

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2018 年 3 月 15 日 (15.03.2018)



(10) 国际公布号
WO 2018/045910 A1

(51) 国际专利分类号:
G06F 17/27 (2006.01)

(21) 国际申请号: PCT/CN2017/100060

(22) 国际申请日: 2017 年 8 月 31 日 (31.08.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201610812853.4 2016 年 9 月 9 日 (09.09.2016) CN

(71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。

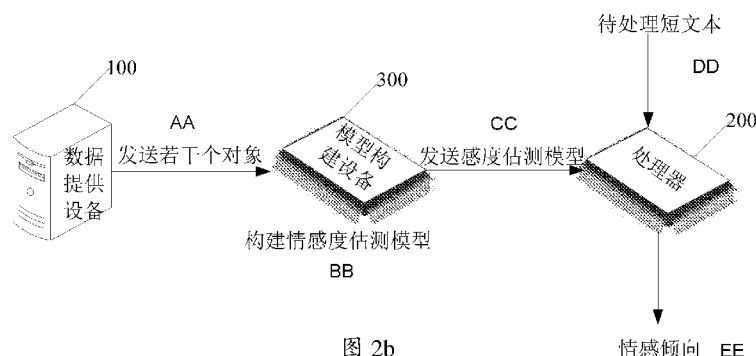
(72) 发明人; 及

(71) 申请人 (仅对 US): 潘林林 (PAN, Linlin) [CN/CN]; 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。赵争超 (ZHAO, Zhengchao) [CN/CN]; 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。林君 (LIN, Jun) [CN/CN]; 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。肖谦 (XIAO, Qian) [CN/CN]; 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。张一昌 (ZHANG, Yichang) [CN/CN]; 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。

(74) 代理人: 北京国昊天诚知识产权代理有限公司 (CO-HORIZON INTELLECTUAL PROPERTY INC.); 中

(54) Title: SENTIMENT ORIENTATION RECOGNITION METHOD, OBJECT CLASSIFICATION METHOD AND DATA PROCESSING SYSTEM

(54) 发明名称: 情感倾向的识别方法、对象分类方法及数据处理系统



- 100 Data provision device
- 200 Processor
- 300 Model constructing apparatus
- AA Sending a plurality of objects
- BB Constructing a sentiment degree estimation model
- CC Sending the sentiment degree estimation model
- DD Short text to be processed
- EE Sentiment orientation

(57) Abstract: Provided by the present application are a sentiment orientation recognition method, object classification method and data processing system. A sentiment degree estimation model, constructed by the present invention in the sentiment ornamentation recognition method, fully considers the category to which a short text belongs. Therefore, sentiment orientation is determined more accurately on the basis of the sentiment degree estimation model. In addition, since the object classification method provided by the present application use text feature information, image feature information and other feature information of an object as the basis of object classification, the object classification method provided by the present application may simultaneously give consideration to the text feature information, image feature information and other feature information, thereby being able to improve classification accuracy.

国北京市朝阳区小关北里甲2号渔阳置业大厦B座605, Beijing 100029 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 本申请提供了一种情感倾向的识别方法、对象分类方法及数据处理系统。本申请在情感倾向的识别方法中构建的情感度估测模型充分考虑了短文本所属的类目, 所以, 基于情感度估测模型确定出情感倾向更加准确。此外, 本申请提供的对象分类方法中, 将对象的文本特征信息、图像特征信息以及其它特征信息一并作为对象分类的依据, 所以本申请提供的对象分类方法、可以兼顾文本特征信息、图像特征信息和其他特征信息, 这样可以提高分类的准确率。

情感倾向的识别方法、对象分类方法及数据处理系统

技术领域

本申请涉及数据处理技术领域，尤其涉及情感倾向的识别方法、对象分类方法及数据处理系统。

背景技术

目前，在很多技术领域都涉及对对象进行分类的问题，通常而言，依据对象的文本对对象进行分类，将对象分为两个类别：第一类别或第二类别。在对象的文本中，按标点符号可以将文本为多个短文本。

由于汉字的词义丰富，在不同的语境下相同的短文本可能对应不同的类别。例如，以对象为衣服用户评价文本为例，第一条用户评价为“衣服颜色暗淡，刚好”，第二条用户评价为“衣服颜色暗淡，不鲜亮”。上述两个对象具有相同的短文本“衣服颜色暗淡”。若按文本进行分类，则会将两个短文本归为一类，可是两者理应对应不同的类别。

可以看出在不同语境中，第一条用户评价中的“衣服颜色暗淡”对应正面情感，理应分为第一类别；第二条用户评价中的“衣服颜色暗淡”对应负面情感，理应分为第二类别。因此，目前通常利用短文本对应的情感倾向来确定对象的类别。

为了确定短文本的情感倾向，传统方式通常为人工查看并确定短文本的情感倾向。虽然人工标注确定短文本的情感倾向的准确率较高，但是效率较低，无法适用于批量短文本的处理。

发明内容

本申请的申请人在研究过程中发现：可以利用处理器自动识别短文本的情感倾向。具体实现过程可以为：

在处理器具体执行之前，先构建情感词库。情感词库包含很多正面词汇，例如，“衣服”“屏幕大”“漂亮”、“快速”、“合适”、“美丽”等，情感词库也包含很多负面词汇，例如，“衣服”“难看”、“慢速”、“屏幕小”等。

为了对待处理对象进行处理，首先对待处理对象按标点符号进行切分，相邻两个标点符号之间为一个短文本，从而将待处理对象切分为若干个待处理短文本。例如，以“衣服很适合，老妈很喜欢”为例，按照标点符号切分后，可以获得两个短文本“衣服很适合”和“老妈很喜欢”。待处理对象的每个短文本，均为待处理短文本。

参见图 1，为处理器确定待处理短文本的情感倾向的流程图，执行过程具体包括以下步骤：

步骤 1：处理器对待处理短文本进行分词，获得分词结果。

按照预设分词规则，将待处理短文本分为若干词语，若干词语均为分词结果。

例如，以待处理短文本为“衣服很适合”为例，在分词后获得的结果为“衣服”、“很”和“合适”。以待处理短文本为“手机屏幕很大”，则分词后获得的分词结果为“手机”、“屏幕”、“很”和“大”。

由于对待处理短文本进行分词，并不是本申请关注的重点，在此不再详细描述预设分词规则的具体实现方式。

步骤 2：将分词结果与情感词库，按情感匹配规则进行匹配。

步骤 3：确定与待处理短文本对应的情感倾向。

将分词结果、与情感词库和情感规则进行匹配，若分词结果中分词均对应正面情感且不包含否定词，则确定短文本对应正面情感。若分词结果中情感词均对应负面情感且不包含否定词，则确定短文本对应负面情感。

处理器可以自动执行图 1 所示的过程，从而可以自动确定待处理短文本的情感倾向。但是，本申请申请人在研究过程中发现：虽然上述自动处理过程在一定程度上可以识别待处理短文本的情感倾向，但是，上述处理过程获得的待处理短文本的情感倾向可能不准确。

例如，以对象为淘宝网的用户评价为例，由于淘宝网上有很多类目（例如服饰类目、电子设备类目、母婴类目等），各个类目的物品均有相应的用户评价。申请人在研究过程中发现：在不同类目下包含相同情感词的短文本可能对应不同的情感倾向。

比如，在电子设备类目下、一个短文本为“屏幕很大”，该短文本的情感倾向为正面情感。在服饰类目下、一个短文本为“衣服很大”，该短文本的情感倾向为负面情感。从上述举例可以看出，在两个不同类目下、两个短文本均有“很大”，所以两个短文本包含相同的情感词，但是这两个短文本却具有不同的情感倾向。

由于上述图 1 中处理器自动确定短文本的情感倾向的过程中，处理器针对所有对象均采用同样的处理方式，即现有的处理过程没有从对象类目的角度、来分别处理短文本的情感倾向，所以，现有技术中确定短文本的情感倾向不准确。

因此，本申请提供一种情感倾向的识别方法，以便可以准确确定待处理短文本的情感倾向。

为了实现上述目的，本申请提供了以下技术特征：

一种情感倾向的识别方法，包括：

确定待处理短文本对应类目标识；其中，一个文本相邻两个标点符号之间文字称为短文本；

确定与所述类目标识对应的情感度估测模型的实现方式；

若所述情感度估测模型的实现方式为所有类目对应一个情感度估测模型，则确定待处理短文本对应的特征集合；其中，所述特征集合中每个特征包括：所述待处理短文本的分词和所述待处理短文本所属的类目标识；依据预先训练的情感度估测模型，结合待处理短文本的特征集合，对待处理短文本进行情感度估测；其中，所述情感度估测模型包括：依据至少两种类目的、带有情感倾向的若干个短文本样本训练后得到的、输出正面情感度和负面情感度的模型；基于所述待处理短文本对应的正面情感度和负面情感度，确定所述待处理短文本对应的情感倾向；

若所述情感度估测模型的实现方式为一个类目对应一个情感度估测模型，确定待处理短文本对应的特征集合；其中，所述特征集合中每个特征包括：所述待处理短文本的分词；依据与所述类目标识对应的情感度估测模型，结合待处理短文本的特征集合，对待处理短文本进行情感度估测；其中，所述情感度估测模型为：依据所述类目标识对应的、带有情感倾向的若干个短文本样本训练后得到的、输出正面情感度和负面情感度的模型；基于所述待

处理短文本对应的正面情感度和负面情感度，确定所述待处理短文本对应的情感倾向。

优选的，在确定所述待处理短文本对应的情感倾向后，还包括：

输出所述待处理短文本对应的情感倾向。

一种情感倾向的识别方法，包括：

确定待处理短文本对应的特征集合；其中，一个文本相邻两个标点符号之间的文字称为短文本；所述特征集合中每个特征包括：所述待处理短文本的分词和所述待处理短文本所属的类目标识；

依据预先训练的情感度估测模型，结合待处理短文本的特征集合，对待处理短文本进行情感度估测；其中，所述情感度估测模型包括：依据至少两种类目的、带有情感倾向的若干个短文本样本训练后得到的、输出正面情感度和负面情感度的模型；

基于所述待处理短文本对应的正面情感度和负面情感度，确定所述待处理短文本对应的情感倾向。

优选的，所述确定待处理短文本对应的特征集合，包括：

获取所述待处理短文本对应的类目标识，以及所述待处理短文本执行分词操作后获得的分词结果；

将所述分词结果中的各个分词和所述类目标识进行组合，获得各个特征；

将各个特征的集合，确定为所述待处理短文本的特征集合。

优选的，所述确定待处理短文本对应的特征集合，包括：

获取所述待处理短文本对应的类目标识，以及所述待处理短文本执行分词操作后获得的分词结果；

将所述分词结果中的各个分词和所述类目标识进行组合，获得各个特征；

利用 n 元语言模型对所述各个特征进行特征组合，获得若干个组合特征；

将各个特征和所述若干个组合特征的集合，确定为所述待处理短文本的特征集合。

优选的，所述利用 n 元语言模型对所述各个特征进行特征组合，获得若干个组合特征，包括：

利用二元语言模型对所述各个特征进行特征组合，获得若干个组合特征。

优选的，所述依据预先训练的情感度估测模型，结合待处理短文本的特征集合，对待处理短文本进行情感度估测，包括：

将所述特征集合输入至所述情感度估测模型；

由所述情感度估测模型估算后，输出待处理短文本对应的正面情感度和负面情感度。

优选的，所述基于所述待处理短文本对应的正面情感度和负面情感度，确定所述待处理短文本对应的情感倾向，包括：

确定所述正面情感度和所述负面情感度两者中的较大情感度；

判断所述较大情感度是否大于预设置信度；

若所述较大情感度大于预设置信度，则确定所述待处理短文本对应的情感倾向与所述较大情感度的情感倾向一致。

优选的，所述情感度估测模型包括：

利用最大熵模型，依据至少两个类目标识对应的若干个短文本的特征集合训练后得到的、输出正面情感度和负面情感度的模型。

优选的，在确定所述待处理短文本对应的情感倾向后，还包括：

输出所述待处理短文本对应的情感倾向。

一种情感倾向的识别方法，包括：

确定待处理短文本对应的特征集合和类目标识；其中，一个文本相邻两个标点符号之间的文字称为短文本；所述特征集合中每个特征包括：所述待处理短文本的分词；

依据与所述类目标识对应的情感度估测模型，结合待处理短文本的特征集合，对待处理短文本进行情感度估测；其中，所述情感度估测模型为：依据所述类目标识对应的、带有情感倾向的若干个短文本样本训练后得到的、输出正面情感度和负面情感度的模型；

基于所述待处理短文本对应的正面情感度和负面情感度，确定所述待处理短文本对应的情感倾向。

优选的，所述确定待处理短文本对应的特征集合，包括：

获取所述待处理短文本执行分词操作后获得的分词结果；

利用 n 元语言模型对各个分词进行分词组合，获得若干个组合分词；

将各个分词和若干个组合分词的集合，确定为所述待处理短文本的特征集合，一个分词对应一个特征。

优选的，所述确定待处理短文本对应的特征集合，包括：

获取所述待处理短文本执行分词操作后获得的分词结果；

将所述分词结果，确定为所述待处理短文本的特征集合，一个分词对应一个特征。

优选的，在确定所述待处理短文本对应的情感倾向后，还包括：

输出所述待处理短文本对应的情感倾向。

一种情感倾向的识别系统，包括：

数据提供设备，用于发送若干个对象；

处理器，用于接收所述数据提供设备送的若干个对象，依据若干个对象的短文本构建情感度估测模型，并利用情感度估测模型确定待处理短文本的情感倾向。

优选的，所述处理器，还用于构建情感度估测模型与对象所属的类目标识的对应关系。

优选的，所述系统还包括接收设备；

所述处理器，还用于输出所述待处理文本的情感倾向；

所述接收设备，用于接收所述待处理文本的情感倾向。

一种情感倾向的识别系统，包括：

数据提供设备，用于发送若干个对象；

模型构建设备，用于接收所述数据提供设备送的若干个对象，依据若干个对象的短文本构建情感度估测模型，并发送所述情感度估测模型；

处理器，用于接收所述情感度估测模型，并利用情感度估测模型确定待处理短文本的情感倾向。

优选的，所述模型构建设备，还用于构建情感度估测模型与对象所属的类目标识的对应关系，并将对应关系发送至所述处理器。

优选的，所述系统还包括接收设备；

所述处理器，还用于输出所述待处理文本的情感倾向；

所述接收设备，用于接收所述待处理文本的情感倾向。

一种对象分类方法，包括：

确定待处理对象的特征信息；其中，所述特征信息包括文本特征信息和图像特征信息，并且，所述文本特征信息包括短文本的情感倾向；

依据预先训练的类别识别模型，对所述待处理对象的特征信息进行类别识别；其中，所述类别识别模型为：依据若干对象样本的特征信息训练后得到的、第一类别和第二类别的分类器。

优选的，所述特征信息还包括：

构建所述对象的第一主体的特征信息；和/或，

所述对象所附属于第二主体的特征信息。

优选的，所述依据预先训练的类别识别模型，对所述特征信息进行类别识别，包括：

将所述特征信息输入至所述类别识别模型；确定所述待处理对象对应的第一类别匹配度和第二类别匹配度；

对所述第一类别匹配度和第二类别匹配度进行比较；

若第一类别匹配度大于第二类别匹配度，则确定所述待处理对象的类别为第一类别；

若第二类别匹配度大于第一类别匹配度，则确定所述待处理对象的类别为第二类别。

优选的，还包括：

在确定所述待处理对象为第一类别之后，将所述待处理对象添加至对象集合中；

发送所述对象集合中的对象。

优选的，还包括：

接收多个对象样本，所述对象样本来源于所述对象集合，且，满足预设规则；

将所述多个对象样本，添加至训练类别识别模型的已有对象样本中；

基于更新后的已有对象样本，重新训练类别识别模型。

一种用户评价的分类方法，包括：

确定待处理用户评价的特征信息；其中，所述特征信息包括用户评价的文本特征信息、用户评价的图像特征信息、卖家的特征信息和买家的特征信息，并且，所述文本特征信息包括短文本的情感倾向；

依据预先训练的梯度提升决策树模型，对所述待处理用户评价的特征信息进行类别识别；其中，所述类别识别模型为：依据若干用户评价样本的特征信息训练后得到的、第一类用户评价和第二类用户评价的分类器。

优选的，还包括：

在确定所述待处理用户评价为第一类用户评价之后，将所述待处理用户评价添加至第一类用户评价集合中；

发送所述第一类用户评价集合。

优选的，还包括：

接收多个第一类用户评价，所述第一类用户评价来源于所述第一类用户评价集合；

将所述多个第一类用户评价，添加至类别识别模型已有的用户评价样本中；

基于更新后的已有的用户评价样本，重新训练类别识别模型。

一种对象分类系统，包括：

数据提供设备，用于发送若干个对象；

处理器，用于接收所述数据提供设备送的若干个对象，依据若干对象的特征信息训练后得到、输出第一类别和第二类别的类别识别模型；用于确定待处理对象的特征信息；其中，所述特征信息包括文本特征信息和图像特征信息，并且，所述文本特征信息包括短文本的情感倾向；依据所述类别识别模型，对所述待处理对象的特征信息进行类别识别；还用于输出第一类别的对象；

数据接收设备，用于接收并使用所述第一类别的对象。

一种对象分类系统，包括：

数据提供设备，用于发送若干个对象；

模型构建设备，用于接收所述数据提供设备送的若干个对象，依据若干个对象的特征信息训练后得到、输出第一类别和第二类别的类别识别模型，并发送所述类别识别模型；

处理器，用于接收所述类别识别模型，并确定待处理对象的特征信息；其中，所述特征信息包括文本特征信息和图像特征信息，并且，所述文本特

征信息包括短文本的情感倾向；依据所述类别识别模型，对所述待处理对象的特征信息进行类别识别；还用于输出第一类别的对象；

数据接收设备，用于接收并使用所述第一类别的对象。

通过以上技术手段，可以实现以下有益效果：

本申请提供一种情感倾向的识别方法，本方法利用与类目对应的若干个带有情感倾向的短文本作为训练样本，获取短文本的特征集合进行训练，并获得情感度估测模型。由于每个特征包含短文本的分词和类目标识，所以，申请构建的情感度估测模型充分考虑了短文本所属的类目。因此，基于情感度估测模型确定出的待处理短文本的情感倾向也更加准确。

附图说明

为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为现有技术确定待处理短文本的情感倾向的流程图；

图 2a-2b 为本申请实施例提供的情感倾向的识别系统的结构示意图；

图 3a-3c 为本申请实施例提供的情感度估测模型与类目的对应关系的示意图；

图 4a-4c 为本申请实施例提供的构建情感度估测模型的流程图；

图 5 为本申请实施例提供的又一构建情感度估测模型的流程图；

图 6a-6b 为本申请实施例提供的又一构建情感度估测模型的流程图；

图 7 为本申请实施例提供的情感倾向的识别方法的流程图；

图 8a-8b 为本申请实施例提供的情感倾向的识别方法的流程图；

图 9 为本申请实施例提供的情感倾向的识别方法的流程图；

图 10 为本申请实施例提供的情感倾向的识别方法的流程图；

图 11a-11b 为本申请实施例提供的情感倾向的识别方法的流程图；

图 12 为本申请实施例提供的对象分类方法的流程图；

图 13 为本申请实施例提供的又一对象分类方法的流程图；

图 14 为本申请实施例提供的又一对象分类方法的流程图；

图 15 为本申请实施例提供的又一对象分类方法的流程图；

图 16 为本申请实施例提供的一种对象分类系统的结构示意图；

图 17 为本申请实施例提供的又一种对象分类系统的结构示意图；

图 18 为本申请实施例提供的对象分类方法的场景实施例的流程图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，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他实施例，都属于本申请保护的范围。

为了准确确定待处理短文本的情感倾向，本申请提出构建情感度估测模型的技术手段，以利用情感度估测模型来估测待处理短文本对应的正面情感度和负面情感度。其中，正面情感度用于表示待处理短文本属于正面情感的程度，同理，负面情感度用于表示待处理短文本属于负面情感的程度。在确定正面情感度和负面情感度之后，可以进一步确定待处理短文本的情感倾向。

为了使本领域技术人员更加清楚了解本申请的应用场景，参见图 2a 或图 2b，为本申请提供了情感倾向的识别系统。

图 2a 提供的情感倾向的识别系统具体包括：数据提供设备 100，与数据提供设备 100 相连的处理器 200。

其中，数据提供设备 100 用于向处理器 200 发送若干个对象。处理器 200，用于依据若干个对象的短文本构建情感度估测模型，并利用情感度估测模型确定待处理短文本的情感倾向。

本申请还提供另一种情感倾向的识别系统（参见图 2b）。

图 2b 提供的情感倾向的识别系统具体包括：数据提供设备 100，与数据提供设备相连的模型构建设备 300，与所述模型构建设备相连的处理器 200。模型构建设备 300 可以为具有处理能力的处理设备。

其中，数据提供设备 100 用于向模型构建设备 300 发送若干个对象。模型构建设备 300，用于依据若干个对象的短文本构建情感度估测模型，并将情

感度估测模型发送至处理器 200。处理器 200，用于利用情感度估测模型确定待处理短文本的情感倾向。

在图 2a 和图 2b 提出的情感倾向的识别系统中，处理器 200 和模型构建设备 300 均可以执行构建情感度估测模型的过程，并且，两者构建情感度估测模型的过程是一致的。因此，将处理器 200 或模型构建设备 300 统称为处理设备，以便在下述介绍构建情感度估测模型的过程中，采用处理设备来统一表示处理器 200 或模型构建设备 300。

在图 2a 和图 2b 所示的系统中还可以包括与处理器相连的接收设备（图中未示出）。在处理器确定待处理短文本的情感倾向后，处理器，还用于输出所述待处理文本的情感倾向；所述接收设备，用于接收所述待处理文本的情感倾向，以便接收设备可以利用待处理文本的情感倾向执行其它处理过程。

下面介绍构建情感度估测模型的过程。由于现有技术确定待处理短文本的情感倾向的过程中未考虑短文本的类目，所以现有技术中确定出的情感倾向不准确。因此，本申请在处理设备构建情感度估测模型的过程中考虑短文本的类目，以便构建的情感度估测模型可以准确确定出待处理短文本的正面情感度和负面情感度。

本申请提出处理设备构建情感度估测模型的三种实现方式，参见图 3a-3c 为三种实现方式中类目与情感度估测模型的示意图。

第一种实现方式：所有类目对应一个情感度估测模型（参见图 3a）。第二种实现方式：每个类目对应一个情感度估测模型（参见图 3b）。第三种实现方式：介于第一种实现方式和第二种实现方式之间的一种实现方式（参见图 3c）；假设有 N 个类目，则第三种实现方式可以构建 M 个情感度估测模型，其中， M 为非零自然数，且， $1 < M < N$ 。

下面详细介绍这三种实现方式的具体实现过程：

第一种实现方式：所有类目对应一个情感度估测模型。

为了准确确定各个类目下的短文本对应的情感倾向，本实现方式为所有类目构建一个对应的情感度估测模型。

参见图 4a, 为所有类目对应的情感度估测模型的过程, 具体包括以下步骤:

步骤 S401: 确定用于构建情感度估测模型的短文本样本。

a) 获取数据提供设备发送的各个类目下的若干个对象, 并对每个对象进行切分, 获得每个对象的短文本集合。

数据提供设备可以向处理设备发送各个类目下的对象, 处理设备可以获取每个类目下的多个对象。为了方便后续处理, 处理设备可以对每个对象按标点符号进行切分, 从而将每个对象切分为多个短文本。

例如, 以对象为淘宝用户评价为例, 在服饰类目下的一个用户评价“衣服很合适, 老妈很喜欢”, 则按照标点符号切分后, 可以获得两个短文本“衣服很合适”和“老妈很喜欢”。目标短文本。例如, 在电子设备类目下的一个用户评价“手机屏幕很大, 外观很漂亮”, 则按照标点符号切分后, 可以获得两个短文本“手机屏幕很大”和“外观很漂亮”。

b) 在所有的短文本中筛选出用于构建情感度估测模型的短文本样本。

经过实验发现, 图 1 所示的执行过程, 确定一个短文本属于正面情感的准确率较高, 确定一个短文本属于负面情感的准确率较低。

因此, 在本步骤中处理设备可以将每个短文本执行如图 1 所示的过程, 若按图 1 所示的过程, 确定出一个短文本对应正面情感。那么, 确定该短文本可以用于构建情感度估测模型, 且, 该短文本对应正面情感。

若按图 1 所示的过程, 确定一个短文本对应负面情感。那么, 再由人工进行进一步的确认。若一个短文本在人工确认后属于负面情感, 则确定该短文本可以用于构建情感度估测模型, 且, 该短文本对应负面情感。

若一个短文本在人工确认后属于正面情感, 则说明该短文本的特点不明显, 不适合作为构建情感度估测模型的短文本。因此则丢弃该短文本。

步骤 S402: 确定每个短文本对应的特征集合。

在步骤 S401 使用图 1 所示的过程中可以获得每个短文本的分词结果 (详见图 1 中步骤 1, 在此不再赘述)。然后, 进一步确定每个短文本对应的特征集合。

本步骤可以有两种执行方式，两种方式的区别在于：第一种方式确定出的特征集合中包含组合特征，而第二种方式中确定出的特征集合不包含组合特征。

由于确定每个短文本对应的特征集合均是一致的，因此，以一个目标短文本为例，对确定目标短文本的特征集合的过程进行详细介绍。

参见图 4b，为确定目标短文本的特征集合的第一种执行方式的具体过程：

步骤 411：获取所述待处理短文本对应的类目标识，以及所述待处理短文本执行分词操作后获得的分词结果。

处理设备在步骤 S301 中已经获得目标短文本的分词结果。由于目标短文本与待处理对象的类目是一致的，因此，处理设备可以将待处理对象的类目标识，确定为目标短文本的类目标识。

以目标短文本属于服饰类目，且为“衣服很大”为例，目标短文本对应的分词结果为“衣服”“很”和“大”，假设服饰类目的标识为“16”，则目标短文本的对应的类目标识为“16”。

以目标短文本属于电子设备类目，且为“屏幕很大”为例，目标短文本对应的分词结果为“屏幕”“很”和“大”，假设电子设备类目的标识为“10”，则目标短文本的对应的类目标识为“10”。

步骤 412：将各个分词和所述类目标识进行组合，获得各个特征。

由于不同类目下的短文本对应的分词可能是一致的，因此，为了充分考虑类目对短文本的影响，本申请将各个分词与类目进行组合，获得各个特征。

由于特征包含了类目标识，并且，不同类目的标识是不同的，所以采用特征可以准确区分不同类目的分词。这样，训练得到的情感度估测模型可以准确区分不同类目下的相同分词。

继续延续上述举例，以目标短文本“衣服很大”为例，则目标短文本对应的各个特征可以为“衣服 16”“很 16”和“大 16”。以目标短文本“屏幕很大”为例，则目标短文本对应的各个特征可以为“屏幕 10”“很 10”和“大 10”。站在特征角度，处理设备可以分辨出分词“大 16”和“大 10”是两个不同的特征，且两个特征属于不同的类目。

在本举例中，分词和类目标识的组合方式为分词在前、类目标标识后，还可以是类目标标识在前、分词在后。当然，分词和类目标标识还可以有其它组合方式，在此不做限定。

步骤 413：对各个特征进行 n 元组合，获得若干个组合特征。

因为，通过研究过程中发现，一些特征具有固定搭配，例如“没有色差”、“没有掉色”、“没有起球”等等。对于这种固定搭配，由于两个词均是负面情感的词汇，但是两者叠加起来表达则为正面情感，所以这样的词汇若分开的话会造成一定的误判。因此，本实施例可以进行特征组合。

具体而言，为利用 n 元语言模型对每个短文本的各个特征进行组合。 n 为非零自然数， n 元语言模型中的一个元对应短文本中的一个分词。 n 元语言模型进行特征组合具体为：将相邻的 n 个特征合并在一起，将 $n-1$ 个特征合并在一起，直到将 2 个特征合并在一起。

以 $n=2$ 为例，若目标短文本的各个特征为“衣服 16”、“很 16”和“大 16”，则利用二元语言模型进行特征组合后，获得组合特征为“衣服 16 很 16”和“很 16 大 16”。

以 $n=3$ 为例，若目标短文本的各个特征为“衣服 16”、“很 16”和“大 16”，则进行三元语言模型进行特征组合后，获得组合特征为“衣服 16 很 16 大 16”、“衣服 16 很 16”和“很 16 大 16”。

步骤 414：将各个特征和若干个组合特征的集合，确定为所述目标短文本的特征集合。

延续上述实施例，则以二元语言模型进行特征组合为例，则最终获得的目标短文本的特征集合包括：“衣服 16”、“很 16”、“大 16”、“衣服 16 很 16”和“很 16 大 16”。

参见图 4c，为确定目标短文本的特征集合的第二种执行方式的具体过程：

步骤 421：获取所述待处理短文本对应的类目标识，以及所述待处理短文本执行分词操作后获得的分词结果。

步骤 422：将各个分词和所述类目标识进行组合，获得各个特征。

图 4c 中的步骤 S421 和步骤 S422 的执行过程与图 4b 中的步骤 S411 和步骤 S412 一致，在不再赘述。

步骤 423: 将各个特征的集合, 确定为所述目标短文本的特征集合。

在图 4c 的执行过程中缺少进行特征组合的步骤, 所以, 可以直接将步骤 S422 中确定的各个特征的集合, 确定为目标短文本的特征集合。

以目标短文本为“衣服很大”为例, 则按图 4c 执行后最终获得的目标短文本的特征集合包括: “衣服 16”、“很 16”、“大 16”。

接着返回图 4a, 进入步骤 S403: 确定各个短文本对应特征集合中每个特征的情感倾向, 以及每个特征的正面情感度和负面情感度, 并将各个特征以及各个特征对应的情感倾向、正面情感度和负面情感度, 作为情感度估测模型的输入参数。

在步骤 S401 执行图 1 实施例的过程中, 已经确定短文本的情感倾向。由于各个特征的情感倾向与短文本的情感倾向是一致的。因此, 在短文本对应正面情感时, 确定特征集合中每个特征对应正面情感; 在短文本对应负面情感时, 确定特征集合中每个特征对应负面情感。

以一个特征为例, 对确定特征的正面情感度和负面情感度的过程进行详细介绍。处理设备可以获得很多数量的同一个特征, 并且, 该特征对应的情感倾向可能相同, 可能不同。

因此, 处理设备可以统计该特征的总数量, 并统计属于正面情感的第一数量, 以及属于负面情感的第二数量。依据第一数量与总数量的比例关系, 确定该特征的正面情感度; 依据第一数量与总数量的比例关系, 确定该特征的负面情感度。

步骤 S404: 按照预设分类器模型进行训练, 并获得训练后得到的情感度估测模型。

预设分类器模型可以包括最大熵模型、支持向量机、神经网络算法等等。有关训练过程已有相关技术手段, 在此不再赘述。

下面介绍处理设备构建情感度估测模型的第二种实现方式, 在第二种实现方式中为每个类目构建一个情感度估测模型, 因此, 由于每个情感度估测模型中只有一个类目, 所以在第二种实现方式中分词即相当于特征, 因此在第二种实现方式中无需将分词和类目标识进行组合。

由于每个类目对应的情感度估测模型的构建过程是一致的。因此，以一个目标类目为例，对构建目标类目对应的目标情感度估测模型的过程进行详细介绍。

参见图 5，构建目标情感度估测模型的过程具体包括以下步骤：

步骤 S501：确定构建目标情感度估测模型的短文本样本。

a) 获取数据提供设备发送的目标类目下的若干个对象，并对每个对象进行切分，获得每个对象的短文本集合。

b) 在所有的短文本中筛选出用于构建情感度估测模型的短文本。

步骤 S501 的具体执行过程与步骤 S401 的执行过程类似，在此不再赘述。

步骤 S502：确定每个短文本对应的特征集合。

在步骤 S501 使用图 1 所示的过程中可以获得每个短文本的分词结果（详见图 1 中步骤 1，在此不再赘述）。然后，进一步确定每个短文本对应的特征集合。本步骤可以有两种执行方式，两种方式的区别在于：第一种方式确定出的特征集合中包含组合特征，而第二种方式中确定出的特征集合不包含组合特征。

由于确定每个短文本对应的特征集合均是一致的，因此，以一个目标短文本为例，对确定目标短文本的特征集合的过程进行详细介绍。

参见图 6a，为确定目标短文本的特征集合的第一种执行方式的具体过程：

步骤 601：获取所述目标短文本对应的分词结果，每个分词对应一个特征。

步骤 602：对所述各个特征进行 n 元组合，获得若干个组合特征。

步骤 603：将各个特征和若干个组合特征的集合，确定为所述目标短文本的特征集合。

以待处理短文本为“衣服很大”，以二元语言模型进行特征组合为例，则本实施例最终获得的目标短文本的特征集合包括：“衣服”、“很”、“大”、“衣服很”和“很大”。

参见图 6b，为确定目标短文本的特征集合的第二种执行方式的具体过程：

步骤 611：获取所述目标短文本对应的分词结果，每个分词对应一个特征。

步骤 612：将分词结果，确定为所述目标短文本的特征集合。

在图 6b 的执行过程中缺少进行特征组合的步骤，所以，可以直接将步骤 S611 中确定的各个特征的集合，确定为目标短文本的特征集合。

以目标短文本为“衣服很大”为例，则按图 6b 执行后最终获得的目标短文本的特征集合包括：“衣服”、“很”、“大”。

接着返回图 5，进入步骤 S503：确定目标类目下各个短文本对应特征集合中每个特征的情感倾向，以及每个特征的正面情感度和负面情感度，并将目标类目下各个特征以及各个特征对应的情感倾向、正面情感度和负面情感度，作为目标情感度估测模型的输入参数。

在步骤 S501 执行图 1 实施例的过程中，已经确定各个短文本的情感倾向。由于各个特征的情感倾向与短文本的情感倾向是一致的。因此，在短文本对应正面情感时，确定特征集合中每个特征对应正面情感；在短文本对应负面情感时，确定特征集合中每个特征对应负面情感。

步骤 S504：按照预设分类器模型进行训练，并获得训练后得到的目标情感度估测模型。

预设分类器模型可以包括最大熵模型、支持向量机、神经网络算法等等。有关训练过程已有相关技术手段，在此不再赘述。

图 5 为构建一个类目的情感度估测模型，图 3 为构建所有类目的情感度估测模型的过程，两者的处理步骤很类似，因此，图 5 的实施例的执行过程，可以参考图 4 的具体执行过程，在此不再赘述。

在第二实现方式中，每个类目对应一个情感度估测模型。因此，为了避免混淆，处理设备在一个情感度估测模型构建完毕之后，还会构建情感度估测模型与类目标识之间的映射，以便后续处理器在使用时，可以准确确定与每个类目对应的情感度估测模型。

下面介绍处理设备构建情感度估测模型的第三种实现方式。

在第三种实现方式中，可以包括：两个或两个以上的类目对应的情感度估测模型，和/或，一个类目对应的情感度估测模型。两个或两个以上类目对应的情感度估测模型的构建过程，可以参考图 4 所示的实施例。一个类目对应的情感度估测模型，可参考图 5 所示的实施例，在此不再赘述。

结合图 2a 和图 2b, 若上述构建情感度估测模型的处理设备为处理器 200 自身的情况下, 处理器 200 完成情感度估测模型后, 便可以直接使用, 以利用情感度估测模型确定待处理短文本的情感倾向。

在处理设备为模型构建设备 300 的情况下, 模型构建设备 300 会将情感度估测模型发送至处理器 200, 以便处理器 200 利用情感度估测模型确定待处理短文本的情感倾向。

下面介绍处理器 200 依据情感度估测模型确定待处理短文本的情感倾向的过程。由于情感度估测模型有三种不同的实现方式, 在不同实现方式下, 处理器 200 的执行过程也不尽相同, 所以, 下面分别介绍在情感度估测模型的不同实现方式下, 处理器的执行过程。

第一种:

在情感度估测模型采用第一种实现方式(所有类目对应一个情感度估测模型)实现的情况下, 处理器 200 采用以下方式来确定待处理的短文本的情感倾向。

参见图 7, 本申请一种情感倾向的识别方法, 具体包括以下步骤:

步骤 S701: 确定待处理短文本对应的特征集合; 其中, 所述特征集合中每个特征包括: 待处理短文本的分词和所述待处理文本所属的类目标识。

假设第一种实现方式在确定情感度估测模型的过程中, 采用第一种执行方式确定短文本的特征集合; 则在本步骤中也采用第一种执行方式确定待处理短文本特征集合。

参见图 8a, 确定待处理短文本对应的特征集合的第一种执行方式, 具体包括以下步骤:

步骤 S801: 获取所述待处理短文本对应的类目标识, 以及所述待处理短文本执行分词操作后获得的分词结果。

步骤 S802: 将分词结果中的各个分词和所述类目标识进行组合, 获得各个特征。

步骤 S803: 对所述各个特征进行 n 元组合, 获得若干个组合特征。

步骤 S804: 将各个特征和若干个组合特征的集合, 确定为所述待处理短文本的特征集合。

图 8a 的执行过程可参见图 4a 的执行过程, 在此不再赘述。

假设第一种实现方式在确定情感度估测模型的过程中，采用第二种执行方式确定短文本的特征集合；则在本步骤中也采用第二种执行方式确定待处理短文本的特征集合。

参见图 8b，确定待处理短文本对应的特征集合的第二种执行方式，具体包括以下步骤：

步骤 S811：获取所述待处理短文本对应的类目标识，以及所述待处理短文本执行分词操作后获得的分词结果。

步骤 S812：将分词结果中的各个分词和所述类目标识进行组合，获得各个特征。

步骤 S813：将各个特征的集合，确定为所述待处理短文本的特征集合。

图 8b 的执行过程可参见图 4b 的执行过程，在此不再赘述。

接着返回图 7，步骤 S702：依据预先训练的情感度估测模型，结合待处理短文本的特征集合，对待处理短文本进行情感度估测；其中，所述情感度估测模型包括：依据至少两个类目、带有情感倾向的若干个短文本样本训练后得到的、输出正面情感度和负面情感度的模型。

处理器将所述特征集合输入至所述情感度估测模型，由所述情感度估测模型估算后输出所述特征集合对应的正面情感度和负面情感度。

步骤 S703：基于所述待处理短文本对应的正面情感度和负面情感度，确定所述待处理短文本对应的情感倾向。

在确定所述待处理短文本对应的情感倾向，还可以输出所述待处理短文本对应的情感倾向，以便进行其它方面的使用。

在步骤 S702 中估测得到待处理短文本属于正面情感的正面情感度，以及待处理文本属于负面情感的负面情感度之后，为了进一步确定待处理短文本的情感倾向，可以将正面情感度与负面情感度进行对比。若正面情感度大于负面情感度，则确定待处理短文本属于对应正面情感；若负面情感度大于正面情感度，则确定待处理短文本对应负面情感。

在一些情况下，正面情感度和负面情感度相差不大。以情感度采用概率表示为例，正面情感度的概率值为 0.51，负面情感度的概率值为 0.49。可以理解的是，由于正面情感度和负面情感度非常接近，所以理论上是无法准确确

定待处理短文本的情感倾向的。但是，在此情况下，仍然按照上段方式确定待处理短文本的情感倾向，则会出现误差。

因此，参见图 9，本申请提供以下方式来待处理短文本的情感倾向。

步骤 S901: 确定所述正面情感度和所述负面情感度两者中的较大情感度。

将正面情感度和负面情感度进行对比，确定两者中的较大情感度。若正面情感度大于负面情感度，则确定正面情感度为较大情感度；若负面情感度大于正面情感度，则确定负面情感度为较大情感度。

步骤 S902: 判断所述较大情感度是否大于预设置信度。

为了判定较大情感度是否可信，本申请预先设定了预设置信度。预设置信度为确定较大情感度可信的程度。然后，判断较大情感度与预设置信度的大小。

步骤 S903: 若所述较大情感度大于预设置信度，则确定所述待处理短文本对应的情感倾向与所述较大情感度的情感倾向一致。

若较大情感度大于预设置信度，则确定较大情感度的可信度较高。因此，可以准确确定待处理短文本的情感倾向。此时，待处理短文本的情感倾向与较大情感度的情感倾向一致。

即，若较大情感度对应正面情感度，则确定待处理短文本属于对应正面情感；若较大情感度对应负面情感度，则确定待处理短文本对应负面情感。

假设较大情感度为 0.8，预设置信度为 0.7，则在此情况下，可以准确确定待处理短文本的情感倾向。

步骤 S904: 若所述较大情感度不大于预设置信度，则执行其它处理过程确定待处理文本的情感倾向。

若较大情感度不大于预设置信度，则确定较大情感度的可信度较低。因此，可以无法准确确定待处理短文本的情感倾向。假设较大情感度为 0.55，预设置信度为 0.7，则在此情况下，无法准确确定待处理短文本的情感倾向。

在此情况下，可以执行一些其它处理过程，以便进一步确定待处理短文本的情感倾向。此过程不是本申请的重点，在此不再赘述。

在图 2a 和图 2b 所示的系统中还可以包括与处理器相连的接收设备（图中未示出）。在处理器确定待处理短文本的情感倾向后，处理器，还用于

输出所述待处理文本的情感倾向；所述接收设备，用于接收所述待处理文本的情感倾向，以便接收设备可以利用待处理文本的情感倾向。

第二种：

在情感度估测模型采用第二种实现方式实现的情况下，处理器 200 采用以下方式来确定待处理的短文本的情感倾向。参见图 10，本申请一种情感倾向的识别方法，具体包括以下步骤：

步骤 S1001：确定待处理短文本对应的特征集合和类目标识。

假设第二种实现方式在确定情感度估测模型的过程中，采用第一种执行方式确定短文本的特征集合；则在本步骤中也采用第一种执行方式确定待处理短文本特征集合。

参见图 11a，为确定待处理短文本的特征集合的第一种执行方式的具体过程：

步骤 1101：获取所述待处理短文本执行分词操作后获得的分词结果。

步骤 1102：利用 n 元语言模型对各个分词进行分词组合，获得若干个组合分词。

步骤 1103：将各个分词和若干个组合分词的集合，确定为所述待处理短文本的特征集合，一个分词对应一个特征。

在图 11a 的执行过程与图 6a 的执行过程类似，具体执行过程可参见图 6a 的执行过程，在此不再赘述。

假设第二种实现方式在确定情感度估测模型的过程中，采用第二种执行方式确定短文本的特征集合；则在本步骤中也采用第二种执行方式确定待处理短文本特征集合。

参见图 11b，为确定待处理短文本的特征集合的第二种执行方式的具体过程：

步骤 1111：获取所述待处理短文本执行分词操作后获得的分词结果。

步骤 1112：将所述分词结果，确定为所述待处理短文本的特征集合，一个分词对应一个特征。

在图 11b 的执行过程与图 6b 的执行过程类似，具体执行过程可参见图 6a 的执行过程，在此不再赘述。

接着返回图 10，进入步骤 S1002：依据与所述类目标识对应的情感度估测模型，结合待处理短文本的特征集合，对待处理短文本进行情感度估测；其中，所述情感度估测模型为：依据所述类目标识对应的、带有情感倾向的若干个短文本样本的特征集合训练后得到的、输出正面情感度和负面情感度的模型。

在第二种实现方式中，具有多个情感度估测模型。为了获得适用于待处理短文本的情感度估测模型，可以依据类目标识在多个情感度估测模型进行查找，从而确定与类目标识对应的情感度估测模型。

处理器将所述特征集合输入至所述情感度估测模型，由所述情感度估测模型估算后输出所述特征集合对应的正面情感度和负面情感度。

步骤 S1003：基于所述待处理短文本对应的正面情感度和负面情感度，确定所述待处理短文本对应的情感倾向。本步骤的执行过程与图 7 的步骤 703 的执行过程一致，在此不再赘述。

在图 2a 和图 2b 所示的系统中，还可以包括与处理器相连的接收设备（图中未示出）。在处理器确定所述待处理短文本对应的情感倾向后，处理器，还用于输出所述待处理文本的情感倾向；所述接收设备，用于接收所述待处理文本的情感倾向。

在情感度估测模型采用第三种实现方式实现的情况下，处理器 200 会预先存储类目标识与情感度估测模型的对应关系，并预先构建每个类目标识和情感度估测模型的构建方式的对应关系。

若处理器 200 接收到一个类目标识后，首先判断与类目标识对应的情感度估测模型的构建方式；

若情感度估测模型采用第一种实现方式构建，则适应性的按图 4 所示的过程确定待处理短文本的情感倾向；即：确定待处理短文本对应的特征集合；其中，所述特征集合中每个特征包括：所述待处理短文本的分词和所述待处理短文本所属的类目标识；依据预先训练的情感度估测模型，结合待处理短文本的特征集合，对待处理短文本进行情感度估测；其中，所述情感度估测模型包括：依据至少两种类目的、带有情感倾向的若干个短文本样本训练后

得到的、输出正面情感度和负面情感度的模型；基于所述待处理短文本对应的正面情感度和负面情感度，确定所述待处理短文本对应的情感倾向。

若情感度估测模型采用第二种实现方式构建，则按适应性的按图 5 所示的过程确定待处理短文本的情感倾向。即：确定待处理短文本对应的特征集合；其中，所述特征集合中每个特征包括：所述待处理短文本的分词；依据与所述类目标识对应的情感度估测模型，结合待处理短文本的特征集合，对待处理短文本进行情感度估测；其中，所述情感度估测模型为：依据所述类目标识对应的、带有情感倾向的若干个短文本样本训练后得到的、输出正面情感度和负面情感度的模型；基于所述待处理短文本对应的正面情感度和负面情感度，确定所述待处理短文本对应的情感倾向。通过图 7 和图 10 所示的实施例，可以看出本申请具有以下有益效果：

本申请提供一种情感倾向的识别方法，本方法利用若干个带情感倾向的短文本进行训练，并获得情感度估测模型。由于每个特征集合包含短文本的分词和类目标识，所以，申请构建的情感度估测模型充分考虑了短文本所属的类目。因此，基于情感度估测模型确定出的待处理短文本的正面情感度和负面情感度，相对于现有技术而言更加准确。进而，利用正面情感度和负面情感度确定出的情感倾向也更加准确。

下面以最大熵模型为例，对本申请构建情感度估测模型的训练过程进行详细介绍：

首先构建两个矩阵：矩阵 A 和矩阵 B，矩阵 A 中包含各个特征和各个特征对应的正面情感度和负面情感度。矩阵 B 包含两个分类结果：正面情感和负面情感。对于矩阵 A 中的任意个特征 a，采用 b 表示其情感倾向。 $f_i(a,b)$ 表示 (a,b) 共同出现情况。

首先计算 $f_i(a,b)$ 在训练样本中的期望，由于训练模型中没有变量，所以在计算完毕后该期望值为一个常数。具体计算公式如下所示：

$$E_{\bar{p}}(f_i) = \sum_{A,B} \hat{p}(a,b) f_i(a,b) \dots \dots \dots \text{公式 (1)}$$

其中， $E_{\bar{p}}(f_i)$ 表示 $f_i(a,b)$ 在训练样本 i 中的期望， $\hat{p}(a,b)$ 表示 $f_i(a,b)$ 在训练样本的经验概率分布。

$f_i(a,b)$ 在模型中的概率分布的公式如下:

$$\hat{p}(a|b) = \arg \max_{p \in P} (-\sum_a \hat{p}(b)p(a|b) \log p(a|b)) \dots \dots \dots \text{公式 (2)}$$

其中, $\hat{p}(b)$ 表示训练样本中短文本对应的情感倾向是 b 的概率, $p(a|b)$ 表示短文本的情感倾向是 b 的前提下, 特征 a 的条件概率。

则 $f_i(a,b)$ 在最大熵模型中的计算公式为:

$$E_p(f_i) = \sum_{A,B} \hat{p}(a)p(b|a)f_i(a,b) \dots \dots \dots \text{公式 (3)}$$

在最大熵模型中, $f_i(a,b)$ 在训练样本中的期望, 与 $f_i(a,b)$ 在模型中的期望应该是一致的。即:

$$E_{\hat{p}}(f_i) = E_p(f_i) \dots \dots \dots \text{公式 (4)}$$

采用拉格朗日乘子法, 在满足公式 (4) 的约束条件下求解目标方程 (2) 的最优解, 最优解如下所示:

$$\hat{p}(a|b) = \frac{1}{Z(b)} \exp(\sum_{i=1}^l w_i \bullet f_i(a,b)) \dots \dots \dots (5)$$

其中, $Z(b) = \sum_A \exp(\sum_{i=1}^l w_i f_i(a,b))$ 为归一化因子, 使得 $\sum_a \hat{p}(a,b) = 1$, w_i 为特征 f_i 的权重。

将公式 (5) 代入到公式 (1) 中从而获得最大熵模型的训练的结果, 也即情感度估测模型。

如图 12 所示, 本申请提供了一种对象分类方法。应用于处理器中, 在本实施例, 可以直接利用待处理对象的短文本的情感倾向来对对象进行分类。具体包括以下步骤:

步骤 S1201: 确定待处理对象的短文本信息, 其中, 所述短文本信息包括短文本的情感倾向。

处理器可以利用标点符号将待处理对象分为若干个短文本, 每个短文本可以按照本申请图 7 或图 10 提供的过程确定其情感倾向, 从而可以确定出待处理对象中每个短文本的情感倾向。此外, 短文本信息还可以包括: 待处理对象中属于正面情感的短文本数量、属于负面情感的短文本数量、正面短文本的所占比例、负面短文本的所占比例等等。

步骤 S1202: 依据预先训练的类别识别模型, 对所述短文本信息进行类别识别; 其中, 所述类别识别特征模型为: 依据若干对象的短文本信息训练得到的、第一类别和第二类别的分类器。

类别识别模型为预先利用若干个对象的短文本信息训练后, 得到的输出第一类别和第二类别的分类器。具体而言, 可以利用最大熵模型、神经网络算法或者支持向量机等分类模型, 对若干个对象的短文本信息进行训练, 从而获得类别识别模型。相关技术手段, 可以采用现有技术中的训练方式, 在此不再赘述。

在获得待处理对象的短文本信息后, 将待处理对象的短文本输入至类别识别模型, 类别识别模型处理后, 可以确定待处理对象的类别。

在实际过程中发现, 针对一个对象而言, 对象除了包括文本之外还可以包括图像。以对象为电商系统的用户评价为例, 用户评价中除了具有文本(字符用户评价)之外, 还可以具有商品的图像。

可以理解的是, 单独通过对象的短文本信息确定出的对象类别不准确, 因为并没有考虑到对象的图像特征信息; 同理, 单独采用对象的图像特征信息确定出的对象类别也不准确, 因为并没有考虑到对象的短文本信息。因此, 本实施例将短文本信息和图像特征信息进行合并, 采用短文本信息和图像特征信息一并确定对象类别, 从而提高对象类别的准确率。

本申请又提供了一种对象分类方法, 在本实施例中利用待处理对象的多个特征来对对象进行分类。如图 13 所示, 具体包括以下步骤:

步骤 S1301: 确定与待处理对象对应的特征信息; 其中, 所述特征信息包括短文本信息和图像特征信息, 并且, 所述短文本信息包括短文本的情感倾向。

处理器可以利用标点符号将待处理对象分为若干个短文本, 每个短文本可以按照本申请图 7 或图 10 提供的过程确定其情感倾向, 从而可以确定出待处理对象中每个短文本的情感倾向。此外, 短文本信息还可以包括: 待处理对象中属于正面情感的短文本数量、属于负面情感的短文本数量、正面短文本的所占比例、负面短文本的所占比例等等。

处理器可以对图像进行处理，从而获得图像特征信息。图像特征信息可以包括下述图像特征中的一个或多个：图像宽度、图像高度、图像中人脸个数、图像包含的子图的个数、图像的背景是否是纯色、图像包含文字区域占比是多少、图像显著区域主颜色个数、图像主颜色个数、图像牛皮癣分数、图像主体质量分数、图像是假人模特的概率得分、图像中是真人模特的概率得分、图像展示的是商品细节的概率得分等等。

步骤 S1302：依据预先训练的类别识别模型，对所述特征信息进行类别识别；其中，所述类别识别特征模型为：依据若干对象的特征信息训练得到的、第一类别和第二类别的分类器。

类别识别模型为预先利用若干个对象的短文本信息和图像特征信息训练后，得到的输出第一类别和第二类别的分类器。具体而言，可以利用最大熵模型、神经网络算法或者支持向量机等分类模型，对若干个对象的短文本信息进行训练，从而获得类别识别模型。相关技术手段，可以采用现有技术中的训练方式，在此不再赘述。

在获得待处理对象的短文本信息后，将待处理对象的短文本发送至类别识别模型，从而确定待处理对象的类别。

可以理解的是，待处理对象的特征信息中的特征种类越多，则最终获得的结果越准确。所以，为了进一步提高待处理对象的类别的准确率，特征信息还可以包括：所述待处理对象所附属于第一主体的特征信息；和/或，所述待处理对象所附属于第二主体的特征信息。当然还可以包括其它特征信息，在此不再一一列举。

例如，以用户评价为例，所述待处理对象所附属于第一主体的特征信息具体为：商品的所附属于卖家（第一主体）特征信息，例如，卖家的信用等级、卖家的销售量等。所述待处理对象所附属于第二主体的特征信息具体为：商品的所附属于买家（第二主体）特征信息，例如，买家的信用等级、发布非默认的用户评价数据量、发布带图的用户评价数量、发布带图的用户评价占比。

在特征信息中增加短文本信息、图像特征信息以及其它特征信息后，对象的特征信息便会具有多个特征信息。为了综合考虑多个特征信息，本实施

例提出采用梯度提升决策树模型对若干个训练样本进行训练，从而获得类别识别模型。

梯度提升决策树模型是以决策树为基函数的提升方法。梯度提升决策树模型包括多棵决策树，之所以采用多棵决策树是考虑对于单棵决策树会因为过度分裂而造成过拟合，失去泛化能力；如果分裂太少，又会造成学习不够充分。

下面介绍梯度提升决策树模型的训练过程：

第一，估计初值 F_0 。

初值 F_0 可以是一个随机的数值，也可以等于 0，具体数值可以根据实际情况而定，在此不做限定。

第二，按照下述方式迭代 M 次，获得 M 棵决策树

A) 利用上一梯度提升决策树更新全部训练样本对应多个特征信息的估计值。

B) 从所有训练样本中随机选择部分样本，作为本次构建决策树的训练样本。

C) 根据样本所包含的特征，计算每种特征的信息增益，选择信息增益最大的特征进行第一次划分，左侧代表第一类别，右侧代表第二类别。计算本次的梯度，结合梯度重新估计样本的特征信息的特征值。

将上段步骤重复 J 次，得到 J 层叶子节点的决策树。

D) 根据获得 M 棵决策树，计算训练样本在该棵决策树上的准确率，将准确率作为该棵决策树的权重。

第三，将 M 棵决策树进行线性组合，得到最终的梯度提升决策树模型。

梯度提升决策树模型包括多棵决策树，可以表示为多棵决策树的加法模型：
$$F(X) = F_0 + \beta_1 T_1(X) + \beta_2 T_2(X) + \cdots + \beta_i T_i(X) + \cdots + \beta_M T_M(X) \dots \dots \text{公式 (6)}$$

其中， F_0 是一个初值， $T_i(X)$ 表示待处理对象的特征信息与一个决策树的匹配度， β_i 表示一个决策树的权重， M 表示决策树的总数量。

梯度提升决策树模型使用多棵决策树正是希望能够在训练精度和泛化能力两个方面都达到较好的结果。梯度提升决策树模型作为一种 boosting 算法，梯度提升决策树模型自然包含 boosting 的思想：将一系列弱分类器组合起来，

构成一个强分类器。它不要求每棵决策树学到太多的东西，每颗树都学一点知识，然后将每个决策树学到的知识累加起来构成一个强大的模型。

本申请又提供了一种对象分类方法，如图 14 所示，具体包括以下步骤：

步骤 S1401：确定与待处理对象对应的特征信息。

其中，所述特征信息包括短文本信息、图像特征信息、待处理对象所附属于第一主体的特征信息、所述待处理对象所附属于第二主体的特征信息。并且，所述短文本信息包括短文本的情感倾向。

以对象对用户评价为例，则本步骤可以为：确定待处理用户评价的特征信息；其中，所述特征信息包括用户评价的文本特征信息、用户评价的图像特征信息、卖家的特征信息和买家的特征信息，并且，所述文本特征信息包括短文本的情感倾向。

步骤 S1402：将所述特征信息与预先训练的梯度提升决策树模型进行识别。

继续以对象为用户评价为例，则本步骤为依据预先训练的梯度提升决策树模型，对所述待处理用户评价的特征信息进行类别识别；其中，所述类别识别模型为：依据若干用户评价样本的特征信息训练后得到的、第一类用户评价和第二类用户评价的分类器。

如图 15 所示，具体而言本步骤包括以下步骤：

步骤 S1501：将所述特征信息输入至所述类别识别模型，也即梯度提升决策树模型。

梯度提升决策树模型有 M 棵树，将特征信息分别与 M 棵树进行匹配，从而获得与每棵树匹配后确定的类别。

步骤 S1502：确定所述待处理对象对应的第一类别匹配度和第二类别匹配度。

按上述公式 6 确定第一类别匹配度和第二类别匹配度。

第一类别匹配度 $F_1(X) = F_0 + \beta_1 T_1(X) + \beta_2 T_2(X) + \cdots + \beta_i T_i(X) + \cdots + \beta_M T_M(X)$ 。其中， $T_i(X)$ 表示特征信息与一棵树的匹配度， β_i 表示该树对应的权重。若一棵树确定特征信息对应第一类别，则权重为 β_i ；若一棵树确定特征信息对应第二类别，则权重为 0。

第二类别匹配度 $F_2(X) = F_0 + \beta_1 T_1(X) + \beta_2 T_2(X) + \cdots + \beta_i T_i(X) + \cdots + \beta_M T_M(X)$ 。其中， $T_i(X)$ 表示特征信息与一棵树的匹配度， β_i 表示该树对应的权重。若一棵树确定特征信息对应第二类别，则权重为 β_i ；若一棵树确定特征信息对应第一类别，则权重为 0。

步骤 S1503: 对所述第一类别匹配度和第二类别匹配度进行比较。若第一类别匹配度大于第二类别匹配度，进入步骤 S1504；若第二类别匹配度大于第一类别匹配度，则进入步骤 S1505。

步骤 S1504: 确定所述待处理对象的类别为第一类别。

继续以对象为用户评价为例，则本步骤为确定待处理用户评价的类别为第一类别。第一类别为优质用户评价，那么本步骤即为确定待处理用户评价的类别为优质用户评价。步骤 S1505: 确定所述待处理对象的类别为第二类别。

继续以对象为用户评价为例，则本步骤为确定待处理用户评价的类别为第二类别。第二类别为劣质用户评价，那么本步骤即为确定待处理用户评价的类别为劣质用户评价。

在确定所述待处理对象为第一类别之后，将所述待处理对象添加至对象集合中；发送所述对象集合中的对象。对象集合可以被其它设备使用，在使用过程中，可以再次经过筛选确定出多个更优的对象样本，然后将对象样本再发送至处理器，以便处理器利用更优的对象样本，重新训练类别识别模型，以便类别识别模型更加准确。即，处理器可以接收多个对象样本，所述对象样本来源于所述对象集合；将所述多个对象样本，添加至训练类别识别模型的已有对象样本中；基于更新后的已有对象样本，重新训练类别识别模型。

继续以对象为用户评价为例，则本过程为：在确定所述待处理用户评价为第一类用户评价之后，将所述待处理用户评价添加至第一类用户评价集合中；发送所述第一类用户评价集合。第一用户评价集合可以对用户进行使用，在使用过程中可以在第一类用户评价集合中确定出更优的用户评价。然后，可以将更优的用户评价发送至处理设备，以便处理设备重新训练类别识别模型。即本系统可以形成闭环系统。

即，处理器接收多个第一类用户评价，所述第一类用户评价来源于所述第一类用户评价集合；将所述多个第一类用户评价，添加至类别识别模型已

有的用户评价样本中；基于更新后的已有的用户评价样本，重新训练类别识别模型。

参见图 16，本申请提供了一种对象分类系统，包括：

数据提供设备 100，用于发送若干个对象。

处理器 200，用于接收所述数据提供设备送的若干个对象，依据若干对象的特征信息训练后得到、输出第一类别和第二类别的类别识别模型；用于确定待处理对象的特征信息；其中，所述特征信息包括文本特征信息和图像特征信息，并且，所述文本特征信息包括短文本的情感倾向；依据所述类别识别模型，对所述待处理对象的特征信息进行类别识别；还用于输出第一类别的对象。

数据接收设备 400，用于接收并使用所述第一类别的对象。

数据接收设备 400 在使用对象集合的过程中，可以再次经过筛选确定出多个更优的对象样本，然后将对象样本再发送至处理器 200，以便处理器利用更优的对象样本，重新训练类别识别模型，以便类别识别模型更加准确。

参见图 17，本申请还提供了一种对象分类系统，包括：

数据提供设备 100，用于发送若干个对象。

模型构建设备 300，用于接收所述数据提供设备送的若干个对象，依据若干个对象的特征信息训练后得到、输出第一类别和第二类别的类别识别模型，并发送所述类别识别模型。

处理器 200，用于接收所述类别识别模型，并确定待处理对象的特征信息；其中，所述特征信息包括文本特征信息和图像特征信息，并且，所述文本特征信息包括短文本的情感倾向；依据所述类别识别模型，对所述待处理对象的特征信息进行类别识别；还用于输出第一类别的对象。

数据接收设备 400，用于接收并使用所述第一类别的对象。

数据接收设备 400 在使用对象集合的过程中，可以再次经过筛选确定出多个更优的对象样本，然后将对象样本再发送至处理器 200，以便处理器利用更优的对象样本，重新训练类别识别模型，以便类别识别模型更加准确。

下面以一个具体场景实施例，来详细描述对象分类方法。

在电商系统中有很多用户评价，如何从众多用户评价中筛选出优质用户评价，是本实施例所要解决的问题。由于电商系统中用户评价数量和种类繁多，商家需要花费很多时间找出店铺中的优质用户评价，这无形中需要花费巨大的人力成本。目前在优质用户评价识别领域，工业界常用的技术主要有两种：第一种，基于短文本的识别技术；第二种，基于图像特征的识别技术。

基于短文本的识别技术相对比较容易实现，但是存在着一些局限性：不关注用户评价中买家发布的图像信息。在实际场景中，比如服饰类，用户不单单关心用户评价中的文字描述部分，还关心商品真实的样子，即图像特征信息。

基于图像特征的识别技术效果显著，但也有一定的局限性。基于图像特征的优质用户评价识别技术仅仅利用用户评价中的图像信息进行识别，并不关心已购买者具体购买后的心得体会，即短文本信息。因此，可以看出用户评价中的短文本信息和图像特征信息同样重要。

此外，申请人发现还有一些其它特征对确定优质用户评价，可以起到辅助作用。例如，卖家特征和买家特征。因此，本实施例将以上特征均作为确定用户评价为优质用户评价或劣质用户评价的依据。为此，本实施例提出基于多种特征融合的机器学习方法，即梯度提升决策树模型，来训练若干个训练样本，从而获得类别识别模型。

如图 18 所示，为本申请提供确定优质用户评价的流程图。从图中可以清晰地整个确定优质用户评价的过程。主要由三部分组成：

(1) 构建用户评价库

在用户评价服务器中获取大量的用户评价，首先利用预处理规则过滤掉一部分劣质用户评价。预处理规则可以为：优质用户评价中图像和文本所需要满足的一些要求，即使用短文本和图像特征中少量维度的特征对大量用户评价进行过滤。

具体而言为，优质用户评价中的短文本不能均为负面情感，基于此，若用户评价中的短文本均对应负面情感，则判定为非优质用户评价。对于优质用户评价中的图像也有基本要求，图像的分辨率达到预设分辨率、图像为非对话截屏、图像中的明显广告宣传语以及水印占比小于预设值，等等。

将用户评价服务器中满足上述短文本要求和图像特征要求的用户评价，将其放入用户评价库中。针对不满足短文本要求和图像特征要求的用户评价，则将这些用户评价判定为优质用户评价，不放入用户评价库中。

通过预处理规则的过滤可以过滤出一些非优质用户评价，这样不仅能够减少优质用户评价识别模型的使用次数，而且，还可以有效地过滤掉非优质用户评价，提升优质用户评价识别模型预测的准确率。

(2) 确定优质用户评价集合

利用优质用户评价识别模型对用户评价库中用户评价进行识别，若识别结果为优质用户评价，则放入到优质用户评价集合中。

(3) 使用优质用户评价集合。

数据接收设备可以从优质用户评价集合中获取优质用户评价，并在实际应用过程中使用优质评价。数据接收设备在使用优质用户评价集合中优质用户评价的过程中，会根据预先设定准则重新对优质评价集合中的优质用户评价进行筛选，从而筛选出符合预先设定准则的优质用户评价。然后，将符合预先设定准则的优质用户评价发送至处理器或模型构建设备，以便处理器或模型构建设备对优质用户评价识别模型进行迭代更新。

(4) 优质用户评价识别模型的迭代更新。

利用符合预先设定准则的优质用户评价，重新对优质用户评价识别模型进行训练，以便优质用户评价识别模型能够尽可能的输出满足用户需求的优质用户评价。

由于在优质用户评价集合中挑选出的优质用户评价，均满足卖家或运行人员的预设规则，所以将这些优质用户评价重新加入用户评价库中，重新对优质用户评价识别模型的更新优化，以便优质用户评价识别模型更好地识别出满足用户期望的优质用户评价。

基于上述过程可以发现：本实施例中用户可以不再需要从原始用户评价库中一条一条去筛选，只需要在优质用户评价集合中进行挑选就能快速期望的优质用户评价，有效地降低人力成本。与此同时，优质用户评价模型能够有效地利用商家提供的优质用户评价进行迭代更新，从而进一步识别出满足商家期望的优质用户评价。

本实施例方法所述的功能如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算设备可读取存储介质中。基于这样的理解，本申请实施例对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来，该软件产品存储在一个存储介质中，包括若干指令用以使得一台计算设备（可以是个人计算机，服务器，移动计算设备或者网络设备等等）执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（ROM, Read-Only Memory）、随机存取存储器（RAM, Random Access Memory）、磁碟或者光盘等各种可以存储程序代码的介质。

本说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其它实施例的不同之处，各个实施例之间相同或相似部分互相参见即可。

对所公开的实施例的上述说明，使本领域专业技术人员能够实现或使用本申请。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的，本文中所定义的一般原理可以在不脱离本申请的精神或范围的情况下，在其它实施例中实现。因此，本申请将不会被限制于本文所示的这些实施例，而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

权利要求书

1、一种情感倾向的识别方法，其特征在于，包括：

确定待处理短文本对应类目标识；其中，一个文本相邻两个标点符号之间文字称为短文本；

确定与所述类目标识对应的情感度估测模型的实现方式；

若所述情感度估测模型的实现方式为所有类目对应一个情感度估测模型，则确定待处理短文本对应的特征集合；其中，所述特征集合中每个特征包括：所述待处理短文本的分词和所述待处理短文本所属的类目标识；依据预先训练的情感度估测模型，结合待处理短文本的特征集合，对待处理短文本进行情感度估测；其中，所述情感度估测模型包括：依据至少两种类目的、带有情感倾向的若干个短文本样本训练后得到的、输出正面情感度和负面情感度的模型；基于所述待处理短文本对应的正面情感度和负面情感度，确定所述待处理短文本对应的情感倾向；

若所述情感度估测模型的实现方式为一个类目对应一个情感度估测模型，确定待处理短文本对应的特征集合；其中，所述特征集合中每个特征包括：所述待处理短文本的分词；依据与所述类目标识对应的情感度估测模型，结合待处理短文本的特征集合，对待处理短文本进行情感度估测；其中，所述情感度估测模型为：依据所述类目标识对应的、带有情感倾向的若干个短文本样本训练后得到的、输出正面情感度和负面情感度的模型；基于所述待处理短文本对应的正面情感度和负面情感度，确定所述待处理短文本对应的情感倾向。

2、如权利要求1所述的方法，其特征在于，在确定所述待处理短文本对应的情感倾向后，还包括：

输出所述待处理短文本对应的情感倾向。

3、一种情感倾向的识别方法，其特征在于，包括：

确定待处理短文本对应的特征集合；其中，一个文本相邻两个标点符号之间的文字称为短文本；所述特征集合中每个特征包括：所述待处理短文本的分词和所述待处理短文本所属的类目标识；

依据预先训练的情感度估测模型，结合待处理短文本的特征集合，对待处理短文本进行情感度估测；其中，所述情感度估测模型包括：依据至少两种类目的、带有情感倾向的若干个短文本样本训练后得到的、输出正面情感度和负面情感度的模型；

基于所述待处理短文本对应的正面情感度和负面情感度，确定所述待处理短文本对应的情感倾向。

4、如权利要求3所述的方法，其特征在于，所述确定待处理短文本对应的特征集合，包括：

获取所述待处理短文本对应的类目标识，以及所述待处理短文本执行分词操作后获得的分词结果；

将所述分词结果中的各个分词和所述类目标识进行组合，获得各个特征；
将各个特征的集合，确定为所述待处理短文本的特征集合。

5、如权利要求3所述的方法，其特征在于，所述确定待处理短文本对应的特征集合，包括：

获取所述待处理短文本对应的类目标识，以及所述待处理短文本执行分词操作后获得的分词结果；

将所述分词结果中的各个分词和所述类目标识进行组合，获得各个特征；
利用n元语言模型对所述各个特征进行特征组合，获得若干个组合特征；
将各个特征和所述若干个组合特征的集合，确定为所述待处理短文本的特征集合。

6、如权利要求5所述的方法，其特征在于，所述利用n元语言模型对所述各个特征进行特征组合，获得若干个组合特征，包括：

利用二元语言模型对所述各个特征进行特征组合，获得若干个组合特征。

7、如权利要求3所述的方法，其特征在于，所述依据预先训练的情感度估测模型，结合待处理短文本的特征集合，对待处理短文本进行情感度估测，包括：

将所述特征集合输入至所述情感度估测模型；

由所述情感度估测模型估算后，输出待处理短文本对应的正面情感度和负面情感度。

8、如权利要求 3 所述的方法，其特征在于，所述基于所述待处理短文本对应的正面情感度和负面情感度，确定所述待处理短文本对应的情感倾向，包括：

确定所述正面情感度和所述负面情感度两者中的较大情感度；

判断所述较大情感度是否大于预设置信度；

若所述较大情感度大于预设置信度，则确定所述待处理短文本对应的情感倾向与所述较大情感度的情感倾向一致。

9、如权利要求 3 所述的方法，其特征在于，所述情感度估测模型包括：

利用最大熵模型，依据至少两个类目标识对应的若干个短文本的特征集合训练后得到的、输出正面情感度和负面情感度的模型。

10、如权利要求 3 所述的方法，其特征在于，在确定所述待处理短文本对应的情感倾向后，还包括：

输出所述待处理短文本对应的情感倾向。

11、一种情感倾向的识别方法，其特征在于，包括：

确定待处理短文本对应的特征集合和类目标识；其中，一个文本相邻两个标点符号之间的文字称为短文本；所述特征集合中每个特征包括：所述待处理短文本的分词；

依据与所述类目标识对应的情感度估测模型，结合待处理短文本的特征集合，对待处理短文本进行情感度估测；其中，所述情感度估测模型为：依据所述类目标识对应的、带有情感倾向的若干个短文本样本训练后得到的、输出正面情感度和负面情感度的模型；

基于所述待处理短文本对应的正面情感度和负面情感度，确定所述待处理短文本对应的情感倾向。

12、如权利要求 11 所述的方法，其特征在于，所述确定待处理短文本对应的特征集合，包括：

获取所述待处理短文本执行分词操作后获得的分词结果；

利用 n 元语言模型对各个分词进行分词组合，获得若干个组合分词；

将各个分词和若干个组合分词的集合，确定为所述待处理短文本的特征集合，一个分词对应一个特征。

13、如权利要求 11 所述的方法，其特征在于，所述确定待处理短文本对应的特征集合，包括：

获取所述待处理短文本执行分词操作后获得的分词结果；

将所述分词结果，确定为所述待处理短文本的特征集合，一个分词对应一个特征。

14、如权利要求 11 所述的方法，其特征在于，在确定所述待处理短文本对应的情感倾向后，还包括：

输出所述待处理短文本对应的情感倾向。

15、一种情感倾向的识别系统，其特征在于，包括：

数据提供设备，用于发送若干个对象；

处理器，用于接收所述数据提供设备送的若干个对象，依据若干个对象的短文本构建情感度估测模型，并利用情感度估测模型确定待处理短文本的情感倾向。

16、如权利要求 15 所述的系统，其特征在于，

所述处理器，还用于构建情感度估测模型与对象所属的类目标识的对应关系。

17、如权利要求 15 所述的系统，其特征在于，所述系统还包括接收设备；

所述处理器，还用于输出所述待处理文本的情感倾向；

所述接收设备，用于接收所述待处理文本的情感倾向。

18、一种情感倾向的识别系统，其特征在于，包括：

数据提供设备，用于发送若干个对象；

模型构建设备，用于接收所述数据提供设备送的若干个对象，依据若干个对象的短文本构建情感度估测模型，并发送所述情感度估测模型；

处理器，用于接收所述情感度估测模型，并利用情感度估测模型确定待处理短文本的情感倾向。

19、如权利要求 18 所述的系统，其特征在于，

所述模型构建设备，还用于构建情感度估测模型与对象所属的类目标识的对应关系，并将对应关系发送至所述处理器。

20、如权利要求 18 所述的系统，其特征在于，所述系统还包括接收设备；

所述处理器，还用于输出所述待处理文本的情感倾向；

所述接收设备，用于接收所述待处理文本的情感倾向。

21、一种对象分类方法，其特征在于，包括：

确定待处理对象的特征信息；其中，所述特征信息包括文本特征信息和图像特征信息，并且，所述文本特征信息包括短文本的情感倾向；

依据预先训练的类别识别模型，对所述待处理对象的特征信息进行类别识别；其中，所述类别识别模型为：依据若干对象样本的特征信息训练后得到的、第一类别和第二类别的分类器。

22、如权利要求 21 所述的方法，其特征在于，所述特征信息还包括：

构建所述对象的第一主体的特征信息；和/或，

所述对象所附属于第二主体的特征信息。

23、如权利要求 21 所述的方法，其特征在于，所述依据预先训练的类别识别模型，对所述特征信息进行类别识别，包括：

将所述特征信息输入至所述类别识别模型；确定所述待处理对象对应的第一类别匹配度和第二类别匹配度；

对所述第一类别匹配度和第二类别匹配度进行比较；

若第一类别匹配度大于第二类别匹配度，则确定所述待处理对象的类别为第一类别；

若第二类别匹配度大于第一类别匹配度，则确定所述待处理对象的类别为第二类别。

24、如权利要求 23 所述的方法，其特征在于，还包括：

在确定所述待处理对象为第一类别之后，将所述待处理对象添加至对象集合中；

发送所述对象集合中的对象。

25、如权利要求 24 所述的方法，其特征在于，还包括：

接收多个对象样本，所述对象样本来源于所述对象集合，且，满足预设规则；

将所述多个对象样本，添加至训练类别识别模型的已有对象样本中；

基于更新后的已有对象样本，重新训练类别识别模型。

26、一种用户评价的分类方法，其特征在于，包括：

确定待处理用户评价的特征信息；其中，所述特征信息包括用户评价的文本特征信息、用户评价的图像特征信息、卖家的特征信息和买家的特征信息，并且，所述文本特征信息包括短文本的情感倾向；

依据预先训练的梯度提升决策树模型，对所述待处理用户评价的特征信息进行类别识别；其中，所述类别识别模型为：依据若干用户评价样本的特征信息训练后得到的、第一类用户评价和第二类用户评价的分类器。

27、如权利要求 26 所述的方法，其特征在于，还包括：

在确定所述待处理用户评价为第一类用户评价之后，将所述待处理用户评价添加至第一类用户评价集合中；

发送所述第一类用户评价集合。

28、如权利要求 26 所述的方法，其特征在于，还包括：

接收多个第一类用户评价，所述第一类用户评价来源于所述第一类用户评价集合；

将所述多个第一类用户评价，添加至类别识别模型已有的用户评价样本中；

基于更新后的已有的用户评价样本，重新训练类别识别模型。

29、一种对象分类系统，其特征在于，包括：

数据提供设备，用于发送若干个对象；

处理器，用于接收所述数据提供设备送的若干个对象，依据若干对象的特征信息训练后得到、输出第一类别和第二类别的类别识别模型；用于确定待处理对象的特征信息；其中，所述特征信息包括文本特征信息和图像特征信息，并且，所述文本特征信息包括短文本的情感倾向；依据所述类别识别模型，对所述待处理对象的特征信息进行类别识别；还用于输出第一类别的对象；

数据接收设备，用于接收并使用所述第一类别的对象。

30、一种对象分类系统，其特征在于，包括：

数据提供设备，用于发送若干个对象；

模型构建设备，用于接收所述数据提供设备送的若干个对象，依据若干个对象的特征信息训练后得到、输出第一类别和第二类别的类别识别模型，并发送所述类别识别模型；

处理器，用于接收所述类别识别模型，并确定待处理对象的特征信息；其中，所述特征信息包括文本特征信息和图像特征信息，并且，所述文本特征信息包括短文本的情感倾向；依据所述类别识别模型，对所述待处理对象的特征信息进行类别识别；还用于输出第一类别的对象；

数据接收设备，用于接收并使用所述第一类别的对象。

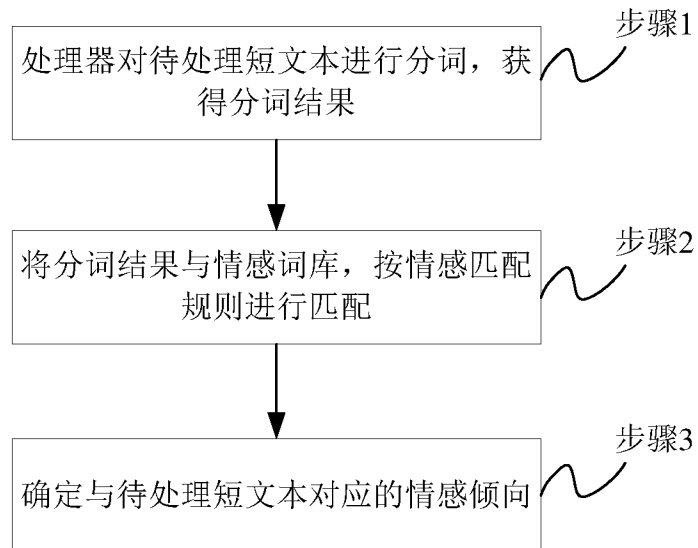


图 1

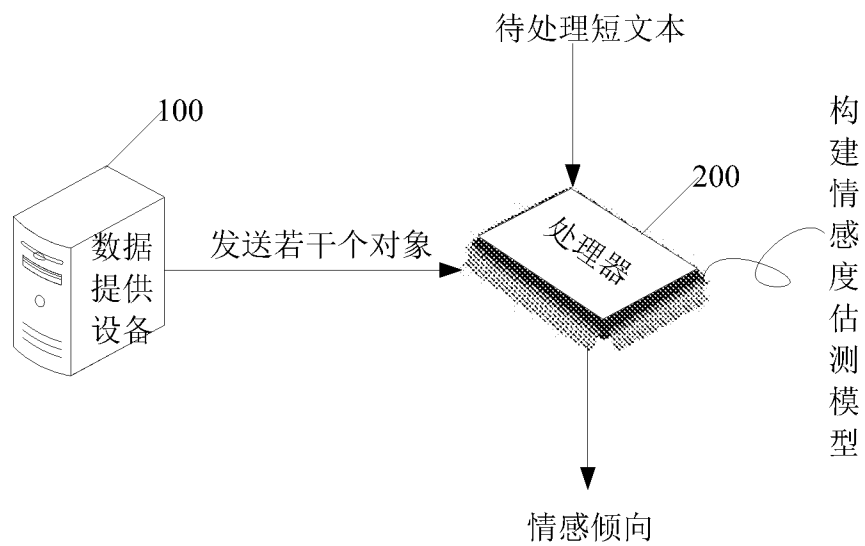


图 2a

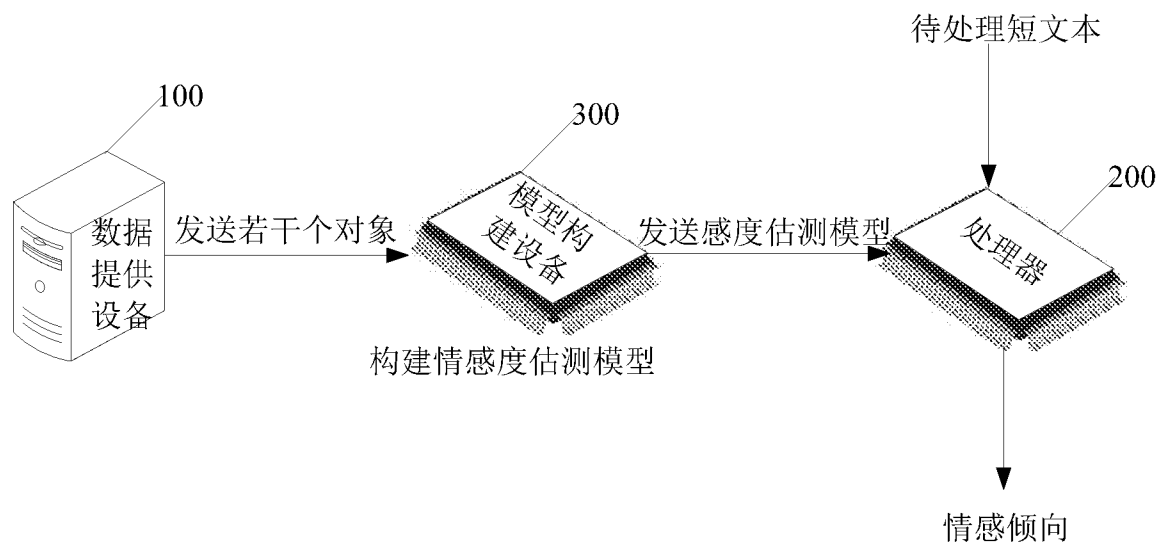


图 2b

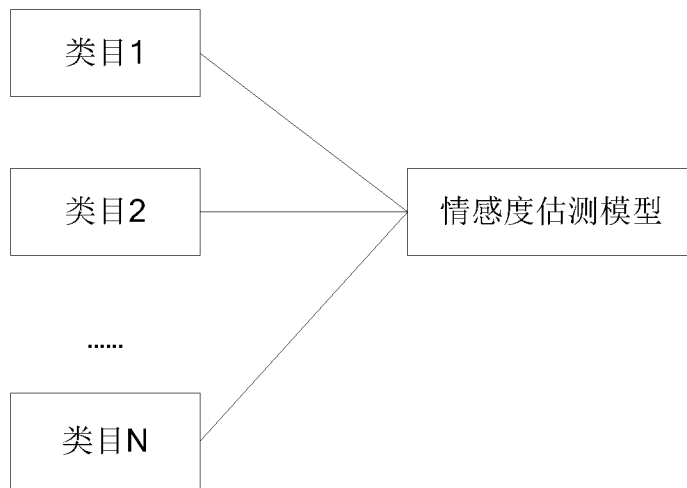


图 3a

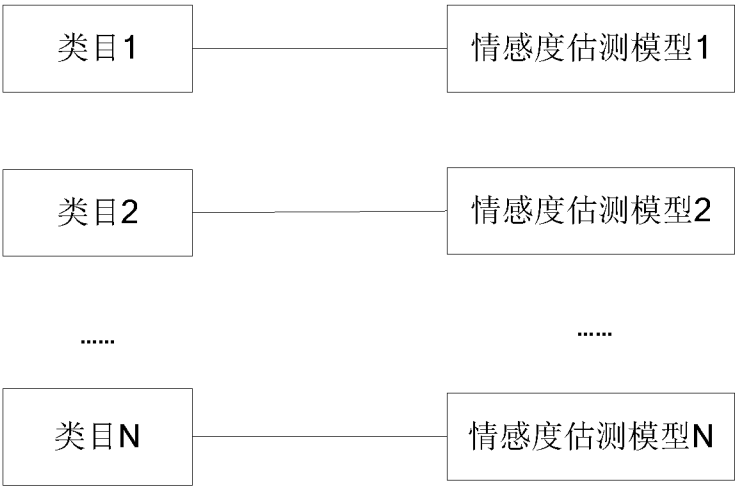


图 3b

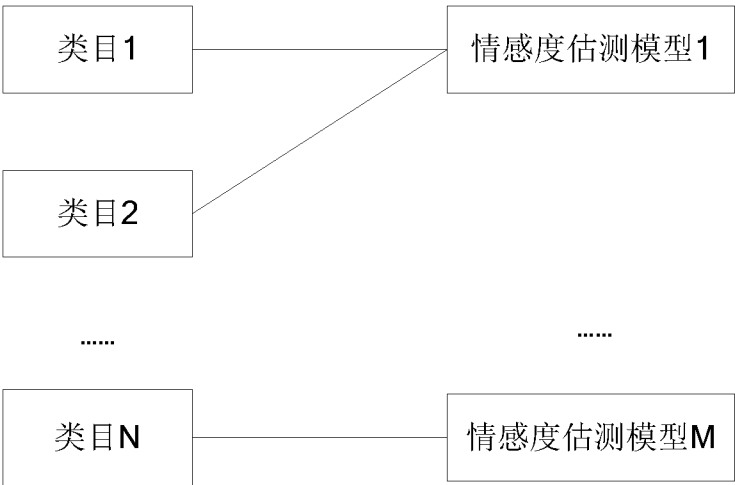


图 3c

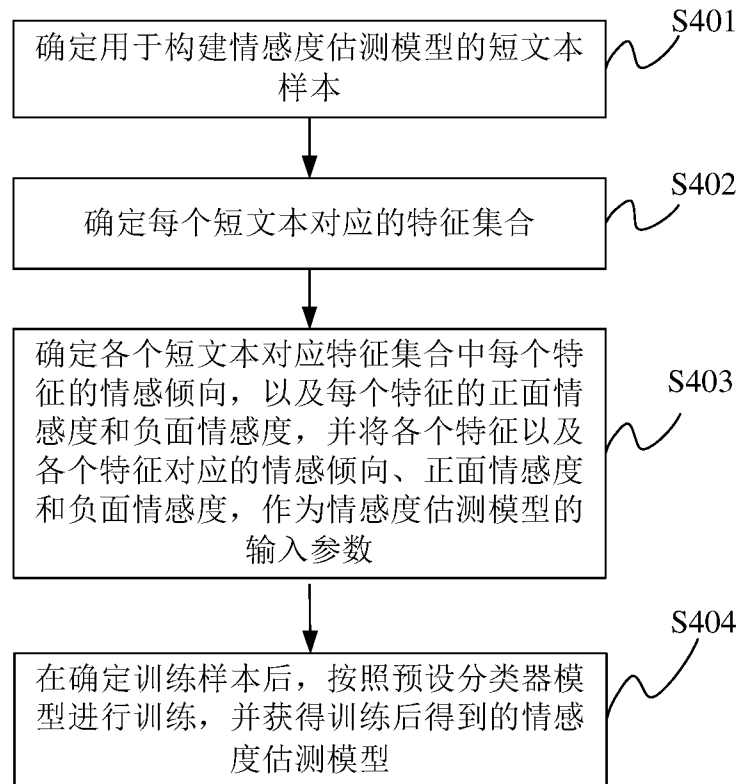


图 4a

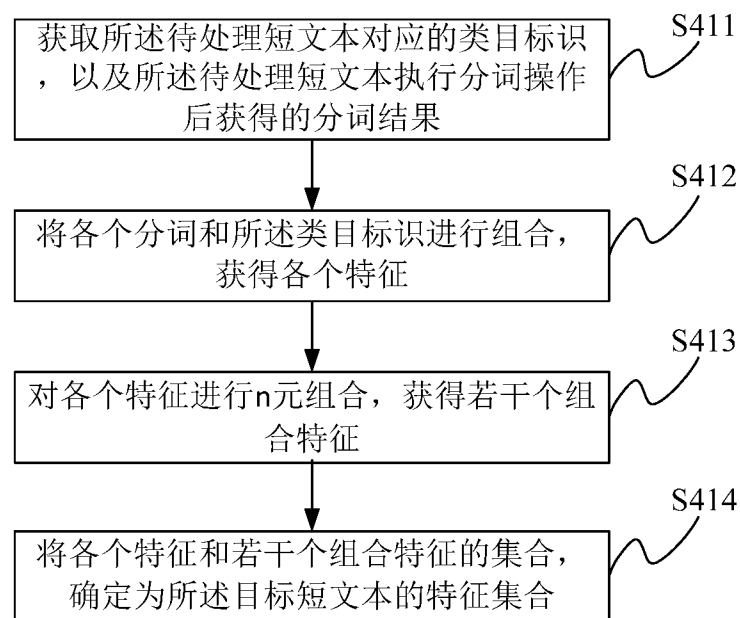


图 4b

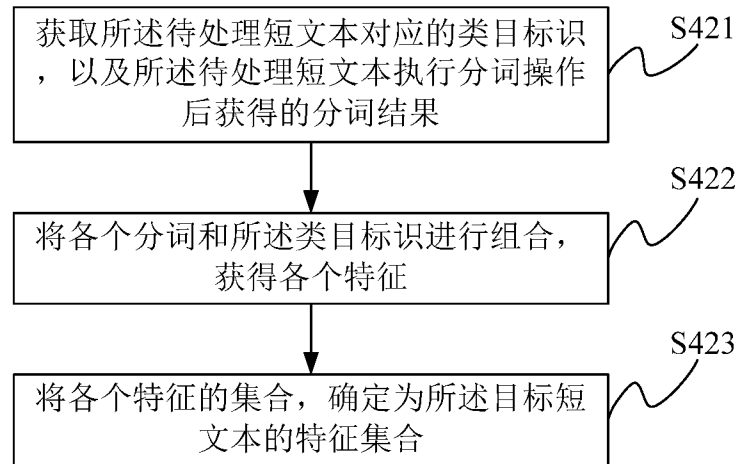


图 4c

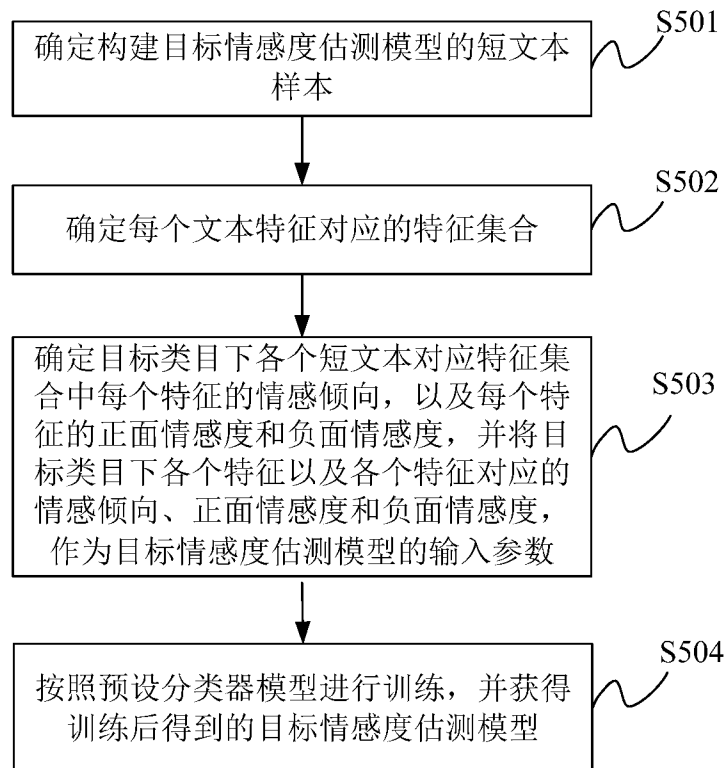


图 5

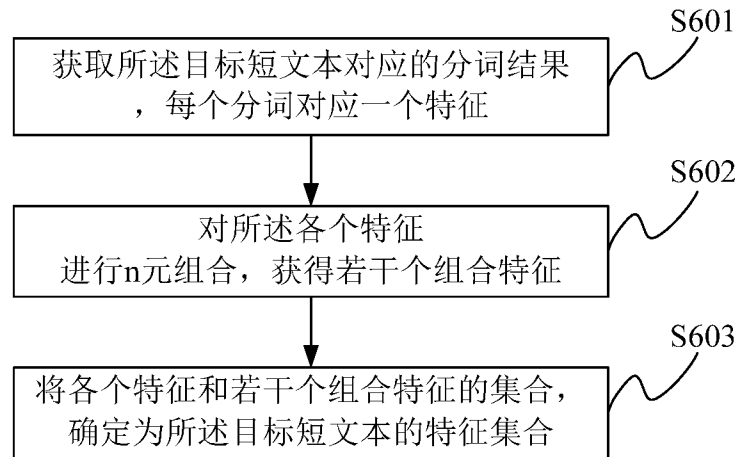


图 6a

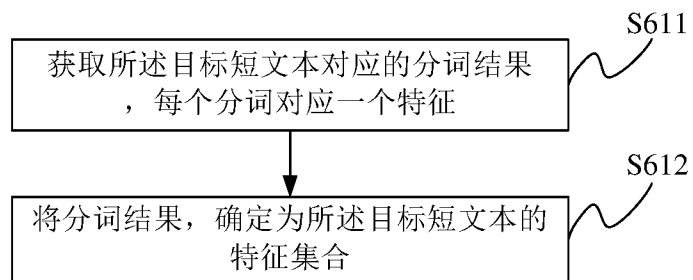


图 6b

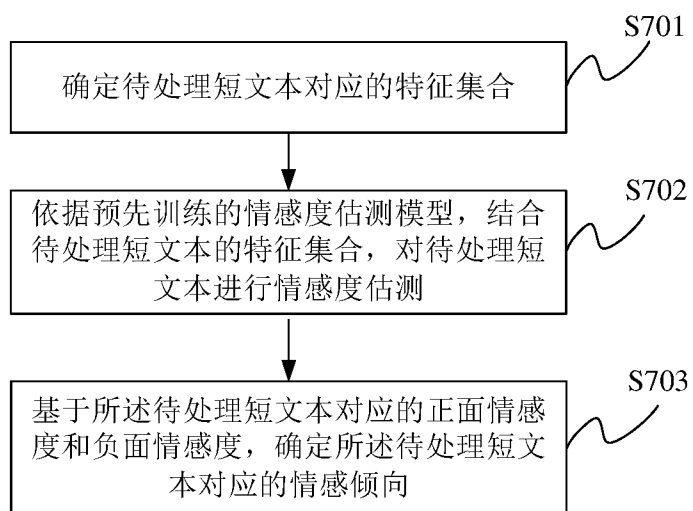


图 7

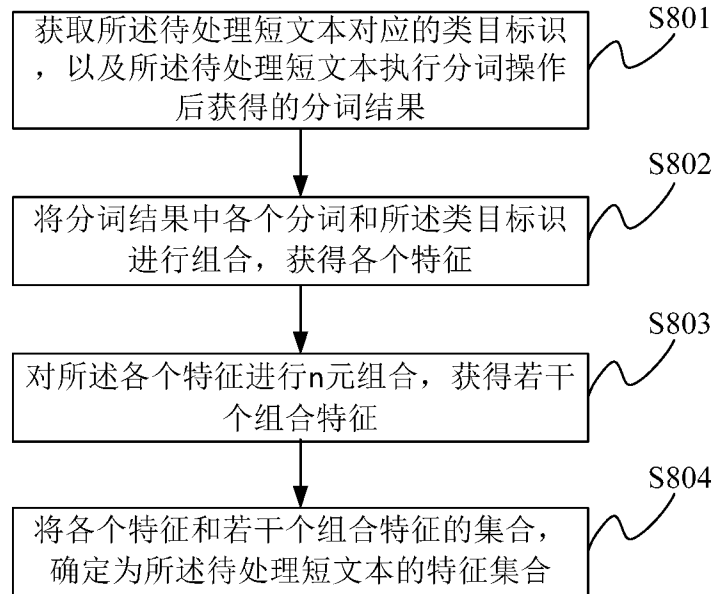


图 8a

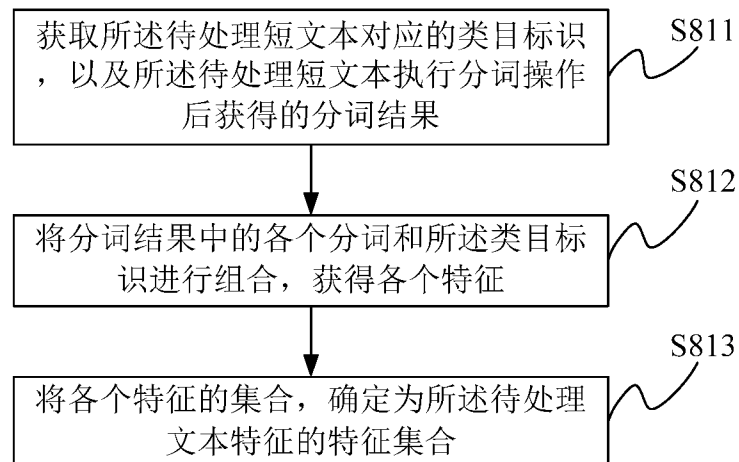


图 8b

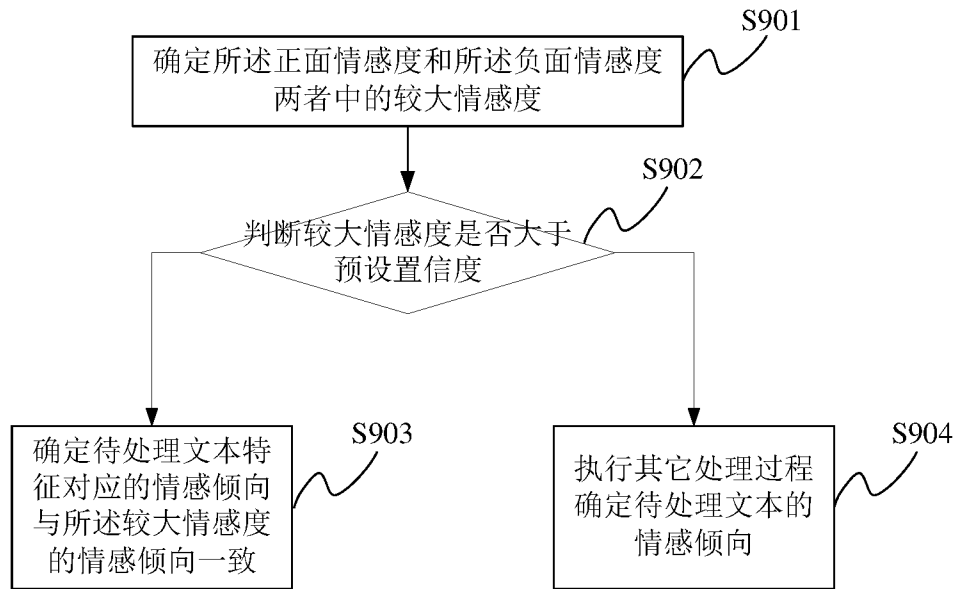


图 9

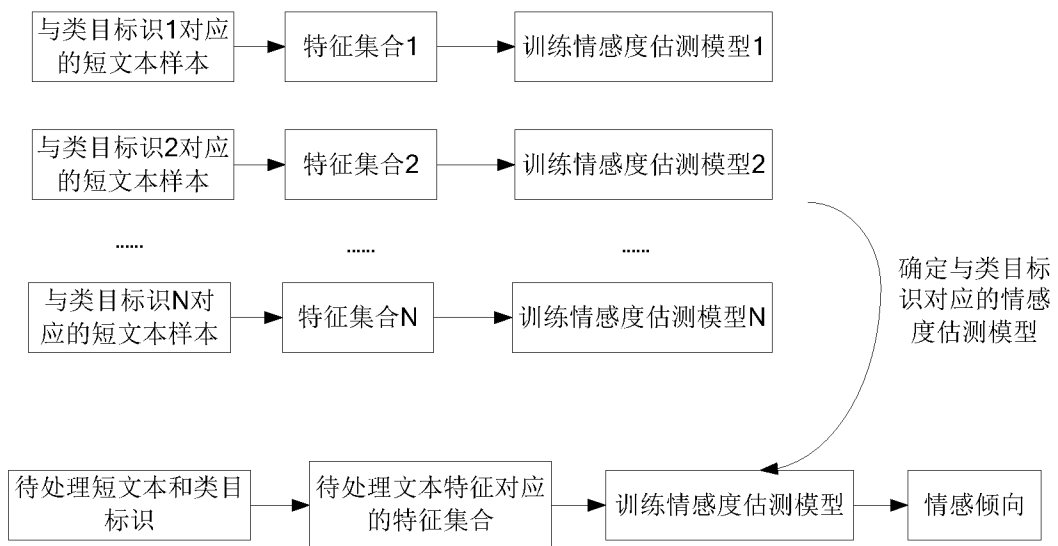


图 10

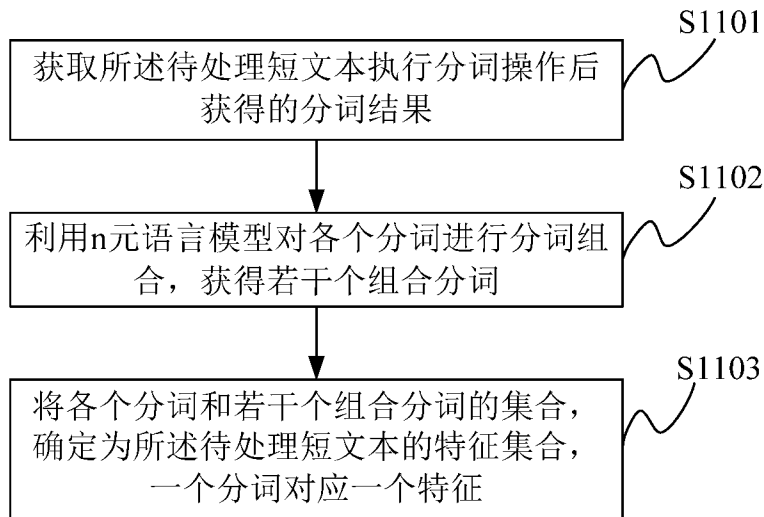


图 11a

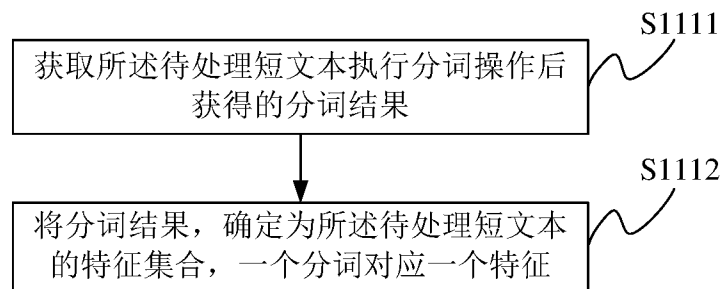


图 11b

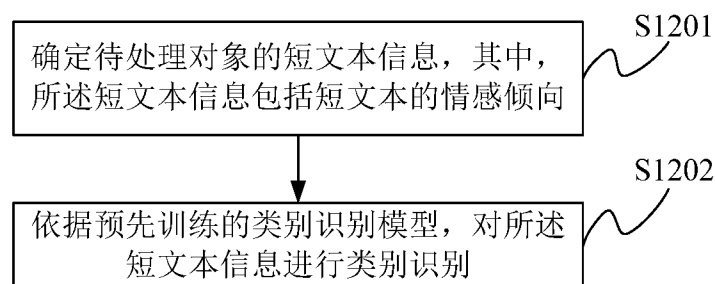


图 12

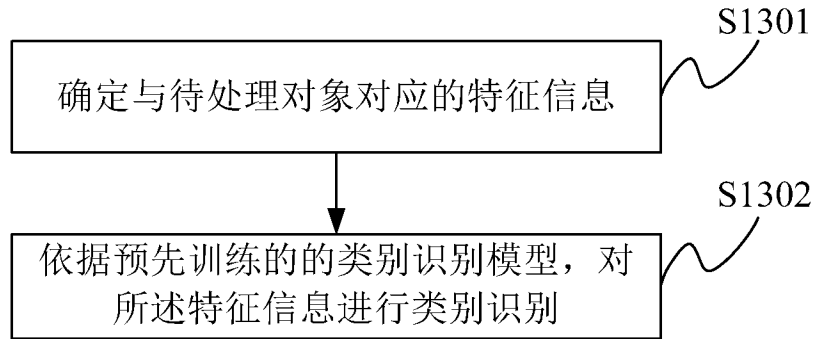


图 13

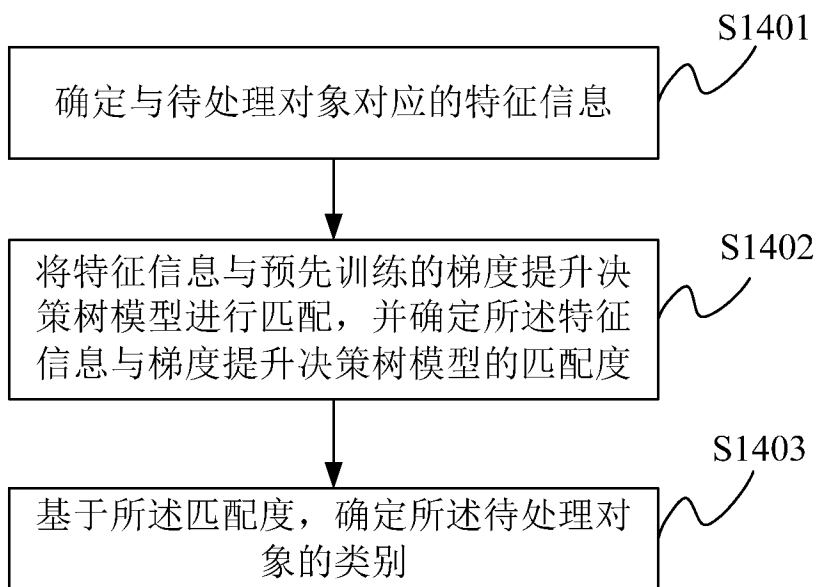


图 14

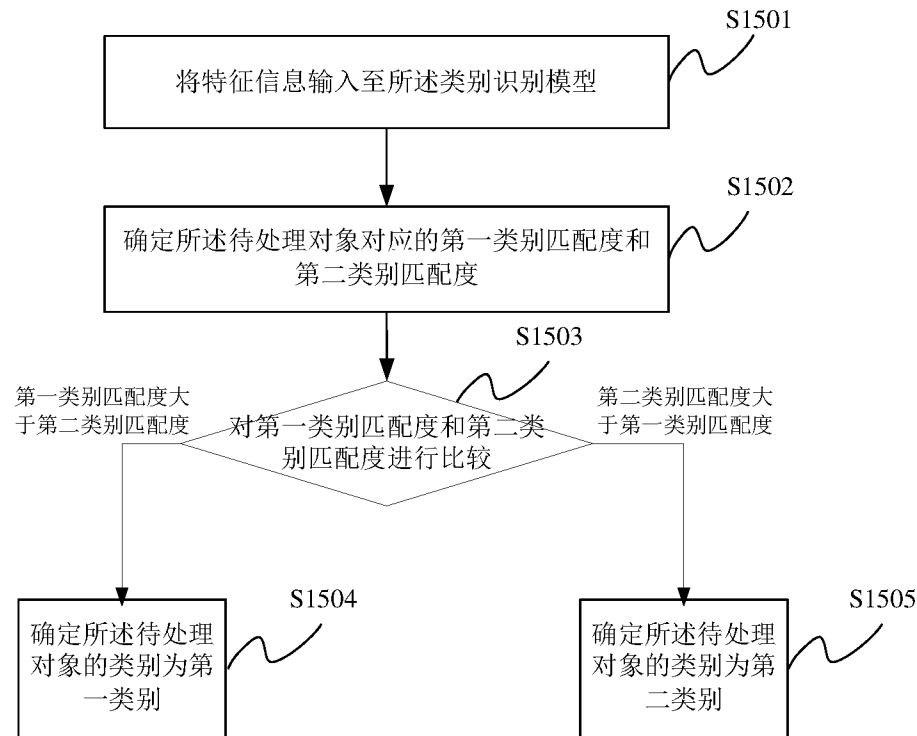


图 15

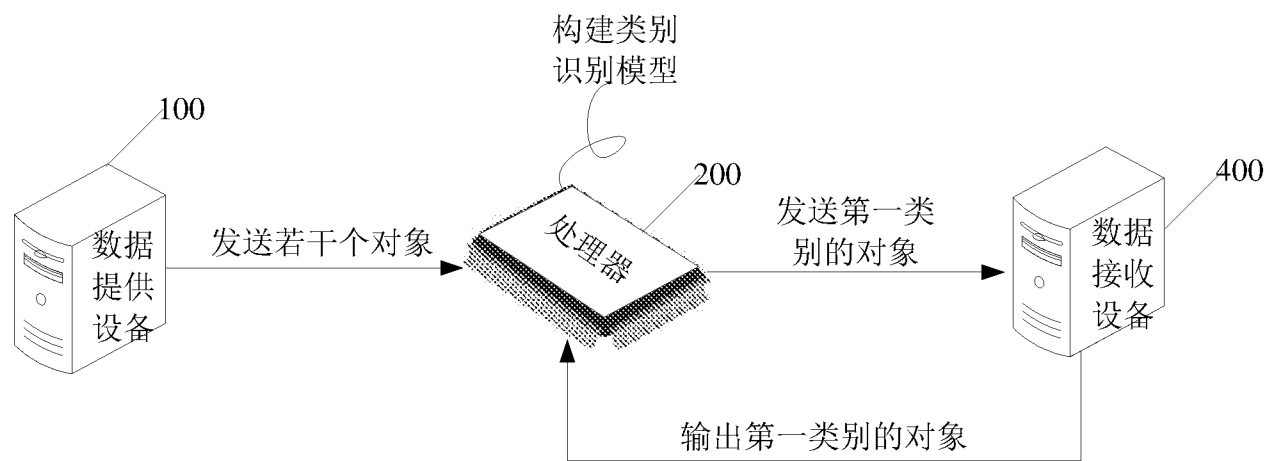


图 16

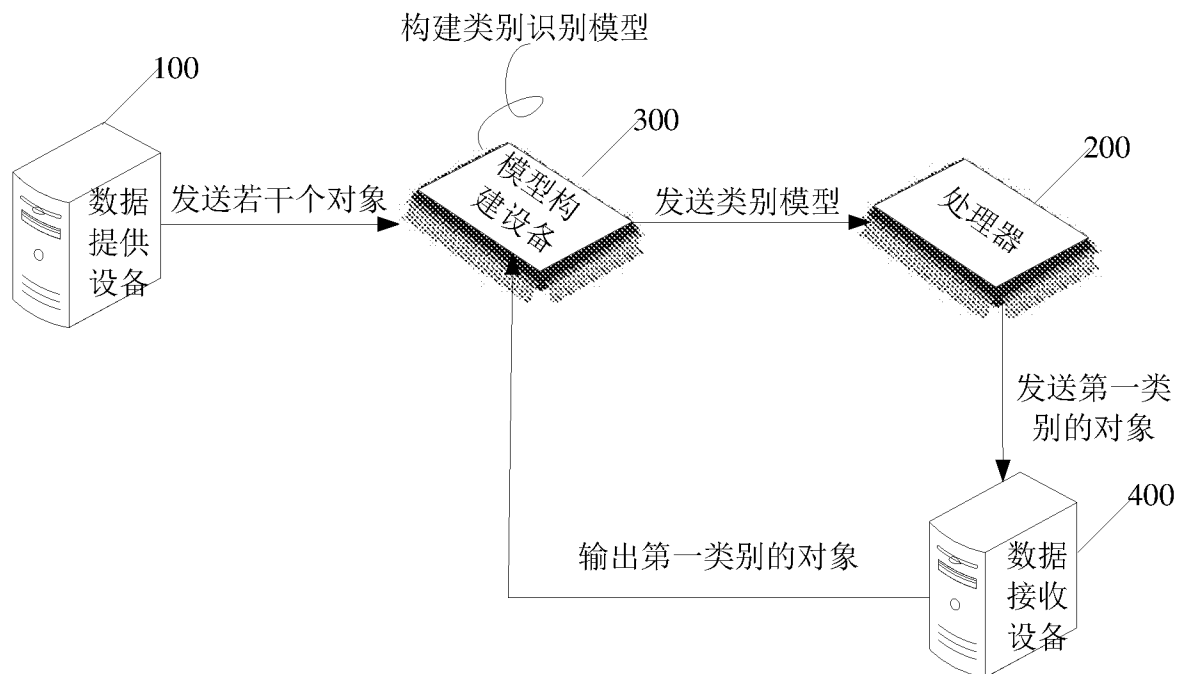


图 17

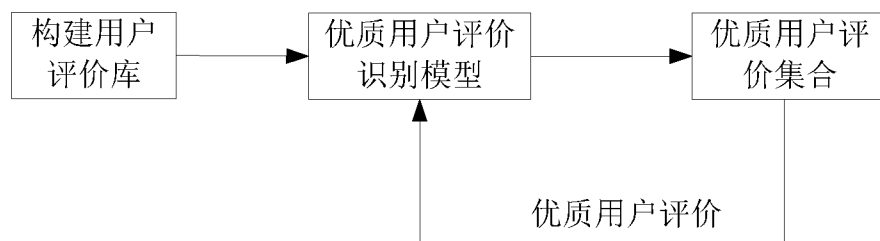


图 18

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/100060

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/27 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F 17/-

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, CNTXT, CNKI, DWPI, SIPOABS: 文本, 评价, 情感, 模型, 类目, 类别, 类型, 分类, 领域, text, evaluation, emotion, model, classif+, sort, type, field

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103365867 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.), 23 October 2013 (23.10.2013), description, paragraphs [0033]-[0148]	1-30
A	CN 103455562 A (XI'AN UNIVERSITY OF ARCHITECTURE AND TECHNOLOGY), 18 December 2013 (18.12.2013), entire document	1-30
A	CN 102968408 A (XIDIAN UNIVERSITY), 13 March 2013 (13.03.2013), entire document	1-30
A	CN 105005560 A (ZHANGJIAGANG INSTITUTE OF INDUSTRIAL TECHNOLOGIES SOOCHOW UNIVERSITY), 28 October 2015 (28.10.2015), entire document	1-30

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 21 November 2017	Date of mailing of the international search report 13 December 2017
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer ZHU, Xiaoli Telephone No. (86-10) 62411672

International application No.
PCT/CN2017/100060

Form PCT/ISA/210 (patent family annex) (July 2009)

国际检索报告

国际申请号

PCT/CN2017/100060

A. 主题的分类 G06F 17/27 (2006.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类																	
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06F 17/- 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNABS, CNTXT, CNKI, DWPI, SIPOABS: 文本, 评价, 情感, 模型, 类目, 类别, 类型, 分类, 领域, text, evaluation, emotion, model, classif+, sort, type, field																	
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 103365867 A (腾讯科技深圳有限公司) 2013年 10月 23日 (2013 - 10 - 23) 说明书第[0033]-[0148]段</td> <td>1-30</td> </tr> <tr> <td>A</td> <td>CN 103455562 A (西安建筑科技大学) 2013年 12月 18日 (2013 - 12 - 18) 全文</td> <td>1-30</td> </tr> <tr> <td>A</td> <td>CN 102968408 A (西安电子科技大学) 2013年 3月 13日 (2013 - 03 - 13) 全文</td> <td>1-30</td> </tr> <tr> <td>A</td> <td>CN 105005560 A (苏州大学张家港工业技术研究院) 2015年 10月 28日 (2015 - 10 - 28) 全文</td> <td>1-30</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 103365867 A (腾讯科技深圳有限公司) 2013年 10月 23日 (2013 - 10 - 23) 说明书第[0033]-[0148]段	1-30	A	CN 103455562 A (西安建筑科技大学) 2013年 12月 18日 (2013 - 12 - 18) 全文	1-30	A	CN 102968408 A (西安电子科技大学) 2013年 3月 13日 (2013 - 03 - 13) 全文	1-30	A	CN 105005560 A (苏州大学张家港工业技术研究院) 2015年 10月 28日 (2015 - 10 - 28) 全文	1-30
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
X	CN 103365867 A (腾讯科技深圳有限公司) 2013年 10月 23日 (2013 - 10 - 23) 说明书第[0033]-[0148]段	1-30															
A	CN 103455562 A (西安建筑科技大学) 2013年 12月 18日 (2013 - 12 - 18) 全文	1-30															
A	CN 102968408 A (西安电子科技大学) 2013年 3月 13日 (2013 - 03 - 13) 全文	1-30															
A	CN 105005560 A (苏州大学张家港工业技术研究院) 2015年 10月 28日 (2015 - 10 - 28) 全文	1-30															
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																	
<table border="0"> <tr> <td style="vertical-align: top;"> * 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 </td> <td style="vertical-align: top;"> “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件 </td> </tr> </table>			* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件													
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																
国际检索实际完成的日期 2017年 11月 21日		国际检索报告邮寄日期 2017年 12月 13日															
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		授权官员 朱晓莉 电话号码 (86-10)62411672															

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/100060

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	103365867	A	2013年 10月 23日	CN 103365867 B	2017年 7月 21日
CN	103455562	A	2013年 12月 18日	无	
CN	102968408	A	2013年 3月 13日	无	
CN	105005560	A	2015年 10月 28日	无	