**Weekly Report**

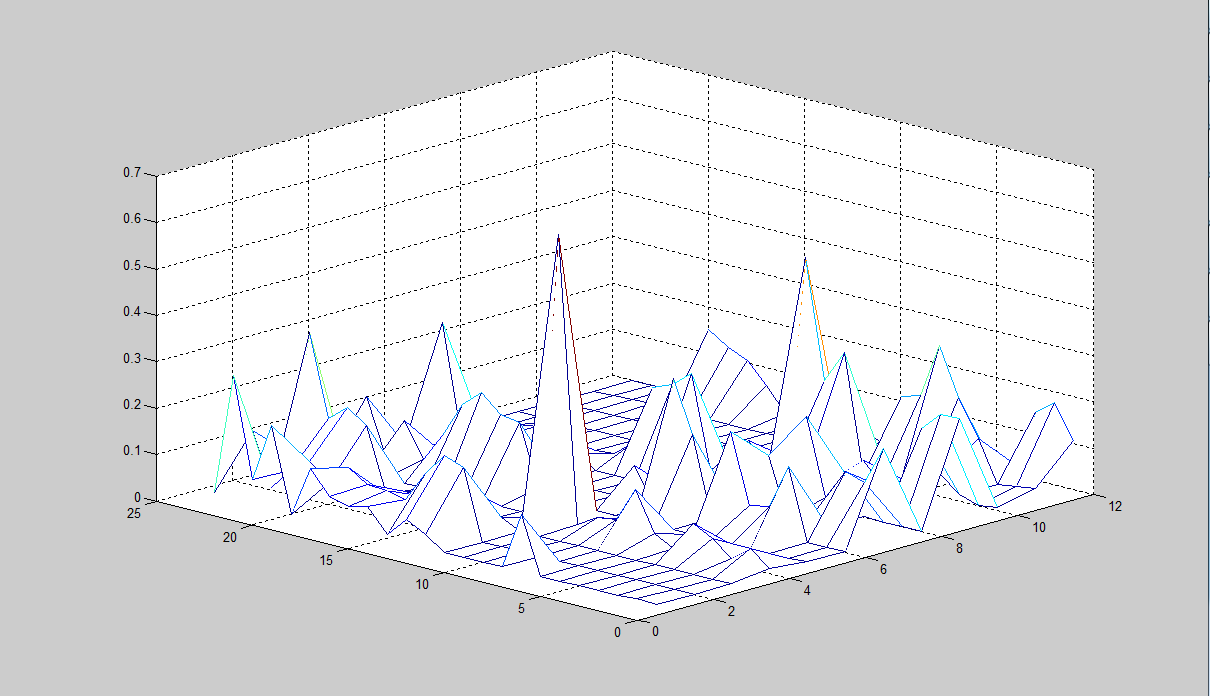
姓名：朱恬骅

时间： 2012-7-23

1. 上周计划及其完成情况：
   1. 工作总结

试了Financial Topic Models一文中的方法；看了几篇论文，主要是考虑利用时间先后顺序的特征来对股票按照价格进行聚类。

* 1. 主要成果



这个是用12支股票（三类各4支）用Financial Topic Models一文中的方法运行出来的结果，24个分别表示12支股票的涨跌，12个表示产生的12个话题。从图中可以看到，同一板块的股票涨跌有一定的相关性。

上网的时候看到一篇有关关键词提取的论文（阅读文献1），他们将一片文挡的标题和内容作为对同一个事物的两个描述，进而将二者看作对同一事物的两种语言描述，从而训练模型提取关键字。在这个过程中，他们抛弃了topic model，转而采用 一种“翻译”的思想，就是将长的文档内容翻译成较短的关键词组合。他们采用的模型（word alignment models，阅读文献2）可以考虑每个单词的前后顺序问题。不过这个模型需要两个视图中提供等长的信息进行训练。

Topics over time这篇论文，关注的是topic随时间的变化问题。可以说把Financial topic models这篇中的方法应用到一个比较小的时间段上去，然后不同时间段同一股票所属topic的变化，是topics over time可以解决的问题。但它不能帮助我们提取股价在时间上的特征。

* 1. 技术报告
  2. 阅读文献

Zhiyuan Liu, Xinxiong Chen, Yabin Zheng, and Maosong Sun. 2011. Automatic keyphrase extraction by bridging vocabulary gap. In Proceedings of the Fifteenth Conference on Computational Natural Language Learning (CoNLL '11)

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. Comput. Linguist. 29, 1 (March 2003), 19-51.

G. Doyle and C. Elkan. Financial topic models. In Working Notes of the NIPS-2009 Workshop on "Applications for Topic Models: Text and Beyond Workshop", 2009.

Xuerui Wang , Andrew Mccallum: Topics over time: a non-Markov continuous-time model of topical trends, In SIGKDD ’06

1. 下一周详细安排计划：
   1. 目标

找一个topics over time的实现看看结果。周四（26号）到下周一（30号）要去台湾旅游。

* 1. 详细安排
  2. 阅读文献

1. 补充说明