長安大學

毕业设计（论文）中英文对照翻译

题　　目　网络安全设备数据可视化发展技术及应用

学 院(部)　信息工程学院

专　　业　计算机科学与技术

班　　级　2016240202

学生姓名　薛硕

学　　号　2016902804

指导教师(签字) ____________

教学院长(签字) ____________

2019 年 12 月 26 日

1. 介绍

高维数据的可视化是许多不同领域中的重要问题，并且处理维度变化很大的数据。与乳腺癌有关的细胞核，例如，用大约 30 个变量来描述（Street 等，1993），而像素用于表示图像的强度矢量或用于表示文档的字数矢量通常具有数千个尺寸。在过去的几十年中，已经提出了这种高维数据的可视化，其中许多已由 Ferreira de Oliveira 和 Levkowitz（2003）。重要技术包括图像显示例如 Chernoff 的面孔（切尔诺夫，1973），基于像素的技术（Keim，2000）以及将数据的维表示为图形中的顶点（Battista 等，1994）。这些大多数技术只是提供工具来显示两个以上的数据维度，并将数据的解释留给观察者，这严重限制了这些技术对包含数千个高维数据点的真实世界数据集的适用性。

与上面讨论的可视化技术相比，降维方法将高维数据集 X = {x1，x2，...，xn}转换为二维或三维数据 Y = {y1，y2，... ，yn}可以显示在散点图中。在本文中，我们指的是低维数据表示 Y 作为地图，而将各个数据点的低维表示 yi 作为地图点。降维的目的是保留尽可能多的意义。在低维地图中，高维数据的结构可能不可行。已经提出了解决该问题的各种技术，它们保留的结构类型不同。传统降维技术，例如主成分分析（PCA； Hotelling）（1933））和经典多维比例缩放（MDS; Torgerson（1952））是线性技术，着重于使相异数据点的低维表示相距较远。对于位于低维非线性流形上或附近的高维数据，通常会更重要的是要使非常相似的数据点的低维表示形式紧密相连，使用线性映射通常是不可能的。

2. 随机邻居嵌入

随机邻居嵌入（SNE）首先将数据点之间的高维欧氏距离转换为表示相似度的条件概率 1。 相似点数据点 xj 到数据点 xi 是条件概率 pj | i，如果在以高斯为中心的情况下按与概率密度成比例的比例选择邻居，则 xi 将选择 xj 作为其邻居。xi。 对于附近的数据点，pj | i 相对较高，而对于广泛分离的数据点，pj | i 将几乎是无限的（对于高斯方差 σ i 的合理值）。 数学上条件概率 pj | i 是（谁）给的

$$p_{j|i} = \frac{\exp\left(-\|x_i - x_j\|^2 / 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\|x_i - x_k\|^2 / 2\sigma_i^2\right)},$$

其中σi 是以数据点 xi 为中心的高斯方差。 确定方法σi 的值在本节稍后介绍。 因为我们只对成对建模感兴趣相似之处，我们设置 pi | i 的值归零。 对于低维的 yi 和 yj 高维数点 xi 和 xj，可以计算出类似的条件概率，我们用 qj | i 表示。 我们 set2 计算中采用的高斯方差条件概率 qj | i 至√12。 因此，我们对映射点 yj 与地图的相似性进行建模

$$q_{j|i} = \frac{\exp\left(-\|y_i - y_j\|^2\right)}{\sum_{k \neq i} \exp\left(-\|y_i - y_k\|^2\right)}.$$

同样，由于我们只对建模成对相似性感兴趣，因此将 qi | i 设置为 0。如果地图点 yi 和 yj 正确地建模了高维数据点之间的相似度 xi 和 xj，条件概率 pj | i 和 qj | i 相等。 出于这一观察的动机，SNE 的目标是找到一种低维数据表示形式，以最小化 pj | i 之间的不匹配和 qj | i。 qj | i 对 pj | i 建模的真实性的自然度量是 Kullback-Leibler 散度（在这种情况下等于交叉熵，直到加法常数）。 SNE 使用梯度下降法将所有数据点的 Kullback-Leibler 散度之和最小化。成本函数 C 由下式给出

$$C = \sum_i KL(P_i \| Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}},$$

其中 Pi 表示在给定数据点 xi 的情况下所有其他数据点的条件概率分布，而 Qi 表示在给定的所有其他地图点上的条件概率分布易点图。 由于 Kullback-Leibler 散度不是对称的，因

此会产生不同类型的误差在低维图中的成对距离中的权重不均等。

特别是那里使用广泛分离的地图点表示附近的数据点（例如，使用较小的 qj | i 建模大 pj | i），但使用附近的地图点到代表广泛分离的数据点。 这么小的成本来自浪费一些可能性相关 Q 分布中的质量。 换句话说，SNE 成本函数侧重于保留地图中数据的局部结构（用于获得高斯方差的合理值）高维空间σi）。其余要选择的参数是高斯方差σi，该方差以每个中心为中心高维数据点，xi。 σi 不可能只有一个值最适合所有人数据集中的数据点，因为数据的密度可能会变化。在人口稠密的地区，较小的σi 值通常比稀疏地区更合适。σi 的任何特定值诱发概率分布，Pi，覆盖所有其他数据点。 该分布具有一个熵随着σi 增加增加。 SNE 对σi 值执行二进制搜索产生一个 Pi 用户指定的固定困惑。 困惑被定义为

$$Perp(P_i) = 2^{H(P_i)},$$

其中 H（Pi）是 Pi 的 Shannon 熵，以位为单位

$$H(P_i) = -\sum_j p_{j|i} \log_2 p_{j|i}.$$

困惑可以解释为邻居有效数量的平滑度量。 的 SNE 的性能对于复杂性的变化相当强大，典型值在 5 之间和 50。使用梯度下降法执行公式 3 中成本函数的最小化。渐变具有令人惊讶的简单形式

$$\frac{\delta C}{\delta y_i} = 2\sum_j (p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j})(y_i - y_j).$$

从物理上讲，坡度可解释为由一组弹簧之间产生的合力地图点 yi 和所有其他地图点 yj。所有弹簧都沿方向（yi yj）施加力。yi 和 yj 之间的弹簧取决于距离是否排斥或吸引地图点地图中两者之间的距离太小或太大，无法代表两者之间的相似性高维数据点。弹簧在 y　 i 和 yj 之间施加的力与它的长度，也与其刚度成正比，即不匹配（pj | i qj | i + pi | j qi | j）数据点和映射点的成对相似性之间的关系。通过从各向同性高斯随机采样图点来初始化梯度下降以原点为中心的小方差。为了加快优化和为了避免较差的局部最小值，可以在梯度中添加相对较大的动量项。换一种说法，当前梯度被添加到先前梯度的指数衰减总和中，以便确定在每次梯度搜索迭代时地图点坐标的变化。数学上，动量项的梯度更新为

$$\mathcal{Y}^{(t)} = \mathcal{Y}^{(t-1)} + \eta\frac{\delta C}{\delta \mathcal{Y}} + \alpha(t)\left(\mathcal{Y}^{(t-1)} - \mathcal{Y}^{(t-2)}\right),$$

3.　t 分布随机邻居嵌入

第二节讨论了 Hinton 和 Roweis（2002）提出的 SNE。尽管 SNE 可以构建合理的可视化效果，但是它受到难以优化的成本函数的阻碍被一个问题称为“拥挤问题”。在本节中，我们介绍一种新技术旨在减轻这些问题的称为“ t 分布随机邻居嵌入”或“ t-SNE”。 t-SNE 使用的成本函数与 SNE 使用的成本函数有两个方面的不同：（1）它使用简要介绍了 SNE 成本函数的对称版本以及更简单的渐变由 Cook 等人撰写。（2007）和（2），它使用 Student-t 分布而不是高斯分布来计算低维空间中两点之间的相似性。 t-SNE 在低维空间中采用了重尾分布，以缓解拥挤问题和优化问题 SNE 的问题。在本节中，我们首先讨论 SNE 的对称版本（3.1 节）。随后，我们讨论拥挤问题（3.2 小节），以及使用重尾分布来解决这个问题（第 3.3 小节）。我们通过描述优化 t-SNE 成本函数的方法（第 3.4 节）来结束本节。

3.1 对称 SNE

作为最小化条件概率 pj | i 和 qj | i 之间的 Kullback-Leibler 散度之和的替代方法，也有可能使单个 Kullback-Leibler 差异最小化高维空间中的联合概率分布 P 与联合概率之间的关系低维空间中的分布 Q：

$$C = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}.$$

同样，我们将 pii 和 qii 设置为零。 我们将这种类型的 SNE 称为对称 SNE，因为它对于∀i，j 具有 pij = pji 和 qij = qji 的性质。 在对称 SNE 中，低维映射 qij 中的成对相似性由下式给出：

$$q_{ij} = \frac{\exp\left(-\|y_i - y_j\|^2\right)}{\sum_{k\neq l} \exp\left(-\|y_k - y_l\|^2\right)},$$

在高维空间 pij 中定义成对相似性的明显方法是

$$p_{ij} = \frac{\exp\left(-\|x_i - x_j\|^2/2\sigma^2\right)}{\sum_{k\neq l} \exp\left(-\|x_k - x_l\|^2/2\sigma^2\right)},$$

但这会在高维数据点 xi 产生问题是一个异常值(即所有成对距离 kxi xjk2 对于 xi 来说很大)。对于这样的离群值，pij 的值对于全部为 j，因此其低维地图点 yi 的位置对成本函数的影响很小。结果，地图点的位置不能很好地由其他地图的位置确定点。我们通过定义高维联合概率 pij 来规避此问题空间作为对称条件概率，即，我们设置 pij = pj | i + pi | j 2n P。这样可以确保 j pij> 12n 对于所有数据点 xi 结果，每个数据点 xi 对成本函数有重大贡献。在低维空间中，对称 SNE 仅使用公式 9。SNE 的对称版本的主要优点是其渐变的简单形式，即计算速度更快。对称 SNE 的梯度与非对称 SNE 的梯度非常相似，并且是（谁）给的

$$\frac{\delta C}{\delta y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j).$$

在初步实验中，我们观察到对称 SNE 似乎会产生仅和不对称 SNE 一样好，有时甚至更好。

3.2 拥挤问题

考虑位于二维曲面流形上的一组数据点，其近似为线性小规模，并嵌入到高维空间中。有可能在二维图中相当好地建模数据点之间的小成对距离通常在玩具示例（例如“瑞士卷”数据集）上进行说明。现在假设玛尼褶皱有十个内在维度 5 并嵌入到更高维度的空间中。

二维映射中的成对距离无法如实地存在的原因有几个模拟十维流形上各点之间的距离。例如，在十个维度中可能有 11 个相互等距的数据点，无法对此建模忠实于二维地图。一个相关的问题是成对的分布非常不同两个空间中的距离。以数据点 i 为中心的球体的体积缩放为 rm，其中 r 为球形的半径和维数。因此，如果数据点大致均匀分布在 10 维流形上 i 周围的区域中，我们尝试对 i 到二维地图中其他数据点的距离进行建模，得到以下“拥挤问题”：可用于容纳中等距离的二维地图区域与可容纳附近的区域相比，数据点的大小将不足以容纳足够的空间数据点。因此，如果我们要在地图中准确地为小距离建模，则与数据点 i 距离适中的大多数点都必须放置在距离数据点太远的位置二维地图。在 SNE 中，弹簧将数据点 i 连接到这些太远的地图中的每一个因此，积分将施加很小的吸引力。尽管这些吸引力很小，如此大量的力量将地图中心的点压在一起，防止自然簇之间形成间隙。注意，拥挤问题不是特定于 SNE，但它也发生在其他用于多维缩放的本地技术中作为 Sammon 映射。

提出了通过向所有弹簧略加排斥来解决拥挤问题的尝试由 Cook 等人撰写。 （2007）。通

过引入统一的背景模型可产生轻微的排斥混合比例ρ很小。因此，无论两个地图点相距多远，qij 都永远不会低于 2ρn（nn 1）（因为均匀的背景分布超过 n（n 1）/ 2 对）。结果，对于在高维空间中相距较远的数据，qij 将始终大于 pij，从而略有排斥。这种技术称为 UNI-SNE，尽管它通常胜过标准 SNE，但 UNI-SNE 成本函数的优化却很繁琐。最佳的优化方法已知的方法是先将背景混合比例设置为零（即通过执行标准 SNE）。使用模拟退火优化 SNE 成本函数后，背景如图所示，可以增加混合比例以允许自然簇之间形成一些间隙由 Cook 等人撰写。（2007）。直接优化 UNI-SNE 成本函数不起作用，因为有两个相距较远的地图点将从统一背景中获取几乎所有的 qij。所以即使如果它们的 pij 大，则它们之间就不会有吸引力，因为它们的变化很小分离对 qij 的比例影响将消失得很小。这意味着，如果集群在优化的早期就被分离了，没有力量将它们拉回一起。

3.3 不匹配的尾巴可以补偿不匹配的尺寸

由于对称 SNE 实际上匹配高维和低维空间中的数据点对的联合概率，而不是它们之间的距离，因此我们有一种自然的方法缓解如下所述的拥挤问题。 在高维空间中，我们使用高斯分布将距离转换为概率。 在低维地图中，我们可以使用尾部比高斯重的概率分布来转换距离转化为概率。 这样可以在高维空间中忠实地保持距离通过地图中更大的距离进行建模，从而消除了不必要的吸引力表示适度不同数据点的地图点之间的力。

在 t-SNE 中，我们采用具有一个自由度的 Student t 分布（与柯西分布）作为低维图中的重尾分布。 使用这个分布，联合概率 qij 定义为

$$q_{ij} = \frac{\left(1 + \|y_i - y_j\|^2\right)^{-1}}{\sum_{k \neq l} \left(1 + \|y_k - y_l\|^2\right)^{-1}}.$$

我们使用具有单一自由度的 Student t 分布，因为它具有特别好的 1 + kyi yjk2 1 的属性接近大成对距离的平方反比定律 kyi yjk 在低维图中。这使得地图可以表示联合概率对于相距较远的地图点，（几乎）不变。这也意味着彼此相距遥远的大型点簇以与单个点相同的方式进行交互，因此除了最佳规模外，优化工作的方式完全相同。我们选择学生 t 分布的理论依据是，它与高斯分布密切相关，因为学生 t 分布是高斯的无限混合。计算方便的属性是在学生 t 分布下评估点的密度要比在学生 t 分布下评估点的密度快得多高斯，因为它不涉及指数，即使学生 t 分布为等效于具有不同方差的高斯的无限混合。

P 和基于 Student-t 的联合概率之间的 Kullback-Leibler 散度的梯度分布 Q（使用公式 12 计算）在附录 A 中得出，由下式给出

$$\frac{\delta C}{\delta y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j) \left(1 + \|y_i - y_j\|^2\right)^{-1}.$$

在图 1（a）到 1（c）中，我们将两个低维数据点 yi 和 yj 之间的梯度显示为在高维和低维中成对欧几里得距离的函数 SNE，UNI SNE 和 t-SNE 的对称版本的空间（即，作为 kxi xjk 和 kyi yjk 的函数）。 在图中，梯度的正值表示低维数据点 yi 和 yj，而负值表示两者之间的排斥两个数据点。 从图中可以看出，t-SNE 梯度相对于 SNE 和 UNI-SNE 的梯度。首先，t-SNE 梯度强烈排斥由小两两模型化的异类数据点低维表示中的距离。SNE 也有这种排斥，但其作用是与梯度中其他地方的强景点相比最小（我们的最大景点梯度的图形表示约为 19，而最大斥力约为 1）。 在 UNI-SNE 中，异类数据点之间的排斥量略大，但是，只有在低维表示中的点之间的成对距离已经很大时，这种排斥才很强（这通常不是这种情况，因为低维表示是通过从高斯样本中进行采样来初始化的，该样本的居中偏差很小围绕原点）。

其次，尽管 t-SNE 在以成对的小距离建模的异类数据点之间引入了强排斥，但这些排斥不会

达到无穷大。在这方面，t-SNE 不同来自 UNI-SNE，其中非常不同的数据点之间的排斥强与它们在低维图中的成对距离成比例，这可能会导致不同的数据点彼此之间的距离太远。总而言之，t-SNE 强调（1）通过大的成对距离对不同的数据点进行建模，以及（2）通过小的成对距离对相似的数据点进行建模。此外，t-SNE 成本函数的这些特征的结果（以及近似规模的结果)学生 t 分布的不变性)，t-SNE 成本函数的优化容易得多而不是优化 SNE 和 UNI-SNE 的成本函数。具体来说，t-SNE 介绍低维地图中的远程力可以将两个（类似的）簇拉回在优化中较早分离的点。 SNE 和 UNI-SNE 没有这么长的作用力，因此 SNE 和 UNI-SNE 需要使用模拟退火来获得合理的解决方案。相反，t-SNE 中的远程力量有助于识别不求助于模拟退火的局部最优。