

尊敬的吴老师，您好

首先有个坏消息：我的复杂网络暑期课程申请没有通过，被拒了。。

在最近的一段时间里，我实现了预测转发链长度的代码。转发链长度预测是在已有的微博关系网络和LIS模型下，预测一条微博的转发用户数量。具体的，假设一条微博 m 从用户 u 开始传播，用户 u 的近邻 v_i 接收到微博 m 后，转发此条微博的概率为

$$1 - \exp(-\lambda I_u^T S_{v_i}) \quad (1)$$

其中 I_u 和 S_{v_i} 分别为用户 u 的influence向量和用户 v_i 的susceptibility向量。我们使用轮盘赌方法来决定用户 v_i 是否转发了此条微博，如果转发了，则将用户 v_i 压入队列。考察完用户 u 的所有近邻后，依次取出队列中的用户进行考察，直至队列为空。在此过程中被激活的用户数量则为微博 m 的转发链长度。

对微博数量为 n 的test集来说，经过上述的预测过程后，记录每一转发链长度对应的微博数量，得到以转发链长度为下标的向量 A 。同样的，根据真实的数据得到向量 B 。经过平滑处理后，最后我们利用mean absolute percentage error(MAPE)来评估预测情况。MAPE的计算公式为

$$M = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (2)$$

其中 A_t 是实际值， F_t 是预测值。

实验测试所用的数据集包含199408个用户、972340条连边、395852条微博。实验结果如图1所示，结果并不好。真实数据在转发链长度为2处有一个波峰，而预测结果的波峰在1处。同时，MAPE值为1.27，比论文中的结果0.140大了一个数量级，原因暂时还不清楚。

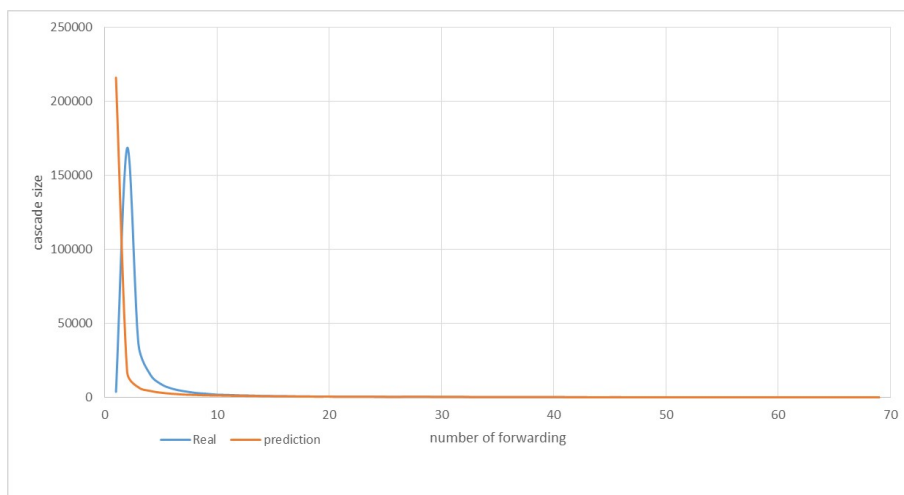


图 1: 预测结果

另外，我读了其他几篇有关微博转发预测的论文，他们解决问题的角度或在于对用户之间的连边权值建模来预测，或结合微博的具体消息内容来预测。

学生王超民，2016年6月20日