

# 基于深度神经网络的社会媒体网络分析

张艳红<sup>1</sup> 王宝会<sup>2</sup>

(首都经济贸易大学信息学院 北京 101500)<sup>1</sup> (北京航空航天大学软件学院 北京 100191)<sup>2</sup>

**摘要** 社会媒体网络中不仅包含了用户、文本、图片和视频等多种模态的数据,还包含了反映不同模态数据之间交互的群体特征。为了更好地描述社会媒体网络,从而为上层应用提供更好的服务,提出了一种基于深度神经网络的社会媒体网络模型。该模型采用深度神经网络对单个模态的数据进行学习,从而得到任意一个模态数据的潜在特征表示方法。对于两种不同模态的数据,利用具有高斯分布的先验矩阵与两个模态数据的后验分布建立反映这两个模态数据间群体特征的生成模型。实验结果表明,提出的模型在网络结构的链接分析中具有更好的预测效果,能有效地描述社会媒体网络的整体特征。

**关键词** 深度学习,神经网络,社会媒体网络,特征学习,链接预测

**中图分类号** TP319 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.4.051

## Analysis of Social Media Networks Based on Deep Neural Networks

ZHANG Yan-hong<sup>1</sup> WANG Bao-hui<sup>2</sup>

(College of Information, Capital University of Economics and Business, Beijing 101500, China)<sup>1</sup>

(School of Software, Beihang University, Beijing 100191, China)<sup>2</sup>

**Abstract** Social media networks include not only multi-modal data, such as users, text, images, video, and so on, but also collective effect between data of different modals. In order to better describe social media networks and provide better services for applications, this paper proposed a deep neural network based social media network model. The proposed model models data of each single modal with a deep neural network, and then gets a latent feature representation for each modal data. For data between two different modals, a Gaussian distributed prior matrix and two posterior distributions of different modals were applied to build a generative model that describes the collective effect between two different modals. The experiments show that the proposed model has better prediction performance in link analysis of networks than related works, and can describe the whole underlying social media network effectively.

**Keywords** Deep learning, Neural networks, Social media network, Feature learning, Link prediction

随着互联网上用户生成的数据的爆炸式增长,社会媒体网络已经成为用户最重要的网络应用之一。在社会媒体中,用户在虚拟的群体中创建、共享并且交换数据,他们之间的交互行为及其在网络中使用的数据一起构成了社会媒体网络<sup>[1]</sup>。社会媒体网络中往往含有文本、图片、音频和视频等多种模态的数据,这些数据可以用来对用户上传的图片和视频进行标注<sup>[2]</sup>、对各种模态数据进行检索<sup>[3]</sup>、对用户进行个性化的推荐<sup>[4]</sup>和精准广告<sup>[5]</sup>等。因此,要使社会媒体网络向用户提供方便快捷并且令人满意的服务,必须综合运用各种数据。

区别于传统的媒体形式,社会媒体网络主要有以下 3 个特征。首先,从数据产生的视角来看,社会媒体数据包含了大量噪声数据<sup>[6]</sup>。由于不同用户的背景不同,并且用户在生成数据时通常具有主观性,因此其对同一事物的判断具有差异性。此外,由于语言中存在着多义词,当用多义词对某一事物进行描述时,很难将其具体化为某一确定的释义。其次,从数据分布的视角来看,社会媒体数据往往是异构的多模态数据<sup>[3,7]</sup>。社会媒体网络中往往同时包含文本、音频、图片和视

频等多种形式的的数据,由于不同数据类型的表示方式和计算方法不同,因此传统的数据挖掘方法不适用于异构的社会媒体网络。最后,从数据交互的视角来看,社会媒体数据之间是相互关联的。数据之间的这种关联反映了用户间的群体效应。群体效应既包括显式的用户与用户之间的交互行为<sup>[8]</sup>,也包括隐式的对图片进行标注的协作行为<sup>[2]</sup>。

为了有效地解决上述问题,研究人员主要从以下 3 个方面展开工作。第一,社会媒体数据预处理。为了应对社会媒体数据中的噪声,研究人员面向特定的应用场景设计了不同的数据清洗和提炼方法来对用户生成的数据进行预处理<sup>[9]</sup>,主要的研究内容包括重复视频的检测<sup>[10]</sup>、图片标签的推荐<sup>[2,11]</sup>和社会媒体数据的组织<sup>[12]</sup>等。第二,社会媒体数据的语义理解。为了应对社会媒体数据的多模态性,通过对不同模态数据的分析提取出每个模态数据的语义信息,并在得到的语义空间中形成对社会媒体数据的整体理解。这类研究内容主要有跨模态的信息检索<sup>[13]</sup>、话题和事件识别<sup>[14-16]</sup>以及社会媒体知识挖掘<sup>[17]</sup>等。第三,社会媒体网络分析。为了分析

到稿日期:2015-12-13 返修日期:2016-01-13 本文受国家自然科学基金资助项目(61303074)资助。

张艳红(1983-),女,硕士,讲师,主要研究方向为软件工程、算法设计;王宝会 男,教授,主要研究方向为软件工程、网络安全。

社交媒体网络中的关联关系,研究人员通常将社交媒体网络表示为社会关系图<sup>[18]</sup>或者将关联关系嵌入到特定的张量空间<sup>[19,20]</sup>,并用社会网络分析方法对社交媒体网络进行分析。这类研究包括社团检测<sup>[21]</sup>和上下文社交媒体分析<sup>[22]</sup>。

现有的社交媒体网络研究主要致力于上层应用模型的设计,而忽略了社交媒体网络自身基本的特征表示方法,即表征学习。深度神经网络采用多层的神经网络模型对数据的表征进行学习,是一种有效的表征学习方法,已经被广泛应用于图像处理<sup>[23]</sup>、语义识别<sup>[24,25]</sup>和自然语言处理<sup>[26]</sup>等应用中。本文首先采用深度神经网络模型对每个模态数据的表征向量进行学习,然后引入表示群体特征的矩阵再加上两个不同模态的表征向量构成群体特征的生成模型,最后应用该模型对社交媒体网络的应用进行分析。

## 1 社交媒体网络的深度神经网络模型

本文的目的是建立一种包含多种模态数据个体特征和模态之间交互信息的社会特征(或群体特征)的联合神经网络模型,并基于该模型进行社会特征的交互预测。如图1所示,本文首先针对不同模态的数据分别建立相应的低层特征向量表示,然后将这些低层特征和社会特征联合起来构建统一的潜在特征表示。

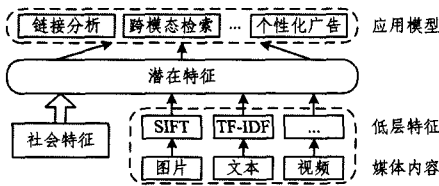


图1 社交媒体网络分析示意图

给定A和B两种模态的数据,这两种模态数据的原始特征矩阵分别为X和Y,其中 $\{x_i\}_{i=1}^M$ 和 $\{y_j\}_{j=1}^N$ 分别为X和Y的第i个和第j个特征向量,M和N分别为模态A和B的数据个数。 $R \in \mathbb{R}^{M \times N}$ 为X和Y之间的链接关系矩阵,其中当 $x_i$ 和 $y_j$ 之间存在着链接关系时, $r_{ij}=1$ ;否则, $r_{ij}=0$ 。

本文提出的A和B两种模态数据之间关联关系的生成模型如图2所示。该模型包括单模态数据的特征表示和模态之间的交互两个部分。对于单模态数据的特征表示,采用多层的神经网络;而对于两种模态数据之间的交互,采用二次线性模型。

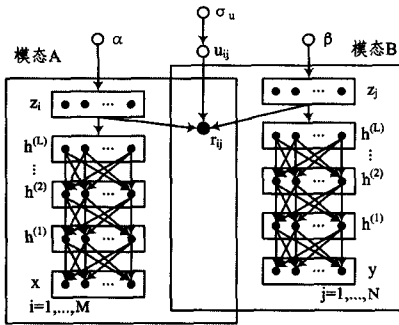


图2 不同模态数据间关联关系的生成模型

在单模态的特征表示方法中, $Z^A$ 和 $Z^B$ 分别为模态A和B的二值特征矩阵,其中 $z_i^A$ 和 $z_j^B$ 分别为 $Z^A$ 和 $Z^B$ 中的第i个和第j个数据对应的特征向量。当Z中第i个数据对应的特征向量 $z_i$ 包含第k个特征时, $z_{ik}=1$ ;否则, $z_{ik}=0$ 。先验分布 $P(Z^A|a)$ 和 $P(Z^B|b)$ 分别指定了模态A和B的特征向量分

布,条件分布 $P(X|Z^A)$ 和 $P(Y|Z^B)$ 为A和B两种模态数据的后验分布。

为了使得具有相似内容特征的数据具有相似的关联关系,本文采用深层的神经网络对单个模态的潜在特征向量表示进行学习。以模态A中的数据为例(模态B的计算方法相似),采用L个隐藏层 $\{h^{(1)}, h^{(2)}, \dots, h^{(L)}\}$ 进行数据的中间表示。在 $\{h^{(1)}, h^{(2)}, \dots, h^{(L)}\}$ 和 $z_i$ 已知的情况下, $x_i$ 的条件概率为

$$P(x_i | h^{(1)}, h^{(2)}, \dots, h^{(L)}, z_i) = P(z_i) P(h^{(L)} | z_i) \cdot \left( \prod_{l=1}^{L-1} P(h^{(l)} | h^{(l+1)}) \right) P(x_i | h^{(1)}) \quad (1)$$

当神经网络中上层的变量 $h^{(l+1)}$ 已知时, $h^{(l)}$ 中各个单元之间的条件概率是相互独立的,其计算方法为

$$P(h_s^{(l)} | h^{(l+1)}) = \frac{1}{1 + \exp(-b_s^{(l)} - \sum_i W_{s,i}^{(l+1)} h_i^{(l+1)})} \quad (2)$$

其中, $h_s^{(l)}$ 为第l层中第s个神经元的激活概率, $b_s^{(l)}$ 为 $h_s^{(l)}$ 的偏差, $W_{s,i}^{(l+1)}$ 为神经元 $h_s^{(l)}$ 和 $h_i^{(l+1)}$ 之间的权重系数。

当 $h^{(L)}$ 已知时, $z_i$ 的计算公式与式(2)相同。如果输入向量 $x_i$ 中每个分量的取值范围为 $[0, 1]$ ,那么 $P(x_i | h^{(1)})$ 的计算公式也如式(2)所示;如果 $x_i$ 中每个分量的取值范围为整个实数空间,那么采用高斯分布计算 $P(x_i | h^{(1)})$ 的概率:

$$P(x_i | h^{(1)}) = N(b_i + \sigma_i \sum_j W_{i,j}^{(1)} h_j^{(1)}, \sigma_i^2) \quad (3)$$

其中, $b_i + \sigma_i \sum_j W_{i,j}^{(1)} h_j^{(1)}$ 和 $\sigma_i^2$ 分别为第s个分量的均值和方差。

在两种模态数据之间的交互表示中,条件分布 $P(R|Z^A, Z^B)$ 为关联关系的后验分布。令U为 $Z^A$ 和 $Z^B$ 的关联关系权重矩阵。给定成对的数据 $z_i^A$ 和 $z_j^B$ , $u_{kk}$ 为 $z_i^A$ 的第k个特征和 $z_j^B$ 的第k'个特征之间建立关联关系的概率权重。假设关联关系依赖于 $Z^A$ 、 $Z^B$ 和U,并且这些关联关系之间是相互独立的,那么有

$$P(R|Z^A, Z^B, U) = \prod_{i,j} P(r_{ij} | Z^A, Z^B, U) \quad (4)$$

其中

$$P(r_{ij} | Z^A, Z^B, U) = \frac{1}{1 + \exp(-\sigma(z_i^A U z_j^B))}, u_{ij} \sim N(0, \sigma_u^2) \quad (5)$$

依据式(1)一式(3)可以得到单个模态数据的生成概率,然后再依据式(4)和式(5)便可以得到两个模态数据之间的关联关系的生成概率模型。于是,可以得到整个社交媒体网络的生成模型。

## 2 模型的参数学习与推导

在上节所述的社交媒体网络深度神经网络模型中,对模型的参数进行准确的推导是不可行的。为了便于计算,先要根据数据的内容特征推导出第L层的特征表达,然后推导出 $Z^A$ 和 $Z^B$ 的后验分布,最后依据 $Z^A$ 、 $Z^B$ 和U,采用马尔可夫链蒙特卡洛算法进行参数的估计。依据数据的内容特征学习高层的特征表示就是对模型中包含的偏差和邻接层的权重矩阵进行学习。本文采用两步训练方法:第一步为逐层预训练的非监督学习;第二步为参数的微调过程。

在逐层的预训练过程中,在邻接的两个层次中构建约束玻尔兹曼机模型。约束玻尔兹曼机包含隐藏层 $h \in \{0, 1\}^F$ 和可见层 $v \in \{0, 1\}^D$ ,其中F和D分别为隐藏层和可见层中神经元的个数。当可见层v已知时,隐藏层中第t个神经元值

为 1 的条件分布为

$$Q(h_t=1|v)=\frac{1}{1+\exp(-b_t-\sum_s v_s W_{s,t})} \quad (6)$$

其中,  $v_s$  为可见层  $v$  中第  $s$  个神经元的状态值,  $h_t$  为隐藏层  $h$  中第  $t$  个神经元的状态值,  $b_t$  为与  $h_t$  的偏差,  $W_{s,t}$  为  $v_s$  和  $h_t$  之间链接的权重, 并且矩阵  $W$  是对称的。根据相似的原理, 当隐藏层  $v$  已知时, 可见层中第  $s$  个神经元值为 1 的条件分布为

$$Q(v_s=1|h)=\frac{1}{1+\exp(-a_s-\sum_t h_t W_{s,t})} \quad (7)$$

约束玻尔兹曼机是没有封闭解的, 但是可以通过交替 Gibbs 采样或者对比分歧采样得到近似解。本文通过对比分歧采样方法对约束玻尔兹曼机进行求解:

$$\Delta W_{s,t}=\epsilon(\langle v_s h_t \rangle_{data}-\langle v_s h_t \rangle_{recon}) \quad (8)$$

其中  $\epsilon$  为学习率,  $\langle \cdot \rangle_{data}$  为数据分布的期望,  $\langle \cdot \rangle_{recon}$  为  $k$  步 Gibbs 采样后重建分布的期望。偏差项参数的更新规则与式(7)相同。

当神经网络的最底层为高斯玻尔兹曼机时, 基于隐藏层数据计算可见层数据的概率公式为式(3), 基于可见层数据计算隐藏层数据的公式为

$$Q(h_t^{(1)}=1|x)=\frac{1}{1+\exp(-\mu_t-\sum_s W_{s,t}^{(1)} \frac{x_s}{\sigma_s})} \quad (9)$$

其中,  $\mu_t$  为隐藏层中第  $t$  个神经元的均值, 参数的更新规则为

$$\Delta W_{s,t}^{(1)}=\epsilon(\frac{1}{\sigma_s} \langle v_s h_t \rangle_{data}-\frac{1}{\sigma_s} \langle v_s h_t \rangle_{recon}) \quad (10)$$

在构建单个模态数据的潜在特征表示时, 神经网络采用至下而上的方式逐层进行, 将每两个相邻接的层表示为约束玻尔兹曼机, 并且令下层为可见层, 上层为隐藏层。

为了对  $Z^A$ 、 $Z^B$  和  $U$  的后验分布进行推导, 本文以模态  $A$  为例进行说明, 模态  $B$  的推导过程相同。给定  $U$  以及  $Z^A$  和  $h^{(L)}$  两层之间的参数  $\theta$  (包含权重和偏差), 采用任意的二值矩阵对  $Z^A$  进行采样, 并且对  $Z^A$  的所有行进行迭代。如果  $Z^A$  的第  $i$  行第  $k$  列的值不为 0, 那么有

$$\begin{aligned} P(z_{ik}=1|Z_{-ik}, z_j, h_i^{(L)}, h_j^{(L)}, R) \\ \propto P(z_{ik}=1|Z_{-ik})P(r_{ij}, h_i^{(L)}|z_{ik}=1, Z_{-ik}, z_j, U, \theta) \\ \propto P(z_{ik}=1|Z_{-ik})P(r_{ij}|z_{ik}=1, Z_{-ik}, z_j, U)P(h_i^{(L)}|z_i, \theta) \\ \propto m_k \cdot P(r_{ij}|z_{ik}=1, Z_{-ik}, z_j, U)P(h_i^{(L)}|z_i, \theta) \end{aligned} \quad (11)$$

其中,  $m_k$  为  $Z^A$  中除第  $k$  列以外第  $i$  行中非 0 元素的个数,  $Z_{-ik}$  为  $Z^A$  第  $k$  列中除了第  $i$  行以外的所有元素。

式(11)所示的模型中包含了 3 个主要部分:  $P(z_{ik}=1|Z_{-ik})$  描述了社交媒体网络的群体效应;  $P(r_{ij}|z_{ik}=1, Z_{-ik}, z_j, U)$  使得潜在特征向量更容易产生关联关系;  $P(h_i^{(L)}|z_i, Q)$  使得潜在特征向量沿着神经网络的层次结构生成数据的原始特征表示。

基于式(11)所示的后验分布推导, 很容易对社交媒体网络中未观察到的关联关系进行预测。当给定  $T$  个训练数据  $\{Z^{A,(t)}, Z^{B,(t)}, U^{(t)}\}_{t=1}^T$  时, 用这些数据的预测分布的平均值对未观察到的关联关系的分布进行估计。当预测  $z_i^A$  和  $z_j^B$  之间的关联关系  $r_{ij}$  ( $r_{ij}=0$  表示未观察到该关联关系) 时,  $r_{ij}$  的近似预测分布为

$$P(r_{ij}=1|R_{train}, x_i, y_j) \approx \frac{1}{T} \sum_{t=1}^T P(r_{ij}=1|z_i^{(t)}, z_j^{(t)}, U^{(t)}) \quad (12)$$

### 3 实验结果与分析

本文提出的基于深度神经网络的社会媒体网络模型既包含了社交媒体网络中各个模态数据自身的内容, 又包含了不同模态数据之间的关联关系, 因而能够对整个社交媒体网络进行很好的描述。为了验证该模型在实际应用场景下的性能, 实验采用社交媒体网络中最常见的应用之一——链接预测, 进行模型性能的分析。在社交媒体网络中, 链接预测可以用来进行相关内容的推荐和网络的动态特征分析等。

#### 3.1 实验数据与环境

实验采用公开的 MIRROR Flicker 数据集<sup>[27]</sup>, 该数据集包含 9862 个用户上传的 25000 张图片, 并且平均每张图片包含 8.94 个用户自定义的文本标签。在算法的性能分析过程中, 采用曲线下面积 (Area Under Curve, AUC) 作为评价指标。AUC 是一种排名质量的度量方法, 如果正样本的排名越靠前, 那么 AUC 的值越大, 说明排名的结果越好。AUC 的计算公式为

$$AUC=\frac{S_0-n_0(n_0+1)/2}{n_0 n_1} \quad (13)$$

其中,  $n_0$  为正样本的个数,  $n_1$  为负样本的个数,  $S_0=\sum Pos(q)$ , 而  $Pos(q)$  为排序列表中第  $q$  个正样本的位置排名。

我们主要通过链接预测实验来分析提出的基于深度神经网络的社会媒体网络模型的有效性, 故采用矩阵分解<sup>[28]</sup>、逻辑回归<sup>[29]</sup>、随机游走<sup>[30]</sup> 和 K 最近邻<sup>[30]</sup> 4 种网络链接预测算法作为对比算法。

#### 3.2 实验结果

我们进行了两组实验。第一组实验是对用户间的链接关系进行预测。在给定的数据集中, 如果两个用户间存在链接关系, 那么令相应的关联关系  $r_{ij}=1$ ; 否则令  $r_{ij}=0$ 。实验的目的是预测那些  $r_{ij}=0$  的关联关系在未来变为  $r_{ij}=1$  的概率。因提出的模型中参数  $\sigma_u$  的取值影响着整个矩阵  $U$ , 故其取值的大小会严重影响链接预测的结果。首先, 就  $\sigma_u$  的取值对链接预测结果 AUC 的性能进行分析, 实验结果如图 3 所示。从图中可以看出, 随着  $\sigma_u$  的不断增大, AUC 的值波动较大, 并且在  $\sigma_u=0.3$  处得到最大值。因此令提出的神经网络模型的参数  $\sigma_u=0.3$ , 并且与其它 4 种算法的 AUC 值进行对比。从图 4 所示的对比结果可以看出, 本文提出的模型的链接预测的 AUC 值明显高于其它 4 种算法的 AUC 值。

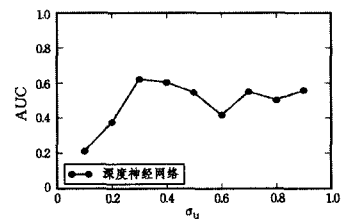


图 3  $\sigma_u$  的取值对 AUC 的影响(用户链接预测)

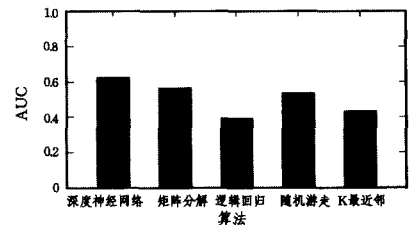


图 4 算法的 AUC 对比(用户链接预测)

第二组实验是对图片的标签进行预测。在给定的数据集中,如果一张图片被标注了某个文本标签,那么令相应的关联关系  $r_{ij}=1$ ; 否则令  $r_{ij}=0$ 。我们随机选取了 20 张图片作为测试数据,删除了它们的标签,并在链接预测中预测这些图片的标签。预测结果的 AUC 值取 20 张图片的平均值。首先,依然对  $\sigma_u$  的取值进行了分析,实验结果如图 5 所示。从图中可以看出,随着  $\sigma_u$  的不断增大,AUC 的值并没有固定的变化规律。令 AUC 为最大值时的参数  $\sigma_u=0.6$  为本文提出的模型的参数取值,同时将此时的 AUC 值与其它 4 种算法的 AUC 值进行对比,对比结果见图 6。从图 6 可以看出,本文提出的模型的链接预测的 AUC 值依然高于其它 4 种算法的 AUC 值。

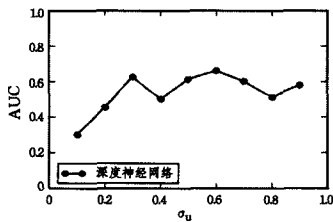


图 5  $\sigma_u$  的取值对 AUC 的影响(图片的标签预测)

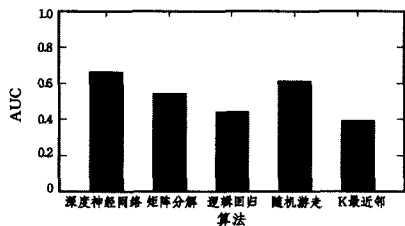


图 6 算法的 AUC 对比(图片的标签预测)

在上述两组实验中,通过用户间的链接预测实验验证了本文提出的模型在同构网络(网络中的节点为用户)中链接预测的效果,通过图片的标签推荐实验验证了提出的模型在异构网络(网络中的节点包含图片和文本标签)中的链接预测的效果。这两组实验的结果表明,本文提出的基于深度神经网络的社会媒体网络模型能更好地描述社会媒体网络的整体信息,因而可以取得更好的链接预测效果。

**结束语** 本文提出了一种基于神经网络的社会媒体网络模型。该模型采用神经网络对单个模态的数据进行学习,从而得到任意一个模态数据的潜在特征表示方法。对于两种不同模态的数据,利用具有高斯分布的先验矩阵与两个模态数据的后验分布建立反映这两个模态数据间群体特征的生成模型。实验结果表明,本文提出的模型在网络结构的链接分析中具有更好的预测效果,因而可以有效地描述社会媒体网络的整体特征。

## 参 考 文 献

[1] Ellison N B. Social network sites: Definition, history, and scholarship[J]. Journal of Computer-Mediated Communication, 2007, 13(1): 210-230

[2] Wei Jian-liang, Ju Chun-hua. Research on User Collaboration Model Based on Social Tagging[J]. Journal of the China Society for Scientific and Technical Information, 2012, 31(3): 281-288 (in Chinese)

魏建良, 琚春华. 基于社会化标注的用户协同模型研究[J]. 情报学报, 2012, 31(3): 281-288

[3] Wang W, Ooi B C, Yang X, et al. Effective multi-modal retrieval based on stacked auto-encoders[J]. Proceedings of the VLDB Endowment, 2014, 7(8): 649-660

[4] Gao Ming, Jin Che-qing, Qian Wei-ning, et al. Real-time and Personalized Recommendation on MicroBlogging Systems[J]. Chinese Journal of Computers, 2014, 37(4): 963-975 (in Chinese)

高明, 金澈清, 钱卫宁, 等. 面向微博系统的实时个性化推荐[J]. 计算机学报, 2014, 37(4): 963-975

[5] Farahat A, Bailey M C. How effective is targeted advertising? [C] // Proceedings of the 21st International Conference on World Wide Web. ACM, 2012: 111-120

[6] Liang Y, Caverlee J, Cao C. A Noise-Filtering Approach for Spatio-temporal Event Detection in Social Media[M] // Advances in Information Retrieval. Springer International Publishing, 2015: 233-244

[7] Song J, Wang Y, Wu F, et al. Multi-modal Retrieval via Deep Textual-Visual Correlation Learning[M] // Intelligence Science and Big Data Engineering. Image and Video Data Engineering. Springer International Publishing, 2015: 176-185

[8] Kim Y, Hsu S H, de Zúñiga H G. Influence of social media use on discussion network heterogeneity and civic engagement: The moderating role of personality traits[J]. Journal of Communication, 2013, 63(3): 498-516

[9] Wang Lian-xi. A Literature Review on Pre-processing and Learning of Microtext[J]. Library and Information Service, 2013, 57(11): 125-131 (in Chinese)

王连喜. 微博短文本预处理及学习研究综述[J]. 图书情报工作, 2013, 57(11): 125-131

[10] Uysal M S, Beecks C, Seidl T. On efficient content-based near-duplicate video detection[C] // 2015 13th International Workshop on Content-Based Multimedia Indexing (CBMI). IEEE, 2015: 1-6

[11] Liu J, Li Z, Tang J, et al. Personalized Geo-Specific Tag Recommendation for Photos on Social Websites[J]. IEEE Transactions on Multimedia, 2014, 16(3): 588-600

[12] Trainor K J, Andzulis J M, Rapp A, et al. Social media technology usage and customer relationship performance: A capabilities-based examination of social CRM[J]. Journal of Business Research, 2014, 67(6): 1201-1208

[13] Costa Pereira J, Coviello E, Doyle G, et al. On the role of correlation and abstraction in cross-modal multimedia retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(3): 521-535

[14] Peng B, Li J, Chen J, et al. Trending Sentiment-Topic Detection on Twitter[M] // Computational Linguistics and Intelligent Text Processing. Springer International Publishing, 2015: 66-77

[15] Chierichetti F, Kleinberg J, Kumar R, et al. Event detection via communication pattern analysis [C] // Eighth International AAAI Conference on Weblogs and Social Media. 2014

[16] Shen Guo-wei, Yang Wu, Wang Wei, et al. Burst Topic Detection Oriented Large-Scale Microblogs Streams[J]. Journal of Computer Research and Development, 2015, 52(2): 512-521 (in Chinese)

## 参考文献

- [1] Tian Sheng, Li You-yong, Wang Jun-mei, et al. ADME Evaluation in Drug Discovery. 9. Prediction of Oral Bioavailability in Humans Based on Molecular Properties and Structural Fingerprints[J]. *Molecular Pharmaceutics*, 2011, 8(3): 841-851
- [2] Hou Ting-jun, Wang Jun-mei, Zhang Wei, et al. ADME Evaluation in Drug Discovery. 6. Can Oral Bioavailability in Humans Be Effectively Predicted by Simple Molecular Property-Based Rules? [J]. *Journal of Chemical Information and Modeling*, 2007, 47(2): 460-463
- [3] Lai Jun, Liao Zheng-gen, Yang Ming-fu, et al. Research and progress of bioavailability[J]. *Chinese Journal of Experimental Traditional Medical Formulae*, 2010, 18: 226-229 (in Chinese)  
赖珺, 廖正根, 杨明福, 等. 生物利用度研究进展[J]. *中国实验方剂学杂志*, 2010, 18: 226-229
- [4] Wang Jun-mei, Hou Ting-jun. Advances in computationally modeling human oral bioavailability[J]. *Advanced Drug Delivery Reviews*, 2015, 86(6): 11-16
- [5] Ma Chang-ying, Yang Sheng-yong, Zhang Hui, et al. Prediction models of human plasma protein binding rate and oral bioavailability derived using GA-CG-SVM method[J]. *Journal of Pharmaceutical and Biomedical Analysis*, 2008, 47: 677-682
- [6] Tian Sheng. Theoretical studies on the prediction of drug-likeness and oral bioavailability[D]. Suzhou: Soochow University, 2011 (in Chinese)  
田盛. 类药性和生物利用度的理论预测研究[D]. 苏州: 苏州大学, 2011
- [7] Ahmed S S S J, Ramakrishnan V. Systems Biological Approach of Molecular Descriptors Connectivity: Optimal Descriptors for Oral Bioavailability Prediction[J]. *Plos One*, 2012, 7(7): e40654
- [8] Zhang Hui, Chen Qing-yi, Xiang Ming-li, et al. In silico prediction of mitochondrial toxicity by using GA-CG-SVM approach [J]. *Toxicology in Vitro*, 2009, 23(1): 134-140
- [9] Liu Jian-wei, Liu Yuan, Luo Xiong-lin. Research and development on deep learning[J]. *Application Research of Computers*, 2014, 31(7): 128-134 (in Chinese)  
刘建伟, 刘媛, 罗雄麟. 深度学习研究进展[J]. *计算机应用研究*, 2014, 31(7): 128-134
- [10] Wang Yong, Zhao Jian-hui, Zhang Deng-yi, et al. Forest fire image classification based on deep neural network of sparse autoencoder [J]. *Computer Engineering and Applications*, 2014, 50(24): 173-177 (in Chinese)  
王勇, 赵俭辉, 章登义, 等. 基于稀疏自编码深度神经网络的林火图像分类[J]. *计算机工程与应用*, 2014, 50(24): 173-177
- [11] Kim M T, Sedykhy A, Suman K, et al. Critical Evaluation of Human Oral Bioavailability for Pharmaceutical Drugs by Using Various Cheminformatics Approaches[J]. *Pharm Res*, 2014, 31: 1002-1014
- 
- (上接第 255 页)
- 申国伟, 杨武, 王巍, 等. 面向大规模微博消息流的突发话题检测 [J]. *计算机研究与发展*, 2015, 52(2): 512-521
- [17] Holzinger A. Social Media Mining and Social Network Analysis: Emerging Research [J]. *Online Information Review*, 2014, 38(1): 157-158
- [18] Wang Li, Cheng Xue-qi. Dynamic Community in Online Social Networks[J]. *Chinese Journal of Computers*, 2015, 38(2): 219-237 (in Chinese)  
王莉, 程学旗. 在线社会网络的动态社区发现及演化[J]. *计算机学报*, 2015, 38(2): 219-237
- [19] Zhou D, Huang J, Schölkopf B. Learning with hypergraphs: Clustering, classification, and embedding [M] // *Advances in Neural Information Processing Systems*. 2006: 1601-1608
- [20] Sang J, Xu C, Liu J. User-aware image tag refinement via ternary semantic analysis [J]. *IEEE Transactions on Multimedia*, 2012, 14(3): 883-895
- [21] Bi J, Huang J, Qin Z. A Relationship Strength-Aware Topic Model for Communities Discovery in Online Social Networks [M] // *Advances in Computer Science and its Applications*. Springer Berlin Heidelberg, 2014: 709-715
- [22] Diakopoulos N, De Choudhury M, Naaman M. Finding and assessing social media information sources in the context of journalism[C] // *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012: 2451-2460
- [23] Li Huan-yu, Bi Du-yan, Yang Yuan, et al. Research on Visual Tracking Algorithm Based on Deep Feature Expression and Learning[J]. *Journal of Electronics & Information Technology*, 2015, 37(9): 2033-2039 (in Chinese)  
李寰宇, 毕彦彦, 杨源, 等. 基于深度特征表达与学习的视觉跟踪算法研究[J]. *电子与信息学报*, 2015, 37(9): 2033-2039
- [24] Chen Shuo. Application research of deep learning neural networks in speech recognition[D]. Guangzhou: South China University of Technology, 2013  
陈硕. 深度学习神经网络在语音识别中的应用研究[D]. 广州: 华南理工大学, 2013
- [25] Deng L, Li J, Huang J T, et al. Recent advances in deep learning for speech research at Microsoft[C] // *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2013: 8604-8608
- [26] He X, Gao J, Deng L. Deep learning for natural language processing and related applications (Tutorial at ICASSP) [C] // *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 2014
- [27] Huiskes M J, Lew M S. The MIR Flickr retrieval evaluation [C] // *Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval*. ACM, 2008: 39-43
- [28] Ermiş B, Acar E, Cemgil A T. Link prediction in heterogeneous data via generalized coupled tensor factorization[J]. *Data Mining and Knowledge Discovery*, 2015, 29(1): 203-236
- [29] Popescul A, Ungar L H. Structural logistic regression for link analysis[J/OL]. [http://repository.upenn.edu/cgi/viewcontent.cgi?article=1139&context=cis\\_papers](http://repository.upenn.edu/cgi/viewcontent.cgi?article=1139&context=cis_papers)
- [30] Lü L, Zhou T. Link prediction in complex networks: A survey [J]. *Physica A: Statistical Mechanics and its Applications*, 2011, 390(6): 1150-1170