**2018年广东省大学生计算机设计大赛**

作品类别：人工智能

类别细目：人工智能应用方案设计

作品题目：基于深度学习的智能推荐系统方案设计

参赛队伍：东莞理工学院40957号卢健豪 罗敏 杨景安

**目录**

[第一章 作品背景 3](#_Toc511489369)

[第二章 设计理念 4](#_Toc511489370)

[第三章 方案设计 5](#_Toc511489371)

[3.1 用户需求 5](#_Toc511489372)

[3.2 可行性分析 6](#_Toc511489373)

[3.3 技术路线 7](#_Toc511489374)

[3.3.1 新型商品配置器技术框图 7](#_Toc511489375)

[3.3.2 技术路线实现过程 8](#_Toc511489376)

[第四章 作品优势 12](#_Toc511489377)

[4.1 技术上的优势 12](#_Toc511489378)

[4.2 用户交互上的优势 13](#_Toc511489379)

[第五章 作品外观设计 15](#_Toc511489380)

# 第一章 作品背景

近年来，众多电子商务平台伴随着互联网技术的日新月异，从而得到了前所未有的迅猛发展。在淘宝 、天猫、京东、苏宁等电子商务龙头企业的带领下，无论是PC端还是移动端的购物体验都趋于便捷和完善，这使得越来越多的消费者热衷在网上进行购物，并在平台发表自己对选购商品的评价意见或是对服务的反馈，逐渐积累了丰富的商品评论数据。根据第41次中国互联网发展状况统计报告显示，截至2017年12月，我国网络购物用户规模达到 5.33 亿，占网民总体的 69.1%，而手机网络购物用户规模达到 5.06 亿，使用比例提升至 67.2%。面对庞大的网络购物用户群体和海量的商品信息，无论对于电子商务平台还是消费者而言，都亟需一种可以提升购物满意度和购物效率的方法，商品推荐系统应运而生。

传统的商品推荐系统的主要方法包括协同过滤、基于内容的推荐、混合推荐和基于流行度的推荐等，并且广泛应用于现有的电子商务平台当中。然而这些传统的推荐系统大多是以商品为中心，利用用户与商品之间的二元关系，基于用户历史购物行为记录或商品相似度关系，从而发现用户可能感兴趣的商品。可是这必须是用户在完成至少一次的购物搜索或购买之后，才能形成的推荐。这种形式的推荐违背购物的客观性，也产生了如下地弊端：首先，用户无法在第一时间得到商品的推荐，使得购物的时效性降低；其次，依赖于商品卖家描述的推荐，往往无法验证商品在实际体验上的优劣，使得平台的可信度降低；再者，在用户完成购物后不断推荐同类的商品，不但使推荐显得多余，还会让用户产生厌烦的购物心理，极大地损坏了购物体验。

# 第二章 设计理念

为了帮助电商平台实现精细化的产品定制，使用户能享受到更高效准确的个性化商品推荐与定制服务，本作品的设计理念主要从三个方面展开：商品评论文本分类、商品情感标签挖掘以及用户搜索的需求识别。

（1）商品评论文本分类：用户评论能反映出用户对商品、服务、物流等方面的客观事实。从中进行粗粒度的情感分析，则能够得到用户的正面评价和负面评价。而在细粒度上，可以将正面评价按照情感属性进行分类，便于供其它用户参考；也可以将负面评论按照业务环节进行分类，便于定位哪个环节需要不断优化。

（2）商品情感标签挖掘：通常电商平台需要对商品的功能或风格加上直观的标签，便于用户查找。那么对于众多的普通用户来说，将海量的商品描述和具体的参数转化为描述商品体验的情感标签，是帮助他们消除选购不同种类商品的专业壁垒的重要一环。在结合人工的标签主题挖掘的基础上，利用机器学习分类模型去预测商品所属的标签的解决方法，不但减轻了人工投入的时间和成本，还保证了分类模型具有极高的准确度。

（3）用户搜索的需求识别：用户搜索行为是电商平台上用户购物的常用入口，是用户需求的强体现。将用户搜索分别归一到具体的品类需求， 并映射到商品库中计算相似度，在第一时间得到基于评论大数据的商品推荐列表，实现个性化的情感推荐。

# 第三章 方案设计

## 3.1 用户需求

 实际上，现有的电子商务平台都为购买并使用了商品的用户提供发表在线评论的功能。根据中国互联网络信息中心在2009年和2015年发表的两份中国网络购物市场研究报告表明，用户评论是影响消费者进行购买决策最关键的因素。而在网络购物用户购买商品时主要考虑的众多因素中，用户评论的占比也从2009年的43.3%急剧提升到2015年的77.5%。由次可知，用户评论积累形成的网络口碑相较于卖家对商品的描述，更具真实性、可靠性和参考性，是消费者进行网购行为中不可获取的一部分。

然而现有的商品评论还是不足以为消费者带来精准快速的购物体验。首先，同类型的商品数量繁多，商品的评论更是成千上万条，这使得大多数的用户智能选择性地浏览前几条最多几十条，不但耗时长而且获取地信息十分有限，无法对购物决策做到完整地判断，加大了购买到真正心仪商品的难度；其次，消费者从整体的评论中难以判断该商品在自己所关注的属性中的口碑好坏，从而造成失败的购物经历；再者，消费者对商品属性的认知度不同，就手机领域而言，普通消费者往往只了解外观、性价比、电池大小等属性，而对专业名词诸如CPU型号之间的强弱认知有所欠缺，所以需要将专业的硬件参数映射到评论反馈的情感属性之上。

因此，用户亟需一套高效可行的智能商品推荐系统，帮助他们理解商品的海量评论信息，并识别他们搜索文本的需求，即时形成准确的个性化推荐。

## 3.2 可行性分析

本作品主要处理的实验数据是电商平台中的在线商品评论大数据，它是由一条条来自不同用户的评论文本构成。经过对评论大数据的调研后发现，对于绝大多数的用户而言，他们在商品过程中的理解程度存在知识上的欠缺，因此不会将商品的规格属性作为评价的对象。而实际出现在商品评论中的描述情感体验的属性词，我们称为商品的相对属性，如智能手机评论中经常出现的属性词：性能、速度、手感等词。



图3-1 京东商城中某款智能手机的评论

就以智能手机为例，用户在线下实体店购物时，常常会对导购员说：“我想要一款运行速度快，手感好而且续航给力的手机”；而在线上电商平台中，常常会有用户在购买智能手机并体验后写下“这款手机性能强，运行速度很快，而且续航杠杠的！”的评论。通过分析和计算后不难发现，用户的需求文本和评论文本在文本相似度上达到较高的程度。所以基于商品评论的新型推荐系统不仅可行，而且将会更贴近用户的需求，相信能达到非常优秀的个性化推荐效果。

## 3.3 技术路线

### 3.3.1 新型商品配置器技术框图

|  |
| --- |
| **训练层** **应用层**  **商品评论数据库**  **用户**  **需求文本**  人工 标注  **商品**  **无关文本**  **商品**  **相关文本**  **分类器M1**  **(Bi-LSTM**  **&**  **CRF)**  **商品需求**  **相关文本**  **商品评论**  **情感词典**  **(n-gram)**  **分类器M2**  **(Bi-LSTM**  **&**  **CRF)**  **商品需求**  **文本向量**  **商品相对属性向量库**  **集成推荐系统**  **商品推荐列表** |

### 3.3.2 技术路线实现过程

1.商品文本分类

（1）商品信息的抓取：经过对京东商城中智能手机的前期调研，为了让研究更具有客观性和代表性，决定选取用户评论超过1000条的热门智能手机作为研究的对象，按型号分类，抓取A款智能手机的评论文本总计B条。由于本作品采用的是基于词向量的“Bi-LSTM+CRF”的三层结构进行模型训练，属于句子级别的标签分类预测技术。所以要将每一条评论文本按标点符号分割为句子，并统计评论句子数量，设为C段评论文本（这里的A值、B值和C值由实验数据统计信息来决定）。

（2）商品评论类别研究：根据评论文本中的相对属性依附的对象不同，可以将商品评论文本归属为以下两类：一是与商品直接相关的评论文本；二是与服务、物流等直接相关的评论文本（可视为与商品本身属性无关的描述）。基于这两个类别标签，就可以在所有商品大数据中选取部分评论文本进行序列标注。

（3）商品评论的预处理：经实验验证，只需要从评论文本中随机选取10%的数量级作为文本分类的样本集，通过人工标注为样本集分类，最终得到与商品相关的评论文本共计E段，与商品无关的评论文本F段（这里的E值和F值由实验数据统计信息来决定）。然后分两方面处理数据：训练数据的处理方面，利用结巴分词工具对样本集的文本进行分词处理，同时保持句子的序列顺序，标注出对应的词性和类别标签后，编写脚本处理样本集为可用于命名实体识别模型分类的格式。为了能在每次迭代训练后检验模型的效能，将样本集按7：1：2的比例随机分割为训练集、验证集、测试集；而在文本分布式表示方面，使用所以的样本集数据，结合GloVe模型生成100维的词向量，用于分类模型的文本表示。

（4）分类模型训练：将经过预处理后的词向量和样本集输入到基于词的Bi-LSTM+CRF三层结构中进行文本分类器的训练。分类模型经过G次的迭代训练后，记录每次迭代训练过程中对训练集、验证集、测试集的评估结果，分析选取综合分类准确效果最优的迭代次数，保存实验模型，得到用于识别商品评论文本中与商品直接相关的文本分类器M1。

2. 商品评论的相对属性（情感）标签挖掘

本作品提出使用基于N元字组（N-gram）的统计分析方法挖掘评论文本中的相对属性标签（即表达情感的商品属性标签）。具体的流程如图3-1所示：



图3-1 商品评论相对属性标签挖掘流程

通过N元组（N-gram）等方法来统计分析样本数据集中与商品直接相关的评论文本段，从中筛选到出现频次在M以上的N元组作为候选商品评论相对属性标签的描述词（这里的M值和N值由实验数据统计信息来决定）；然后，通过语义词典对这些N元组描述词进行过滤，并加入人工审查，保留具有现实情感意义的描述词；最后，通过人工聚类的方法，将若干个含义相近或相同的N元组数据合并到一个商品评论相对属性标签上，构成商品评论相对属性标签词典D1。

3. 用户搜索的需求识别

首先，通过读取D1，假设词典中有H个相对属性标签，那么就对与商品直接相关的E段评论文本标注上对应的相对属性标签（一段评论文本可对应多个相对属性标签），用于训练分类模型（这里的H值由实验数据统计信息来决定）。比如：“手机用起来很流畅”、“这款手机用起来速度不错”这两个文本段，都将被标为“手机流畅”这个商品评论相对属性标签；

然后，同样采用“词向量+Bi-LSTM+CRF”的三层结构来训练分类模型。其中，模型训练的策略是采用构造每个相对属性标签的二元分类模型，运用这H个模型依次预测评论文本段是否含有该相对属性标签的方法，让每个模型进行J次迭代训练（这里的J值由实验数据统计信息来决定）。记录每次迭代训练过程中对训练集、验证集、测试集的评估结果，分析选取综合分类准确效果最优的迭代次数，保存分类效果最优迭代次数的模型作为各自相对属性的分类模型，并编写脚本集成这H个分类模型，构建用于实现用户搜索词需求识别的文本分类器M2。

4.基于商品评论的推荐系统构建

首先，本作品提出了将由文字构成的相对属性，运用向量的形式来进行归一化处理。对于相对属性来说，构成商品相对属性词典的H个相对属性标签可以被用于描述商品之间差异的H个维度，接着通过统计每款型号的智能手机在各个维度下，含有该相对属性标签的文本在该款手机与商品直接相关评论文本中的占比L作为权值，，可作为衡量该相对属性在这款手机中表现的优秀程度，得到由组成的H维相对属性向量，经过对每款智能手机的依次统计和计算后，得到A款智能手机的相对属性向量库。

然后，系统会收集到用户的需求文本，通过第三步构建的集成文本分类器M2的预测，得到用户的H维的需求向量。

最后，将需求向量逐一与A款智能手机的商品相对属性向量库中的向量计算余弦相似度，按照相似度从高到低，输出商品推荐列表，实现个性化的商品推荐。实现流程如图3-2所示：



图3-2 基于商品评论的推荐系统流程图

# 第四章 作品优势

## 4.1 技术上的优势

本文主要采用基于深度学习的文本分类方法，其中的基于词向量的“Bi-LSTM+CRF”的三层结构模型对于文本大数据具有优秀的分类效果。模型的第一层是look-up层，利用利用预训练或随机初始化的embedding矩阵将句子中的每个词由one-hot向量映射为低维稠密的词向量。在输入下一层之前，设置dropout以缓解过拟合；模型的第二层是双向LSTM层，自动提取句子特征。其实现过程是将一个句子的各个词的embedding序列作为双向LSTM各个时间步的输入，再将正向LSTM输出的隐状态序列与反向LSTM在各个位置输出的隐状态进行按位置拼接，得到完整的隐状态序列。在设置dropout后，接入一个线性层，将隐状态向量从 m 维映射到 k 维（k 是标注集的标签数）从而得到自动提取的句子特征；模型的第三层是CRF层，进行句子级的序列标注。模型构造如图4-1所示：

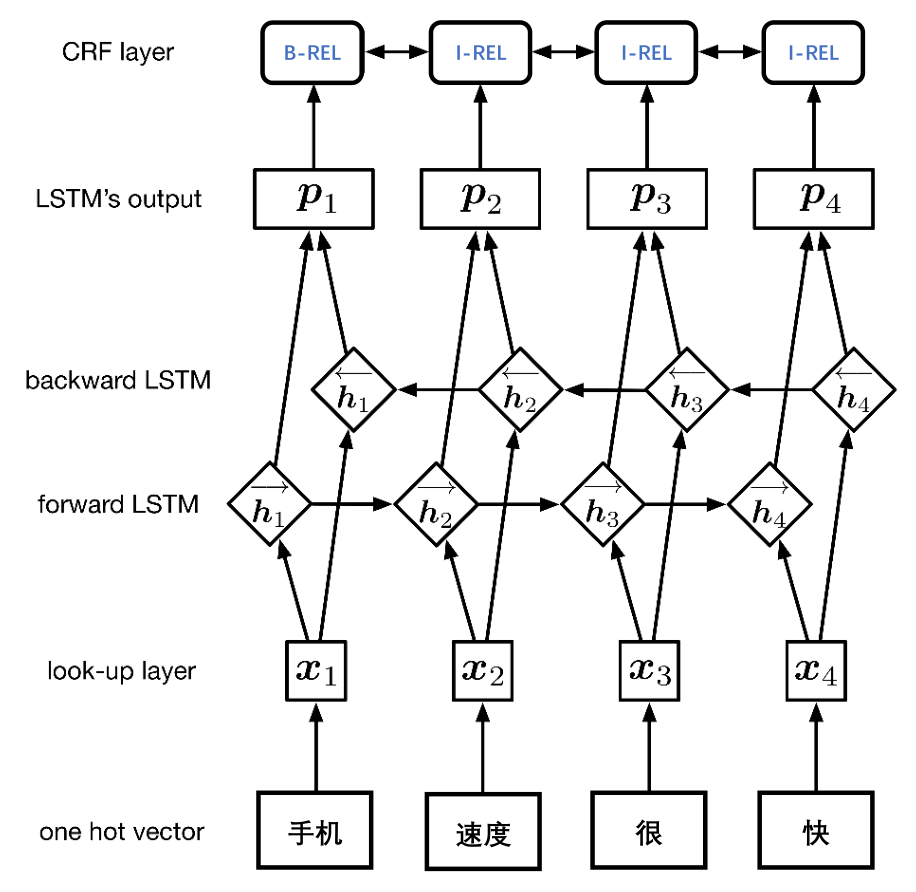


图4-1 “词向量+Bi-LSTM+CRF”三层模型构造图

## 4.2 用户交互上的优势

为了实现用户在购物时的实时推荐，本作品提出了对用户搜索词的需求识别。在日常的线下实体店购物时，消费者通常会向导购员说出自己要购买的商品类别和期望商品能实现的功能，这种描述功能的文本就是消费者的需求文本。接着，导购员会去理解消费者的需求文本，并会以现有商品的价格，配置，外观等方面的因素，给消费者推荐相应的商品。

而基于商品评论的推荐系统就是通过模仿导购员的推荐流程，构建个性化的商品推荐系统。首先，与常用的电子商务平台不同，用户需要将自己的需求用一段文字表达出来，其中必须包括一种商品品类和至少一种需求，比如：“我想要一款运行速度流畅，电池用量大的智能手机”；然后，系统会收集到用户的需求文本，通过集成的推荐系统分析计算之后，输出商品推荐列表，实现个性化的商品推荐。

# 第五章 作品外观设计

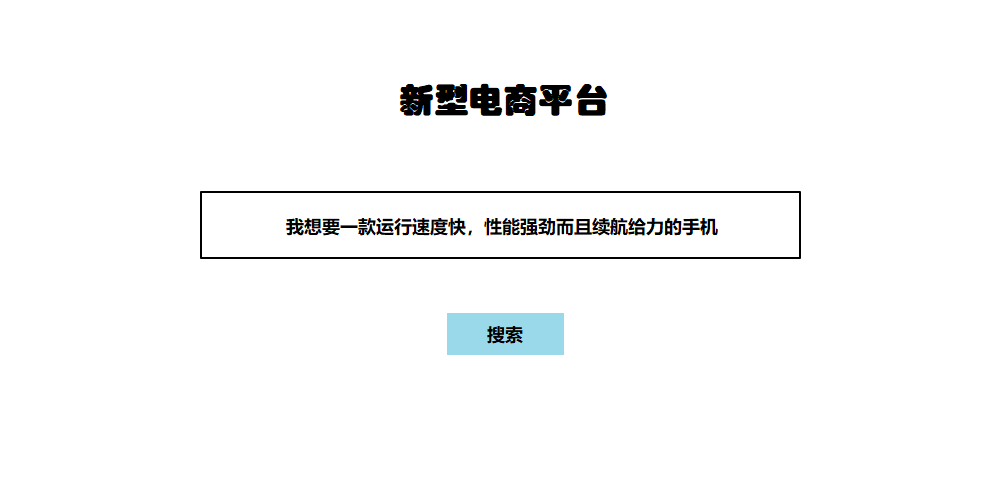


图5-1 用户搜索入口



图5-2 商品推荐展示