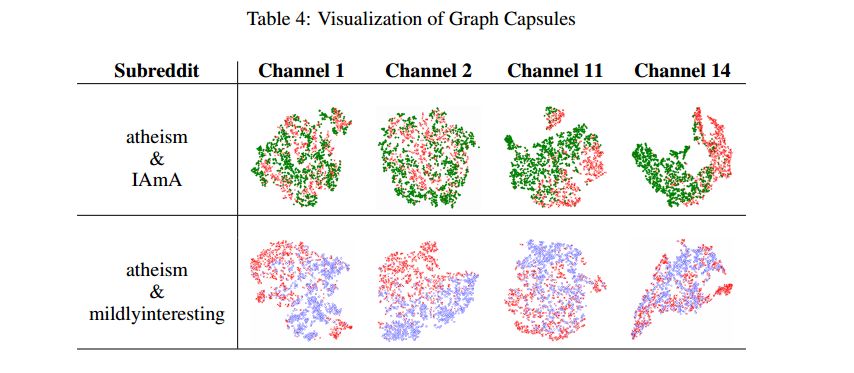
实验的目的是检测 capsules 在 encodegraph 特征的效率。高效是指在相同数量的神经元上特征编码得到更多的信息。作者为每个 CapsGNN 构建一个 scalar-based 神经网络，然后通过比较训练准确度和测试准确度，来比较CapsGNN和它的scalar-base 网络。



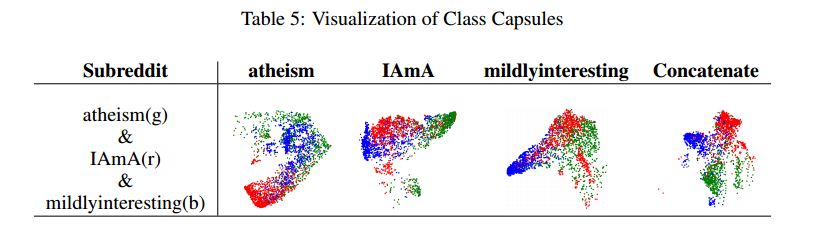
从下图可以看到，CapsGNN 的测试准确度在任何设定下，均高于对应的scalar-based结构。再看训练准确度，当FC的维度稍微高于 CapsGNN 的graph embedding 维度时，CapsGNN 任然达到了很高的准确度。这表明，CapsGNN 表示整个数据集的能力更加强大。



CapsGNN利用capsules得到每个graph的多个embedding，所以可以有效的捕获更复杂的潜在信息。为了探索 graph/calss capsules 的这种潜在属性，作者对从各个不同 channel 提取的 capsules 的分布做了 t-SNE 可视化



从 graph capsules 的第1、第2、第14 channel提取出类别分别为atheism、IAmA 和 mildlyinteresting 的 capsules，不同的 capsules 代表 graph 属性的不同方面。atheism and IAmA 两类可以从11和14 channel 提取出的 capsules 很好的区分，但是很难从1和 2 channel的 capsule 区分。相反atheism和mildlyinteresting能用1和2 channel提取的capsules区分开。这种情况同样出现再其他多类别数据上。虽然很难指出这些capsules聚焦在那些具体的方面，但是比较scalar-base结构，使用多个embedding对对象建模，能够探索每个channel的意义，这能够知道模型学习到更多可解释的embedding。



不同的class capsules专注于不同的类别相关的图属性，如图5，代表atheism的capsules可以很好的区分athesism和其他另外两个类别。他们在这个capsules上是混淆无法区分的。这种现象在其他class capsules上也出现了。如果将这三个capsules concatenate在一起，三个不同类型的graph都能通过这个concat capsules区分，它直接反映了分类表现。

**论文链接:**

https://openreview.net/pdf?id=Byl8BnRcYm