1尊敬的评委老师，下午好。我是来自网络健康信息推荐模型项目的王润霖。

2我将从研究背景、研究方案、成果应用、预期进度四个方面介绍我们的项目。

3首先是研究背景。

4由于社会公众对健康信息关注度的提高，在面对健康风险时，我们往往会及时浏览和搜寻健康相关问题，关注健康信息。

5但是，社交媒体语境下，人人都是麦克风，以往充当“把关人”角色的审核机制和专业人员作用被不断弱化消减，乃至缺位。

6虚假健康信息泛滥，大量未加证实的、虚假错误的“伪健康”信息得到广泛传播，严重扰乱用户健康认知。

7如何在海量网络信息中为用户推荐高质量的健康信息，满足用户个性化需求，提高健康信息消费和利用效率，解决“信息迷航”和“伪健康”信息问题，成为当前社交媒体信息推荐和网络健康信息传播和利用领域的研究热点。

8我们的研究意义分为两个方面。首先，我们有志于在荣誉课所学知识基础上，深入研究推荐算法、进行算法创新。

9丰富和完善网络信息推荐的理论和方法体系。利用丰富的用户行为数据、社交内容数据、社交关系数据，解决传统推荐方法存在的推荐质量和精准性不高，数据稀疏性和冷启动等问题。

10提高网络健康信息推荐服务质量。建立一个兼顾准确性、实时性与多样性的推荐系统模型。在分析大量用户行为数据的基础上，深入挖掘用户潜在健康信息需求，为构建相应的推荐模型和方法、提升网络健康信息推荐的准确性与有效性提供重要依据。

11下面是我们的研究方案。

12我们的实现方案主要是基于用户行为，确定网络健康信息推荐方法。它分为三个平行的方面。

13首先，针对用户标注行为，研究基于社会化标签挖掘的网络健康信息内容推荐方法，以提高健康信息推荐的精准性和实时性。

14第二，针对用户关系行为，研究基于主题和多重信任关系的网络健康信息协同过滤推荐方法，以提高协同过滤推荐的质量和准确性。

15针对用户发布行为和交互行为，研究基于语义和用户间交互关系的多样性与个性化推荐方法，以提高推荐结果的多样性与个性化。

16最后，以网络信息发布平台为数据采集对象，通过采集健康领域信息构建实验数据集，通过对比实验进行相关方法的有效性验证。

17我们的特色与创新点在于提高科学、准确、符合用户需求的健康信息利用效率，这当然包括防止虚假信息传播，同样重要的是把健康信息科学准确地推荐给需要的人。

18建立面向用户行为分析的网络健康信息个性化推荐方法体系，满足用户的个性化信息需求。

19然后是我们的成果应用。

20我们计划校企联合，开发健康信息推荐原型系统。将用户需求与健康信息进行匹配，实现网络健康信息精准推荐。

21此外，推荐模型有机会应用于辅助诊疗、医学教育与就诊引导等多个方面。

22最后介绍我们的预期进度。

23从去年十二月开始，我们已经进行了推荐系统算法的学习，本着脚踏实地的原则，详细制定了项目研究方案。

24在过去的一个月里，我们已经基于电影信息数据集，初步开展了推荐系统算法的编程实现工作。

25我们计划在未来三个月构建基于网络健康信息的完整推荐模型，在第四个月进行有效性验证工作。然后，努力进行算法创新，尝试应用深度学习与知识图谱算法，改进现有的推荐系统模型，力争提出我们的解决方案。最后，让项目得以实际应用。

26我们团队一直严谨求实、脚踏实地。在过去取得了很多优秀成果和奖项，包括创新创业类竞赛的一二三等奖、软件著作和论文发表。我们很希望获得国创项目的立项机会，借助这个平台做出我们的创新与贡献。

27我的答辩到此结束，恳请评委老师批评指正。

**具体实现方案**

1【用户标注】

**首先要基于信息内容推荐。**

首先，分析用户在某段时间内发布、转发和浏览的所有文本信息内容，通过对这些文本信息内容的特征抽取与统计分析，获取用户的兴趣偏好；

其次，通过相似度计算公式来计算待推荐信息内容与用户兴趣偏好的相似度，判断待推荐信息与用户兴趣偏好的契合度；

最后，设置相似度阀值，将满足预设阀值的待推荐信息推荐给用户。

**网络平台中用户主动自发标注的标签，可以反映出用户的兴趣偏好和主观认知。通过挖掘用户标签，获取深层次的用户兴趣偏好，是提高网络健康信息内容推荐质量和精准性的关键。**

2【用户关系】

**首先要进行基于协同过滤的推荐算法。**

基于协同过滤的推荐方法是依据用户之间或者物品之间的关联性进行的，这种关联性来源于用户对信息内容的兴趣偏好的相似度计算。

**基于文章主题和用户关系。**

进行目标用户及其关注用户的主题挖掘，获取目标用户及其关注用户个体兴趣偏好；

其次，通过计算目标用户及其关注用户的相似度和多重信任关系强度，获取目标用户群体兴趣偏好；

然后，通过将目标用户个体兴趣偏好与群体兴趣偏好进行线性加权，得到目标用户对徼博主题的综合兴趣度；

最后，根据新发布文章的主题概率分布和目标用户对文章主题的综合兴趣度，计算目标用户对新文章的兴趣度并降序排列，得到前若干个推荐结果。

不足：数据稀疏、冷启动等问题较为突出。例如，当用户评分数据或评分信息数量过少时，将导致计算无法进行或是计算结果不准确。

3【用户发布和交互】

**基于多样化的推荐方法。**

一方面利用协同过滤算法进行评分，另一方面是通过对计算预测评分方法的改进，采取一些特殊的手段来提高低流行度项目的优先级。

**基于用户发布的文章和用户的交互关系。**

首先，根据文章之间的语义关系和浏览文章的用户之间的交互关系，构建文章主题-交互关系相关度矩阵，实现文章语义与用户间交互关系的融合；

其次，在文章主题-交互关系相关度矩阵的基础上，利用k-means聚类方法进行文章多样性划分，得到若干具有差异性的博文聚类簇；

最后，利用矩阵分解类算法进行文章选取与个性化排序，从各聚类簇中选出满足用户偏好的博文推荐列表。

**4【基于信息内容推荐】**

它的核心思想是根据推荐物品或内容的元数据，发现物品或者内容的相关性，然后基于用户以往的喜好记录，推荐给用户相似的物品。

**5【协同过滤推荐算法】**

基于用户的协同过滤（User CF）：（物品）

（1）基本原理：通过用户对不同内容（物品）的行为，来评测用户之间的相似性，找到“邻居”基于这种相似性做出推荐。相当于，给相似的用户推荐其他用户喜欢的内容。

（2）基本步骤：找到和目标用户兴趣相似的用户集合找到这个集合中的用户喜欢的，并且目标用户没有听说过的物品推荐给目标用户。

（3）用户相似度的计算指标：皮尔逊相关系数、余弦相关系数等等。

协同过滤细节

1.收集用户信息

收集可以代表用户兴趣的信息。一般的网站系统使用评分的方式或是给予评价，这种方式被称为“主动评分”。另外一种是“被动评分”，是根据用户的行为模式由系统代替用户完成评价，不需要用户直接打分或输入评价数据。电子商务网站在被动评分的数据获取上有其优势，用户购买的商品记录是相当有用的数据。

2.最近邻搜索(Nearest neighbor search, NNS)

以用户为基础（User-based）的协同过滤的出发点是与用户兴趣爱好相同的另一组用户，就是计算两个用户的相似度。例如：查找n个和A有相似兴趣用户，把他们对M的评分作为A对M的评分预测。一般会根据数据的不同选择不同的算法，较多使用的相似度算法有Pearson Correlation Coefficient、Cosine-based Similarity、Adjusted Cosine Similarity。

3.产生推荐结果

有了最近邻集合，就可以对目标用户的兴趣进行预测，产生推荐结果。依据推荐目的的不同进行不同形式的推荐，较常见的推荐结果有Top-N 推荐和关系推荐。Top-N 推荐是针对个体用户产生，对每个人产生不一样的结果，例如：通过对A用户的最近邻用户进行统计，选择出现频率高且在A用户的评分项目中不存在的，作为推荐结果。关系推荐是对最近邻用户的记录进行关系规则(association rules)挖掘。