

# 基于复杂网络学习框架的客户精准营销研究

冯玥<sup>1</sup> 苏正洋<sup>2</sup>

1. 南京银行博士后科研工作站，南京大学博士后流动站；2. 中国电信股份有限公司江苏分公司

**摘 要：**企业积累了大量客户数据资源，包括客户属性数据、业务相关数据及行业衍生数据。本文从移动运营商和商业银行角度，探讨了如何在大数据背景下构建统一的研究框架，深度挖掘企业所积累的客户行为数据，助力企业实施精准营销战略。本文创新地将先进的网络表示学习算法运用在基于行业衍生数据搭建的异构网络中，借助网络结构的便利，将客户行为模式的学习转化为向量表征，将结果带入分类器进行训练，获得精准营销基础数据。

**关键词：**大数据；异构网络；精准营销；机器学习

## 0 引言

在大数据实践中，企业挖掘客户数据资源的目标是为了更好地了解客户，将有限的营销资源投放到目标客户群，进而更好地服务客户及拓展业务，开展精准营销（precise marketing），而实现客户特征识别和分类，准确地识别潜在客户，是实现精准营销在技术层面的关键环节。数据分析难度伴随数据维度的增长而递增，如何在高维稀疏的数据中找到规律和共性是数据挖掘的重点，而机器学习和人工智能算法为大数据分析提供了技术基础。本文试图在高维海量数据中构建统一的分析框架，结合最先进的机器学习算法，来分析客户行为模式规律，进而实现以精准营销为代表的一系列基于大数据的行业应用。从隐私保护角度来看，本文通过群体

行为特征识别形成的综合指标定位客户类型，有效地杜绝了客户具体信息泄露的问题。

## 1 研究现状

企业所积累的客户相关数据包括三大类：客户基础属性数据、业务相关数据及行业衍生数据。客户基础属性数据是最基本的客户分类指标，如性别、年龄、地址、工作等。业务相关数据反映了客户与企业现阶段的直接业务关系，如电信运营商与个人客户的移动套餐交易，商业银行与个人客户的金融产品交易等。行业衍生数据是企业所属行业独特的数据资产。例如，电信运营商所掌握的客户活动轨迹数据，及商业银行所拥有的客户交易流水数据等。如表 1 所示。

表 1 客户数据分类

数据类别	内容	移动运营商	商业银行
基础属性	基本信息：性别、年龄、工作、收入等	客户基本信息	客户基本信息
业务相关	参与企业直接业务的信息	移动套餐购买历史，业务统计等	金融产品购买历史，业务统计等
衍生数据	不同行业的独特数据资产	移动定位轨迹数据	银联交易流水数据

目前基于大数据的精准营销领域研究多集中在对基础属性数据及业务数据的挖掘，而这两类数据提供的大数据分析维度有限，无法全面刻画客户特征。衍生数据涉及客户的行为属性，为客户行为分析提供了全方位的视角。运营商所独有的移动轨迹数据能够反映个人的活动区域、工作区域、生活场所等。金融机构所独有的客户交易流水数据反映了客户交易行为特征、消费能力、消费偏好等。生活工作区域类似的群体或消费对象类似的群体在行为特征上更容易存在趋同性。

对这些衍生数据的分析，将有助企业全面深入地挖掘客户行为模式和偏好，进而对其开展针对性营销。所以，对衍生数据的挖掘是区分企业大数据应用能力的关

键之一。衍生数据的分析难度在于对象的复杂性以及数据的稀疏性，例如，通过银行渠道的交易流水记录，涉及交易对象、交易金额、交易时间、交易频率等多维度信息；通过移动基站实时反馈记录的个人行动轨迹存在随机性，很难从实时更新的数据层面有效挖掘群体特性。如何设计研究框架，将一

定时间内积累的大量复杂行为数据投射到低维空间,形成个人行为模式的表征,是本文试图解决的技术难题。本文创新地从网络视角重新构建数据,通过构建网络的方式,将收集到的轨迹数据网络化之后,通过最先进的网络表示机器学习算法应用于客户行为研究之中,全面刻画客户特征。

## 2 网络研究框架

具有网络结构的数据可以自然地表达微观主体之间的关联,在日常活动中无处不在,例如,移动定位数据反映个人与地点之间的位置关系,金融交易数据反映个人与交易对象之间的交易关系,个人属性数据反映了个人与其标签的从属关系。基于关联关系所搭建的网络,不仅可以直观地反映对象间的直

接关联关系,也具备进一步探索间接关联关系及多重关联关系的潜力,易于洞察节点间关联关系的层次性。通过构建合适的网络学习框架,我们能将复杂关系中的多重群体特征通过对微观主体组成的网络结构学习剥离出来,进而对微观主体的特征进行分类,通过数据挖掘实现各种现实目标。

图1简单地描述了网络学习的示意图。首先,构造具备网络结构的主体。第二步,按照固定规则提取轨迹样本,进而通过网络表示学习算法获得每一个节点对应的嵌入向量,将结果带入一系列学习任务中,完成应用目标。本文的内容包括三部分,首先讨论了基于网络视角的数据挖掘及机器学习理论方法和步骤,第二部分从移动运营商、金融机构及两者结合的角度讨论了应用方法,最后对金融行业进行了案例分析。

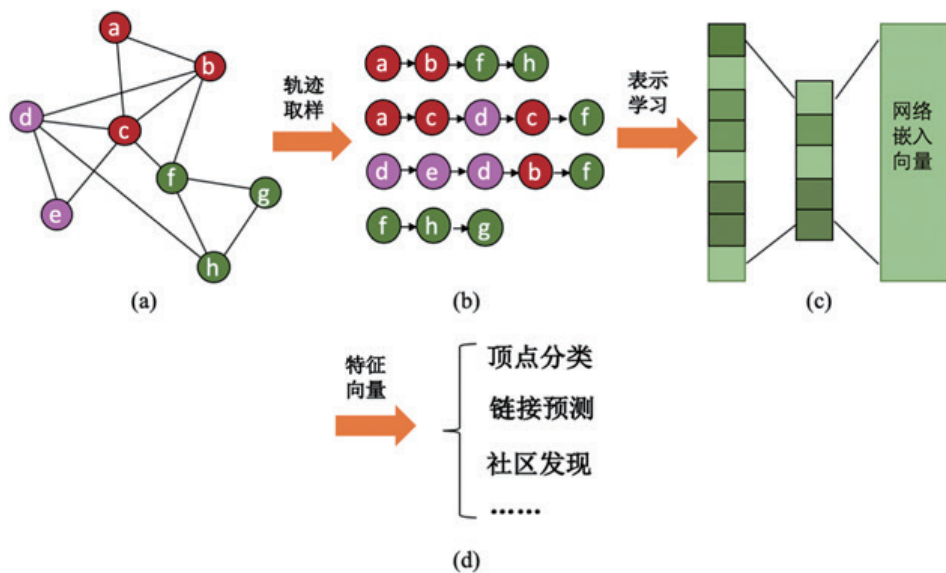


图1 网络学习流程示意图

### 2.1 网络构建

网络是由各类别顶点及通过某些关系构成连接边所组成的结构。原始网络  $G := \{V, E, A, R, \phi, \varphi\}$  是由顶点集合  $V$ , 边集合  $E$ , 顶点类别  $A$ , 边类别  $R$ , 顶点类别映射  $\phi: V \rightarrow A$ , 以及边类别映射  $\varphi: E \rightarrow R$  组成的集合。其中, 对任意的边  $e \in E$ , 存在顶点  $v_1, v_2 \in V$ , 满足  $(v_1, v_2) = e$ , 且  $\phi(v_1), \phi(v_2) \in A, \varphi(e) \in R$ 。

根据不同的目标设定,网络可基于原始网络或重构网络。如果研究侧重于网络结构本身,则沿用原始网络,例如社交网络的结构识别和判定。若研究的目标是利用网络结构的便利,去挖掘深层次的关联关系,则需在原始数据基础之上重新设计节点规则。本文需根据具体衍生数据某些统计指标的结果对顶点设置进行进一步细分。节点的设置也可通过经验判断。

对于移动轨迹数据,在网络构建的时候对基站节点的周平均时长进行统计,通过这个维度的统计指标聚类结果对节点进行细分。对于金融交易数据,从消费金额和频率对交易对象节点进行统计,通过聚类结果对节点进行细分。

#### 案例一：移动网络构建

移动轨迹数据表示为手机用户途径基站的时间序列数据,以及每个基站的地理坐标信息。每条记录的属性包括手机用户、基站编号、经过时间、所属区域等维度。网络构建基于移动轨迹样本(手机用户,基站编号,轨迹细节)。由手机用户节点,接收基站节点及连接边构成,具体构成规则如下:

首先,根据移动轨迹样本数据确定手机用户点集。手机用户节点——对应于数据集中所有手机用户,通过手机号唯一

识别。其次，根据轨迹中样本的统计结果，构建接收基站点集。基于原始数据的重构移动网络应能全面地反映用户移动行为特征，因此网络中节点的定义应包含一定移动信息。但网络结构的设计不能过度稀疏或过度稠密。过于稀疏的网络结构，意味着网络结构中公共节点很少，无法很好地识别群体特征。过于稠密的网络结构意味着大量节点拥有公共节点，无法很好地区别个体特征。因此，需对网络结构中基站对象的细分程度进行取舍。接收基站对象节点的设置依照以下规则，首先依照基站对应的区域对接收基站节点进行初步分类。其次，获得所有与该基站有过定位记录的手机用户集。根据统计期内移动用户集中所有客户对于该对象的周平均时长属性特征的密度分布，对该对象节点进一步细分。对于涉及工作日上午时间、晚间、及周末等与工作生活区域关系较大的基站节点进一步细分到对应基站标签。因此，重构网络交易对象节点是基于移动轨迹数据分布特征分布获得的分类结果，并不是原始的交易对象的简单复制。最后，根据移动用户对应的交易轨迹样本，构成连接用户节点和基站对象节点的公共边。最终形成的重构交易网络由用户节点，基站节点，及两点间的连接边构成。

案例二：金融交易网络

金融交易流水信息表示为卡号（card\_no）、金额（tran\_amt）、交易时间（tran\_date）、行业分类（merch\_type）、商户名称（merch\_name）组成的五维坐标。网络构建基于交易轨迹样本（交易发起方，交易对象，交易细节），由交易发起节点，交易对象节点及连接边构成，具体构成规则如下：

首先，根据交易轨迹样本数据确定交易发起节点集。交易发起节点——对应于数据集中所有客户。其次，根据交易细节的统计结果，构建交易对象点集。交易对象节点的设置依照以下规则，首先依照交易对象对应的行业分类对交易对象节点进行初步分类。其次，获得所有与该行业分类有过交易历史的客户集。根据统计期内客户集中所有客户对于该对象的平均交易金额及交易频率两个维度属性特征的密度分布，对该对象节点进一步细分。对于涉及商场、金融产品和服务、房地产、物业、保险、珠宝、信用卡还款等与消费能力关系较大的对象节点进一步细分到对应商户名称标签。其中，金融产品和服务对应的细节标签包括：理财、结构性存款、大额存单、定期存款、贵金属、普通基金等。因此，重构网络交易对象节点是基于交易数据分布特征获得的分类结果，并不是原始的交易对象的简单复制。最后，根据交易发起点对应的交易轨迹样本，构成连接交易发起节点和交易对象节点的公共边。如图2所示。

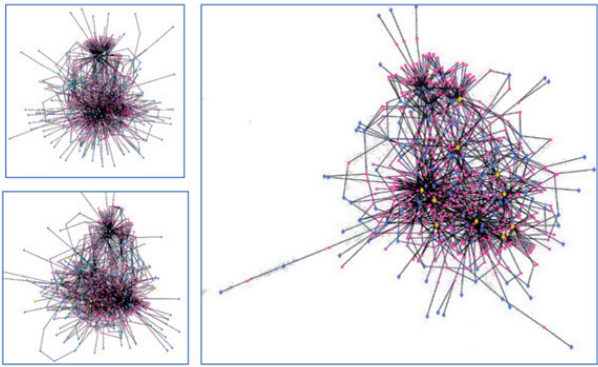


图2 客户交易网络示意图

图中，粉色为客户节点，蓝色为交易对象节点，随交易对象节点进一步细分，由左至右，重构网络的稠密程度降低。

案例三：综合异构网络

将移动子网和金融交易子网通过客户节点相连，构成具备两种特质的重构网络。在重构网络的基础上，在网络中加入了客户基本属性分类节点。如图3所示，节点类型a代表客户属性分类，c代表客户节点，d代表交易对象节点，t代表接收基站节点分类。图中黄色子网络为金融重构网络，红色子网络为客户属性子网，蓝色子网络为移动子网。本质上，异构网络将更多的外界属性分类指标加入到原始网络结构中，形成综合网络结构框架。将客户节点对应的客户基础属性节点（按年龄、性别两个维度进行分类）加入网络结构中，构成客户属性分类子网。

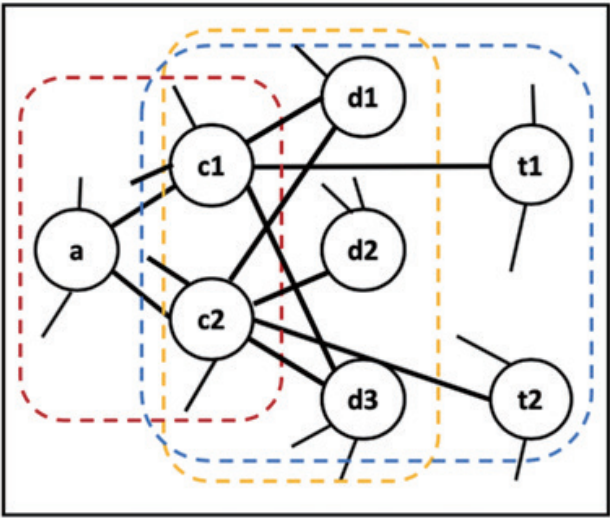


图3 异构综合网络



## 2.2 网络学习算法

基于同构交易网络的研究框架属于网络表示学习范畴。本文为网络表示学习中的节点表示学习算法在客户信息网络中的应用。这个领域在近几年得到了深度学习和数据挖掘领域学者的广泛关注。目前该领域技术主要包括三大类：基于矩阵分解、基于生成模型和基于深度学习算法。对 10 种最通用的网络表示算法在 8 个网络上进行的多标签分类实验结果表明，Deepwalk 及 Node2vec 这两种深度学习算法在各种类型的同构网络中均具有较为突出且稳定的效果。Bryan Perozzi (2014) 从自然语言处理领域的深度学习算法技术 word2vec 受到启发，将简单随机游走生成的网络路径类比为文本语料，得到每个节点的向量表示算法 Deepwalk，获得了很好的结果。Node2vec 算法 (2016) 则在 Deepwalk 基础之上，改进了其随机游走策略，提高了随机游走路径的质量，并利用负采样技术获得节点表示。Deepwalk 和 Node2vec 等常见网络学习算法是针对同构网络结构设计的网络表示学习算法，并不适用于具有更多节点类别和边类别的异构网络。本文采用 metapath2vec 算法 (Yu Xiaodong, 2017) 对异构网络进行节点表示学习。该算法基于元路径构建采样机制，在挖掘不同类型节点的关联关系方面在多类公开数据集上的表现优于 Deepwalk 及 Node2vec 等同构网络表示算法。

对于本研究而言，重构交易网络具备基本的网络结构。交

易发起节点及交易接收节点组成了网络中的顶点集合。边的构成方式依据边存在交易即存在连接边的原则。基于深度学习的模型目标是找到一组映射  $f: V \rightarrow R^d$ ，使得该映射最大化地保留网络中每一个节点出现在其领域的概率如公式 (1)：

$$\max_f \sum_{u \in V} \log \Pr(N_S(u) | f(u)) \quad (1)$$

理论上来说，交易行为类似的客户节点连接的公共交易对象节点集也十分类似，则其成为彼此邻域的概率更高。有效的网络表示算法应能保证客户节点对应的嵌入表示坐标更加接近。网络学习算法的核心是将网络结构转化为节点序列带入神经网络模型进行训练。每一个节点对应领域的样本采样方式十分关键。Deepwalk 采用的是简单随机游走机制，Node2vec 采用的是可以调节的随机游走机制，而 Metapath2vec 则采用元路径限制下的采样方式。得到采样数据后，三个方法采用类似的训练步骤，将采样轨迹作为 skip-gram 模型的数据源，获得每一个顶点的嵌入表征，并通过 negative sampling 进行参数迭代更新。

Deepwalk 网络学习框架通过简单随机游走机制，将网络结构转化为节点序列。Node2vec 在 Deepwalk 基础之上，在轨迹采样方式上考虑了网络结构的差异。假设路径中已获得  $(c_0, \dots, c_{i-1}, c_i)$  的采样服从以下概率分布：

$$P(c_i = x | c_{i-1}) = \begin{cases} \frac{\pi_{c_{i-1}x}}{z} & (c_{i-1}, x) \in E \\ 0 & \text{other wise} \end{cases}, \text{其中}, \pi_{c_{i-1}x} = \begin{cases} \frac{1}{p} & \text{if } d_{c_{i-2}x} = 0 \\ 1 & \text{if } d_{c_{i-2}x} = 1 \\ \frac{1}{q} & \text{if } d_{c_{i-2}x} = 2 \end{cases} \quad (2)$$

其中  $p$  和  $q$  为调整采样节点遍历网络的深度和广度的超参数。

异构网络学习框架在 Node2vec 的基础之上，考虑了节点

类别的差异，采取了基于元路径的随机采样方式。对于异构网络  $G = (V, E, T)$  以及长度为  $l$  的元路径方案  $P: V_1 \xrightarrow{R_1} V_2 \xrightarrow{R_2} \dots V_l \xrightarrow{R_l} V_{l+1} \dots \xrightarrow{R_{l+1}} V_p$ ， $c_i$  的采样服从以下概率分布：

$$P(c_i = x | c_{i-1}^{t-1}, p) = \begin{cases} \frac{1}{|N_t(c_{i-1}^{t-1})|} & (c_i, c_{i-1}^{t-1}) \in E, \varphi(c_i) = t \\ 0 & \text{other wise} \end{cases} \quad (3)$$

## 2.3 网络学习框架

本文提出了基于大数据的精准营销机器学习框架 (见图 4)。通过客户基础属性数据及行业衍生数据搭建异构网络，进行异构网络表示学习，模型输出为每一个客户节点所对应

的客户综合属性表征。将客户综合属性作为模型输入，对具有价值分类标签的客户数据进行分类模型训练。



图 4 精准营销机器学习框架

本研究通过一组可视化导图（见图5），阐释了网络学习框架。左上为一个原始网络，这个网络结构包含52个顶点及随机产生的连接边。右上图指出，这组网络结构具备明显的4个聚集结构（clusters）。有效的网络学习算法应具备识别聚集的能力，即为每一个节点分配对应的向量化表达，进而识别网络中的聚集结构。通过 node2vec 算法，获得的网络节点

对应的向量化坐标，左下方图为对应顶点坐标降维后投射在二维坐标系中的结果。将这组坐标通过聚类算法进行聚类，结果被分为四类。将这四类与对应的顶点匹配，可以从网络结构图（右下图）看出，网络识别算法的结果可以很好地识别网络聚集结构。

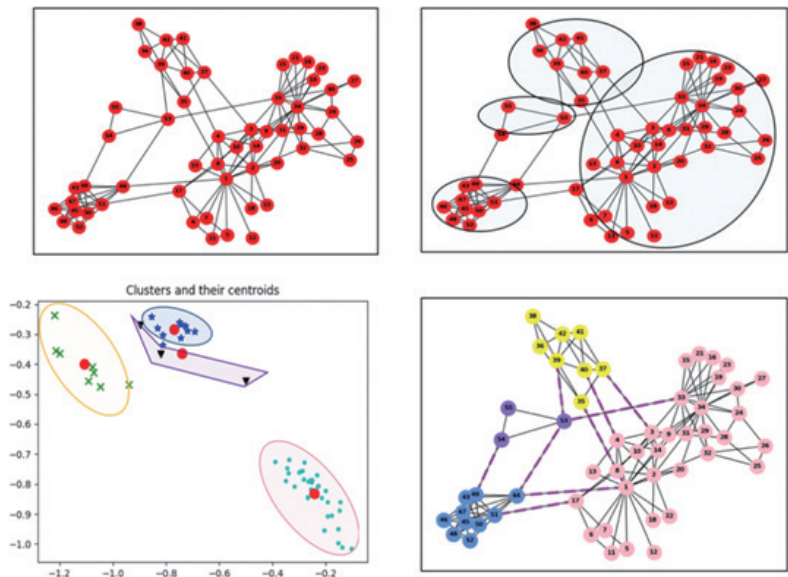


图5 网络分析示意图

### 3 试验结果

基于前述中阐述的研究方法和框架，本研究通过真实数据集验证方法的有效性。研究数据来自一家中资中等规模的城市商业银行2019年（2019年1月1日至2019年12月31日）的客户借记卡银联交易流水表单。选取的客户标准是活跃客户，即每月银联交易均超过5笔的客户。从满足条件的活跃客户集中随机选取的8000名客户。根据8000名客户在银行客户管理系统所对应的客户分类价值标签，客户集被分为高价值客户（金融资产大于50万），价值客户（金融资产大于20万），普通客户（在实验期内有过金融产品购买但未晋升至价值客户），未知客户（并未有金融产品购买记录）。其中，高价值客户共1645户，价值客户2292户，普通客户2387户，未知客户1676户。

根据金融交易网络的重构规则，通过对交易轨迹样本对应信息的分布统计确定网络对应的交易对象节点。依照交易轨迹样本，构成连接交易发起节点和交易对象节点的公共边。最终形成的重构交易网络由交易发起节点，交易对象节点，及两点间的连接边构成。网络由8478个节点，179691条边组成，

网络的密度为0.005。

本实验汇报了在不同算法下，由不同时间段数据集构成的，具备不同属性的子网络结构，对于客户节点表征在分类任务中的效率结果。首先，从网络构成角度考虑了网络研究的有效性。除了完整网络结构之外，本文考虑了基础信息子网及金融交易子网及这两个子网络所对应的网络学习结果对于分类效率的影响。其次，从数据角度考虑基于不同数据采集时间段构成的网络结构与基于全年数据的完整网络，在客户行为识别方面的差异。我们在一年数据的基础之上，也考虑了上半年和下半年数据构成的网络结构。最后，本实验进一步测试了不同网络识别算法在不同网络结构下的结果对应的分类效率。采用 Deepwalk 及 Node2vec 算法分别对同构网络（基础信息子网及金融交易子网）进行网络表示学习，采用 Metapath2vec, Node2vec 及 Deepwalk 分别对于异构网络（综合信息网络）进行网络表示学习。

对 Deepwalk 进行以下设置：节点表示的维度为128，每个节点开始的路径数为10，路径长度为100，窗口大小为5。Node2vec 在 Deepwalk 基础之上基础上添加两个参数 p、q 的设置，根据网格搜索的结果从 {0.25, 0.5, 1, 2, 4} 中选择最优结果。

将 Metapath2vec 的元路径为“ACDCTCDCA”，其余参数与 Deepwalk 一致。实验采用的训练数据与测试数据比为 0.7，分类器使用  $L_2$  正则技术的 one-vs-rest 的逻辑回归分类器，对具

有有效标签的客户节点数据，重复实验 10 次，并报告 10 次结果的平均微观  $F_1$  值。如表 2 所示。

表 2 实验分类效果：平均微观  $F_1$  值

	基础信息子网		金融交易子网		综合信息网络		
	Deepwalk	Node2vec	Deepwalk	Node2vec	Deepwalk	Node2vec	Metapath2vec
上半年	0.213	0.272	0.413	0.403	0.426	0.403	0.602
下半年	0.186	0.204	0.382	0.417	0.479	0.464	0.621
全年	0.231	0.227	0.401	0.432	0.538	0.507	0.778

实验结果表明：首先，基于完整网络结构（综合信息网络）的网络学习结果对应的分类效率，要明显高于基础信息子网及金融交易子网的分类结果。结果说明，具备多维度信息的异构网络相对于具备单一信息的网络结构，能够突出各类群体间的行为特征差异。其次，基于全年数据搭建的网络表示结果相对于半年数据，更能反应客户的行为属性特征，进而对客户价值分类的判断更有帮助。第三，通过使用针对异构网络设计的算法 Metapath2vec 得到的嵌入坐标的分类效果明显优于在 Deepwalk 及 Node2vec 学习框架下结果。这说明针对异构网络表示学习的算法更适用于异构网络。结果说明了基于客户基础属性和交易属性异构网络及异构网络表示学习算法的学习框架，能更好地识别客户行为及基础属性上的异同，将不同价值分类的客户区分开来。结合网络表示学习的获得的客户属性指标能显著增强了分类任务效率。

## 4 结束语

企业在日益激烈的市场竞争中若想立于不败，需要准确定位价值客户并将有限的营销资源向这一部分客群倾斜。在大数据的背景下，企业也积累了大量客户数据资源，包括客户属性数据、业务相关数据及行业衍生数据。本研究为大数据挖掘技术在企业精准营销领域的应用探索，并从移动运营商及商业银行角度做了具体分析。在方法方面，本研究创新地将客户行为模式的研究转化为网络结构的识别，基于客户移动轨迹数据及日常交易流水数据搭建重构交易网络，并借

助网络表示学习算法，将对应节点的网络表示结果作为交易行为属性指标纳入客户价值分类模型，从而解决了直接运用轨迹数据带来的数据维度高且稀疏的问题。

通过针对实验分析证明，将客户基础属性和交易行为属性放入统一的研究框架之下并利用异构网络学习算法获得的价值客户分类效率最高。各类具备大规模客户衍生数据的企业在实际业务中均可以采用本研究提供的研究框架，构建基于客户数据的网络，将未知客户的对应嵌入坐标导入分类器，获得相应客户分类，为营销团队实施营销策略提供了极大便利。本文采用的方法可以用于第三方支付公司、电信企业、政府部门等具备构建网络数据的实体，进行基于网络的个人行为研究。未来的研究方向包括：优化网络构建方式；针对不同规模数据集可能产生的研究结果差异进行进一步分析论证；加强异构网络分类任务研究框架的理论机制研究等。

（收稿日期：2020-06-24；技术审稿：丁飞；

见习责任编辑：韩菁菁）