**2019年北京高校数学建模校际联赛题目**

（请先阅读“全国大学生数学建模竞赛论文格式规范”）

**B题:和谐宜居之都**

十八大以来，我国政府高层决策相继把建设和谐宜居城市提升到前所未有的战略高度，《国家新型城镇化规划（2014—2020）》、中央城市工作会议、十三五规划纲要等均对建设和谐宜居城市作出专门论述，明确指出要提高城市发展宜居性，并把建设和谐宜居城市作为城市发展的主要目标。

北京作为现代化的国际大都市，其城市战略定位以及未来发展方向在全国具有示范作用，因此北京和谐宜居之都的建设更是备受重视。2014年和2017年习近平总书记在两次视察北京工作时强调“要明确城市战略定位，坚持和强化首都全国政治中心、文化中心、国际交往中心、科技创新中心的核心功能，深入实施人文北京、科技北京、绿色北京战略，努力把北京建设成为国际一流的和谐宜居之都”，这一战略目标为北京未来城市发展指明了方向。

北京为了 “和谐宜居”的目标，实施了大量城市建设的工作。城市客观的发展状况是建设和谐宜居之都的基础，但其最终目标是提升居民主观的居住感受。因此城市建设的规划者非常关心广大北京居民是否感受到了和谐宜居水平的变化，并希望通过对居民主观感受的分析，进一步了解在建设中需要加以改进或修正的方面。

大众的主观居住感受可以从社交网络上存储的各种形式的留言和评论信息获得。请你的团队根据在某社交网络上收集的2015年到2018年带有与和谐宜居相关的关键词的评论（附件1），完成以下任务。

1. 判断每条评论中包含的对城市“和谐宜居”看法的情感语意，尝试为评论包含的情感打分。这一打分应当涉及情感的正负向和强烈程度两方面，例如“最近的天气不错。”为正向情感；“今天的天气实在是太好了！”也为正向情感，且是相比“最近的天气不错。”更强烈的正向情感。

【备注】在我们所提供的评论数据中，存在很多没有意义的评论。你们团队需要判断出哪些评论是有用的，进而用这些有用的评论再去进一步进行研究。如何判断一个评论有没有用呢？使用人工判断当然是最准确的方法，但是这样做效率太低。幸运的是随着大数据技术的发展，现在有很多机器学习方法可以帮助我们进行这样的判断（事实上就是将评论分为有意义和无意义两类），在R语言和Python语言平台上，就有很多成熟的机器学习程序包可以供你们使用。当然这仅是一种可行性方案，你们团队也可以选择其他方法来解决这个问题。

1. 基于问题1得到的数据， 构建一个反映居民对城市“和谐宜居”的满意度综合指数,分析影响居民对“和谐宜居”主观感受的关键因素。
2. 请结合北京市2015-2018年的经济、环境、基础设施建设等客观数据，分析客观现实与主观感受之间的关联；研究怎样合理规划北京市城市建设才能有效提高民众对和谐宜居的主观感受。
3. 根据你们团队的研究结果，给北京市政府写一封信，为北京的城市规划和建设提出建议。

【注】 需要说明的是，一个城市是否“和谐宜居”，包含了很多内在含义。在本题所附的数据中，通过“关键词”这一列数据，体现了我们对“和谐宜居”的理解（当然，你的团队也可以在此基础上进一步加入你们的理解）。同时，所有数据我们都做了脱敏处理，保证了评论者的隐私不被泄露。

参考材料：

1. 蔡奇：更加紧密团结在以习近平同志为核心的党中央周围 为建设国际一流的和谐宜居之都而努力奋斗http://cpc.people.com.cn/n1/2017/0627/c64102-29365001.html
2. 北京城市总体规划(2016年—2035年)（附表 建设国际一流的和谐宜居之都评价指标体系）<http://www.beijing.gov.cn/gongkai/guihua/2841/6640/1700220/1532470/index.html>
3. 北京统计局（北京统计年鉴）http://www.bjstats.gov.cn/

“和谐宜居之都评论数据” 见附件1。

1. 应该是从微博爬下来的数据，所以有很多@回复这样的东西。
2. 会有一些空数据和一些表情包以及无意义字符
3. 对和谐宜居的指标进行主成分分析？从而找到关键因素？分词的作用？是否可以使用情感分数的离散度来衡量🡪代表情感波动？