

doi:10.1631/FITEE.1500385

题目: 基于稀疏噪声数据的家的位置推断：模型与应用

目的: 家，是人们生活的中心。由于家的特殊意义，在对于人类活动的研究中，确定家的位置就显得尤为重要。本文旨在从一个人的签到记录上准确预测家的具体位置（精度在 100 米以内）。

创新点: 由于家的位置属于隐私，我们无法，也不能直接使用用户的隐私数据来进行研究。因此数据的采集和近似是第一个难题。本文的解决方法是认为人们在家里说的话跟在外面说的话不一样。由于人们在家里签到会说一些特点的词汇，比如“睡觉”、“洗澡”，等等。我们收集了带有这样词汇的签到，然后把这样的签到句子经由多人筛选。如果所有人都认为某一条签到是来自家里的，我们就认为这个签到的位置是发送者的家的位置。

方法: 从人们的签到中抽取一些关键的特征，再把这些特征经由数据挖掘的算法提炼得出一个综合的判断。我们考虑的特征包括，人们出现在某地点的频率、时间，以及是否在夜间出现等等。

结论: 实验证明，可以以 70%+的准确率预测 70%+的活跃社交网络用户，而且精度是 100 米以内。

关键词: 家的位置；移动模式；医疗保健