



(21) 申请号 202410085455.1

(22) 申请日 2024.01.22

(71) 申请人 北京嘉华铭品牌策划有限公司

地址 102488 北京市房山区拱辰街道月华
大街1号八层815

(72) 发明人 张渴欣

(74) 专利代理机构 北京天奇智新知识产权代理
有限公司 11340

专利代理师 龙涛

(51) Int. Cl.

G06Q 30/0203 (2023.01)

G06Q 30/0202 (2023.01)

G06Q 30/0207 (2023.01)

G06Q 30/0251 (2023.01)

G06Q 10/0631 (2023.01)

G06F 16/9535 (2019.01)

G06F 16/332 (2019.01)

G06F 16/33 (2019.01)

G06F 40/289 (2020.01)

G06F 40/30 (2020.01)

G06F 16/35 (2019.01)

G06F 18/2415 (2023.01)

G06F 18/2433 (2023.01)

G06N 3/0464 (2023.01)

G06N 3/047 (2023.01)

G06N 3/09 (2023.01)

权利要求书2页 说明书12页 附图1页

(54) 发明名称

一种个性化商户营销智能推荐方法和系统

(57) 摘要

本发明公开了一种个性化商户营销智能推荐方法和系统,所述方法包括:步骤S1,数据收集与集成。步骤S2,无效数据筛选。步骤S3,对有效数据进行情感分析。步骤S4,用户满意度分析。步骤S5,智能确定营销策略。步骤S6,智能进行个性化广告的推送。本发明通过更准确的满意度指标,可以更好地理解顾客需求,通过预测分析和优惠策略的智能制定,可以更有效地分配营销资源。



1. 一种个性化商户营销智能推荐方法,包括以下步骤:

步骤S1,建立数据管道,集成来自多个数据源的顾客反馈和互动数据;

步骤S2,无效数据筛选,包括使用数据清洗工具识别和删除无效数据;

步骤S3,对有效数据进行情感分析,包括基于自然语言处理NLP技术来处理文本数据,应用情感分析算法来确定用户评论的正面和负面倾向的得分;

步骤S4,结合用户评分分值和该用户评分对应的情感分析结果,确定一个顾客满意度指标;

以及对预设时间段内所有用户满意度进行细粒度分析,包括分析不同维度的满意度数据,所述不同维度包括产品类别和顾客群体;

其中,将用户评分作为满意度指标的基础,将基于NLP技术进行情感分析得到的情感得分与用户评分结合;

其中,分析不同维度的满意度数据,包括将顾客根据年龄、性别分成不同的群体,并分析每个群体的满意度;

步骤S5,基于POS系统当前销售数据和历史销售数据,预测下一预设时间段内不同产品的购买率;

根据购买率预测结果生成优惠策略,所述优惠策略包括在预测到购买率会增加的时间段内提供折扣促销策略策略;

步骤S6,基于细粒度的用户满意度分析结果、下一预设时间段内不同产品的购买率以及折扣促销策略策略,对不同顾客群体定制个性化广告推送内容。

2. 如权利要求1所述的一种个性化商户营销智能推荐方法,其特征在于,所述步骤S2中进行无效数据筛选,包括:

对于每一条评论,确定评论内容长度L;

当评论内容长度大于阈值L0时,进行评论文本分析;

使用文本分析算法中的使用相似度检测算法Jaccard相似度,当两条评论的相似度大于阈值时,确定评论为内容高度类似的评论。

3. 如权利要求2所述的一种个性化商户营销智能推荐方法,其特征在于,使用数据比较队列的方式存储有效评论,所述队列将维护待比较的评论对,包括对队列中的评论进行一次筛查,包括:

将每一个第一评论与队列中现有的每条评论配对并计算相似度;

当第一评论已与比较队列中的所有其他评论比较过后,如果评论与比较队列中的所有其他评论的Jaccard相似度都不超过阈值S1,则将第一评论加入比较队列;

如果第一评论与比较队列中的至少一个其他评论的Jaccard相似度都等于或大于阈值S1,则将第一评论删除。

4. 如权利要求3所述的一种个性化商户营销智能推荐方法,其特征在于,对于队列中剩余的所有评论,再根据文字与图片的匹配程度来进行二次筛查,包括:

识别附图评论中图片内的所有对象和所述对象的类型;

将文本输入进行标准化后进行分词,将句子分解为单独的词;

对图像和文本分别使用嵌入向量表示;

通过Jaccard相似度来计算图像向量和文本向量之间的匹配度;

当通过Jaccard相似度来计算文字向量和图片向量之间的匹配度低于匹配阈值时,判断该评论符合刷单评论通常存在长评大量内容重复和带图评论和文字内容无关的情形,在所述队列中将该条评论进行删除。

5.如权利要求4所述的一种个性化商户营销智能推荐方法,其特征在于,经过一次筛查和二次筛查后,队列中剩余的评论为有效评论。

6.如权利要求1所述的一种个性化商户营销智能推荐方法,其特征在于,所述步骤S3中应用情感分析算法来确定用户评论的正面和负面倾向的得分,包括在情感分析算法中的损失函数中为正面情绪和负面情绪分配不同的权重。

7.如权利要求1所述的一种个性化商户营销智能推荐方法,其特征在于,在使用深度学习进行情感分析时,模型的输出每个类别的概率,包括正面情感的概率 P_{positive} 和负面情感的概率 P_{negative} ;

基于正面情感的概率 P_{positive} 和负面情感的概率 P_{negative} 转换为一个情感得分S:

$$S = P_{\text{positive}} - P_{\text{negative}},$$

情感得分S取值在-1到1之间,其中0代表中性情感。

8.如权利要求1所述的一种个性化商户营销智能推荐方法,其特征在于,所述步骤S4中,结合用户评分分值和该用户评分对应的情感分析结果,确定一个顾客满意度指标,其中,将用户评分作为满意度指标的基础,将基于NLP技术进行情感分析得到的情感得分与用户评分结合,包括:

设R为用户主动给出的评;

设S为通过情感分析得到的情感得分;

顾客满意度指标C的计算公式如下:

$$C = \alpha \cdot f(R) + \theta \cdot g(S) + \gamma \cdot h(R, S),$$

其中, $f(R)$ 是关于用户评分R的函数; $g(S)$ 是关于情感得分S的函数, $h(R, S)$ 是一个混合函数,体现评分和情感得分的相互作用;

α, θ, γ 是权重参数,用于控制 $f(R), g(S), h(R, S)$ 在最终满意度指标中的相对重要性。

9.如权利要求1所述的一种个性化商户营销智能推荐方法,其特征在于,步骤S5,智能确定营销策略,包括:

基于POS系统当前销售数据和历史销售数据,预测下一预设时间段内不同产品的购买率,包括:使用简单移动平均SMA模型来预测下一时间段的购买率;

对于每个产品,其在时间t的预测购买率 P_t 可以根据过去N个时间点的实际购买率 $P_{t-N+1}, P_{t-N+2}, \dots, P_t$ 来计算:

$$P_t = \frac{1}{N} \sum_{i=0}^{N-1} P_{t-i},$$

其中, P_t 是时间t的预测购买率, P_{t-i} 是时间t-i的实际购买率,N是移动平均的窗口大小,即考虑最近N个时间点的数据。

10.一种个性化商户营销智能推荐系统,所述系统包括存储器和处理器,所述存储器中存储有计算机程序代码,所述处理器用于执行所述计算机程序代码以实现权利要求1-9的个性化商户营销智能推荐方法。

一种个性化商户营销智能推荐方法和系统

技术领域

[0001] 本发明属于计算机技术领域,尤其涉及一种个性化商户营销智能推荐方法和系统。

背景技术

[0002] 随着信息技术的飞速发展,尤其是大数据、云计算、人工智能等现代信息技术的蓬勃兴起,市场营销模式正经历着翻天覆地的变化。传统的营销策略由于缺乏对消费者个性化需求的理解和精准的数据支撑,正逐渐显现出其局限性。消费者行为数据的广泛可用和计算技术的进步为商户提供了前所未有的个性化营销机会。

[0003] 当前,尽管商户积累了大量的消费者交易数据,但在如何有效地利用这些数据以提供个性化的服务和产品方面,仍存在一定的挑战。许多商户缺乏有效的工具来分析和理解这些数据,以至于无法从中提取出有价值的业务洞见。此外,随着消费者偏好的不断变化和市场竞争的加剧,商户亟需一种能够快速响应市场变化并提供个性化推荐的智能营销方法,能够利用跨域数据为商户制定更加精准和个性化的营销策略。通过综合利用消费者的交易数据、行为数据、社交网络数据等多源信息。本发明旨在实现对消费者需求的深入理解和对市场变化的快速响应,从而帮助商户实现销售增长和客户满意度提升。

发明内容

[0004] 针对上述现有技术中存在的缺陷,本发明提供一种个性化商户营销智能推荐方法,包括以下步骤:

[0005] 步骤S1,建立数据管道,集成来自多个数据源的顾客反馈和互动数据;步骤S2,无效数据筛除,包括使用数据清洗工具识别和删除无效数据;步骤S3,对有效数据进行情感分析,包括基于自然语言处理NLP技术来处理文本数据,应用情感分析算法来确定用户评论的正面和负面倾向的得分;

[0006] 步骤S4,结合用户评分分值和该用户评分对应的情感分析结果,确定一个顾客满意度指标;

[0007] 以及对预设时间段内所有用户满意度进行细粒度分析,包括分析不同维度的满意度数据,所述不同维度包括产品类别和顾客群体;

[0008] 其中,将用户评分作为满意度指标的基础,将基于NLP技术进行情感分析得到的情感得分与用户评分结合;

[0009] 其中,分析不同维度的满意度数据,包括将顾客根据年龄、性别分成不同的群体,并分析每个群体的满意度;

[0010] 步骤S5,基于POS系统当前销售数据和历史销售数据,预测下一预设时间段内不同产品的购买率;

[0011] 根据购买率预测结果生成优惠策略,所述优惠策略包括在预测到购买率会增加的时间段内提供折扣促销策略策略;

[0012] 步骤S6,基于细粒度的用户满意度分析结果、下一预设时间段内不同产品的购买率以及折扣促销策略策略,对不同顾客群体定制个性化广告推送内容。

[0013] 其中,所述步骤S2中进行无效数据筛除,包括:

[0014] 对于每一条评论,确定评论内容长度L;

[0015] 当评论内容长度大于阈值L0时,进行评论文本分析;

[0016] 使用文本分析算法中的使用相似度检测算法Jaccard相似度,当两条评论的相似度大于阈值时,确定评论为内容高度类似的评论。

[0017] 其中,使用数据比较队列的方式存储有效评论,所述队列将维护待比较的评论对,包括对队列中的评论进行一次筛查,包括:

[0018] 将每一个第一评论与队列中现有的每条评论配对并计算相似度;

[0019] 当第一评论已与比较队列中的所有其他评论比较过后,如果评论与比较队列中的所有其他评论的Jaccard相似度都不超过阈值S1,则将第一评论加入比较队列;

[0020] 如果第一评论与比较队列中的至少一个其他评论的Jaccard相似度都等于或大于阈值S1,则将第一评论删除。

[0021] 其中,对于队列中剩余的所有评论,再根据文字与图片的匹配程度来进行二次筛查,包括:

[0022] 识别附图评论中图片内的所有对象和所述对象的类型;

[0023] 将文本输入进行标准化后进行分词,将句子分解为单独的词;

[0024] 对图像和文本分别使用嵌入向量表示;

[0025] 通过Jaccard相似度来计算图像向量和文本向量之间的匹配度;

[0026] 当通过Jaccard相似度来计算文字向量和图片向量之间的匹配度低于匹配阈值时,判断该评论符合刷单评论通常存在长评大量内容重复和带图评论和文字内容无关的情形,在所述队列中将该条评论进行删除。其中,经过一次筛查和二次筛查后,队列中剩余的评论为有效评论。其中,所述步骤S3中应用情感分析算法来确定用户评论的正面和负面倾向的得分,包括在情感分析算法中的损失函数中为正面情绪和负面情绪分配不同的权重。

[0027] 其中,在使用深度学习进行情感分析时,模型的输出每个类别的概率,包括正面情感的概率 $P_{positive}$ 和负面情感的概率 $P_{negative}$;

[0028] 基于正面情感的概率 $P_{positive}$ 和负面情感的概率 $P_{negative}$ 转换为一个情感得分S:

[0029] $S = P_{positive} - P_{negative}$,

[0030] 情感得分S取值在-1到1之间,其中0代表中性情感。

[0031] 其中,所述步骤S4中,结合用户评分分值和该用户评分对应的情感分析结果,确定一个顾客满意度指标,其中,将用户评分作为满意度指标的基础,将基于NLP技术进行情感分析得到的情感得分与用户评分结合,包括:

[0032] 设R为用户主动给出的评;

[0033] 设S为通过情感分析得到的情感得分;

[0034] 顾客满意度指标C的计算公式如下:

[0035] $C = \alpha \cdot f(R) + \theta \cdot g(S) + \gamma \cdot h(R, S)$,其中, $f(R)$ 是关于用户评分R的函数; $g(S)$ 是关于情感得分S的函数, $h(R, S)$ 是一个混合函数,体现评分和情感得分的相互作用;

[0036] α, θ, γ 是权重参数,用于控制 $f(R), g(S), h(R, S)$ 在最终满意度指标中的相对重要性。

[0037] 其中,步骤S5,智能确定营销策略,包括:

[0038] 基于POS系统当前销售数据和历史销售数据,预测下一预设时间段内不同产品的购买率,包括:使用简单移动平均SMA模型来预测下一时间段的购买率;

[0039] 对于每个产品,其在时间 t 的预测购买率 P_t 可以根据过去 N 个时间点的实际购买率 $P_{t-N+1}, P_{t-N+2}, \dots, P_t$ 来计算:

$$[0040] \quad P_t = \frac{1}{N} \sum_{i=0}^{N-1} P_{t-i},$$

[0041] 其中, P_t 是时间 t 的预测购买率, P_{t-i} 是时间 $t-i$ 的实际购买率, N 是移动平均的窗口大小,即考虑最近 N 个时间点的数据。

[0042] 本发明还公开了一种个性化商户营销智能推荐系统,所述系统包括存储器和处理器,所述存储器中存储有计算机程序代码,所述处理器用于执行所述计算机程序代码以实现前述的个性化商户营销智能推荐方法。

[0043] 本发明通过API自动化数据收集过程,提高了数据收集的效率和准确性。本发明使用数据清洗工具确保数据质量,为后续分析提供可靠的数据基础。本发明通过NLP技术处理复杂的文本数据,基于情感分析模型,更准确地评估顾客的情绪倾向。本发明联合用户评分与情感得分,形成更全面的满意度指标。本发明分析不同维度的满意度数据,结合POS系统数据预测不同产品的购买率,为制定营销策略提供科学依据。根据细粒度分析结果,定制个性化的广告推送内容。

[0044] 本发明通过更准确的满意度指标,可以更好地理解顾客需求,通过预测分析和优惠策略的智能制定,可以更有效地分配营销资源。

附图说明

[0045] 通过参考附图阅读下文的详细描述,本公开示例性实施方式的上述以及其他目的、特征和优点将变得易于理解。在附图中,以示例性而非限制性的方式示出了本公开的若干实施方式,并且相同或对应的标号表示相同或对应的部分,其中:

[0046] 图1是示出根据本发明实施例的一种个性化商户营销智能推荐方法的流程图。

具体实施方式

[0047] 为了使本发明的目的、技术方案和优点更加清楚,下面将结合附图对本发明作进一步地详细描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其它实施例,都属于本发明保护的范围。

[0048] 在本发明实施例中使用的术语是仅仅出于描述特定实施例的目的,而非旨在限制本发明。在本发明实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式,除非上下文清楚地表示其他含义,“多种”一般包含至少两种。

[0049] 应当理解,尽管在本发明实施例中可能采用术语第一、第二、第三等来描述……,但这些……不应限于这些术语。这些术语仅用来将……区分开。例如,在不脱离本发明实施

例范围的情况下,第一……也可以被称为第二……,类似地,第二……也可以被称为第一……。

[0050] 应当理解,本文中使用的术语“和/或”仅仅是一种描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。另外,本文中字符“/”,一般表示前后关联对象是一种“或”的关系。

[0051] 取决于语境,如在此所使用的词语“如果”、“若”可以被解释成为“在……时”或“当……时”或“响应于确定”或“响应于检测”。类似地,取决于语境,短语“如果确定”或“如果检测(陈述的条件或事件)”可以被解释成为“当确定时”或“响应于确定”或“当检测(陈述的条件或事件)时”或“响应于检测(陈述的条件或事件)”。

[0052] 还需要说明的是,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的商品或者装置不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种商品或者装置所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的商品或者装置中还存在另外的相同要素。

[0053] 本发明提出了一种系统化的方法来提升顾客满意度和优化营销策略。通过集成和分析顾客反馈数据,本发明旨在提供更精准的顾客满意度指标,并使用这些指标来指导营销决策。

[0054] 如图1所示,本发明公开了一种个性化商户营销智能推荐方法,包括:步骤S1,数据收集与集成。

[0055] 建立数据管道,集成来自多个数据源的顾客反馈和互动数据。使用这些平台的API来自动化数据收集过程,并基于统一的数据库系统来存储和管理收集的数据。

[0056] 步骤S2,无效数据筛除,包括使用数据清洗工具识别和删除无效数据。步骤S3,对有效数据进行情感分析。包括基于自然语言处理NLP技术来处理文本数据,包括分词、词性标注等。应用情感分析算法来确定用户评论的正面和负面倾向的得分,所述情感分析算法可以为情感得分模型,其中包括在情感分析算法中的损失函数中为正面情绪和负面情绪分配不同的权重。

[0057] 步骤S4,用户满意度分析,包括:

[0058] 结合用户评分分值和该用户评分对应的情感分析结果,确定一个顾客满意度指标。

[0059] 以及对预设时间段内所有用户满意度进行细粒度分析,包括分析不同维度的满意度数据,所述不同维度包括产品类别和顾客群体。

[0060] 其中,将用户评分作为满意度指标的基础,将基于NLP技术进行情感分析得到的情感得分与用户评分结合。

[0061] 其中,分析不同维度的满意度数据,包括将顾客根据年龄、性别分成不同的群体,并分析每个群体的满意度。

[0062] 步骤S5,智能确定营销策略,包括:

[0063] 基于POS系统当前销售数据和历史销售数据,预测下一预设时间段内不同产品的购买率。

[0064] 根据购买率预测结果生成优惠策略,所述优惠策略包括在预测到购买率会增加的

时间段内提供折扣促销策略策略。

[0065] 步骤S6,智能进行个性化广告的推送,包括:

[0066] 基于细粒度的用户满意度分析结果、下一预设时间段内不同产品的购买率以及折扣促销策略策略,对不同顾客群体定制个性化广告推送内容。

[0067] 本发明通过API自动化数据收集过程,提高了数据收集的效率和准确性。本发明使用数据清洗工具确保数据质量,为后续分析提供可靠的数据基础。本发明通过NLP技术处理复杂的文本数据,基于情感分析模型,更准确地评估顾客的情绪倾向。本发明联合用户评分与情感得分,形成更全面的满意度指标。本发明分析不同维度的满意度数据,结合POS系统数据预测不同产品的购买率,为制定营销策略提供科学依据。根据细粒度分析结果,定制个性化的广告推送内容。

[0068] 可选地,在步骤S1中,需要使用ETL工具(如Apache NiFi,Talend)构建自动化的数据管道来抽取、转换和加载数据。之后与多个数据源平台的API接口对接,自动化数据抓取过程。最后使用MongoDB数据库系统存储数据。其中,多个数据源可以为美团、大众点评、口碑、自有点单平台等。

[0069] 其中,数据管道是数据从源头到目的地(如分析数据库、数据湖或业务应用)的全过程。建立数据管道通常包括选择适合需求的ETL工具如Apache NiFi用于流数据的实时处理,Talend做复杂的数据集成任务。根据业务需求设计管道流程,明确数据源、中间处理步骤和最终输出。配置ETL工具从数据源(如APIs、数据库、文件系统等)抽取数据。之后,对数据转换逻辑,如数据清洗、格式转换、数据合并等。

[0070] 将转换后的数据加载到目标储存系统中,如数据库或数据湖。

[0071] 其中,API集成是指将外部服务的数据接口(APIs)与自己的系统相连通,实现数据的自动流动。通过数据源平台提供的API文档,理解数据结构、认证方式及调用限制。编写脚本或使用ETL工具中的API连接器,定期调用API接口抓取数据。

[0072] 选择MongoDB等NoSQL数据库,适合存储半结构化和非结构化数据,提供高性能和易扩展性。

[0073] 可选地,在步骤S2中,使用数据清洗工具识别和删除无效数据。无效数据包括重复数据和错误数据。

[0074] 可选地,使用SQL查询或数据处理库(如Pandas)检查和删除完全重复的记录,包括用户ID和评论内容同时完全相同的情况,避免数据库处理错误的情形。

[0075] 所述错误数据包括虚假数据,如刷单评论、假评论。假评论(刷单评论)是指那些不真实反映用户意见的评论,这些评论可能会扭曲产品的真实反馈和评分。刷单评论通常存在长评大量内容重复和带图评论和文字内容无关的情形。

[0076] 可选地,使用基于机器学习的方式来筛出假评论,包括:

[0077] 对于每一条评论,确定评论内容长度L。当评论内容长度大于阈值L0时,进行评论文本分析。

[0078] 除了内容完全相同的评论以外,刷单者可能会轻微修改评论内容以避免直接被检测。

[0079] 可选地,使用文本分析算法中的使用相似度检测算法Jaccard相似度,当两条评论的相似度大于阈值时,确定评论为内容高度类似的评论。可选地,Jaccard相似度衡量的是

两个集合交集的大小与它们并集的大小之间的比例。在文本分析的情况下,这两个集合可以是两段文本中的词汇集合。使用Jaccard相似度来比较两个评论内容相似度,包括:对评论文本进行预处理以提取词汇集合。包括将每条评论分解成单词或短语。移除常见的停用词(如“和”、“是”、“在”等)。基于两个评论的词汇集合,计算它们的Jaccard相似度 $S_J:S_J = (A \cap B) / (A \cup B)$;

[0080] 其中,A和B分别是两个评论的词汇集合, $A \cap B$ 是两集合的交集, $A \cup B$ 是两集合的并集。

[0081] 为了比较大量评论的相似度,可以使用数据比较队列的方式存储有效评论。这个队列将维护待比较的评论对,包括:

[0082] 将新的评论与队列中现有的每条评论配对并计算相似度。当评论A已与比较队列中的所有其他评论比较过后,如果评论A与比较队列中的所有其他评论的Jaccard相似度都不超过阈值 S_1 ,则将评论A加入比较队列;如果评论A与比较队列中的至少一个其他评论的Jaccard相似度都等于或大于阈值 S_1 ,则将评论A及对应的其他信息(如主动评分)删除。

[0083] 阈值 S_1 设置得过低可能会导致很多正常评论被错误地标记为刷单,而设置得过高则可能会加入过度的刷单评论。本发明中,将阈值 S_1 设置为0.9。

[0084] 可选地,对于队列中剩余的所有评论,再根据文字与图片的匹配程度来进行二次筛查。具体地:

[0085] 识别附图评论中图片内的所有对象和所述对象的类型。如使用Fast R-CNN, Faster R-CNN、YOLO、SSD(Single Shot MultiBox Detector)或EfficientDet模型同时识别图像中的多个目标及其位置,并分类每个目标的类型。

[0086] 将文本输入进行标准化后进行分词,将句子分解为单独的词或短语。识别文本中的关键词和实体,如“松鼠桂鱼”、“鱼香肉丝”、“用餐环境”等等。

[0087] 对图像和文本分别使用嵌入向量表示后,通过Jaccard相似度来计算它们之间的匹配度。

[0088] 当通过Jaccard相似度来计算文字向量和图片向量之间的匹配度低于匹配阈值时,判断该评论符合刷单评论通常存在长评大量内容重复和带图评论和文字内容无关的情形,在队列中将该条评论进行删除。

[0089] 经过一次筛查和二次筛查后,队列中剩余的评论为有效评论。在后续步骤中,只对有效评论进行情感分析并进行细粒度分析统计。

[0090] 可选地,在步骤S3中,使用NLP库进行文本预处理,包括分词和词性标注。使用情感分析算法来计算情感倾向得分。

[0091] 可选地,Word2Vec模型会将文本中的每个单词表示成一个多维空间中的向量。这些向量捕获了单词的语义信息,即在高维空间中语义相似的单词会被映射到相近的点上。

[0092] 选择出现频率最高的N个向量,向量 X_i 和该向量对应的频率 f_i 组成了N个输入对,将N个输入对组成输入矩阵,包括按照从 f_i 高到低的顺序整理输入矩阵,将所述输入矩阵输入到预先训练的神经网络模型中,输出样本正面情绪的概率 $P_{positive}$ 和负面情绪的概率 $P_{negative}$ 。具体地,对于训练集中的文本数据,统计每个单词出现的频率。选取出现频率最高的N个词构建输入对。对于每个高频词,Word2Vec生成的对应词向量 X_i 和该词的出现频率 f_i ,创建输入对,每个输入对由词向量 X_i 和其频率 f_i 组成。

[0093] 将这些输入对按照频率 f_i 从高到低的顺序排列,形成一个输入矩阵。每一行代表一个高频词,包含该词的词向量和频率。

[0094] 将输入矩阵输入到神经网络,将输入矩阵输入到一个预先训练好的神经网络模型中。

[0095] 神经网络模型通过学习如何将这个词向量和频率信息与情感标签相关联,输出正面情绪的概率 P_{positive} 和负面情绪的概率 P_{negative} 。假设的语料库中有三个分词“好”,“差”,“超赞”,它们出现的频率分别是100,50,80次。使用Word2Vec得到每个词的词向量,假设的词向量是3维的(可以比3维更高维度更高),得到:“好”的向量是 $[0.5, 1.2, -0.3]$,频率是100。“差”的向量是 $[-1.0, 0.3, 0.8]$,频率是50。“超赞”的向量是 $[0.7, 1.5, -0.5]$,频率是80。

[0096] 按照频率从高到低排列这些词,构建输入矩阵,输入矩阵将是:

| 0.5 | 1.2 | -0.3 | 100 |

| 0.7 | 1.5 | -0.5 | 80 |

| -1.0 | 0.3 | 0.8 | 50 |

[0097] 这个矩阵每一行代表一个输入对,包含一个单词的词向量和它的频率。将所述输入矩阵直接输入到一个预训练的神经网络模型中,模型将根据这些数据预测整个文本的情感倾向,输出正面情绪的概率 P_{positive} 和负面情绪的概率 P_{negative} ,其中,正面情绪的概率 P_{positive} 和负面情绪的概率 P_{negative} 的取值都在0到1之间。

[0098] 情感得分是用来表示文本情感倾向的一个量化值。在情感分析中,这个得分可以是一个介于某个范围内的数值,例如从-1到1,-1代表完全负面,1代表完全正面,0可能代表中性。

[0099] 损失Loss是一个在机器学习模型训练中使用的,在监督学习中用来衡量模型的预测输出与实际数据之间的差异。在情感分析模型中,用损失函数指导模型的训练,通过最小化损失来调整模型的参数,使模型能更准确地预测情感得分。

[0100] 在使用深度学习进行情感分析时,模型的最终输出通常不是直接的情感得分,而是类别的概率分布。例如,对于二分类问题(正面或负面),模型会输出每个类别的概率(P_{positive} 和 P_{negative})。这些概率可以通过如下方式转换为一个情感得分S:

[0101] $S = P_{\text{positive}} - P_{\text{negative}}$,

[0102] 情感得分S将会在-1(完全负面)到1(完全正面)之间,其中0表示无限趋近于中性情感,可以理解为中性情感。该得分直接反映了模型对文本情感极性的估计。

[0103] 神经网络,包含嵌入层、卷积层、循环层或全连接层。前向传播包括输入文本 x ,得到两个类别的logits: z_{positive} 和 z_{negative} 。使用softmax函数计算概率后计算情感得分。

[0104] 在神经网络模型中,情感得分可以从输出层得出,使用softmax函数为每个情感类别生成概率。以二分类模型为例,其中 P_{positive} 和 P_{negative} 分别是正面和负面类别的预测概率,而 z_{positive} 和 z_{negative} 是对应类别的logits(softmax激活函数之前的激活值)。两个类别的softmax函数可以表示为:

[0105] $P_{\text{positive}} = \exp(z_{\text{positive}}) / (\exp(z_{\text{positive}}) + \exp(z_{\text{negative}}))$,

[0106] $P_{\text{negative}} = \exp(z_{\text{negative}}) / (\exp(z_{\text{positive}}) + \exp(z_{\text{negative}}))$,

[0107] P_{positive} 是文本为正面的概率。

[0108] P_{negative} 是文本为负面的概率。

[0109] \exp 是指数函数。

[0110] z_{positive} 是对应正面类别的logits。

[0111] z_{negative} 是对应负面类别的logits。

[0112] z_{positive} 和 z_{negative} 是神经网络模型中最后一个全连接层(通常称为输出层)的两个输出值,代表了正面和负面情感类别的logits。其中,输入文本 x 需要通过词嵌入将单词转换为向量。在嵌入层,文本 x 的数值表示通过嵌入层进行处理,以捕捉词汇的语义信息。在卷积层/循环层,通过卷积神经网络(CNN)层来捕获局部相关性。在全连接层,经过一系列中间层处理后,数据最终传递到全连接层。全连接层负责将网络中提取的特征映射到目标类别上。

[0113] 输出层为softmax激活函数的全连接层,用于多分类问题。对于softmax激活函数的二分类问题,输出层会有两个节点,分别输出 z_{positive} 和 z_{negative} 。

[0114] 计算logits, z_{positive} 和 z_{negative} 是网络的直接输出,不经过任何激活函数的处理。两个类别的概率可以表示为 P_{positive} 和 P_{negative} ,由 z_{positive} 和 z_{negative} 计算得到。

[0115] 可选地,使用标准交叉熵损失函数作为损失函数,公式为: $\text{Loss}_i = -y_i \cdot \log(P_{\text{positive}}) - (1 - y_i) \cdot \log(P_{\text{negative}})$ 。

[0116] 可选地,引入负面情绪的惩罚因子 β 来调整损失函数,公式为: $\text{Loss}_i = -y_i \cdot \log(P_{\text{positive}}) - \beta \cdot (1 - y_i) \cdot \log(P_{\text{negative}})$ 。

[0117] 其中, $\beta > 1$ 是一个惩罚乘数,它增加了来自负面例子的损失贡献。 Loss_i 是交叉熵损失函数在单个样本上的计算结果,用于二分类问题。这个函数的目的是量化模型预测概率分布与实际标签的差异。这里是每个参数的含义:

[0118] y_i :真实标签。如果文本 x_i 是正面情绪, y_i 为1;如果是负面情绪, y_i 为0。

[0119] P_{positive} :模型预测文本 x_i 为正面情绪的概率。由softmax函数计算得到。

[0120] P_{negative} :模型预测文本 x_i 为负面情绪的概率。由softmax函数计算得到。

[0121] \log :自然对数函数。

[0122] Loss_i :单个样本 x_i 的损失值。

[0123] $-y_i \cdot \log(P_{\text{positive}})$ 表示当真实标签 y_i 为1(即正面情绪)时,这部分损失是模型预测为正面情绪的概率的负对数。如果 P_{positive} 接近1,这部分的损失接近0;如果 P_{positive} 接近0,这部分的损失增大。模型预测正确时(P_{positive} 接近1),这一项接近0;预测错误时(P_{positive} 接近0),这一项趋向无穷大。

[0124] $-(1 - y_i) \cdot \log(P_{\text{negative}})$ 表示当真实标签 y_i 为0(即负面情绪)时,这部分损失是模型预测为负面情绪的概率的负对数。如果 P_{negative} 接近1,这部分的损失接近0;如果 P_{negative} 接近0,这部分的损失增大。这一项计算模型预测为负面的概率的负对数。模型预测正确时(P_{negative} 接近1),这一项接近0;预测错误时(P_{negative} 接近0),这一项趋向无穷大。

[0125] 整个损失函数是这两部分的总和,反映了模型预测概率分布与真实标签之间的总差异。

[0126] 交叉熵损失函数惩罚了预测概率与真实标签不一致的情况。当模型的预测概率接近真实标签的概率分布时,损失值最小。通过训练过程中最小化损失函数,模型学习调整神经网络节点权重,以更准确地预测数据的真实类别。

[0127] 通过在训练过程中对负面情绪的加强处罚,增加了负面情绪对结果影响的敏感度。

[0128] 可选地,对于数据集中的每个样本,计算损失Loss_i并汇总得到总损失Loss_{total}。使用反向传播算法来计算参数相对损失的梯度。应用梯度下降(或其他优化算法)更新CNN模型的权重。

[0129] 步骤S4,用户满意度分析中,结合用户评分分值和该用户评分对应的情感分析结果,确定一个顾客满意度指标,其中,将用户评分作为满意度指标的基础,将基于NLP技术进行情感分析得到的情感得分与用户评分结合,具体包括:

[0130] 设R为用户主动给出的评分,通常这个评分在一个预定的范围内,比如1到5。设S为通过情感分析得到的情感得分,它可能是一个介于-1(非常负面)到+1(非常正面)的值。

[0131] 顾客满意度指标C的计算公式如下:

[0132] $C = \alpha \cdot f(R) + \theta \cdot g(S) + \gamma \cdot h(R, S)$,其中, $f(R)$ 是关于用户评分R的函数,可以用来放大或者变换原始评分的影响。

[0133] $g(S)$ 是关于情感得分S的函数,用于调整情感得分的影响。

[0134] $h(R, S)$ 是一个混合函数,考虑了评分和情感得分的相互作用。

[0135] α, θ, γ 是权重参数,用于控制 $f(R), g(S), h(R, S)$ 在最终满意度指标中的相对重要性。

[0136] 可选地, $f(R)$ 选择指数函数, $g(S)$ 选择基于对数函数增加情感得分的极端值的影响, $h(R, S)$ 选择乘积体现相互作用。

[0137] 可选地,顾客满意度指标C的计算公式如下:

$$[0138] \quad C = a \cdot e^{k(R-\bar{R})} + \theta \cdot \ln(|S| + 1) \cdot \text{sign}(S) + \gamma \cdot (R - \bar{R}) \cdot S,$$

[0139] 其中, \bar{R} 是所有评分的平均值,k是一个调节参数。R为用户主动给出的评分,S为通过情感分析得到的情感得分, α, θ, γ 是权重参数,用于控制对应部分在最终满意度指标中的相对重要性。

[0140] 通过调整 $\alpha, \theta, \gamma, k$ 参数,对满意度指标进行准确调整,以最佳地拟合顾客满意度。

[0141] sign函数是符号符号。其定义如下:

[0142] 如果输入值为正数,sign函数返回1。

[0143] 如果输入值为负数,sign函数返回-1。

[0144] 如果输入值为零,sign函数返回0。

[0145] 在某一实施例中,在步骤S4中,对预设时间段内所有用户满意度进行细粒度分析,包括分析不同维度的满意度数据,所述不同维度包括产品类别和顾客群体。

[0146] 其中,分析不同维度的满意度数据,包括将顾客根据年龄、性别分成不同的群体,并分析每个群体的满意度,包括:

[0147] 按不同产品,将用户满意度数据按产品分别统计。

[0148] 进一步将每个产品下的数据根据用户的年龄、性别等人口统计特征进行分组。

[0149] 对每个维度进行分析,具体包括使用描述性统计分析每个群体的满意度(对顾客满意度指标C归一化后的平均值)。首先对对顾客满意度指标C归一化得到 c_n :

$$[0150] \quad c_n = \frac{c - cm_{i:n}}{c_{max} - cm_{i:n}} \times 5,$$

[0151] 其中, $cm_{i:n}$ 和 c_{max} 分别是所有计算出的顾客满意度指标C的最小值和最大值。转化为五分制标准分。

[0152] 示例性地,假设有两个产品(产品A和产品B)和两个顾客群体(根据性别分为男性和女性),并且考虑的年龄段分为18-25岁和26-35岁。将展示每个群体的平均满意度评分(基于5分制):

产品类别 / 顾客群体	男性	女性	男性	女性
	18-25	18-25	26-35	26-35
产品 A	4.2	3.8	4.0	4.5
产品 B	3.5	4.3	3.8	4.1

[0154] 步骤S5,智能确定营销策略,包括:

[0155] 基于POS系统当前销售数据和历史销售数据,预测下一预设时间段内不同产品的购买率,包括:使用简单移动平均(SMA)模型来预测下一时间段的购买率。对于每个产品,其在时间t的预测购买率 P_t 可以根据过去N个时间点的实际购买率 $P_{t-N+1}, P_{t-N+2}, \dots, P_t$ 来计算:

$$[0156] \quad P_t = \frac{1}{N} \sum_{i=0}^{N-1} P_{t-i},$$

[0157] 其中, P_t 是时间t的预测购买率, P_{t-i} 是时间t-i的实际购买率,N是移动平均的窗口大小,即考虑最近N个时间点的数据。

[0158] 确定想要考虑的时间点数量,记为N。这个N值称为窗口大小。找出最近N个时间点的购买率数据。例如,如果N是5,就需要查找最近5个时间点的购买率,记为P1,P2,P3,P4,和P5。计算这些购买率的总和,把P1,P2,P3,P4,和P5加起来。将总和除以时间点的数量N。这个结果就是时间点t的预测购买率。预测购买率等于过去N个时间点的购买率之和除以N。

[0159] 例如,如果在过去五个时间点的销售数据分别是100,120,130,150,和140的购买次数,那么时间点t的预测购买率将会是这五个数值加起来的总和,也就是640,然后除以5(因为使用了5个数据点),得到128,预测了在时间点t的购买率为128。

[0160] 可选地,分析得出的购买率预测数据,确定哪些产品或服务在下一时间窗口的时间段内预测购买率会有显著增加,例如超过预设的销售折扣阈值时判断该产品在下一时间窗对应的时间段内的购买率会有显著增加。

[0161] 根据预测数据,选择购买率预计会增加的时间段作为产品促销活动的时间,对产品价格进行折扣活动,所述折扣活动的力度可以为基于产品的利润率预设的折扣力度。

[0162] 在某一实施例中,步骤S6中智能进行个性化广告的推送,包括:基于细粒度的用户满意度分析结果、下一预设时间段内不同产品的购买率以及折扣促销策略策略,对不同顾客群体定制个性化广告推送内容,具体地:

[0163] 根据每个用户群体对所述产品的平均满意度确定推送内容,包括确定所有折扣产品中,该用户群体的平均满意度最高的M项产品,如 $M=3$,在该用户群体的用户登录首页展

示该M项产品的促销广告信息。本发明通过API自动化数据收集过程,提高了数据收集的效率和准确性。本发明使用数据清洗工具确保数据质量,为后续分析提供可靠的数据基础。本发明通过NLP技术处理复杂的文本数据,基于情感分析模型,更准确地评估顾客的情绪倾向。本发明联合用户评分与情感得分,形成更全面的满意度指标。本发明分析不同维度的满意度数据,结合POS系统数据预测不同产品的购买率,为制定营销策略提供科学依据。根据细粒度分析结果,定制个性化的广告推送内容。

[0164] 本发明通过更准确的满意度指标,可以更好地理解顾客需求,通过预测分析和优惠策略的智能制定,可以更有效地分配营销资源。

[0165] 需要说明的是,本公开上述的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本公开中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本公开中,计算机可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读信号介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:电线、光缆、RF(射频)等等,或者上述的任意合适的组合。

[0166] 上述计算机可读介质可以是上述电子设备中所包含的;也可以是单独存在,而未装配入该电子设备中。

[0167] 可以以一种或多种程序设计语言或其组合来编写用于执行本公开的操作的计算机程序代码,上述程序设计语言包括面向对象的程序设计语言——诸如Java、Smalltalk、C++,还包括常规的过程式程序设计语言——诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)——连接到用户计算机,或者,可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。

[0168] 附图中的流程图和框图,图示了按照本公开各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,该模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注

意的是,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0169] 描述于本公开实施例中所涉及到的单元可以通过软件的方式实现,也可以通过硬件的方式来实现。其中,单元的名称在某种情况下并不构成对该单元本身的限定。

[0170] 以上介绍了本发明的较佳实施方式,旨在使得本发明的精神更加清楚和便于理解,并不是为了限制本发明,凡在本发明的精神和原则之内,所做的修改、替换、改进,均应包含在本发明所附的权利要求概括的保护范围之内。



图1