Good morning, distinguished professors. I am Zhaoyang Tian. My paper’s topic is Emotional Analysis & Evaluation of Pros and Cons Regarding Students’ Activities ---- Based on Comprehensive Education Evaluation System in Tsinghua University High School.

早上好，尊敬的教授们。我是田肇阳。本文的主题是综合素质评价数据集的情感分析和活动正负面选择——基于清华大学附属中学综合素质评价数据集。

这是我们演讲的议程。我将主要讨论以下三个部分。

首先，我将介绍研究背景、目的和一般过程。清华大学高中在北京市教育委员会的资助下，为北京普通高中学生开发了全面的质量评估平台，旨在让学生自愿上传照片和文字描述，展示他们在课堂上或成绩上无法体现的特殊活动和力量。CQE系统已经实施了一年多，对CQE中的数据还没有进行研究。

以往的研究主要集中在方法论和少数研究者的独特实例上，但数据量小，实验具体。针对这一情况，本文进行了研究，希望能以统计的方式提供一个切实可行的解决方案。

本文以一种全新的方式展示了数据集，评估了数据背后的积极和消极情绪，并推荐了供学校参考的活动。

这是我们研究的一般过程。从我们学校的售后记录来看，我先把句子分成单词。然后我们继续对数据进行处理，并对词频进行分析，以获得大致的印象。接下来，我们根据频率组装一些wordclouds，以一种新的方式演示数据。在此之后，我们创造了一种原始的方法来分析积极和消极的感觉，并使用snownlp来支持这种方法。最后，我们运用支持向量机来提取那些对学生产生强烈情感的活动至关重要的单词。最后，我手动搜索流行活动的积极和消极方面。

本人于2018年2月至2018年6月在清华大学高中G17年级CQE系统中获取数据。每一段数据都是学生上传的活动。本文有6个维度需要研究。

本文应用分词法和词频分析法对数据进行分析，找出最受欢迎的活动。利用这种方法，本文找到了最常被提及的活动，它们代表了受欢迎的活动。本文将这些数据映射成一个条形图来演示这些数据。我们可以看到运动会、篮球赛和足球赛是最频繁的活动。

本文借助于wordcloud模块，通过参数集生成wordclouds，并根据上面计算的频率加载数据。

本文还提出了一种评价句子情感量的原始方法。首先，本文列举了一个正字典和一个负字典。正面情绪词有：好、好、华丽、华丽等；负面情绪词有：坏、坏、坏等。本文设计了一套定义正负关系的方法。如果有阳性词，阳性分数的值加1。如果有否定词，则否定分数的值加1。本文在情感词之前找到度词。然后，本文设定了度词的权重，并将其与情感值相乘。如果句子中有感叹号，则正数加2。本文在情感词语之前发现了反义词。如果反词的数目是奇数，则该词的分数将乘以-1。如果是偶数，单词的分数将保持不变。

本文通过对Snownlp评分与原情绪评价法阳性评分之间的相关系数的检验，对该方法进行了验证。

由于信息技术的数量不多，每个数据量约为10-20条，本文决定手工查找活动的正反两个方面。

“紧张”（neury）是一个经常在一些数据中提到的词，这些数据在体育活动中更可能被视为否定词，包括运动会、足球赛和篮球赛。

如果一段数据中的字太少，则会被视为否定字，因为它包含的信息太少，这意味着数据的主机可能较少参与活动。例如，如果一个人只是参加了一个运动会或只是作为一名工作人员工作，那么数据很可能被视为负面数据。

如果一个人把一次经历描述为“访问”（visit）、“观察”（observe），那么他更可能被视为消极的，而不是被描述为“参与”（participate）、“参与”（involution）。

与创新成就、远足体验、俱乐部活动相关的数据很可能被视为积极的。

一段描述性数据比一段叙述性数据更有可能是正面的。