针对资讯的用户建模和个性推荐系统

## 【命题企业介绍】

浙江网新恒天软件有限公司（恒天软件） 成立于2004年，是一家立足于为中外企业提供可靠的、专业的IT服务与产品的公司。恒天软件总部位于杭州，在上海、合肥，波士顿等地建立了研发和营销网络。目前公司员工2000多人，并且96%以上获得大学本科及以上学位。截止2014年，恒天软件积累了100多个客户，覆盖金融、制造、医疗、零售、物流、高科技、新能源等行业，遍布在中国、美国、加拿大、澳大利亚、英国、爱尔兰以及日本等国家。

恒天软件目前已取得ISO 27001国际信息安全管理体系认证和CMMI软件质量管理五级资质，并获得“2013-2014年度国家规划布局内重点软件企业”、“中国服务外包成长型企业100强”、“中国软件和信息服务业最具潜力奖”、“中国软件和信息服务业突出贡献奖”、“中国软件外包十大龙头企业”、“浙江最佳创新软件企业”、“全球最佳服务外包供应商ITO中国二十强”等荣誉称号；恒天软件多次获得中央电视台、中国日报、人民网、Gartner、Reuters、ComputerWorld、美国CIO杂志等主流媒体的关注和报道。

依托浙江大学、浙大网新和美国道富银行的战略合作联盟，恒天软件秉承“让信息系统更简单、让金融活动更简单”的愿景，在实践中不断总结，逐步积累了深厚的业务知识、领先的技术服务水平与成熟的项目管理能力。恒天软件专注于为中外企业提供应用开发、测试与质量保证、云与大数据、移动应用、系统重构、架构咨询、研究与开发等多项技术服务。

命题企业将对本项目获二等奖以上的优秀团队颁发奖金。

## 1.项目背景说明

**【整体背景】**

近几年，全球数据量出现爆炸式增长，数据成了当今社会增长最快的资源之一。如此增长迅速、庞大繁杂的数据资源，给传统的数据分析、处理技术带来了巨大的挑战。大数据时代的出现简单的讲是海量数据同完美计算能力结合的结果，大数据计算技术完美地解决了海量数据的收集、存储、计算、分析的问题。大数据时代开启人类社会利用数据价值的另一个时代。

**2.项目说明**

**【问题说明】**

大数据时代，数据量够大，涵盖范围够广，新数据源在不断产生并持续产生中。每个人都有很多信息接收渠道，却无法很好地从这些信息中找到真正自己感兴趣的内容。为此，有很多基于个性推荐的应用脱颖而出，这里面最出名的，就是今日头条。通过对用户的兴趣建模，这些应用能有效地发现用户关心的内容，并推送给用户。

**【用户期望】**

设计并开发一个资讯网站平台，通过对用户兴趣的追踪和建模，为用户推荐符合其阅读兴趣的内容。命题单位会提供相关资讯数据。要求：

1. 要识别优质的和低质的资讯内容，优先推送优质内容。
2. 只有用户仔细阅读的资讯才能作为兴趣来源；
3. 不能持续推送过于相似的资讯，以免影响阅读体验；
4. 能通过用户兴趣进行隐性的用户分组，同组用户的内容可以组内传播；
5. 能识别用户强烈不感兴趣的内容，并主动避推；
6. 通过训练的模型对资讯进行实时分类；
7. 资讯有美观的展示，阅读体验佳。

**3.任务要求**

**（1）总体功能要求**

**用户功能：**

支持用户注册，登录，修改用户信息及密码，找回密码

**资讯浏览功能：**

浏览资讯列表，可以动态加载，列表项可以点赞或点踩。

浏览资讯详情，可以点赞或点踩。

显示资讯的分类信息。

可以查看本人兴趣模型。

**（2）项目技术要求**

项目前端为HTML5展示，通过restful调用与后台通讯。后台开发语言不限，平台及操作系统不限，数据存储形式不限。

训练语料由参赛团队自行搜集。

开发期间与评选期间命题单位提供的数据内容可能不一致。开发用数据为JSON格式的静态资讯数据；评选期间，利用命题单位提供相同JSON格式的实时数据接口进行资讯兴趣匹配、推送与展示，该接口接收当前offset (数字，初始是0，按已接收资讯量递增) 作为参数。数据接口地址可作为网站配置参数。