

移动社交网络中的用户行为预测模型

唐杰 杨洋
清华大学

关键词：移动社交网络 用户行为建模

移动社交网络已经成为沟通真实物理世界和虚拟互联网空间的桥梁。人们在互连网络中的行为直接反映了人们在真实世界中的活动和情感。为了对个人的网络行为和情感进行预测，本文将介绍如何利用互连网络构建个人社交场景，并提出动态平滑概率因子图模型。模型同时考虑了网络结构、用户属性和网络用户的偏好，对用户的动态行为进行建模和预测。

介绍

与传统的社交网络服务相比，移动社交网络具有允许用户通过手机、平板电脑等移动终端工具实时互相关注及联系的独特优势。如今，Facebook、Twitter和人人网等热门社交网络都在逐步转向移动端。同时，移动社交网络如Foursquare、Gowalla和Path等也相继出现并走红于市场。移动社交网络所拥有的一些自然特性与传统的、基于Web的社交网络大为不同：首先，移动社交网络中的所有用户都采取实名制；其次，移动社交网络中用户之间的关系更贴近于他们的真实关系；再次，移动社交网络中用户的一切行为（如发短信、打电话、位置信息等）都是物理社交网络中用户行为的真

实反应。

研究表明^[1]，用户在移动社交网络中的行为会被众多微妙的因素所影响。那么，能否通过记录用户的行为和在网络中关系的动态变化对其行为进行预测呢？

图1是用户行为预测问题的一个样例。图的左侧展示了用户约翰（John）及其朋友的社交网络，每一位用户在时间 t 上都有相应的行为记录，目的是预测用户在时间 $t+1$ 的行为，例如在微博或者Twitter网络中，约翰会转发他朋友发出的帖子吗？这需要对用户行为进行建模，建模过程不仅要考虑用户的个人属性（如心情、位置、电话、短信

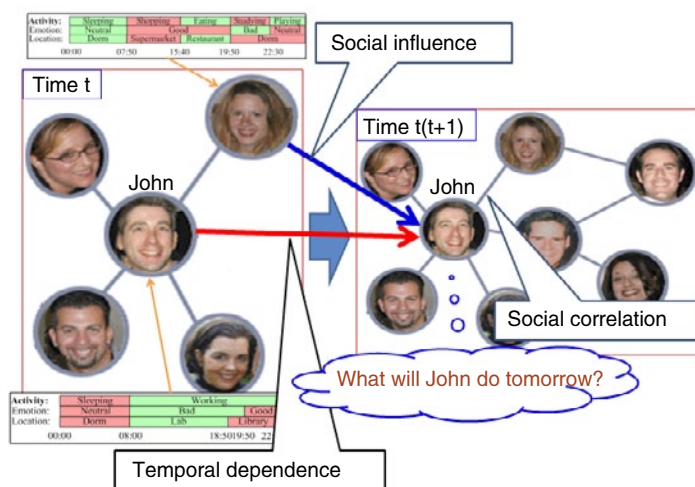


图1 用户行为预测问题示例

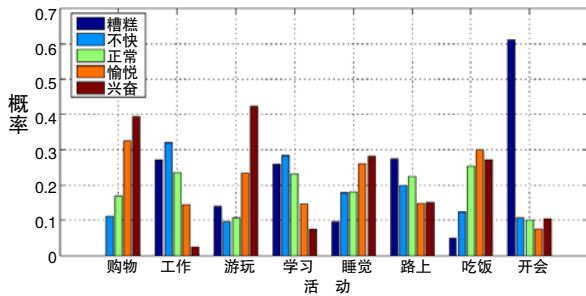


图2 (a) 用户心情与其活动的关联分析

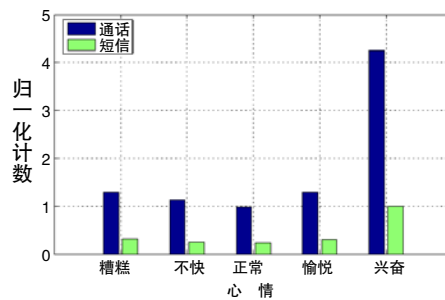


图2 (b) 用户心情与其通话/发信息记录的关联分析

等), 还要考虑用户的行为(社交影响力)和兴趣模式。

用户情感预测

本节通过对移动社交网络数据的分析, 向读者展现我们目前在用户情感分析方面得到的一些有趣的结果。该移动社交网络数据以清华大学在校学生为用户群, 记录了30位用户在2010年5~7月期间, 连续36000个小时的行为及状态。

在以下分析中, 我们集中讨论几个方面:

属性关联 一个人所处的环境及其行为会影响心情;

时间关联 一个人过去的心情会影响当前的心情;

社交关联 朋友的心情会影响本人的心情。

属性关联分析

每一位用户在社交网络中都有其特定的行为模式, 这些模式在一定程度上决定该用户未来的属性状态。在移动社交网络中, 用户的主要属性包括: 活动、位置、通话记录、短信等。我们首先分析每一个属性对用户心情状态的影响。

图2(a)展现了用户所参与的活动和当时心情的关联关系。可以看到当一个人在玩或者逛街的时候, 心情状态趋向于“正心情”(觉得快乐、开心); 参加会议的时候心情最差(感到烦躁或难过)。同时还发现一个有趣的现象: 用户通常在

路上的时候, 心情比较糟糕, 这可能和北京拥堵的交通状况有关。图3展示了用户所在位置与其心情状态的关联关系。用户的心情状态隐性地构成了“快乐区域”(在图中用红圈表示), 和“悲伤区域”(在图中用绿圈表示)。从图中可以看出, 用户(由学生构成)在宿舍或操场附近时, 大多是快乐的心情, 而在教室或图书馆附近时, 则趋向于“负心情”。

再看对用户通话及短信的观察分析, 图2(b)展现了分析结果。总体来说, 64%的用户倾向于和朋友分享他们的快乐, 36%的用户倾向于在

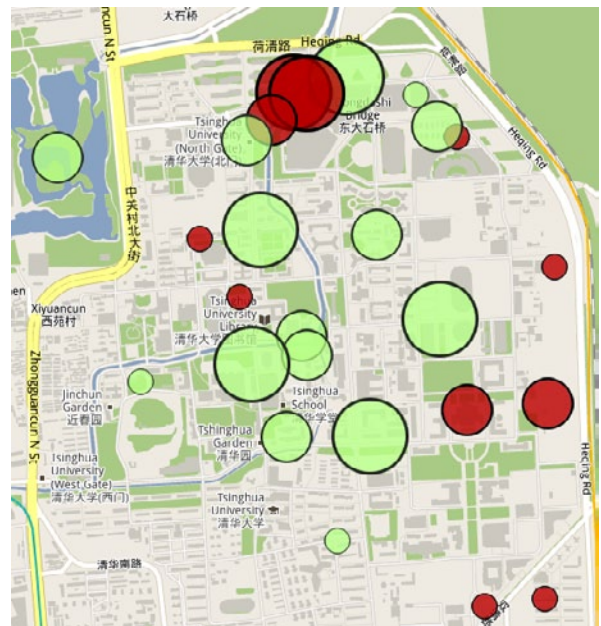


图3 用户心情与其位置的关联分析

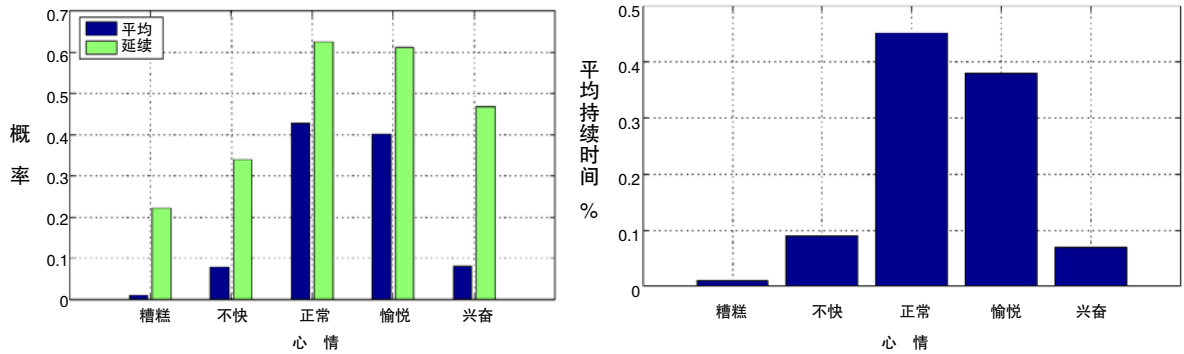


图4 (a) 用户当前的心情状态受以前心情的影响 (b) 不同心情状态的平均持续时间分析 (“平均”代表用户不同心情状态的似然值, “延续”代表已知用户以前的心情状态, 维持同样心情至当前的似然值。)

不开心时和朋友们聊天。另一个观察结果是, 根据用户的心情状态, 其好友隐性地组成了两个团体, “开心的朋友”和“不开心的朋友”。为了更好地说明这一点, 我们从数据集中选择出了一些比较活跃的用户 (如打电话次数较多的用户), 并统计了他们在两个月内联系过的人。为了防止偏差, 我们移除了与当前用户在两个月内通话不足10分钟的联系者。然后根据用户给联系人打电话时拥有“正心情”的比例, 将联系人进行分组。出乎意料的是, 用户的朋友们往往能被清楚地分到不同的组中。

时间关联分析

用户当前的心情状态与过去的心情有很大关系。图4 (a) 展现了用户在移动社交网络中心情状态与时间的关联关系。一名用户在此之前拥有“正心情” (图中以愉悦或兴奋标识), 对比其他用

户, 他更有可能在当前依然拥有“正心情”。图4 (b) 展现了不同心情状态的平均持续时间。总体上, “正心情”比“负心情”持续的时间更长久。

社交关联分析

用户的心情状态很可能被朋友影响, 若一名用户感到快乐, 他很可能会感染他的朋友们, 令他们也感到快乐。该类社交情绪的影响问题被广泛地研究过, 如克里斯坦基斯和福勒 (Christakis and Fowler)^[2]。而怀特菲尔德 (Whitfield)^[3]则研究过一种特殊的情绪传播案例。他们发现在社交网络中, 情感具有“三度” (一度: 我们的朋友; 二度: 我们朋友的朋友; 三度: 我们朋友的朋友的朋友) 影响力: “快乐”情感往往会传播给三度关系内的人群, 即当一位用户感到快乐时, 其在三度以内的朋友也会感到快乐。据此, 我们尝试验证情绪

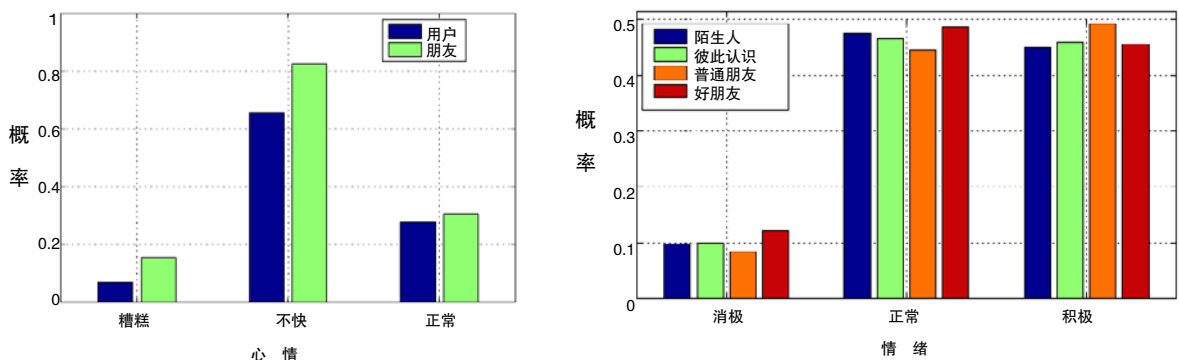


图5 (a) 用户的朋友对用户心情状态的影响 (b) 不同类型的朋友对用户心情的影响

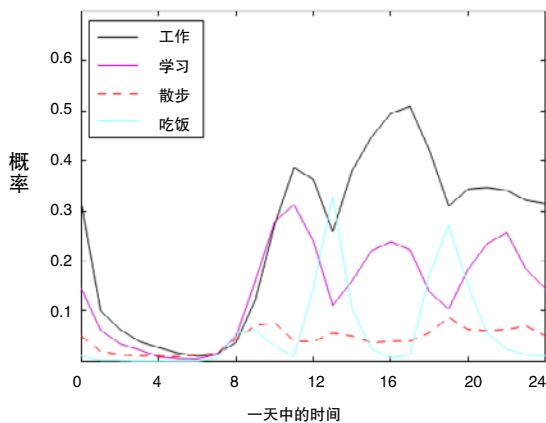
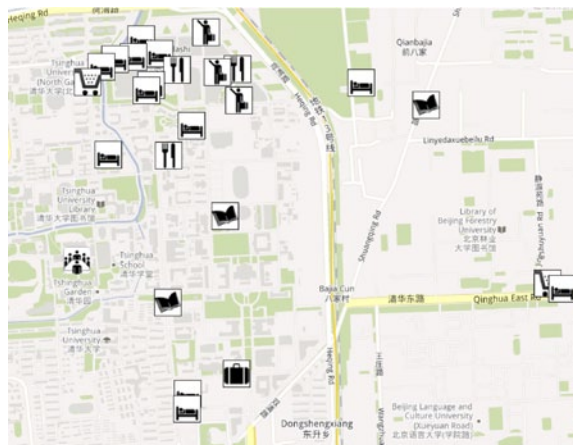


图6 (a) 用户在不同时间段不同行为的似然值



(b) 用户在不同地点的行为标注

状态是否在移动社交网络中也存在类似的情况。

图5(a)展现了用户的心情状态是如何被朋友的情绪所影响的。研究发现用户和朋友有可能拥有相同的情绪状态。“负心情”在社交网络中更有感染力，当一位用户心情糟糕(terrible)时，他朋友的心情也糟糕的可能性翻了一番(18% Vs. 9%)。存在社交关系的用户可分为“好朋友”、“普通朋友”以及“彼此认识”。对拥有不同社交关系的用户进行分析，结果表明用户倾向于和“普通朋友”分享快乐，和“好朋友”分享不快，见图5(b)。

用户行为预测

采用与上节同样的数据，对用户的行为进行分析，并验证用户的行为是否有和情绪状态类似的情况。

我们主要讨论以下几个方面：

属性关联 用户的属性是如何影响行为的；

时间关联 用户以前的行为模式是如何影响当前行为的；

社交关联 用户的行为是如何影响好友的行为的。

属性关联分析

用户的行为与情绪状态有紧密的联系。由于用户的行为存在周期性，所以时间是影响用户行为的

重要属性。图6(a)展示了用户在不同的时间段可能进行的活动；图6(b)则表示，地理位置也是影响用户行为的重要属性之一。用户的行为隐性地构成了“吃饭”和“学习”两个类别。

时间关联分析

用户当前的行为受较早时间段内行为的影响。用户可能会持续前一个时间段的行为，图7(a)证实了这一时间关联性。

社交关联分析

图7(b)表明在移动社交网络中，用户的行为受好友的影响。对比陌生人，用户更有可能和好友做相同的事情。“学习”在学生的社交网络中拥有很强的传播性，当一名用户学习时，好友学习的可能性比陌生人高出了一倍。

经分析，用户的行为和情绪状态均存在属性关联、时间关联以及社交关联。那么，是否可以用相同或相似的模型和方法来预测用户的行为和情绪状态呢？

模型与方法

基于以上分析，我们利用概率因子图(factor graph)将问题模型化，在移动社交网络中使用基于梯度下降的学习算法训练模型来预测用户的行

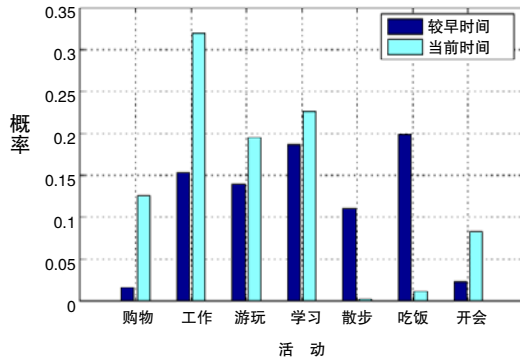
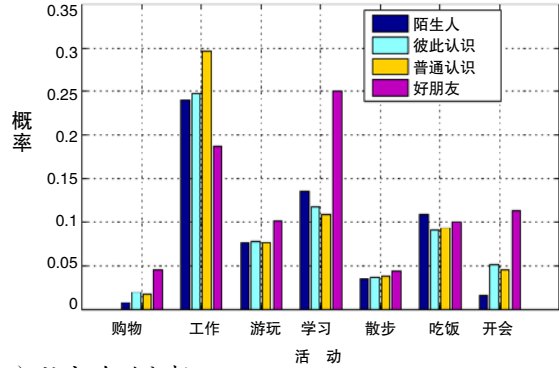


图7 (a) 时间关联分析



(b) 社交关联分析

为。模型的基本思想是用因子函数 (factor function) 定义关联关系, 将所有因子函数的联合概率定义为问题的目标函数。因此, 问题就转化为根据输入的动态网络和学习模型参数, 使用户行为状态最大化的后验概率问题。

针对三种关联关系, 定义了三种因子函数:

属性关联因子函数 $\{f(x_{ij}^t, y_i^t)\}_j$ 表示用户 i 在时间 t 的属性值与行为 y_i^t 的依赖关系。

时间关联因子函数 $h(y_i^{t'}, y_i^t)$, 其中 $t' < t$, 表示用户在时间 t 的行为和近期行为的依赖关系。

社交关联因子函数 $g(y_i^t, y_j^{t'})$, 其中 $t' < t$, 表示用户 i 在时间 t 的行为对用户 j 在时间 t' 的行为的影响。

根据不同的应用背景以及对问题的分析, 这三种因子函数可以用不同的方法实例化。如何定义因子函数请参见文献[4]。

基于因子函数定义问题的目标函数, 对目标函数的求解是为了学习概率因子图中的未知参数集合 θ 。

$$L(\theta) = \log p(Y|G^t)$$

其中 Y 表示用户行为的集合; G^t 表示在时间 t 的社交网络结构; $p(Y|G^t)$ 表示所有因子函数的乘积。

在概率图模型中, 计算归一化因子, 需要对所有 Y 对应的 $p(Y|G^t)$ 求和, 因此计算量与网络中的结点数呈指数关系递增。谭 (Tan) 等人^[3]将因子函数定义为二次函数, 利用改造的多元高斯分布计算归一化因子, 联合树 (Junction Tree)^[7]和可信度传播 (Belief Propagation)^[8]等算法计算归一化因子的近似值。本文的算法是: 对特定的用户

及其行为标记, 计算相应的边缘概率, 用梯度下降的方法更新每个参数的值。得到模型参数后, 抽样出一组 Y 的可能取值, 计算相应的目标函数的值, 多次迭代后得到的最优解即为预测的用户行为标记。

结语

本文研究了移动社交网络中的用户行为预测问题。基于移动社交网络数据, 分析了用户的情绪及行为状态, 提出了概率因子图模型, 并对模型中的概率推理、参数学习、预测算法等问题进行了讨论。■



唐杰

CCF高级会员。清华大学副教授。主要研究方向为社会网络分析和数据挖掘。
jietang@tsinghua.edu.cn



杨洋

清华大学计算机科学与技术系博士生。主要研究方向为机器学习与社会网络分析。SherlockBourne@gmail.com

参考文献

- [1] M. Granovetter. 1973. The strength of weak ties. American Journal of Sociology 78(6):1360~1380

- [2] J. H. Fowler and N. A. Christakis, Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. *British Medical Journal*, 2008
- [3] C. Tan, J. Tang, J. Sun, Q. Lin, and F. Wang, Social action tracking via noise tolerant time-varying factor graphs. *KDD' 10*, 2010, 1049~1058
- [4] J. Tang, Y. Zhang, J. Sun, J. Rao, W. Yu, Y. Chen, and ACM Fong. Quantitative Study of Individual Emotional States in Social Networks. *IEEE Transactions on Affective Computing (TAC)*
- [5] J. Tang, T. Lou, and J. Kleinberg. Inferring Social Ties across Heterogeneous Networks. *WSDM'12*, 2012, 743~752
- [6] J. Whitfield, The secret of happiness: grinning on the internet. *Nature*, 2008
- [7] W. Wiegerinck, Variational approximations between mean field theory and the junction tree algorithm. *UAI' 00*, 2000, 626~633
- [8] J. S. Yedidia, W. T. Freeman, and Y. Weiss, Generalized belief propagation. *NIPS' 01*, 2001, 689~695