**会议纪要 – 2017.4.8**

**每周工作要求：**

* 每周五下班之前将本周所有工作资料（含：工作报告、实验数据、程序、论文、项目资料等）上传到FTP服务器上做备份，所有人员必须严格执行，以防资料因故遗失；
* 每天离开实验室，关闭电脑(包括显示器电源)，关闭显示器电源，关闭电灯，关闭实验室空调，饮水机电源，以保证实验室用电安全;
* 本学期每周需要提交的会议纪要人员：
  + 研究生：李佳、贺星伟
  + 本科生：周建宇、蒋浩（物理系基科班）、李浩冉（自动化系，第九周加入）
  + 外校研究生：高成亮，朱龙霞，柴瑜晗
  + 外校本科生：暴宇翔（北理工）、王雅轩（北邮）、谢文彬（北航）、林哲宇（北林实验班）、莫林立（北林实验班）

**关于纪要的问题：**

1. 会议纪要是对上周工作的总结，是对下周工作的计划；
2. 目前学生会议纪要存在如下突出问题，需要尽快改进：
   1. 总结部分太简略；
   2. 计划部分过于简单，起不到指导下周工作的作用；
   3. 工作内容规划不合理，工作量偏少；
3. 每周五需对前一阶段所有研究工作做备份，在FTP服务器上完成备份。
4. 人走关机，离开实验室务必关闭桌面电脑，显示器电源，以免火灾。
5. **李佳**

* **上周工作**
  + 项目：去华夏银行部署上线研发内容三代码；分析日志的新数据；调研序列模式算法；
  + 论文：开始毕业论文的写作
* **本周工作**
  + 论文：撰写毕业论文，**目标4月23日提交第一稿**；
  + 项目：得到日志规则挖掘的第一版结果；去华夏银行部署上线研发内容二代码；**本周末例会上讨论金融数据挖掘研究内容1的结果**。

1. **贺星伟**

* **上周工作**
  + 使用origin重新绘制了论文中的图片，增大了字体以及分辨率；
  + 运行实验计算了BTM,FastBTM,LDA三个主题模型在Twitter2011 collection上对应K=200, 500的PMI score.
  + 书写完成了前期ijcai 论文专利，已发给刘静。
* **本周工作**
  + KBS的论文修改：
    - 依据主编意见，为论文添加interactive plot。
    - 对论文中的设计的超参数进行解释，回复审稿人意见。
    - 打磨，润色论文的语言，在本周完成一审修改的提交。
    - 目标：本周五之前提交论文的修改和回复信。
  + **周六，周日**：完成两台服务器深度学习平台的搭建，务必完成 + 目标：下周一可以开始使用。

1. **周建宇**

* **上周工作**
  + 实现了初步的中文 word embedding层，并进行了可视化效果评估
* **本周工作**
  + 目前的两条结束路线：
    - 直接基于CNN模型的答案提取，模型参数调整和实验效果提升（调节参数，数据预处理，争取复现论文中关于word embedding 的效果）；
    - 基于机器翻译的答案提取算法参数调整和效果提升；
  + 补充工作：整理中期科研笔记，尽快完成。

1. **高成亮**

* **上周工作**
  + 对已切分的文本进行繁体简化，去停用词的操作。
  + 训练最终实验使用的词向量表，搜索并整合用于测试的微博数据。
  + 进行第一轮实验测试。
* **本周工作**
  + 基于句向量的模型调整：
    - 调试句向量模型实验参数，以便实现最好的试验状态
    - 对中文微博文本进行依存句法分析，缩减模型的输入长度（句子长度）。**优化向量的转化信息。**
  + **开始尝试段向量的实验的工作**
  + **对比几类向量的实验结果**

1. **朱龙霞**

* **上周工作**
  + 更改中文微博语料，用BTM进行测试
  + 更换语料后，发现实验聚类效果较好
  + 为每一条微博进行标签发现文档主题概率小于50%的标签准确率较低(人工评判)
* **本周工作**
  + **相关论文实验的设计，**用BTM对有标签的微博语料进行兴趣抽取
  + **基于前期微博用户兴趣的语料数据，**将抽取的结果与标签结果进行准确度分析
  + 查找论文，找文档主题概率小于50%的解决方法

1. **柴瑜晗**

* **上周工作**
  + 通过给定术语集，完成词共现、lda两种构建图的方法实现。
* **本周工作**
  + 确认和分析术语抽取的技术路线
  + 在确认上述研究技术路线的基础上，考虑术语抽取-》概念抽取的实施过程的每一步细节

1. **林哲宇**

* **本周工作**
  + 继续学习JAVA的语法知识，
  + 继续研究liubing的论文，并对不懂的知识做相关的学习
  + 研究jieba分词的操作和调用

1. **莫林立**

* **本周工作**
  + 阅读完《Sentiment Analysis and Opinion Mining》
  + 搜狗细胞词库破解，**手动扩展Jieba的分词词库**

1. **王雅轩**

* **上周工作**
  + 完成有限问题域下的QA系统的问题匹配工作，正确率达到97.53%
* **本周工作**
  + 完成对之前工作代码的整理以及报告，周五之前完成，并在服务器上压缩提交。
  + 基于问题和相关文档（篇章）的答案自动抽取研发：
    - 进行文献调研，了解近2-3年国际上对于答案抽取的最新研究成果
    - 归纳总结答案提取的相关技术及算法，在中文条件下考虑可行性，提出自己的初步想法

1. **谢文彬**

* **本周工作**
  + 根据问题的检索结果的排序：
    - 提出了新的短期目标--应当实现搜索结果的重排序,优先级的搜索结果应当蕴含答案.
    - 查看复旦大学相关团队的工作

1. **蒋浩**

* **本周工作**
  + 相关词表的扩充工作：
    - 下载或找学长要到 word2vec 词转化成向量
    - 运行原有的相关词表程序，将语料库在原有（4.3G）100万条的基础上扩充
    - 扩充后的相关词词表进行数据清洗