



## 发明技术交底书

\* 标注为必填项

交底书名称*	推荐系统模型：基于双自编码器度量学习的兴趣流形重建		
发明人*	Ziyangding (丁子扬) Wenyema (马文晔)	撰写人*	Ziyangding (丁子扬)
所在部门*	IEG 增值服务部数据挖掘应用中心	涉及产品和技术*	《掌上英雄联盟》
竞争对手产品	CML 推荐算法、协同过滤	紧急联络方式	微信：dzy326230684（最推荐） 手机1：13811177122（推荐） 手机2：+1 9199499396（不推荐）
技术交底书是将创新技术形成发明文献的基本素材，请全面、清楚地进行填写，使一般技术人员/发明代理人能够理解和实施，撰写方面有疑问，请登录 <a href="http://patent.oa.com/patent/">http://patent.oa.com/patent/</a> 找负责的发明工程师咨询。			

## 缩略语和关键技术定义

**深度学习：**深度学习是基于深度神经网络，通过梯度递减等优化方式，而最终获得从输入数据和目标数据之间一对一映射函数的学习过程。E.g.给定人的年龄、性别等信息，希望能够通过这些信息来预测这个人是否喜欢宠物，那么输入数据即是人的年龄、性别等信息，目标数据即是“这个人是否喜欢宠物”。我们需要的是构建一个正确的映射函数，这个函数可以将输入数据映射为目标数据。这样，给定另一个人的信息，我们就能够得知这个人是否喜欢宠物。深度学习就是基于大量数据来求解这个映射函数的过程，并用深度的神经网络来模拟这个函数。

**协同过滤：**协同过滤是一类用于推荐系统的模型算法，其内容核心是通过以往的数据，而推断出哪一些用户（Resp. 商品）之间更加相像，因此，该类模型可以被简单分为‘user-based’和‘item-based’，前者寻找哪一些用户更相像，后者寻找哪一些商品更加相像。对于前者，在得知了相像的用户后，即可把用户之前所购买过的商品，推荐给与该用户相像的其他用户。后者即是在得知了相像的商品后，即可通过某用户购买了这一商品，而推断该用户喜欢与该商品类似的其他商品，从而把其他类似的商品推荐给该用户。

**度量：**一个自定义的，需要符合度量所满足的 4 个要求的，用于测量定义域空间中任意 2 个点之间距离的映射函数。简单而言，即定义一个新的度量，



就是定义了一个新的距离。

**度量学习：**广义的度量学习有太多情况，这里只简述其中一种情况，即为：已知所有对象所在的空间的距离定义（e.g. 欧氏距离，余弦距离），同时已知对象与对象之间连接和远近关系（e.g. 两个物品距离很近/很远/相聚 2 单位长度），最终尝试获得每一个对象在这个空间之中位置。对于这种任务，通常采取的办法是刚开始随机初始化所有已知对象在这一空间中的位置，之后通过深度学习的方式，不断移动和更新对象在空间中的位置，直到所有对象之间相互的距离关系都几乎符合已知的对象与对象之间的距离关系位置。此时，再基于现在所有的对象的位置，来推断那些曾经不知道的对象间距离关系。此算法与图算法有异曲同工之处，可以理解为对于图 Graph 的重新构建。

**自编码器：**Autoencoder 是深度学习中的一种非常常见的神经网络结构，它的输出端和输入端数据是一样的，即输入  $x$  也会输出  $x$ 。这类结构的神经网络往

往会把高维度的  $x$  向量单一映射成一个低维度向量  $z$ ，之后再建立一个反函数，

将  $z$  返回来映射到  $x$ 。而  $z$  就是对于  $x$  的一个低维度表示(Representation)。一般

我们可以取低维度的  $z$  来表示原数据并进行下一步操作。

**流形结构：**（非正式数学定义）流形是局部具有欧几里得空间性质的空间，是欧几里得空间中的曲线、曲面等概念的推广。一个数据集在向量空间中所呈现的空间结构，是一种简单的流形结构。

**嵌入（embedding）：**（非正式数学定义）给定一个连续的映射函数，一个源数据点由这个函数映射至它的像，这个像就是源数据在其值域中的一个嵌入。若该函数为单射函数，那么该嵌入也可称之为对源数据点的嵌入表示。

## 1、\*发明点概述

当今世界的产业界对于物品的销售和推荐正在逐渐全数据化，即基于用户所返还的描述用户行为、特质等数据和商品的特征、特质等数据来进行商家向用户定向投放物品推荐（广告）的推荐模式。在这种基于数据的推荐系统之上，



俨然已经产生了不少优秀的推荐模型：例如矩阵分解（FM），协同过滤（CF）。这些传统的机器学习算法在大数据的情境下，获得了深度化的变体，使得模型进一步复杂并能够捕捉到更加细微和复杂的用户和商品的关系。也正是得益于深度化，这些模型的推荐性能表现再一次上升到新的高度。然而，深度化的模型面临以下几个问题：

- 1) 深度化神经网络过大规模的参数量直接导致了与传统数学以及统计学相悖的理论基础，即模型参数存在无限种可能。因此数学上可以直接否定训练后模型参数的正确性。
- 2) 深度化模型因为过于庞大的模型复杂度导致其无法解释“为什么用户会喜欢物品”诸如此类的问题。因此，深度化模型在得到不错的结果的同时，也一直饱受其工作理论的质疑
- 3) 不论是传统的协同过滤算法还是深度化的协同过滤算法，都一直停留在要么寻找 user 间相似度或者 item 间相似度的其目标之上。这种方式的确能够帮助产业实践商品或内容的推荐。然而，这种推荐只能够在厂家已经知道商品或用户所有的特征之后才能进行推荐。也就是说，协同过滤不能够单方面仅仅通过知道用户的画像，就直接推断出**什么样子的商品**会被该用户喜欢，而是在告知一个商品的画像后，来推断这个商品**是否会被该用户喜欢**。因此，算法并不能学习到用户客观的、绝对的喜好，而是学习到了用户和一些商品之间的“倾睐关系”，并由用户所喜欢 0 的这些商品来间接“表示”用户潜在的绝对喜好。

考虑到以上，当今世界产业需要一个模型来满足以上的需求。这就是本发明所提出的新算法：**基于双自编码器度量学习的兴趣流形重建**。该模型在训练时同时将一对用户数据和商品数据利用自编码器的编码器进行编码，同时映射到一个低维度的度量空间之上。因为已知这对是否用户与商品是否呈“正”关系，即用户喜欢了商品，我们希望调解编码器中的参数，使得正关系的用户与商品在度量空间中的 2 个嵌入尽可能近，负关系的用户与商品在度量空间中的 2 个嵌入尽可能远。之后，我们在利用自编码器的解码器，将用户和商品的低维嵌入反向复原到他们原来的数据格式，解码器的参数也需要通过传统的优化方式进行训练和优化。



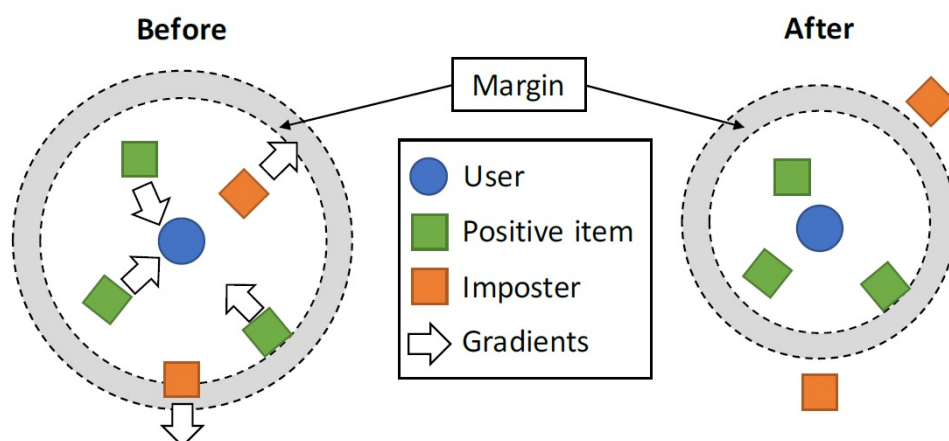
这样的模型可以实现 1) 得到用户用户，用户商品，商品商品之间的关系，而不需要借由用户和商品以及该商品和另一个用户的 2 个关系，来间接猜测用户和另一个用户的关系 2) 绝对化用户的兴趣点（或商品的兴趣嵌入点），更明晰确定用户的兴趣 3) 在无候选商品时，可以实现对于用户所喜欢的商品的画像推测。

具体详细内容，我会在下面详细陈述

## 2、\*与本方案最相近的现有技术

### 2.1、 CML (Collaborative Metric Learning) ， 协同度量学习

该方法与本文的方法最为类似，都会把用户和商品统一在一个低维度量空间中用一些点来进行表示，并根据源数据中用户和商品的关系，来修改用户表示点和商品表示点的位置，从而使其距离关系能够尽可能达到源数据中的距离关系。然而这里有一个问题：不同于本文的模型，该模型并没有建立用户和商品到低维度量空间中的直接映射，也就是说，如果引进一个新的用户，我们没有办法对这个用户的喜好做推断，也就无法对其推荐商品。CML 模型的用途，在于给定固定的用户群以及他们所喜爱和不喜爱的商品，我们如何为他们做已有物品和新物品的推荐。对于新用户，CML 模型无能为力，更不能够进一步实现在无候选商品时，于用户所喜欢的商品的画像推测。

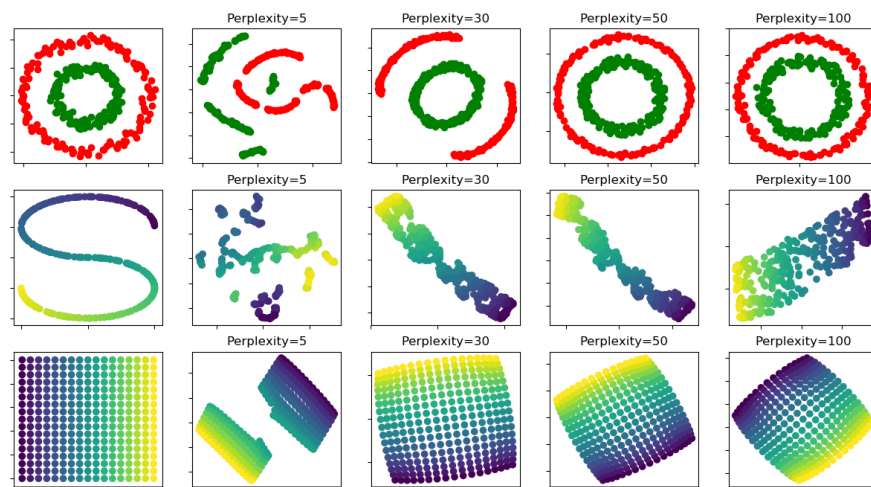


### 2.2、 t-SNE (student-t Stochastic Neighborhood Embedding) ，

该算法于 2008 年被提出。事实上 t-SNE 不是一个推荐系统的算法，但是与



我们的推荐模型有类似之处。t-SNE 算法用于在低维度重现高维度的数据结构，从而实现高维数据的可视化 (Visualization)，其使用的是流形学习的方法。这种算法的假设是：高维度中数据点与点之间的距离关系，要与低维数据中的点与点的关系类似，假设高维中有 2 个点  $a, b$ ，他们在高维空间中相聚很远，那么在低维中，用于表示这两个点的两个表示点  $\hat{a}, \hat{b}$ ，相聚也应该很远，反之亦然。当原高维度空间有很多，比如说  $n$  个数据点的时候，那么低维就会有  $n$  个对于原高维每一个点的表示。因为知道原来高维度中  $n$  个数据点间所有相互的关系，因此我们会有  $n^2$  个距离关系。模型目标，就是通过按一定方式来移动随机初始化的低维空间中的这  $n$  个表示点的位置，来使它们之间的距离关系尽可能与高维空间中的  $n^2$  个距离关系相近。这样，就等同于在低维空间中重新构建了高维空间  $n$  个数据点的流形结构，即在低维中实现对高维数据流形结构的重建。这种流形结构重建的方式，是本发明提出模型的核心内容之一。



### 2.3、VaeCF (Variational Autoencoder Collaborative Filtering),

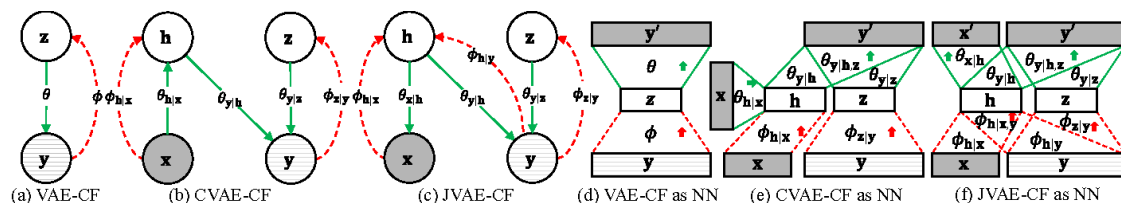
利用变分自编码器对协同过滤的深度化模型。效果出众，然而和一般的协同过滤一样，没有统一用户与物品，更不能够捕捉绝对的用户兴趣嵌入点。与此同类的有使用各种不一样的 Autoencoder 自编码器进行编译，原理都是大同小异。





发明保护创新，积累无形资产

腾讯商业秘密请勿泄露



### 3、\*本技术方案的详细阐述

#### 3.1、产品侧

该算法诞生于掌上英雄联盟的内容推荐情景。英雄联盟中存在很多类似论坛的文章区域供玩家和用户观看。而该模型的应用场景则是输入用户的各类画像特征，以及文章的各类画像特征，从而判定用户是否会阅读这一篇文章，即输出 2 进制推断结果 {True,False}.



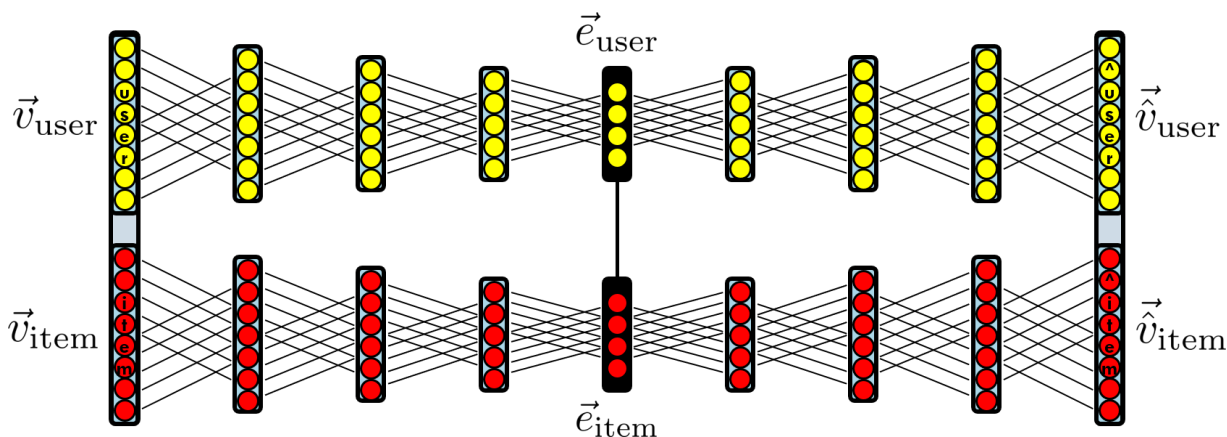


### 3.2、数据结构

模型所需要的数据结构要求，需要每 1 条样本中（1 行数据），称之为 $\vec{v}$ ，要包括关于一个用户的所有特征，我们称之为 $\vec{v}_{user}$ ，是 $\vec{v}$ 的一个亚向量；同时，还要包括一个商品的所有特征，我们称之为 $\vec{v}_{item}$ ，是 $\vec{v}$ 的另一个亚向量；最终还要包括一个 label，这个 label 要么是 0 要么是 1，用于表示这个用户是否喜欢了这个物品。因此， $\vec{v}$ 由 $\vec{v}_{user}$ ， $\vec{v}_{item}$ 和 label 组成。

### 3.3、模型结构

模型基本结构如下图。



### 3.4、基本流程

（注：以下仅为流程基本介绍，其中必然会遇到一些数值（比如“预先设置好的阈值  $a$ ”等）。可能会难以理解，若有该现象，请向下至 3.4 数学原理阅读，会有助于理解。）

**训练：**模型同时将一对用户数据 $\vec{v}_{user}$ 和商品 $\vec{v}_{item}$ 数据利用自编码器的编码器进行编码，通过两个不同的函数 $f_{user}$ ， $f_{item}$ 同时映射到一个低维度的度量空间之上，成为两个嵌入表示，分别为 $\vec{e}_{user}$ ， $\vec{e}_{item}$ 。根据已知用户与商品所呈的正或负的关系，来确定这个用户的表示和商品的表示在低维度的度量空间上应该离得近还是远，并且用相应的已定义好的嵌入空间度量:uniform metric，使用预先设置好的阈值  $p$  作为 hinge loss 的中心设置损失函数，进



行优化第一步。之后，我们再利用自编码器的解码器，将用户和商品的低维嵌入  $\vec{e}_{user}$ ,  $\vec{e}_{item}$  分别用不同的解码函数  $f_{user}$ ,  $f_{item}$  复原到他们原来的数据格式  $\vec{v}_{user}$ ,  $\vec{v}_{item}$ ，解码器的参数也需要通过比较复原后的  $\vec{v}_{user}$ ,  $\vec{v}_{item}$  与原本的  $\vec{v}_{user}$ ,  $\vec{v}_{item}$  产生的 MSE 损失进行优化。

**预测：**训练好的模型可以投入预测作业。输入一个新的向量  $\vec{v}$ ，我们可以得到  $\vec{e}_{user}$  和  $\vec{e}_{item}$ 。算出这两个向量之间的距离  $\|\vec{e}_{user} - \vec{e}_{item}\|$ 。如果这个距离大于一个之前设好的阈值  $a$ ，模型预测为负，反之，预测为正。

**解读：**该模式应用于单独分析某用户的购物喜好（或者单一分析某商品的目标用户）。若是分析用户，首先获得一个用户的画像向量  $\vec{v}_{user}$ ，映射得到用户的嵌入表示  $\vec{e}_{user}$ 。再以  $\vec{e}_{user}$  为中心取  $\vec{e}_{user}$  的半径为预设阈值  $p$  开领域（开球），成为  $B(\vec{e}_{user}, p)$ ，再利用商品的解码器，将  $B(\vec{e}_{user}, p)$  映射为  $\bar{B}(\vec{e}_{user}, p)$ ，其为一个集合。模型将推断所有这个集合内的商品都应该会是用户所喜欢的商品。分析商品的目标购买用户的话，只需要将以上所有的有关 user 和 item 下角标向量和映射函数完全调换，就可以实现。

### 3.5、数学原理

i) **损失方程部分**，我将以模型优化的损失方程的每一部分展开细节描述。

首先，损失方程为以下，共计 3 部分。其中瓶颈距离损失因为涉及到很多细节，我会将其放在 3.4.3 中，也就是本节的最后一部分来讲解。

损失方程：

$$\begin{aligned}\mathcal{L}_{neck} &= \max\{0, \lambda_{margin} + (\vec{y}_{label} \|\vec{e}_{user} - \vec{e}_{item}\| - 0.5)\} \\ \mathcal{L}_{reconstruct} &= \|\hat{\vec{v}}_{user,item} - \vec{v}_{user,item}\| \\ \mathcal{L}_{cov} &= \frac{1}{N} (\|\text{Cov}(E)\|_f - \|\text{diag}(\text{Cov}(E))\|_2^2) \\ \mathcal{L} &= \alpha \mathcal{L}_{neck} + \beta \mathcal{L}_{reconstruct} + \gamma \mathcal{L}_{cov}\end{aligned}$$



### 3.4.i.1、重建损失 (reconstruct loss, 第二个)

这一损失很好理解，就是经过编码器编码解码后再次还原回来的数据和原来的数据相差多少。此处，我们使用的是标准的 MSE 损失即可。

### 3.4.i.2、协方差损失 (covariance loss, 第三个)

该损失旨在让  $\vec{v}_{user}$ ,  $\vec{v}_{item}$  在低维度量空间中的嵌入表示,  $\vec{e}_{user}$ ,  $\vec{e}_{item}$ , 更好地利用空间。假设嵌入空间的维度为 2 (实际上维度比这个会高不少, 但为了方便展示原理, 此处以 2 维作图示范), 那么嵌入的用户和商品可能会出现图 3.4.i.2a 的情况。这种情况不难看出, 嵌入的表示并没有非常好的充盈整个 2 维的空间, 由于巧合而导致的高共线性的嵌入会极度浪费空间, 因而导致调参时不断调高嵌入空间的维度, 以获得更大的嵌入空间, 最终一定会造成严重的过拟合状况。因此, 此处的损失方程, 就是为了使更多的嵌入点能够尽可能均匀分布, 呈现图 3.4.i.2b 的状况。其中,  $\mathbf{E}$  为嵌入空间后所有的  $\vec{e}_{user}$  和  $\vec{e}_{item}$  形成全嵌入数据矩阵, 其协方差矩阵用多变数平均中心化后, 再使用矩阵自乘转置矩阵的到。

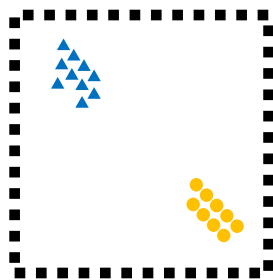


图 3.4.i.2a

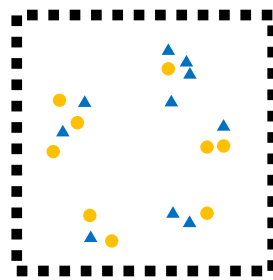


图 3.4.i.2b

### 3.4.i.3、嵌入距离损失 (bottleneck loss, 第一个)

这一部分损失发生于嵌入的低维度量空间, 既然是度量空间, 这一空间中的我们会相对应定义一个度量来测量各个  $\vec{e}_{user}$ ,  $\vec{e}_{item}$  之间的距离关系。此时, 定义一个适合该任务的度量对于模型更好的表现是非常关键的, 首先考虑到以下几点: 1) 定义的度量必须要满足正式数学中对度量



的各个要求 2) 定义的度量要尽可能在低维空间中，在保证目标点与点距离关系的前提下，嵌入更多的点 3) 定义的度量不可是图 Graph 类的距离（比如说汉明距离），其必须由欧式度量通过计算得到（神经网络计算向量的像位于欧式空间），因为需要计算深度学习反向传播过程中的提读数值。因此，这里对于设计合适度量的要求就是非常苛刻的了。传统的  $l_1$ ,  $l_2$ , 甚至  $l_\infty$ ,  $l_p$  距离会面临一个非常严峻的问题，那就是在高维度的时候，其能够嵌入更多的点（例如：使用最简单的  $l_2$  欧式距离，在 1 维上只能嵌入 2 个相互距离相等的点，在 2 维上能嵌入 3 个相互距离相等的点，在 3 维上能嵌入 4 个相互距离相等的点……嵌入能力随维度增加而线性增加），然而就像之前所提到的，在增大维度的时候会遇到严峻的过拟合问题，于是导致模型无法正常工作。然而在一般的图算法中所提到的“最短路径距离”直接突破维度的限制，因此才能有非常好的嵌入能力。于是，我们为了提供一种嵌入能力强，尽可能脱离维度限制的，来源于欧式空间的距离，选择设计并使用一致度量 uniform metric。

**定义：**度量是一个函数  $d$ ，作用在一个空间  $M$  之中，用于将  $M$  中 2 个任意的点映射到  $R$  上得到 2 个点间的“距离”。定义的度量满足对于  $M$  之中任何的 3 个点  $x, y, z$ ，以下 4 点都成立：

1.  $d(x, y) \geq 0$  (非负性)
2.  $d(x, y) = 0$  当且仅当  $x = y$  (不可区分者的同一性)
3.  $d(x, y) = d(y, x)$  (对称性)
4.  $d(x, z) \leq d(x, y) + d(y, z)$  (三角不等式)

**定义：**定义一致度量 (uniform metric)  $d'(\vec{x}, \vec{y})$  为：

$$d'(\vec{x}, \vec{y}) = \min\{a, d(\vec{x}, \vec{y})\} \quad a > 0$$

其中， $d(\vec{x}, \vec{y})$  为我们所经常使用的  $l_\infty$  距离，其定义为：

$$d(\vec{x}, \vec{y}) = \sup_i \{|x_i - y_i|\}$$



**证明：**一致度量是一个明确定义（正确定义）的度量：

1. 关于非负性：因为  $d'(\vec{x}, \vec{y}) = \min\{a, d(\vec{x}, \vec{y})\}$ ，而  $a > 0, d(x, y) \geq 0$ 。因此，二者取小值一定可以保证其度量非负。
2. 关于不可区分者的同一性：假设  $\vec{x} = \vec{x}$ ，那么  $d(\vec{x}, \vec{x}) = 0 \Rightarrow d'(\vec{x}, \vec{x}) = 0$ 。反之，如果  $d'(\vec{x}, \vec{y}) = 0 \Rightarrow d(\vec{x}, \vec{y}) = 0 \Rightarrow \sup_i \{|x_i - y_i|\} = 0 \Rightarrow x_i - y_i = 0 \forall i \Rightarrow \vec{x} = \vec{y}$ 。因此，不可区分者的同一性成立。
3. 关于对称性：因为  $d(\vec{x}, \vec{y})$  是一个明确定义的度量，因此其具有对称性，即  $d(\vec{x}, \vec{y}) = d(\vec{y}, \vec{x}) \Rightarrow \min\{a, d(\vec{x}, \vec{y})\} = \min\{a, d(\vec{y}, \vec{x})\} \Rightarrow d'(\vec{x}, \vec{y}) = d'(\vec{y}, \vec{x})$ 。因此对称性成立。
4. 关于三角不等式。假设我们有 3 个点  $\vec{x}, \vec{y}, \vec{z}$

$$\begin{aligned} d'(\vec{x}, \vec{y}) + d'(\vec{y}, \vec{z}) &= \min\{a, d(\vec{x}, \vec{y})\} + \min\{a, d(\vec{y}, \vec{z})\} \\ &= \min\{a, \sup_i \{|x_i - y_i|\}\} + \min\{a, \sup_i \{|y_i - z_i|\}\} \\ &= \sup_i \left\{ \min(|x_i - y_i|, a) \right\} + \sup_i \left\{ \min(|y_i - z_i|, a) \right\} \end{aligned}$$

此处，我们可以定义一个新的度量  $\bar{d}(x, y) = \min(|x - y|, a)$ 。其非负性，不可区分者的同一性和对称性显而易见。关于三角不等式性质：

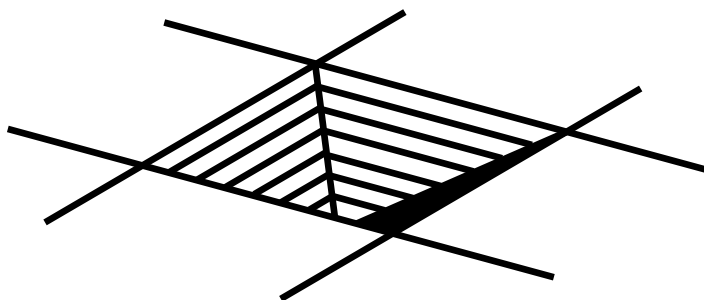
$$\begin{aligned} \bar{d}(x, y) + \bar{d}(y, z) &= \min(|x - y|, a) + \min(|y - z|, a) \\ &\geq \min(|x - z|, a) \end{aligned}$$

可以通过简单的分来讨论的方式得以迅速证明。因此  $\bar{d}(x, y) = \min(|x - y|, a)$  是一个定义明确的度量。因此，以上的等式可以写为：

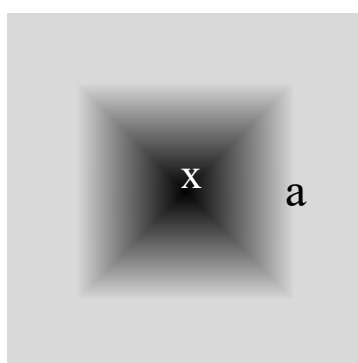
$$\begin{aligned} &\sup_i \left\{ \min(|x_i - y_i|, a) \right\} + \sup_i \left\{ \min(|y_i - z_i|, a) \right\} \\ &= \sup_i \left\{ \bar{d}(x_i, y_i) \right\} + \sup_i \left\{ \bar{d}(y_i, z_i) \right\} \\ &\geq \sup_i \left\{ \bar{d}(x_i, y_i) + \bar{d}(y_i, z_i) \right\} \\ &\geq \sup_i \left\{ \bar{d}(x_i, z_i) \right\} \\ &= d'(\vec{x}, \vec{z}) \end{aligned}$$

故  $d'(\vec{x}, \vec{y})$  是一个定义明确的度量。

既然已经证明了  $d'(\vec{x}, \vec{y})$  是一个明确定义的度量，以下解释使用一致度量  $d'$  的好处：1) 一致度量  $d'$  是基于欧式空间的度量，因此在神经网络的背景之下可以得以数值实现 2) 一致度量不同于其他欧式空间度量，因为在以该度量诱导产生的度量拓扑空间中，尽管在低维空间中，也能够嵌入很多相等距离的点。在 2 维中以视图表示的以点  $x$  为中心和一致度量为度量的距离表示如下图所示：



方便起见，我们使用俯视图，其中最中心纯黑色点（即  $x$  点本身）判定与点  $x$  距离为 0，点所在区域颜色越浅，判定与  $x$  点的距离越远：

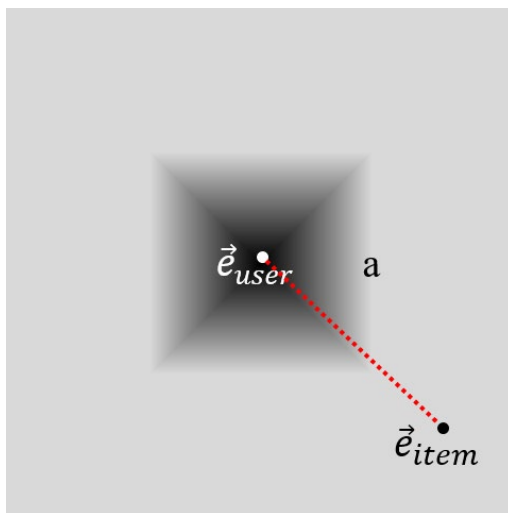


不难看出，在远离以  $l_2$  为度量，距离  $x$  点大于  $a$  的所有的点都判定为与  $x$  点距离为  $a$ 。因此，即使在低维空间中，也有足够的空间来嵌入更多的点，从而应对苛刻的用户商品间距离关系。如果想要调解其嵌入能力，只需要调解  $a$  的大小； $a$  越小，嵌入能力越强； $a$  越大，嵌入能力越弱； $a$  等于 0，其度量空间完全等价于以  $l_2$  为度量诱导出的度量拓扑空间。考虑到以上 2 点，选用一致度量可以为模型调试提供更大的灵活性。

在决定使用一致度量之后，接下来为使用 hinge 损失方程做进一步详细解释。基于度量的学习本应使用基本的定义度量空间中的向量距离差

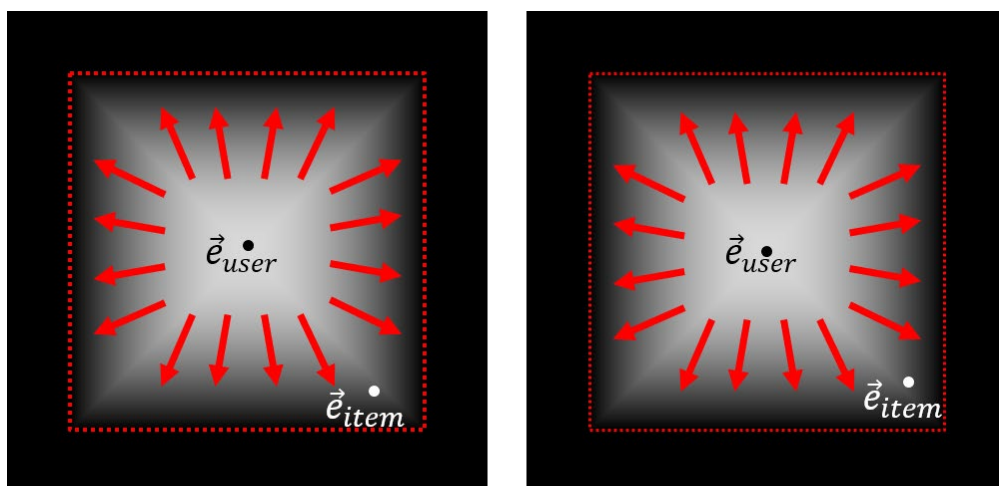
（如：在欧氏距离度量空间中使用欧氏距离差，即  $l_2$  距离）。在此度量空间中，也应该使用在一致度量下引导出的用户和商品 2 个向量间的度量差乘以对应的 label 做为损失函数。然而存在以下问题：1）由一致度量诱导的距离差乘以 label 的损失方程在编程中直接撰写非常繁琐不易，出错率较高且不具有数学美感 2）考虑一种特殊情形：训练样本中某用户

和商品呈现正关系，还未经训练的编码器映射函数  $f_{user}$ ,  $f_{item}$ ，分别将用户和商品映射入度量空间中的 2 个向量，分别记为  $\vec{e}_{user}$  和  $\vec{e}_{item}$ 。计算这两个向量的一致度量距离，却发现他们处于如下这种情况：



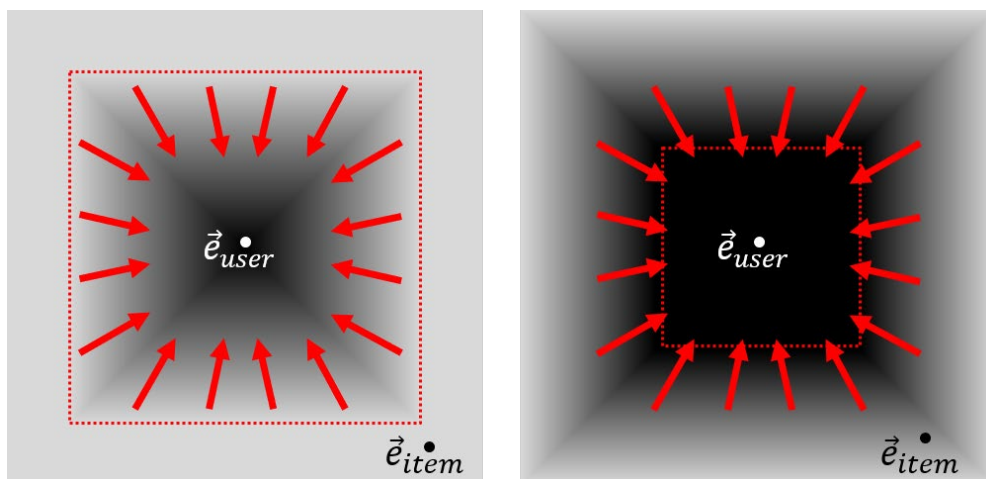
此时，训练无法得以继续进行。因为在  $\vec{e}_{user}$  和  $\vec{e}_{item}$  的邻域，不会产生对二者之间距离的影响，因此梯度为 0；故直接使用度量空间的样本距离差，无法有效对正样本进行充分优化。考虑到以上的 2 个问题，我们提出使用 hinge loss 作为对原度量距离优化的近似。

下图为负样本情况，黑色表示损失方程值小的区域，白色为较高的区域。左图为直接优化度量空间样本距离差，边框为距离中心距离为  $a$  的点，右图为使用 hinge loss，边框为负样本目标安全界限。最后发现，二者完全等价。





下图为正样本情况，黑色表示损失方程值小的区域，白色为较高的区域。左图为直接优化度量空间样本距离差，边框为距离中心距离为  $a$  的点，右图为使用 hinge loss，边框为负样本目标安全界限。最后发现，二者不完全等价，左图为上图所阐述的  $\vec{e}_{item}$  超出了可优化范围的情况，而右图中 hinge loss 则能完全消除左图中所存在的不可优化范围。



综上所述，我们为这里所适用 hinge loss 给出数学上的解答：即基于使用  $l_2$  的度量所使用的 hinge loss，其实际数学意义等价于基于一致度量的向量距离差直接优化，其意义在于解决正样本训练中可能产生的用户或商品嵌入向量超出可优化范围的梯度消失问题（提供优化所需的非零梯度）。因此，这里使用 hinge loss 的真实意义，有别于传统的使用 hinge loss 的意义（制造安全边界），而是帮助优化一致度量。出于此原因，我们对于模型用于预测用户商品对关系的阈值  $p$ ，其实不是  $l_2$  距离中的距离阈值，而是一致度量中的距离阈值，尽管在嵌入距离损失中所计算的距离是  $l_2$  距离。

## ii) 映射函数部分

所有映射函数完全符合深度神经网络的 MLP 结构。但是要注意一下 2 点：

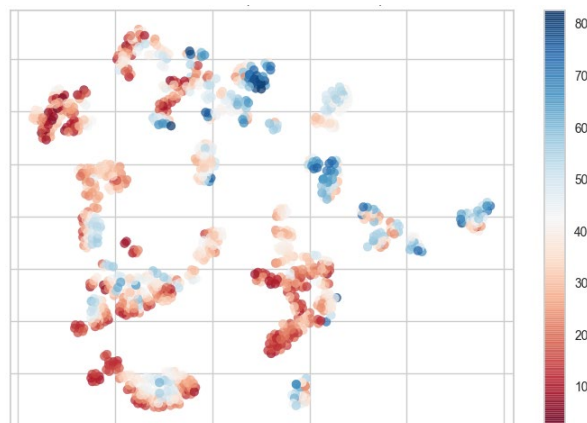
- 1) 编码器最后一层进入嵌入层的激活函数不可以使用 Relu 等无界激活函数。因为无界函数的值域无界，嵌入向量会获得无限的嵌入空间，因此度量学习和流形结构的重建会因此不能够实行。

- 2) 解码器最后一层进入输出层的激活函数需要是 Sigmoid 函数。现实情况中，用户和商品的向量中会既包括数值特征，也包括 2 进制特征。数据与处理的方法是将其全部特征标准化，即放缩在  $I=[0, 1]$  之间。因此，输出层激活函数的值域也应与其值域所相对应。

### iii) 流形结构重建

#### 3.4. iii. 1、嵌入空间的兴趣流形结构重建

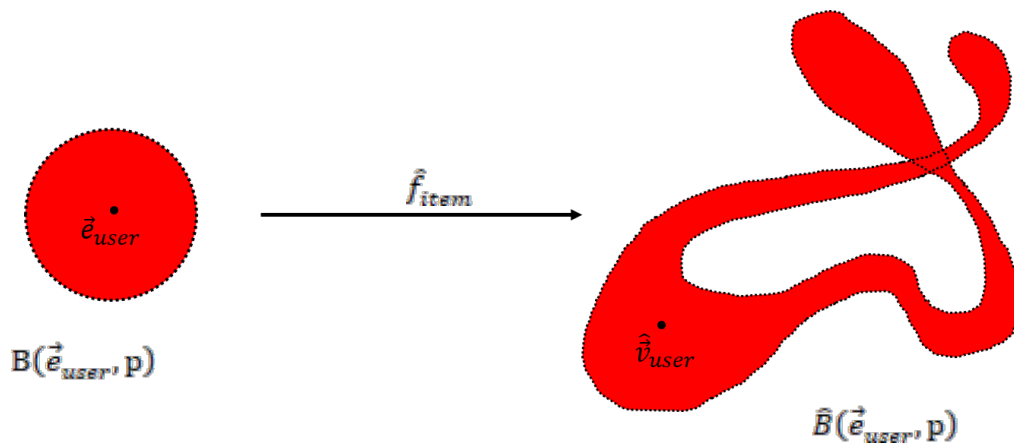
训练后的模型，是基于训练样本的用户商品对信息，在由一致度量所诱导的度量拓扑空间上，重新构建出来的兴趣流形结构。形成这个结构的是所有可以被画像的用户以及商品数据经过映射函数后形成的嵌入向量。其中，任意 2 个嵌入点，不论他们都是用户嵌入点、都是商品嵌入点、还是用户和商品嵌入点各一个，2 点之间的一致度量距离直接表示着这 2 个点的“倾睐”关系，距离越近，倾睐关系越强。当距离低于阈值  $p$ ，则推断为正。这里的流形结构与上文 2.2 中所介绍的 t-SNE 高维数据可视化算法是很类似的。而由于所有的用户、商品都被嵌入在同一个空间中，借助度量本身存在的三角不等式的性质，会对用户和物品无形中起到聚类作用：兴趣相近的用户和用户会被聚合在一起，目标用户群类似的商品会被聚合在一起，以及用户与用户所喜欢的商品也会被聚合在一起。（用户 1 与用户 2 之间的距离，一定小于等于用户 1 与商品  $i$  间距离+用户 2 与商品  $i$  间距离。物品  $i$  为任何一个可画像的物品）这样就形成了一个最基本的流形结构，我们可以称其为“兴趣流形”。



(此图不是该算法产生图，仅作说明使用)

### 3.4. iii. 2、兴趣流形的解码与翻译

通过解码器，将兴趣流形通过解码，翻译成可以解释的用户或商品画像。因为神经网络函数的可导性，可以直接判断其一致连续性，因此嵌入的兴趣模型结构不会因为解码函数而丢失其极值点： $B(\vec{e}_{user}, p)$  为兴趣流形中一个用户的领域，他/她喜欢这个红色区域内所有的商品。这个区域经过商品解码函数  $f_{item}$  解码， $B(\vec{e}_{user}, p)$  被映射为  $\hat{B}(\vec{e}_{user}, p)$ ，为兴趣商品画像（其中可能存在交叉点，因为  $f_{item}$  不需要为单射函数，但不会存在断点）。



## 4、该模型可作出的进一步优化

该发明提出的是一个模型结构，然而其真正意义不局限于模型结构本身，而是将用户和物品同时嵌入同一空间，作度量学习和流形重建的基本思想。因此，模型本身有很多可以进一步改进的可能，以下给出几个例子（但不局限于）：

- 1) 当下模型使用最基本的自编码器，其实可以基于不同的数据样本采用不同的自编码器形式，如采用 VAE 变分自编码器，Contrastive 自编码器。使用不同的编码器所要遵循的原则，是其需要将源数据映射入一个拥有明确定义度量的有界的空间。如使用 VAE，那么嵌入空间则是一个高维高斯分布的概率分布空间，因此需要使用具有明确定义的 KL 散度 (Shannon Entropy) 作为分布于分布之间的距离定义进行嵌入。若有需要，可以模仿 3.4. i. 3 中设计度量的过程。
- 2) 当下模型面对不同的数据稀疏度可以考虑加入嵌入层 (embedding layer)，甚至可以采用推荐系统中常用的 wide & deep 模型的神经网络



结构来代替最基础的 MLP，其意义是设计相对应的神经网络结构以更好地捕捉稀疏数据中的信息。

- 3) 对于具有时序的数据，可以将整个网络“序列化”，引入随时间而变化的神经网络结构。其中，在深度学习上的层面上可以将 MLP 改变为 GLR 或者 LSTM；在统计学习上的改变可以采用贝叶斯先验后验的时序更新，甚至可以使用卡尔曼滤波器进行时间序列数据的学习（需要使用 VAE）。

## 5、本发明的技术关键点（欲保护点）

- （1）将协同过滤改进为协同度量学习并统一用户及商品的嵌入空间的创新。
- （2）设计并应用一致度量以及对其优化的方法以及损失方程与优化由  $L_{\text{KL}}$  诱导产生的 hinge loss 间的等价关系。
- （3）解码器的设计对于兴趣流形的解码和翻译的作用和意义。
- （4）全文中所指出的进行开展协同度量学习对于重建兴趣流形的数学意义。

## 6、参考文献

- [1] Cheng-Kang Hsieh, Longqi Yang, Yin Cui, Tsung-Yi Lin, Serge Belongie, and Deborah Estrin. 2017. Collaborative Metric Learning. In Proceedings of the 26th International Conference on World Wide Web (WWW '17)
- [2] Laurens van der Maaten, Geoffrey Hinton, 2008. Visualizing Data using t-SNE, Journal of Machine Learning Research 11/08
- [3] Liang, Dawen, et al. "Variational Autoencoders for Collaborative Filtering." Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW 18, 2018, doi:10.1145/3178876.3186150.
- [4] Park, Chanyoung, et al. "Collaborative Translational Metric Learning." 2018 IEEE International Conference on Data Mining (ICDM), 2018, doi:10.1109/icdm.2018.00052.
- [5] scikit-learn t-SNE: The effect of various perplexity values on the shape, [https://scikit-learn.org/stable/auto\\_examples/manifold/plot\\_t\\_sne\\_perplexity.html#](https://scikit-learn.org/stable/auto_examples/manifold/plot_t_sne_perplexity.html#)
- [6] Munkres, James Raymond. Topology. Prentice Hall, 2000.



## 6、发明交底书 Checklist

请撰写人写完交底书后，按以下项目进行自检，所有项目都是“是”后再提交，如有两项不满足，将不颁发撰写奖金。如果是产品/设计人员主导交底书，建议与技术人员合作完成。

序号	评审项目	评审结论（打√）
1	是否按照模版完整填写，格式正确	是√/否□
2	解决的问题是否为技术性问题，而并非个人体验等	是√/否□
3	现有技术是否与本发明相关，描述是否恰当	是√/否□
4	是否描述产品侧的应用场景，必要时结合 UI 界面	是√/否□
5	是否描述实现的硬件环境，必要时提供网络拓扑图	是√/否□
6	是否描述技术侧的实现逻辑、对数据的处理流程，必要时结合流程图、时序图等	是√/否□
7	发明点概要是否简要、清晰地指出发明的关键点	是√/否□
8	是否还有其他技术实现方式，或应用于其他领域	是√/否□