

无监督模型评估方法、装置、服务器及可读存储介质

技术领域

[01] 本发明涉及计算机技术领域，尤其涉及一种无监督模型评估方法、装置、服务器及可读存储介质。

5 背景技术

[02] 随着科学技术的不断发展，无监督学习算法得到了广泛应用。无监督学习模型可以对图数据进行计算，将图数据中的每个结点都映射到一个向量空间中，得到计算后的结点向量。在对计算后得到的结点向量进行评估时，通常采用有监督算法评估，再训练一个机器学习模型。但训练机器学习模型需要耗费更多的数据和时间。

10 发明内容

[03] 本说明书实施例提供及一种无监督模型评估方法、装置、服务器及计算机可读存储介质。

[04] 第一方面，本说明书实施例提供一种无监督模型评估方法，包括：

[05] 在经过无监督模型得到的 N 个结点向量中确定出与 S 个结点对应的 S 个结点向量；

15 [06] 确定所述 S 个结点中每个结点的邻居结点，以及所述每个结点的非邻居结点；

[07] 根据所述每个结点的邻居结点的结点向量，以及所述每个结点的结点向量，确定所述每个结点的邻居结点与所述每个结点的相似度作为正样本的预测值；

[08] 根据所述每个结点的非邻居结点的结点向量，以及所述每个结点的结点向量，确定所述每个结点的非邻居结点与所述每个结点的相似度作为负样本的预测值；

20 [09] 根据所述正样本的预测值以及所述负样本的预测值，对所述无监督模型进行评估。

[10] 第二方面，本说明书实施例提供一种无监督模型评估装置，包括：

[11] 第一确定模块，用于在经过无监督模型得到的 N 个结点向量中确定出与 S 个结点对应的 S 个结点向量， N 和 S 均为正整数， N 大于等于 S ；

25 [12] 第二确定模块，用于确定所述 S 个结点中每个结点的邻居结点，以及所述每个结点的非邻居结点；

[13]正样本预测值确定模块，用于根据所述每个结点的邻居结点的结点向量，以及所述每个结点的结点向量，确定所述每个结点的邻居结点与所述每个结点的相似度作为正样本的预测值；

5 [14]负样本预测值确定模块，用于根据所述每个结点的非邻居结点的结点向量，以及所述每个结点的结点向量，确定所述每个结点的非邻居结点与所述每个结点的相似度作为负样本的预测值；

[15]评估模块，用于根据所述正样本的预测值以及所述负样本的预测值，对所述无监督模型进行评估。第三方面，本说明书实施例提供一种服务器，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，所述处理器执行上述任一项所述方法的
10 的步骤。

[16]第四方面，本说明书实施例提供一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现上述任一项所述方法的步骤。

[17]本说明书实施例有益效果如下：

[18]在本说明书实施例提供的无监督模型评估方法中，在经过无监督模型得到的N个结点向量中确定出与S个结点对应的S个结点向量，N和S均为正整数；确定所述S个结点中每个结点的邻居结点，以及所述每个结点的非邻居结点；根据所述每个结点的邻居结点的结点向量，以及所述每个结点的结点向量，确定所述每个结点的邻居结点与所述每个结点的相似度作为正样本的预测值；根据所述每个结点的非邻居结点的结点向量，以及所述每个结点的结点向量，确定所述每个结点的非邻居结点与所述每个结点的相似
15 度作为负样本的预测值；根据所述正样本的预测值以及所述负样本的预测值，对所述无监督模型进行评估。本说明书实施例中的方案，将每个结点与邻居结点的相似度作为正样本，以及将每个结点与非邻居结点的相似度作为负样本，通过对正样本以及负样本进行评估来确定无监督模型的效果，不需要进行额外的模型训练，使无监督模型的评估过程更加容易实现。
20

25 附图说明

[19]通过阅读下文优选实施方式的详细描述，各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的，而并不认为是对本发明的限制。而且在整个附图中，用相同的参考符号表示相同的部件。在附图中：

[20]图 1 为本说明书实施例第一方面提供的一种无监督模型评估方法的流程图;

[21]图 2 为本说明书实施例示出的根据二元组样本对无监督模型进行评估的方法流程图;

[22]图 3 为本说明书实施例第二方面提供的无监督模型评估装置的示意图;

[23]图 4 为本说明书实施例第三方面提供的服务器的示意图。

5 具体实施方式

[24]为了更好的理解上述技术方案,下面通过附图以及具体实施例对本说明书实施例的技术方案做详细的说明,应当理解本说明书实施例以及实施例中的具体特征是对本说明书实施例技术方案的详细的说明,而不是对本说明书技术方案的限定,在不冲突的情况下,本说明书实施例以及实施例中的技术特征可以相互组合。

10 [25]第一方面,本说明书实施例提供一种无监督模型评估方法,如图 1 所示,为本说明书实施例提供的无监督模型评估方法的流程图,该方法包括以下步骤:

[26]步骤 S11:在经过无监督模型得到的 N 个结点向量中确定出与 S 个结点对应的 S 个结点向量, N 和 S 均为正整数;

15 [27]本说明书实施例中, N 个结点向量可以是无监督模型对 N 个输入数据进行计算得到的。其中,可以将 N 个输入数据看作为 N 个结点。输入数据可以根据实际情况确定的数据,在一个实施例中,输入数据可以是一定数量的用于表示词语的向量,可以将每个词语看作是一个结点,每个词语可以用一个初始向量来表示,将词语的初始向量作为无监督模型的输入,通过无监督模型的计算,输出与每个结点对应的 embedding (嵌入向量),即输出的结点向量。应理解的是,embedding 可以看作是将结点由原始空间映射到另一个空间中,并且在另一个空间内保持结点在原始空间的结构信息和距离信息。

20 [28]无监督模型可以根据实际需要进行选择,在一个实施例中,无监督模型可以为 Word2Vec 模型,在对词语进行处理时,通过 Word2Vec 模型将每个词语映射到一个低维向量空间,得到与每个词语结点对应的结点向量。

[29]应理解的是,在无监督模型的计算过程中,最后输出的结点向量可以由损失函数来
25 确定。例如, NCE (Noise Contrastive Estimation, 噪音对比估计) 损失函数,通过 NCE 损失函数得到的结点向量能够使该结点具有与其邻居结点在向量空间中的距离更近,与非邻居结点距离更远的特点。

[30]在确定 S 个结点时，可以在 N 个结点中选择 S 个不同的结点，也可以在 N 个结点中随机选取 S 个结点。S 个结点可以是一次性选取的，也可以是多次选取的，这里不做限定。

5 [31]步骤 S12: 确定所述 S 个结点中每个结点的邻居结点，以及所述每个结点的非邻居结点；

[32]本说明书实施例中，针对 S 个结点的每个结点来说，该结点可以包括多个邻居结点以及多个非邻居结点，可以在多个邻居结点中随机选出任意个数的邻居结点以及在多个非邻居结点中随机选出任意个数的非邻居结点。或者，可以设定选出的邻居结点的个数以及设定选出非邻居结点的个数。这里不做限定。

10 [33]一个结点的邻居结点可以根据预设的邻居结点生成规则进行生成，也可以将 N 个结点根据结点之间的关系构成一个图，在构成的图中进行邻居结点的选择。在一个实施例中，可以根据各个结点之间的关系进行边连接，形成包括有结点之间连接关系的图结构，在该图结构中查找 S 个结点中每个结点的邻居结点和非邻居结点。在另一个实施例中，可以根据预设的邻居生成算法生成每个结点的邻居结点。当然，还可以采用其他方式
15 来确定邻居结点，本说明书实施例不做限定。非邻居结点的确定与邻居结点的确定类似，这里就不再赘述了。

[34]步骤 S13: 根据所述每个结点的邻居结点的结点向量，以及所述每个结点的结点向量，确定所述每个结点的邻居结点与所述每个结点的相似度作为正样本的预测值；

20 [35]步骤 S14: 根据所述每个结点的非邻居结点的结点向量，以及所述每个结点的结点向量，确定所述每个结点的非邻居结点与所述每个结点的相似度作为负样本的预测值；

[36]本说明书实施例中，在确定了每个结点的邻居结点以及非邻居结点之后，可以将邻居结点作为正样本，将非邻居结点作为负样本。针对每个结点，分别计算该结点与各个邻居结点之间的相似度，作为邻居结点的预测值，以及分别计算该结点与各个非邻居结点之间的相似度，作为非邻居结点的预测值。相似度的计算可以根据实际需要来选择，
25 在一个实施例中，可以通过计算向量之间的内积，并对内积做归一化处理得到。

[37]应理解的是，相似度可以是 0 到 1 之间的数，如果两个结点之间的相似度越接近 1，表明两个结点互为邻居结点的概率越大。

[38]步骤 S15: 根据所述正样本的预测值以及所述负样本的预测值，对所述无监督模型进行评估。

[39]本说明书实施例中，邻居结点的预测值越接近 1，非邻居结点的预测值越接近 0 则表示无监督模型的效果越好。反之，邻居结点即正样本的预测值越接近 0，非邻居结点即负样本的预测值越接近 1，无监督模型的效果越差。因此，可以通过正样本的预测值以及负样本的预测值来评估无监督模型的好坏，这种方式更加通用以及易实现。

- 5 [40]可选地，所述在经过无监督模型得到的 N 个结点向量中确定出与 S 个结点对应的 S 个结点向量，包括：确定结点集合，所述结点集合包含 N 个结点，所述 N 个结点与所述 N 个结点向量相对应；在所述结点集合中随机选取一个结点作为采样结点，共计选取 S 次，获得所述 S 个结点。

- 10 [41]本说明书实施例中，可以将经过无监督模型计算得到的 N 个结点向量对应的 N 个结点构成结点集合。在选择 S 个结点时，可以在结点集合中有放回的选取 S 次，每次在 N 个结点中随机选取一个作为采样结点。

- [42]可选地，所述确定所述 S 个结点中每个结点的邻居结点，以及所述每个结点的非邻居结点，包括：针对所述 S 个结点中的每个结点，执行以下步骤：根据预设邻居生成算法，在 N 个结点中确定该结点的邻居结点集合，以及该结点的非邻居结点集合；在所述
15 邻居结点集合中确定一个或多个邻居结点；在所述非邻居结点集合中确定一个或多个非邻居结点。

- [43]本说明书实施例中，针对 S 个结点中的每个结点，都可以采用预设邻居生成算法来获取该结点的邻居结点，预设邻居生成算法可以根据实际需要进行选择，如一阶邻居算法、随机游走邻居算法等。在一个实施例中，为了保证生成的邻居结点能够与无监督模型的
20 计算过程中采用的结点关系保持一致，预设邻居生成算法与无监督模型计算过程中采用的邻居生成算法相同。

- [44]在一个实施例中，当采用预设邻居生成算法生成一个结点的邻居结点集合时，那么非邻居结点可以通过在结点集合中减去邻居结点集合得到。应理解的是，对于每个结点来说，对应的邻居结点集合所包含的元素个数可能不同，例如，第一结点拥有 5 个邻居
25 结点，第二结点拥有 8 个邻居结点。在进行邻居结点的选择时，邻居结点的个数可以是固定的，也可以是随机的。

[45]仍以上面第一结点和第二结点为例，在一个实施例中，针对每个结点选取相同个数的邻居结点，例如，均选取 1 个，则可以在第一结点的 5 个邻居结点中随机选出 1 个邻居结点，在第二结点的 8 个邻居结点中随机选出 1 个邻居结点。在另一个实施例中，针

对每个结点随机确定邻居结点的个数，如在第一结点中的 5 个邻居结点中随机选取 2 个邻居结点，在第二结点中的 8 个邻居结点中随机选取 3 个邻居结点等。当然，针对每个结点，不论使用哪种方式确定邻居结点，确定的邻居结点个数均小于或等于该结点实际对应的邻居结点的总数。

- 5 [46]对于每个结点对应的非邻居结点，确定方式与上述邻居结点的确定方式类似，非邻居结点的数量可以根据实际需要进行设定，在一个实施例，在每个非邻居结点集合中可以随机选取 4~20 个结点作为该结点的非邻居结点。非邻居结点的确定方式请参考上述确定邻居结点的描述，这里就不再赘述了。

- 10 [47]在一种可选的实现方式中，所述根据所述正样本的预测值以及所述负样本的预测值，对所述无监督模型进行评估，包括：根据所述正样本的预测值，以及正样本标签值，构建第一类二元组样本；根据所述负样本的预测值，以及负样本标签值，构建第二类二元组样本；根据所述第一类二元组样本以及所述第二类二元组样本，对所述无监督模型进行评估。

- 15 [48]本说明书实施例中，可以构建一个验证集，通过验证集来确定评估无监督模型的效果。在一个实施例中，验证集初始可以设置为空集，在根据正样本和负样本构建二元组样本后，将二元组样本添加至初始的空集中，得到最终的验证集。

- 20 [49]为了对邻居结点和非邻居结点进行区分，本说明书实施例中可以通过设置正样本标签值以及负样本标签值来实现。在一个实施例中，正样本标签值设置为 1，负样本标签值为 0，那么对于结点的一个邻居结点来说，可以根据邻居结点的正样本的预测值以及正样本标签值来确定一个第一类二元组样本，例如，一个邻居结点的正样本的预测值为 0.9，则对应的第一类二元组样本为 (1,0.9)。对于结点的一个非邻居结点来说，可以根据非邻居结点的负样本的预测值以及负样本标签值来确定一个第二类二元组样本，例如，一个非邻居结点的负样本的预测值为 0.3，则对应的第二类二元组样本为 (0,0.3)。

- 25 [50]根据上述构建的验证集，可以对无监督模型进行评估，如图 2 所示，为本说明书实施例中根据第一类二元组样本以及所述第二类二元组样本，对所述无监督模型进行评估的方法流程图，该方法包括以下步骤：

[51]步骤 S21：根据所述第一类二元组样本以及所述第二类二元组样本，确定目标受试者工作特征曲线；

[52]步骤 S22：获取所述目标受试者工作特征曲线下的面积，以评估所述无监督模型。

[53]在该实施例中，可以对第一类二元组样本以及第二类二元组样本组成的验证集进行二分类，并通过确定目标受试者工作特征曲线下的面积来评价二分类的优劣。

[54]应理解的是，在该实施例中，预测值可以用来表示两个结点互为邻居结点的概率。

在绘制目标受试者工作特征（Receiver Operating Characteristic，ROC）曲线时，可以根据

5 据正样本的预测值以及负样本的预测值按照预测值的大小进行排序。通过选取不同的阈值来对排序好的样本进行二分类，即，预测值大于或等于阈值的二元组样本预测为正样本，预测值小于该阈值的二元组样本预测为负样本。再根据二元组样本中实际对应的标签值来计算该阈值下的真正例率（True Positive Rate，TPR）和假正例率（False Positive Rate，FPR）。

10 [55]在计算真正例率，假正例率时会涉及到以下四种情况：真正例（True positive，TP），假正例（False positive，FP），假反例（False negative，FN），真反例（True negative，TN）。其中，TP是指预测是正样本，且预测对了，FP是指预测是正样本，且预测错了，FN是指预测是负样本，且预测错了，TN是指预测是负样本，且预测对了。以上述二元组样本为例，如果一个二元组样本，预测值大于或等于阈值，则表明该二元组样本被预

15 测是正样本，再查看该二元组样本的标签值，如果是正样本标签值，则表明该二元组样本为 TP，如果标签值是负样本标签，则表明该二元组样本为 FP。如果一个二元组样本，预测值小于阈值，则表明该二元组样本被预测是负样本，再查看该二元组样本的标签值，如果标签值是正样本标签值，则该二元组样本为 FN，如果标签值是负样本标签值，则该二元组样本为 TN。

20 [56]计算真正例率 $TPR=TP/(TP+FN)$ ，以及假正例率 $FPR=FP/(TN+FP)$ 。其中，真正例率代表预测为正样本且预测对了的可能性，假正例率代表预测为正样本但是预测错了的可能性。

[57]将真正例率作为纵轴，假正例率作为横轴，得到一个构成 ROC 曲线的坐标。在一个实施例中，可以将每个二元组样本的预测值作为一个阈值，均计算该阈值下的真正例

25 率以及假正例率。例如，如果二元组样本的个数为 M，则对应应有 M 个预测值，将 M 个预测值按照大小顺序进行排列，并将每个预测值作为一个阈值，当二元组样本的预测值大于或等于这个阈值时，预测该二元组样本为正样本，当二元组样本的预测值小于这个阈值时，预测该二元组样本为负样本。每选取一个阈值，就可以得到一组 TPR 和 FPR，即 ROC 曲线上的点。在该实施例中，一共得到 M 组 TPR 和 FPR，根据这 M 个点可

30 以绘制 ROC 曲线。

[58]AUC 曲线 (Area Under ROC Curve) 为 ROC 曲线下的面积, 能够反映 ROC 曲线表达的分类能。通常来讲, AUC 的值越大, 表明分类的效果越好, 因此可以通过 AUC 来评估无监督模型的效果。应理解的是, 除了使用 AUC 来评估无监督模型的效果, 还可以使用其他方法, 例如 F1 值等, 这里不做限定。

5 [59]为了更好的理解本说明书实施例提供的无监督模型评估方法, 下面以经过无监督模型得到的结点向量为例, 来对本说明书实施例提供的无监督模型评估方法来进行说明。在该实施例中, 经过无监督模型计算得到的结点集合为 V , 与结点集合对应的结点向量集合为 E , 结点的采样数为 S , 验证集 T 初始为空集, 针对 S 个结点中的每个结点, 均确定 1 个邻居结点以及 N 个非邻居结点构建目标验证集。

10 [60]步骤一, 获取结点集合 V , 结点向量集合 E , 目标验证集 $T=\emptyset$, 结点的采样数 S ;

[61]步骤二, 设置循环次数 i , i 从 1 到 S , 在每次循环过程中, 执行以下步骤:

[62]从集合 V 中随机采样出一个结点, 记为 v_i ;

[63]根据预设邻居生成算法, 生成 v_i 的邻居结点集合以及 v_i 的非邻居结点集合, 并从邻居结点集合中随机采样 1 个邻居结点, 即为 p_i ;

15 [64]从集合 E 中查询得到 v_i 的向量, 以及 p_i 的向量, 并计算两个向量之间的相似度, 即为 sam_{vp} ;

[65]构建二元组样本 $(1, sam_{vp})$, 加入验证集 T ;

[66]从非邻居结点集合中随机采样 N 个非邻居结点, 记为集合 N_i ;

20 [67]从集合 E 中查询得到结点 v_i 的向量, 以及 N_i 中各结点的向量, 然后计算 v_i 的向量与 N_i 中各结点向量之间的相似度, 将得到的 N 个相似度值记为集合 sam_{vn} ;

[68]对于集合 sam_{vn} 中的各元素 sam_{vnj} , 构造 N 个二元组 $(0, sam_{vnj})$, 加入验证集 T , 其中, j 的值为 $1\sim N$ 。

[69]步骤三: 集合 T 中的每个二元组可视为由标签值和预测值构成, 因此可使用二分类指标, 如 AUC、F1 值等对集合 T 进行评估, 评估结果可以作为衡量该无监督模型的学习效果的指标。

25

- [70]第二方面，基于同一发明构思，本说明书实施例提供一种无监督模型评估装置，请参考图 3，包括：
- [71]第一确定模块 31，用于在经过无监督模型得到的 N 个结点向量中确定出与 S 个结点对应的 S 个结点向量，N 和 S 均为正整数；
- 5 [72]第二确定模块 32，用于确定所述 S 个结点中每个结点的邻居结点，以及所述每个结点的非邻居结点；
- [73]正样本预测值确定模块 33，用于根据所述每个结点的邻居结点的结点向量，以及所述每个结点的结点向量，确定所述每个结点的邻居结点与所述每个结点的相似度作为正样本的预测值；
- 10 [74]负样本预测值确定模块 34，用于根据所述每个结点的非邻居结点的结点向量，以及所述每个结点的结点向量，确定所述每个结点的非邻居结点与所述每个结点的相似度作为负样本的预测值；
- [75]评估模块 35，用于根据所述正样本的预测值以及所述负样本的预测值，对所述无监督模型进行评估。
- 15 [76]在一种可选实现方式中，评估模块 35，用于：
- [77]根据所述正样本的预测值，以及正样本标签值，构建第一类二元组样本；
- [78]根据所述负样本的预测值，以及负样本标签值，构建第二类二元组样本；
- [79]根据所述第一类二元组样本以及所述第二类二元组样本，对所述无监督模型进行评估。
- 20 [80]在一种可选实现方式中，评估模块 35，用于：
- [81]根据所述第一类二元组样本以及所述第二类二元组样本，确定目标受试者工作特征曲线；
- [82]获取所述目标受试者工作特征曲线下的面积，以评估所述无监督模型。
- [83]在一种可选实现方式中，第二确定模块 32，用于：
- 25 [84]针对所述 S 个结点中的每个结点，执行以下步骤：
- [85]根据预设邻居生成算法，在 N 个结点中确定该结点的邻居结点集合，以及该结点的非邻居结点集合；

[86]在所述邻居结点集合中确定一个或多个邻居结点;

[87]在所述非邻居结点集合中确定一个或多个非邻居结点。

[88]在一种可选实现方式中, 第一确定模块 31 用于:

5 [89]结点集合获取模块, 用于确定结点集合, 所述结点集合包含 N 个结点, 所述 N 个结点与所述 N 个结点向量相对应;

[90]采样结点获取模块, 用于在所述结点集合中随机选取一个结点作为采样结点, 共计选取 S 次, 获得所述 S 个结点。

[91]关于上述装置, 其中各个模块的具体功能已经在本发明实施例提供的无监督模型评估方法的实施例中进行了详细描述, 此处将不做详细阐述说明。

10 [92]第三方面, 基于与前述实施例中无监督模型评估方法同样的发明构思, 本发明还提供一种服务器, 如图 4 所示, 包括存储器 404、处理器 402 及存储在存储器 404 上并可在处理器 402 上运行的计算机程序, 所述处理器 402 执行所述程序时实现前文所述无监督模型评估方法的任一方法的步骤。

15 [93]其中, 在图 4 中, 总线架构(用总线 400 来代表), 总线 400 可以包括任意数量的互联的总线和桥, 总线 400 将包括由处理器 402 代表的一个或多个处理器和存储器 404 代表的存储器的各种电路链接在一起。总线 400 还可以将诸如外围设备、稳压器和功率管理电路等之类的各种其他电路链接在一起, 这些都是本领域所公知的, 因此, 本文不再对其进行进一步描述。总线接口 406 在总线 400 和接收器 401 和发送器 403 之间提供接口。接收器 401 和发送器 403 可以是同一个元件, 即收发机, 提供用于在传输介质上
20 与各种其他装置通信的单元。处理器 402 负责管理总线 400 和通常的处理, 而存储器 404 可以被用于存储处理器 402 在执行操作时所使用的数据。

[94]第四方面, 基于与前述实施例中基于无监督模型评估方法的发明构思, 本发明还提供一种计算机可读存储介质, 其上存储有计算机程序, 该程序被处理器执行时实现前文所述基于无监督模型评估方法的任一方法的步骤。

25 [95]本说明书是参照根据本说明书实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器, 使得通过计算机或其他可编程数据处理设备的处理器执

行的指令产生用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的设备。

[96]这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令设备的制造品,该指令设备实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能。

[97]这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的步骤。

[98]尽管已描述了本发明的优选实施例,但本领域内的技术人员一旦得知了基本创造性概念,则可对这些实施例作出另外的变更和修改。所以,所附权利要求意欲解释为包括优选实施例以及落入本发明范围的所有变更和修改。

[99]显然,本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样,倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内,则本发明也意图包含这些改动和变型在内。

权利要求书

- 1、一种无监督模型评估方法，所述方法包括：

在经过无监督模型得到的N个结点向量中确定出与S个结点对应的S个结点向量，
N和S均为正整数；
- 5 确定所述S个结点中每个结点的邻居结点，以及所述每个结点的非邻居结点；

根据所述每个结点的邻居结点的结点向量，以及所述每个结点的结点向量，确定所述每个结点的邻居结点与所述每个结点的相似度作为正样本的预测值；

根据所述每个结点的非邻居结点的结点向量，以及所述每个结点的结点向量，确定所述每个结点的非邻居结点与所述每个结点的相似度作为负样本的预测值；
- 10 根据所述正样本的预测值以及所述负样本的预测值，对所述无监督模型进行评估。
- 2、根据权利要求1所述的无监督模型评估方法，所述根据所述正样本的预测值以及所述负样本的预测值，对所述无监督模型进行评估，包括：

根据所述正样本的预测值，以及正样本标签值，构建第一类二元组样本；

根据所述负样本的预测值，以及负样本标签值，构建第二类二元组样本；
- 15 根据所述第一类二元组样本以及所述第二类二元组样本，对所述无监督模型进行评估。
- 3、根据权利要求2所述的无监督模型评估方法，所述根据所述第一类二元组样本以及所述第二类二元组样本，对所述无监督模型进行评估，包括：

根据所述第一类二元组样本以及所述第二类二元组样本，确定目标受试者工作特征
- 20 曲线；

获取所述目标受试者工作特征曲线下的面积，以评估所述无监督模型。
- 4、根据权利要求1所述的无监督模型评估方法，所述确定所述S个结点中每个结点的邻居结点，以及所述每个结点的非邻居结点，包括：

针对所述S个结点中的每个结点，执行以下步骤：
- 25 根据预设邻居生成算法，在N个结点中确定该结点的邻居结点集合，以及该结点的非邻居结点集合；

在所述邻居结点集合中确定一个或多个邻居结点；

在所述非邻居结点集合中确定一个或多个非邻居结点。
- 5、根据权利要求1所述的无监督模型评估方法，所述在经过无监督模型得到的N
- 30 个结点向量中确定出与S个结点对应的S个结点向量，包括：

确定结点集合，所述结点集合包含N个结点，所述N个结点与所述N个结点向量

相对应;

在所述结点集合中随机选取一个结点作为采样结点, 共计选取 S 次, 获得所述 S 个结点。

6、一种无监督模型评估装置, 所述装置包括:

5 第一确定模块, 用于在经过无监督模型得到的 N 个结点向量中确定出与 S 个结点对应的 S 个结点向量, N 和 S 均为正整数;

第二确定模块, 用于确定所述 S 个结点中每个结点的邻居结点, 以及所述每个结点的非邻居结点;

10 正样本预测值确定模块, 用于根据所述每个结点的邻居结点的结点向量, 以及所述每个结点的结点向量, 确定所述每个结点的邻居结点与所述每个结点的相似度作为正样本的预测值;

负样本预测值确定模块, 用于根据所述每个结点的非邻居结点的结点向量, 以及所述每个结点的结点向量, 确定所述每个结点的非邻居结点与所述每个结点的相似度作为负样本的预测值;

15 评估模块, 用于根据所述正样本的预测值以及所述负样本的预测值, 对所述无监督模型进行评估。

7、根据权利要求 6 所述的无监督模型评估装置, 所述评估模块, 用于:

根据所述正样本的预测值, 以及正样本标签值, 构建第一类二元组样本;

根据所述负样本的预测值, 以及负样本标签值, 构建第二类二元组样本;

20 根据所述第一类二元组样本以及所述第二类二元组样本, 对所述无监督模型进行评估。

8、根据权利要求 7 所述的无监督模型评估装置, 所述评估模块, 用于:

根据所述第一类二元组样本以及所述第二类二元组样本, 确定目标受试者工作特征曲线;

25 获取所述目标受试者工作特征曲线下的面积, 以评估所述无监督模型。

9、根据权利要求 6 所述的无监督模型评估装置, 所述第二确定模块, 用于:

针对所述 S 个结点中的每个结点, 执行以下步骤:

根据预设邻居生成算法, 在 N 个结点中确定该结点的邻居结点集合, 以及该结点的非邻居结点集合;

30 在所述邻居结点集合中确定一个或多个邻居结点;

在所述非邻居结点集合中确定一个或多个非邻居结点。

10、根据权利要求 6 所述的无监督模型评估装置，所述第一确定模块，用于：

确定结点集合，所述结点集合包含 N 个结点，所述 N 个结点与所述 N 个结点向量相对应；

在所述结点集合中随机选取一个结点作为采样结点，共计选取 S 次，获得所述 S 个
5 结点。

11、一种服务器，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，所述处理器执行所述程序时实现权利要求 1-5 任一项所述方法的步骤。

12、一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现权利要求 1-5 任一项所述方法的步骤。