

(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
2020 年 11 月 26 日 (26.11.2020)



(10) 国际公布号
WO 2020/233344 A1

- (51) 国际专利分类号:
G06F 16/9535 (2019.01)
- (21) 国际申请号: PCT/CN2020/086677
- (22) 国际申请日: 2020 年 4 月 24 日 (24.04.2020)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201910421974.X 2019年5月21日 (21.05.2019) CN
- (71) 申请人: 深圳壹账通智能科技有限公司(ONE CONNECT SMART TECHNOLOGY CO., LTD. (SHENZHEN)) [CN/CN]; 中国广东省深圳市前海深港合作区前湾一路1号A栋201室 (入驻深圳市前海商务秘书有限公司), Guangdong 518000 (CN)。
- (72) 发明人: 刘利 (LIU, Li); 中国广东省深圳市前海深港合作区前湾一路1号A栋201室 (入驻深圳市前海商务秘书有限公司), Guangdong 518000 (CN)。
- (74) 代理人: 广州三环专利商标代理有限公司 (SCIHEAD IP LAW FIRM); 中国广东省广州市越秀区先烈中路80号汇华商贸大厦1508室, Guangdong 510070 (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,

(54) Title: SEARCHING METHOD AND APPARATUS, AND STORAGE MEDIUM

(54) 发明名称: 一种搜索方法、装置及存储介质

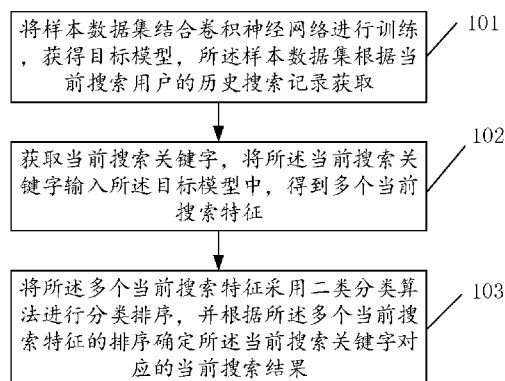


图 1A

- 101 Train a sample data set by combining a convolutional neural network to obtain a target model, the sample data set being obtained according to history searching records of the current searching user
- 102 Obtain the current searching keyword, and input the current searching keyword into the target model to obtain a plurality of current searching features
- 103 Classify and sort the plurality of current searching features by using a binary class classification algorithm, and determine the current searching result corresponding to the current searching keyword according to the sorting of the plurality of current searching features

(57) Abstract: A searching method and apparatus. The searching method comprises: training a sample data set by combining a convolutional neural network to obtain a target model, the sample data set being obtained according to history searching records of the current searching user (101); obtaining the current searching keyword, and inputting the current searching keyword into the target model to obtain a plurality of current searching features (102); and classifying and sorting the plurality of current searching features by using a binary class classification algorithm, and determining the current searching result corresponding to the current searching keyword

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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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))。

according to the sorting of the plurality of current searching features (103). By using the method and apparatus, the current searching features can be obtained by inputting the current searching keyword into the target model, and then the searching features are classified and sorted to determine the searching result. The step of manually obtaining the searching features is avoided in the process, so that the searching process is simplified and the searching efficiency is improved.

(57) 摘要: 一种搜索方法和装置, 其中搜索方法包括: 将样本数据集结合卷积神经网络进行训练, 获得目标模型, 所述样本数据集根据当前搜索用户的历史搜索记录获取(101); 获取当前搜索关键字, 将所述当前搜索关键字输入目标模型中, 得到多个当前搜索特征(102); 将多个所述当前搜索特征采用二类分类算法进行分类排序, 并根据所述多个当前搜索特征的排序确定所述当前搜索关键字对应的当前搜索结果(103)。采用上述方法和装置能够通过将当前搜索关键字输入目标模型中, 获得当前搜索特征, 然后对搜索特征进行分类排序, 确定搜索结果, 这个过程中避免了手动获取搜索特征的过程, 简化了搜索过程, 提升了搜索效率。

一种搜索方法、装置及存储介质

本申请要求于 2019 年 5 月 21 日提交中国专利局、申请号为 201910421974.X, 发明名称为“一种搜索方法、装置及存储介质”的中国专利申请的优先权, 其全部内容通过引用结合在本申请中。

技术领域

本申请涉及大数据的数据处理领域, 具体涉及一种搜索方法、装置及存储介质。

背景技术

个性化搜索是指基于用户之前的搜索记录为其定制搜索结果。通过个性化搜索可以根据用户的历史搜索记录、历史浏览情况、点击情况或交互行为, 为用户的下一次搜索行为提供搜索结果。在这个过程中, 采用传统的个性化搜索方法,

需要手动提取适合个性化搜索行为的特征, 发明人意识到针对不同领域的搜索行为特征的提取需要花费大量的时间, 并且需要大量的相关领域的经验, 这使得在获取搜索结果时需要耗费大量的时间成本和运算成本, 亟待发现一种优化方法以便更高效地进行个性化搜索。

发明内容

本申请实施例提供一种搜索方法、装置及存储介质, 能够通过将当前搜索关键字输入目标模型中, 获得当前搜索特征, 然后对搜索特征进行分类排序, 确定搜索结果, 这个过程中避免了手动获取搜索特征的过程, 简化了搜索过程, 提升了搜索效率。

本申请实施例的第一方面提供了一种搜索方法, 所述方法包括:

将样本数据集结合卷积神经网络进行训练, 获得目标模型, 所述样本数据集根据当前搜索用户的历史搜索记录获取;

获取当前搜索关键字, 将所述当前搜索关键字输入所述目标模型中, 得到多个当前搜索特征;

将所述多个当前搜索特征采用二类分类算法进行分类排序, 并根据所述多个当前搜索特征的排序确定所述当前搜索关键字对应的当前搜索结果。

本申请实施例的第二方面提供了一种搜索装置, 所述搜索装置包括:

训练单元, 用于将样本数据集结合卷积神经网络进行训练, 获得目标模型, 所述样本数据集根据当前搜索用户的历史搜索记录获取;

获取单元, 用于获取当前搜索关键字, 将所述当前搜索关键字输入所述目标模型中, 得到多个当前搜索特征;

搜索单元, 用于将所述多个当前搜索特征采用二类分类算法进行分类排序, 并根据所述多个当前搜索特征的排序确定所述当前搜索关键字对应的当前搜索结果。

本申请实施例第三方面提供了一种电子装置, 包括处理器、存储器、通信接口, 以及一个或多个程序, 所述一个或多个程序被存储在所述存储器中, 并且被配置由所述处理器执行, 所述程序包括执行一种搜索方法, 其中, 所述方法包括:

将样本数据集结合卷积神经网络进行训练, 获得目标模型, 所述样本数据集根据当前搜索用户的历史搜索记录获取;

获取当前搜索关键字, 将所述当前搜索关键字输入所述目标模型中, 得到多个当前搜索特征;

将所述多个当前搜索特征采用二类分类算法进行分类排序, 并根据所述多个当前搜索特征的排序确定所述当前搜索关键字对应的当前搜索结果。

本申请实施例第四方面提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，其中，所述计算机程序被处理器执行以实现一种搜索方法，其中，所述方法包括：

将样本数据集结合卷积神经网络进行训练，获得目标模型，所述样本数据集根据当前搜索用户的历史搜索记录获取；

获取当前搜索关键字，将所述当前搜索关键字输入所述目标模型中，得到多个当前搜索特征；

将所述多个当前搜索特征采用二类分类算法进行分类排序，并根据所述多个当前搜索特征的排序确定所述当前搜索关键字对应的当前搜索结果。

本申请实施例中提供的搜索方法和装置，首先将样本数据集结合卷积神经网络进行训练，获得目标模型；然后获取当前搜索关键字，将搜索方法当前搜索关键字输入搜索方法目标模型中，得到多个当前搜索特征；最后将搜索方法多个当前搜索特征采用二类分类算法进行分类排序，并根据搜索方法多个当前搜索特征的排序确定搜索方法当前搜索关键字对应的当前搜索结果。因为搜索方法样本数据集根据当前搜索用户的历史搜索记录获取的，因此目标模型是当前搜索用户对应的个性化目标模型，将当前搜索关键字输入目标模型后，可以自动提取出多个当前搜索特征而不需要人工参与，简化了特征获取的步骤，提升了特征获取的效率，最后将多个当前搜索特征进行分类排序获得当前搜索结果，提升了搜索效率。

附图说明

图 1A 为本申请实施例提供的一种搜索方法流程示意图；

图 1B 为本申请实施例提供的一种卷积神经网络结构示意图；

图 2 为本申请实施例提供的另一种搜索方法的流程示意图；

图 3 为本申请实施例提供的另一种搜索方法的流程示意图；

图 4 为本申请实施例提供的另一种搜索方法的流程示意图；

图 5 为本申请实施例提供的一种电子装置的结构示意图；

图 6 为本申请实施例提供的一种搜索装置的结构框图。

具体实施方式

本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。

在本文中提及“实施例”意味着，结合实施例描述的特定特征、结构或特性可以包含在本申请的至少一个实施例中。在说明书中的各个位置展示该短语并不一定均是指相同的实施例，也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是，本文所描述的实施例可以与其它实施例相结合。

下面对本申请实施例进行详细介绍。

请参阅图 1A，图 1A 为本申请实施例中一种搜索方法流程示意图，如图 1A 所示，所述搜索方法包括：

101、将样本数据集结合卷积神经网络进行训练，获得目标模型，所述样本数据集根据当前搜索用户的历史搜索记录获取。

用户在进行搜索的过程中，将产生一系列的历史搜索记录，包括用户在搜索界面输入的搜索关键字，根据搜索关键字获得的搜索结果，用户点击搜索页面的概率、次数，用户

点击链接的路线,在每个页面停留的时间,在页面是否进行消费、评论等互动行为等。收集这些历史搜索记录,然后进行预处理或向量化,即可作为样本数据集,用于进行卷积神经网络运算,训练出目标模型,用于进行后续的搜索任务。

请参阅图 1B,图 1B 为本申请实施例提供的一种卷积神经网络的结构示意图,卷积神经网络的训练过程中,采用反向传播算法和随机梯度下降方法,根据前向传播的 loss 值的大小,来进行反向传播迭代更新每一层的权重,直到模型的 loss 值趋向于收敛时,停止训练模型,得到深度学习模型。如图 1B 所示,卷积神经网络中包括卷积层、池化层和全连接层,经过不同层对样本数据集的训练,获得卷积神经网络中的参数,得出目标模型,即为样本数据集对应的目标模型。样本数据集为用户历史搜索记录,将历史输入关键字作为输入值,将历史搜索结果作为输出值,对历史搜索行为结合卷积神经网络进行特征提取,进而训练出历史搜索记录对应的目标模型。

可选的,在将样本数据集结合卷积神经网络进行训练之前,所述方法还包括:根据当前搜索用户的历史搜索记录采集样本数据,历史搜索记录包括历史输入关键字,历史搜索行为,历史搜索结果;对样本数据进行预处理,得到清洁样本数据;对清洁样本数据进行向量化,得到文本向量数据,文本向量数据组成样本数据集。

具体地,用户的历史搜索记录是多种多样的,包括搜索关键字,搜索行为,搜索结果等,这些数据的种类繁多,并不能直接用于进行目标模型的训练。因此首先要对这些数据进行标准化。标准化的方法包括预处理和向量化,其中,因为搜索过程中涉及的都是网页,而网页中包括大量文本内容,因此涉及文本的预处理和向量化。文本的预处理方法包括分词,取出停用词,过滤低频词,编码归一化等;文本向量化即使用向量空间模型 VSM 或者概率统计模型对文本进行表示,使计算机能够理解计算,用的方法基于集合论模型、基于代数轮模型、基于频率统计模型等等。此外,除了对网页的标准化,还包括对搜索行为的标准化,例如根据用户点击搜索页面的概率、次数,确定网页的搜索有效性,根据用户在搜索页面停留的时间,在页面是否进行消费、评论等互动行为,确定用户对网页的满意度等。

可选的,对清洁样本数据进行向量化,得到文本向量数据,包括:获取清洁样本数据对应的样本文本;对样本文本采用文本加权技术进行加权统计,获得样本文本评分;将样本文本评分作为文本向量数据。

具体地,清洁样本数据表示进行预处理后去除无效数据的样本数据。假设样本数据为网页,其中无效数据包括无效浏览网页,网页中的无效文本等。在用户的历史搜索记录中,可能输入一个搜索关键字,但是在这个过程中进行了多次点击链接,那么就涉及到很多网页,对这些网页进行向量化,然后提取特征,并根据提取的特征进行排序,在用户下次进行搜索的时候,可以直接根据特征的排序将排序靠前的特征对应的网页设定为用户输入关键字的搜索结果,可以提升用户的搜索效率。

在这个过程中,假设清洁样本数据为网页,一个网页对应一个样本文本,对网页进行向量化的方法包括对样本文本的向量化,可以通过文本挖掘的方法来实现。TF-IDF (term frequency - inverse document frequency) 是一种用于信息检索与数据挖掘的常用加权技术。TF 意思是词频(Term Frequency),IDF 意思是逆文本频率指数(Inverse Document Frequency)。其中,词频表示词语在样本文本中出现的频率,记为

$$TF_{ij} = \frac{n_{i,j}}{\sum_k n_{k,j}},$$

其中 $n_{i,j}$ 表示文档 j 中第 i 个词语 A 出现的次数, $n_{k,j}$ 表示文档 j 中第 k 个词语出现的次

数。逆文本频率指数是指一个词语普遍重要性的度量。记为： $IDF = \lg \frac{|D|}{1 + |\{j: t_i \in d_j\}|}$,

其中 $|D|$ 表示输入搜索关键字后用户浏览的所有网页个数， $|\{j: t_i \in d_j\}|$ 表示输入搜索关键字后用户浏览的所有网页中，包含词语A的文档个数。

样本文本评分为TF-IDF值，记做：

$$R_i = TF * IDF \quad (1)$$

最后将样本文本评分作为清洁样本数据对应的文本向量数据，完成清洁样本数据的向量化。

可选的，对清洁样本数据进行向量化，得到文本向量数据，包括：获取清洁样本数据对应网页的链接关系，链接关系包括链接对象或链接数量；采用网页排名算法计算清洁样本数据对应网页的排名值；将排名值作为文本向量数据。

具体地，用户在输入搜索关键字后，搜索引擎会给出搜索结果，用户根据搜索结果进行点击，然后进行链接点击，在这个过程中，用户点击的每个网页的重要程度不同，即用户点击的每个网页，引用该网页的网页数量和引用该网页的网页重要程度不同，因此造成每个网页的重要程度也不同。可以对网页进行排名，然后根据排名值对网页文本进行向量化。网页排名算法为PageRank算法，根据该算法得到的网页向量值为排名值，即PageRank值。一个网页的PageRank值计算公式如下：

$$PR(pi) = a \sum_{pj \in M_{pi}} \frac{PR(pj)}{L(pj)} + \frac{(1-a)}{N} \quad (2)$$

其中 $PR(pi)$ 表示网页 pi 的PageRank值， M_{pi} 是所有对网页 pi 有出链的网页集合， $L(pj)$ 是网页 pj 的出链数目， N 是网页总数， a 是阻尼系数。

另外，可选的，对所述清洁样本数据进行向量化，得到文本向量数据还包括：根据用户点击网页的相关数据，确定网页的有效性，相关数据包括点击概率和停留时长；将搜索有效性采用二值法进行表示；将二值法表示结果作为文本的向量数据。

可选的，对所述清洁样本数据进行向量化，得到文本向量数据还包括：根据用户在搜索页面的操作数据，确定页面满意度，操作数据包括页面点击次数，页面评论或页面付费。

具体地，用户点击网页的相关数据包括页面点击概率和停留时长，点击概率可以根据页面显示次数和用户点击次数进行确定，例如 $P = m/k * 100\%$ ，其中 k 表示页面显示次数， m 表示用户点击次数，设置第一概率阈值 $p1$ ，将 p 与 $p1$ 的大小对比结果进行二值化，当 $p < p1$ 时， $y1 = 0$ ，当 $p \geq p1$ 时， $y1 = 1$ 。用户点击了页面后，如果在页面停留时间过短，则可能是刷机操作，因此要确保用户在页面的停留时间达到一定的时长。设置第一时长阈值为 $t1$ ，用户在页面停留的时长为 t ，当 $t < t1$ 时， $y2 = 0$ ，当 $t \geq t1$ 时， $y2 = 1$ 。页面B对应的文本的向量数据 $R2 = y1 \& y2$ ，即只有当 $y1$ 和 $y2$ 同时为1时，文本的向量数据为1。

用户在搜索页面的操作数据包括页面点击次数，页面评论或页面付费。页面点击次数为 $s1$ ，页面评论数为 $s2$ ，页面付费次数为 $s3$ ，然后对这三个参数赋予对应的权值 α ， β ， γ ，并进行加权求和，可获得 $R3 = \alpha * s1 + \beta * s2 + \gamma * s3$ ， $R3$ 即为页面满意度。其中因为付费的可能性最低，其次评论，再次点击，因此权值 $\gamma > \beta > \alpha$ ，或者直接设置为 $\gamma = 10\beta = 100\alpha$ 。

根据上述实施例中确定的文本向量数据，可以直接根据上述每一个文本向量数据结合卷积神经网络进行模型训练，也可以结合上述所有文本向量数据进行模型训练。上述所有文本向量数据可以组成一个文本向量，如表1所示：

表1 网页对应的文本向量表

项目	R1	$PR(pi)$	R2	R3
----	----	----------	----	----

值	TF*IDF	$a \sum_{pj \in M_{pi}} \frac{PR(pj)}{L(pj)} + \frac{(1-a)}{N}$	y1&y2	$\alpha * s1 + \beta * s2 + \gamma * s3$
---	--------	---	-------	--

将上述文本向量结合卷积神经网络进行训练，确定模型中的各个参数，即可获得目标模型。

可见，在本申请实施例中，通过对用户历史搜索记录进行向量化，获得网页对应的文本向量，然后将本文向量作为样本数据集结合卷积神经网络进行训练，获得目标模型，使得模型训练过程数值化，提升了模型训练的效率 and 可靠性。

102、获取当前搜索关键字，将所述当前搜索关键字输入所述目标模型中，得到多个当前搜索特征。

在目标模型训练的过程中，将样本数据集进行卷积神经网络时，卷积层需要进行特征提取，对于文本数据，提取的特征即为各类表现文本内容的数值，以及数值之间的关联关系，数值权重等。对于输入的每一个样本数据，都可以获得其对应的特征，而对于输入的当前搜索关键字，在与样本数据集对应的历史输入关键字进行匹配后，可以获得样本数据集对应的特征作为当前搜索关键字的特征。

可选的，将当前搜索关键字输入目标模型中，得到多个当前搜索特征，包括：将当前搜索关键字进行语义分析，得到至少一个目标分词；将多个目标分词输入目标模型，进行一次或多次卷积运算；获取最后一次卷积运算的特征作为多个目标分词对应的多个当前搜索特征。

具体地，用户输入的当前搜索关键字可能是一个长句，历史输入关键字可能不包括能够与该长句直接匹配的内容，因此要对长句进行语义分析，然后将长句拆分为多个目标分词。将多个目标分词一次输入目标模型，即可匹配到历史输入关键字对应的文本，亦即可获得历史输入关键字对应的文本特征，获取所有目标分词对应的文本特征，即为当前搜索关键字对应的文本特征。其中，目标模型中可能包括多个卷积层，每个卷积层对应一次卷积运算，最后一个卷积层进行的最后一次卷积运算得到的文本特征最高级，因此作为最后的当前搜索特征。

103、将所述多个当前搜索特征采用二类分类算法进行分类排序，并根据所述多个当前搜索特征的排序确定所述当前搜索关键字对应的当前搜索结果。

获得当前搜索关键字对应的当前搜索特征后，对多个搜索特征进行分类排序，即可确定搜索特征对应的搜索结果的排序，进而获得当前搜索关键字对应的搜索结果推荐顺序。

可选的，二类分类算法为支持向量机 SVM 算法，将多个当前搜索特征采用二类分类算法进行分类排序，包括：将多个当前搜索特征中的每两个当前搜索特征进行成对组合，获得多个特征组；对多个特征组中的每个特征组采用 SVM 算法进行评分排序，确定每个特征组中的两个当前搜索特征的相对大小；根据每个特征组中的两个当前搜索特征的相对大小确定多个当前搜索特征的排序。

具体地，采用二分类算法，可以将多个当前搜索特征进行两两组合，获得多个特征组，假设一个特征组中的两个当前搜索特征分别为 d_j 和 d_i ，当 $d_j > d_i$ 时，为其赋值为 1，否则为其赋值为 -1，那么对多个当前搜索特征的排序问题转化为分类问题。支持向量机 (Support Vector Machine, SVM) 广泛用于二元分类问题，可以完成每个特征组的大小对比，进而完成所有当前搜索特征的大小排序问题。

确定所有当前搜索特征的大小排序后，选择排序最前的 TOP-K 作为目标特征，然后获取目标特征对应的输出值，即历史搜索结果，作为当前搜索关键字对应的搜索结果推荐。其中 TOP-K 中的 K 可以是 1，也可以是其他正整数，即搜索结果可以是一个搜索结果，也可以是多个搜索结果。

可见，在本申请实施例中，首先将样本数据集结合卷积神经网络进行训练，获得目标

模型；然后获取当前搜索关键字，将搜索方法当前搜索关键字输入搜索方法目标模型中，得到多个当前搜索特征；最后将搜索方法多个当前搜索特征采用二类分类算法进行分类排序，并根据搜索方法多个当前搜索特征的排序确定搜索方法当前搜索关键字对应的当前搜索结果。因为搜索方法样本数据集根据当前搜索用户的历史搜索记录获取的，因此目标模型是当前搜索用户对应的个性化目标模型，将当前搜索关键字输入目标模型后，可以自动提取出多个当前搜索特征而不需要人工参与，简化了特征获取的步骤，提升了特征获取的效率，最后将多个当前搜索特征进行分类排序获得当前搜索结果，提升了搜索效率。

请参阅图 2，图 2 是本申请实施例提供的另一种搜索方法的流程示意图，如图 2 所示，本实施例中的搜索方法包括：

201、根据当前搜索用户的历史搜索记录采集样本数据，所述历史搜索记录包括历史输入关键字，历史搜索行为，历史搜索结果；

202、对所述样本数据进行预处理，得到清洁样本数据；

203、获取所述清洁样本数据对应的样本文本，并对所述样本文本采用文本加权技术进行加权统计，获得样本文本评分；

204、将所述样本文本评分作为所述清洁样本数据的文本向量数据，所述文本向量数据组成所述样本数据集；

205、将样本数据集结合卷积神经网络进行训练，获得目标模型，所述样本数据集根据当前搜索用户的历史搜索记录获取；

206、获取当前搜索关键字，将所述当前搜索关键字输入所述目标模型中，得到多个当前搜索特征；

207、将所述多个当前搜索特征采用二类分类算法进行分类排序，并根据所述多个当前搜索特征的排序确定所述当前搜索关键字对应的当前搜索结果。

其中，上述步骤 201-步骤 207 的具体描述可以参照图 1A 所描述的搜索方法的相应描述，在此不再赘述。

在本申请实施例中，通过获取当前搜索用户的历史搜索记录，对历史搜索记录进行预处理，并通过 TF-IDF 方法对历史搜索记录中的网页文本进行加权统计，获得网页文本对应的文本向量数据，再将文本向量数据作为样本数据集进行卷积神经网络训练，获得目标模型并进行搜索推荐，这个过程中，通过 TF-IDF 方法对网页文本进行向量化，通过词频统计确定网页文本的重要性，提升了目标模型构建的针对性和可靠性。

请参阅图 3，图 3 是本申请实施例提供的另一种搜索方法的流程示意图，如图 3 所示，本实施例中的搜索方法包括：

301、根据当前搜索用户的历史搜索记录采集样本数据，所述历史搜索记录包括历史输入关键字，历史搜索行为，历史搜索结果；

302、获取所述清洁样本数据对应网页的链接关系，所述链接关系包括链接对象或链接数量；

303、采用网页排名算法计算所述清洁样本数据对应网页的排名值；

304、将所述排名值作为所述文本向量数据，所述文本向量数据组成所述样本数据集；

305、将样本数据集结合卷积神经网络进行训练，获得目标模型，所述样本数据集根据当前搜索用户的历史搜索记录获取；

306、获取当前搜索关键字，将所述当前搜索关键字输入所述目标模型中，得到多个当前搜索特征；

307、将所述多个当前搜索特征采用二类分类算法进行分类排序，并根据所述多个当前搜索特征的排序确定所述当前搜索关键字对应的当前搜索结果。

其中，上述步骤 301-步骤 307 的具体描述可以参照图 1A 所描述的搜索方法的相应描述，在此不再赘述。

在本申请实施例中，通过获取当前搜索用户的历史搜索记录，对历史搜索记录进行预处理，并通过 PageRank 算法对历史搜索记录中的网页文本重要程度评分，获得网页文本对应的文本向量数据，再将文本向量数据作为样本数据集进行卷积神经网络训练，获得目标模型并进行搜索推荐，这个过程中，通过 PageRank 算法对网页文本进行向量化，通过网页链接关系确定网页文本的重要性，提升了目标模型构建的针对性和可靠性。

请参阅图 4，图 4 是本申请实施例提供的另一种搜索方法的流程示意图，如图 4 所示，本实施例中的搜索方法包括：

401、将样本数据集结合卷积神经网络进行训练，获得目标模型，所述样本数据集根据当前搜索用户的历史搜索记录获取；

402、将所述当前搜索关键字进行语义分析，得到至少一个目标分词；

403、将所述多个目标分词输入所述目标模型，进行一次或多次卷积运算；

404、获取最后一次卷积运算的特征作为所述多个目标分词对应的多个当前搜索特征；
405、将所述多个当前搜索特征中的每两个当前搜索特征进行成对组合，获得多个特征组；

406、对所述多个特征组中的每个特征组采用 SVM 算法进行评分排序，确定所述每个特征组中的两个当前搜索特征的相对大小；

407、根据所述每个特征组中的两个当前搜索特征的相对大小确定所述多个当前搜索特征的排序，并根据所述多个当前搜索特征的排序确定所述当前搜索关键字对应的当前搜索结果。

其中，上述步骤 401-步骤 407 的具体描述可以参照图 1A 所描述的搜索方法的相应描述，在此不再赘述。

在本申请实施例中，在训练获得目标模型后，将当前搜索关键字输入目标模型中进行特征提取，在将当前搜索关键字输入目标模型之前，将其进行语义分析并拆分得到至少一个目标分词，然后将目标分词依次输入目标模型，得到与构建目标模型的历史输入关键字对应的特征作为当前搜索关键字对应的当前搜索特征，对当前搜索特征采用 SVM 算法进行评分排序，进而获得搜索推荐结果。这个过程中，对当前搜索关键字进行拆分并依次获得对应的历史输入关键字的特征组成当前搜索特征，提升了当前搜索特征提取的准确性，采用 SVM 算法对当前搜索特征进行评分排序进而获得推荐结果，简化了人工获取特征的步骤，提升了获得搜索推荐结果的效率，进而提升了搜索效率。

图 5 是本申请实施例提供的一种电子装置的结构示意图，如图 5 所示，该电子装置包括处理器、存储器、通信接口以及一个或多个程序，其中，上述一个或多个程序被存储在上述存储器中，并且被配置由上述处理器执行，上述程序包括用于执行以下步骤的指令：

将样本数据集结合卷积神经网络进行训练，获得目标模型，所述样本数据集根据当前搜索用户的历史搜索记录获取；

获取当前搜索关键字，将所述当前搜索关键字输入所述目标模型中，得到多个当前搜索特征；

将所述多个当前搜索特征采用二类分类算法进行分类排序，并根据所述多个当前搜索特征的排序确定所述当前搜索关键字对应的当前搜索结果。

可以看出，在本申请实施例中，电子装置首先将样本数据集结合卷积神经网络进行训练，获得目标模型；然后获取当前搜索关键字，将搜索方法当前搜索关键字输入搜索方法目标模型中，得到多个当前搜索特征；最后将搜索方法多个当前搜索特征采用二类分类算

法进行分类排序,并根据搜索方法多个当前搜索特征的排序确定搜索方法当前搜索关键字对应的当前搜索结果。因为搜索方法样本数据集根据当前搜索用户的历史搜索记录获取的,因此目标模型是当前搜索用户对应的个性化目标模型,将当前搜索关键字输入目标模型后,可以自动提取出多个当前搜索特征而不需要人工参与,简化了特征获取的步骤,提升了特征获取的效率,最后将多个当前搜索特征进行分类排序获得当前搜索结果,提升了搜索效率。

在一个可能的示例中,所述将所述当前搜索关键字输入所述目标模型中,得到多个当前搜索特征,包括:

将所述当前搜索关键字进行语义分析,得到至少一个多个目标分词;

将所述多个目标分词输入所述目标模型,进行一次或多次卷积运算;

获取最后一次卷积运算的特征作为所述多个目标分词对应的多个当前搜索特征。

在一个可能的示例中,所述二类分类算法为支持向量机 SVM 算法,所述将所述多个当前搜索特征采用二类分类算法进行分类排序,包括:

将所述多个当前搜索特征中的每两个当前搜索特征进行成对组合,获得多个特征组;

对所述多个特征组中的每个特征组采用 SVM 算法进行评分排序,确定所述每个特征组中的两个当前搜索特征的相对大小;

根据所述每个特征组中的两个当前搜索特征的相对大小确定所述多个当前搜索特征的排序。

在一个可能的示例中,在将样本数据集结合卷积神经网络进行训练之前,所述方法还包括:

根据当前搜索用户的历史搜索记录采集样本数据,所述历史搜索记录包括历史输入关键字,历史搜索行为,历史搜索结果;

对所述样本数据进行预处理,得到清洁样本数据;

对所述清洁样本数据进行向量化,得到文本向量数据,所述文本向量数据组成所述样本数据集。

在一个可能的示例中,所述对所述清洁样本数据进行向量化,得到文本向量数据,包括:

获取所述清洁样本数据对应的样本文本;

对所述样本文本进行 TF-IDF 加权统计,获得样本文本评分;

将所述样本文本评分作为所述清洁样本数据的文本向量数据。

在一个可能的示例中,所述对所述清洁样本数据进行向量化,得到文本向量数据,包括:

获取所述清洁样本数据对应网页的链接关系,所述链接关系包括链接对象或链接数量;

采用 PageRank 算法计算所述清洁样本数据对应网页的 PageRank 值;

将所述 PageRank 值作为所述文本向量数据。

图6是本申请实施例中所涉及的搜索装置600的功能单元组成框图。该搜索装置600应用于业务系统,业务系统包括服务器和客户端,所述搜索装置包括:

训练单元 601,用于将样本数据集结合卷积神经网络进行训练,获得目标模型,所述样本数据集根据当前搜索用户的历史搜索记录获取;

获取单元 602,用于获取当前搜索关键字,将所述当前搜索关键字输入所述目标模型中,得到多个当前搜索特征;

搜索单元 603,用于将所述多个当前搜索特征采用二类分类算法进行分类排序,并根据所述多个当前搜索特征的排序确定所述当前搜索关键字对应的当前搜索结果。

在此需要说明的是，上述训练单元 601、获取单元 602 和搜索单元 603 的具体工作过程参见上述步骤 101-103 的相关描述。在此不再赘述。

可以看出，本申请实施例中，搜索装置首先将样本数据集结合卷积神经网络进行训练，获得目标模型；然后获取当前搜索关键字，将搜索方法当前搜索关键字输入搜索方法目标模型中，得到多个当前搜索特征；最后将搜索方法多个当前搜索特征采用二类分类算法进行分类排序，并根据搜索方法多个当前搜索特征的排序确定搜索方法当前搜索关键字对应的当前搜索结果。因为搜索方法样本数据集根据当前搜索用户的历史搜索记录获取的，因此目标模型是当前搜索用户对应的个性化目标模型，将当前搜索关键字输入目标模型后，可以自动提取出多个当前搜索特征而不需要人工参与，简化了特征获取的步骤，提升了特征获取的效率，最后将多个当前搜索特征进行分类排序获得当前搜索结果，提升了搜索效率。

在可选情况下，所述获取单元 602 具体用于：

将所述当前搜索关键字进行语义分析，得到至少一个多个目标分词；

将所述多个目标分词输入所述目标模型，进行一次或多次卷积运算；

获取最后一次卷积运算的特征作为所述多个目标分词对应的多个当前搜索特征。

在可选情况下，所述二类分类算法为支持向量机 SVM 算法，所述搜索单元 603 具体用于：

将所述多个当前搜索特征中的每两个当前搜索特征进行成对组合，获得多个特征组；

对所述多个特征组中的每个特征组采用 SVM 算法进行评分排序，确定所述每个特征组中的两个当前搜索特征的相对大小；

根据所述每个特征组中的两个当前搜索特征的相对大小确定所述多个当前搜索特征的排序。

在可选情况下，所述搜索装置还包括数据处理单元 604，具体用于：

根据当前搜索用户的历史搜索记录采集样本数据，所述历史搜索记录包括历史输入关键字，历史搜索行为，历史搜索结果；

对所述样本数据进行预处理，得到清洁样本数据；

对所述清洁样本数据进行向量化，得到文本向量数据，所述文本向量数据组成所述样本数据集。

在可选情况下，所述数据处理单元 604 还具体用于：

获取所述清洁样本数据对应的样本文本；

对所述样本文本进行 TF-IDF 加权统计，获得样本文本评分；

将所述样本文本评分作为所述清洁样本数据的文本向量数据。

在可选情况下，所述数据处理单元 604 还具体用于：

获取所述清洁样本数据对应网页的链接关系，所述链接关系包括链接对象或链接数量；

采用 PageRank 算法计算所述清洁样本数据对应网页的 PageRank 值；

将所述 PageRank 值作为所述文本向量数据。

本申请实施例还提供一种计算机存储介质，所述存储介质为易失性介质和非易失性介质，其中，该计算机存储介质存储用于电子数据交换的计算机程序，该计算机程序使得计算机至少执行以下步骤：

将样本数据集结合卷积神经网络进行训练，获得目标模型，所述样本数据集根据当前搜索用户的历史搜索记录获取；

获取当前搜索关键字，将所述当前搜索关键字输入所述目标模型中，得到多个当前搜索特征；

将所述多个当前搜索特征采用二类分类算法进行分类排序，并根据所述多个当前搜索

特征的排序确定所述当前搜索关键字对应的当前搜索结果。

本申请实施例还提供一种计算机程序产品，上述计算机程序产品包括存储了计算机程序的非瞬时性计算机可读存储介质，上述计算机程序可操作来使计算机执行如上述方法实施例中记载的任一方法的部分或全部步骤。该计算机程序产品可以为一个软件安装包，上述计算机包括移动终端。

在本申请所提供的几个实施例中，应该理解到，所揭露的装置，可通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如上述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性或其它的形式。

上述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

另外，在本申请各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

上述集成的单元若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，可以存储在一个计算机可读取存储器中。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储器中，包括若干指令用以使得一台计算机设备（可为个人计算机、服务器或者网络设备）执行本申请各个实施例上述方法的全部或部分步骤。而前述的存储器包括：U 盘、只读存储器（Read-Only Memory ,ROM）、随机存取存储器（Random Access Memory ,RAM）、移动硬盘、磁碟或者光盘等各种可以存储程序代码的介质。

本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通程序来指令相关的硬件来完成，该程序可以存储于一计算机可读存储器中，存储器可以包括：闪存盘、ROM、RAM、磁盘或光盘等。

以上对本申请实施例进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想；同时，对于本领域的一般技术人员，依据本申请的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本申请的限制。

权利要求

1、一种搜索方法，其中，所述方法包括：

将搜索用户的样本数据集结合卷积神经网络进行训练，获得目标模型，所述样本数据集根据所述搜索用户的历史搜索记录获取；

5 获取当前搜索关键字，将所述当前搜索关键字输入所述目标模型中，得到多个当前搜索特征；

将所述多个当前搜索特征采用二类分类算法进行分类排序，并根据所述多个当前搜索特征的排序确定所述当前搜索关键字对应的当前搜索结果。

10 2、根据权利要求1所述的方法，其中，所述将所述当前搜索关键字输入所述目标模型中，得到多个当前搜索特征，包括：

将所述当前搜索关键字进行语义分析，得到至少一个目标分词；

将所述多个目标分词输入所述目标模型，进行一次或多次卷积运算；

获取最后一次卷积运算的特征作为所述多个目标分词对应的多个当前搜索特征。

15 3、根据权利要求2所述的方法，其中，所述二类分类算法为支持向量机 SVM 算法，所述将所述多个当前搜索特征采用二类分类算法进行分类排序，包括：

将所述多个当前搜索特征中的每两个当前搜索特征进行成对组合，获得多个特征组；

对所述多个特征组中的每个特征组采用 SVM 算法进行评分排序，确定所述每个特征组中的两个当前搜索特征的相对大小；

20 根据所述每个特征组中的两个当前搜索特征的相对大小确定所述多个当前搜索特征的排序。

4、根据权利要求1所述的方法，其中，在将样本数据集结合卷积神经网络进行训练之前，所述方法还包括：

根据当前搜索用户的历史搜索记录采集样本数据，所述历史搜索记录包括历史输入关键字，历史搜索行为，历史搜索结果；

25 对所述样本数据进行预处理，得到清洁样本数据；

对所述清洁样本数据进行向量化，得到文本向量数据，所述文本向量数据组成所述样本数据集。

5、根据权利要求4所述的方法，其中，所述对所述清洁样本数据进行向量化，得到文本向量数据，包括：

30 获取所述清洁样本数据对应的样本文本；

对所述样本文本采用文本加权技术进行加权统计，获得样本文本评分；

将所述样本文本评分作为所述清洁样本数据的文本向量数据。

6、根据权利要求4或5所述的方法，其中，所述对所述清洁样本数据进行向量化，得到文本向量数据，包括：

35 获取所述清洁样本数据对应网页的链接关系，所述链接关系包括链接对象或链接数量；采用网页排名算法计算所述清洁样本数据对应网页的排名值；

将所述排名值作为所述文本向量数据。

7、一种搜索装置，其中，所述搜索装置包括：

40 训练单元，用于将样本数据集结合卷积神经网络进行训练，获得目标模型，所述样本数据集根据当前搜索用户的历史搜索记录获取；

获取单元，用于获取当前搜索关键字，将所述当前搜索关键字输入所述目标模型中，得到多个当前搜索特征；

搜索单元，用于将所述多个当前搜索特征采用二类分类算法进行分类排序，并根据所述多个当前搜索特征的排序确定所述当前搜索关键字对应的当前搜索结果。

8、根据权利要求 7 所述的装置，其中，所述获取单元具体用于：

将所述当前搜索关键字进行语义分析，得到至少一个目标分词；

将所述多个目标分词输入所述目标模型，进行一次或多次卷积运算；

获取最后一次卷积运算的特征作为所述多个目标分词对应的多个当前搜索特征。

5 9、一种电子装置，包括处理器、存储器、通信接口，以及一个或多个程序，所述一个或多个程序被存储在所述存储器中，并且被配置由所述处理器执行，所述程序包括用于执行一种搜索方法，其中所述方法包括：

将搜索用户的样本数据集结合卷积神经网络进行训练，获得目标模型，所述样本数据集根据所述搜索用户的历史搜索记录获取；

10 获取当前搜索关键字，将所述当前搜索关键字输入所述目标模型中，得到多个当前搜索特征；

将所述多个当前搜索特征采用二类分类算法进行分类排序，并根据所述多个当前搜索特征的排序确定所述当前搜索关键字对应的当前搜索结果。

15 10、根据权利要求 9 所述的电子装置，其中，所述将所述当前搜索关键字输入所述目标模型中，得到多个当前搜索特征，包括：

将所述当前搜索关键字进行语义分析，得到至少一个目标分词；

将所述多个目标分词输入所述目标模型，进行一次或多次卷积运算；

获取最后一次卷积运算的特征作为所述多个目标分词对应的多个当前搜索特征。

20 11、根据权利要求 10 所述的电子装置，其中，所述二类分类算法为支持向量机 SVM 算法，所述将所述多个当前搜索特征采用二类分类算法进行分类排序，包括：

将所述多个当前搜索特征中的每两个当前搜索特征进行成对组合，获得多个特征组；

对所述多个特征组中的每个特征组采用 SVM 算法进行评分排序，确定所述每个特征组中的两个当前搜索特征的相对大小；

25 根据所述每个特征组中的两个当前搜索特征的相对大小确定所述多个当前搜索特征的排序。

12、根据权利要求 9 所述的电子装置，其中，在将样本数据集结合卷积神经网络进行训练之前，所述方法还包括：

根据当前搜索用户的历史搜索记录采集样本数据，所述历史搜索记录包括历史输入关键字，历史搜索行为，历史搜索结果；

30 对所述样本数据进行预处理，得到清洁样本数据；

对所述清洁样本数据进行向量化，得到文本向量数据，所述文本向量数据组成所述样本数据集。

13、根据权利要求 12 所述的电子装置，其中，所述对所述清洁样本数据进行向量化，得到文本向量数据，包括：

35 获取所述清洁样本数据对应的样本文本；

对所述样本文本采用文本加权技术进行加权统计，获得样本文本评分；

将所述样本文本评分作为所述清洁样本数据的文本向量数据。

14、根据权利要求 12 或 13 所述的电子装置，其中，所述对所述清洁样本数据进行向量化，得到文本向量数据，包括：

40 获取所述清洁样本数据对应网页的链接关系，所述链接关系包括链接对象或链接数量；采用网页排名算法计算所述清洁样本数据对应网页的排名值；

将所述排名值作为所述文本向量数据。

15、一种计算机可读存储介质，其中，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行以实现一种搜索方法，其中所述方法包括：

将搜索用户的样本数据集结合卷积神经网络进行训练, 获得目标模型, 所述样本数据集根据所述搜索用户的历史搜索记录获取;

获取当前搜索关键字, 将所述当前搜索关键字输入所述目标模型中, 得到多个当前搜索特征;

5 将所述多个当前搜索特征采用二类分类算法进行分类排序, 并根据所述多个当前搜索特征的排序确定所述当前搜索关键字对应的当前搜索结果。

16、根据权利要求 15 所述的计算机可读存储介质, 其中, 所述将所述当前搜索关键字输入所述目标模型中, 得到多个当前搜索特征, 包括:

将所述当前搜索关键字进行语义分析, 得到至少一个目标分词;

10 将所述多个目标分词输入所述目标模型, 进行一次或多次卷积运算;

获取最后一次卷积运算的特征作为所述多个目标分词对应的多个当前搜索特征。

17、根据权利要求 16 所述的计算机可读存储介质, 其中, 所述二类分类算法为支持向量机 SVM 算法, 所述将所述多个当前搜索特征采用二类分类算法进行分类排序, 包括:

将所述多个当前搜索特征中的每两个当前搜索特征进行成对组合, 获得多个特征组;

15 对所述多个特征组中的每个特征组采用 SVM 算法进行评分排序, 确定所述每个特征组中的两个当前搜索特征的相对大小;

根据所述每个特征组中的两个当前搜索特征的相对大小确定所述多个当前搜索特征的排序。

18、根据权利要求 15 所述的计算机可读存储介质, 其中, 在将样本数据集结合卷积神经网络进行训练之前, 还包括:

20 根据当前搜索用户的历史搜索记录采集样本数据, 所述历史搜索记录包括历史输入关键字, 历史搜索行为, 历史搜索结果;

对所述样本数据进行预处理, 得到清洁样本数据;

25 对所述清洁样本数据进行向量化, 得到文本向量数据, 所述文本向量数据组成所述样本数据集。

19、根据权利要求 18 所述的计算机可读存储介质, 其中, 所述对所述清洁样本数据进行向量化, 得到文本向量数据, 包括:

获取所述清洁样本数据对应的样本文本;

对所述样本文本采用文本加权技术进行加权统计, 获得样本文本评分;

30 将所述样本文本评分作为所述清洁样本数据的文本向量数据。

20、根据权利要求 18 或 19 所述的计算机可读存储介质, 其中, 所述对所述清洁样本数据进行向量化, 得到文本向量数据, 包括:

获取所述清洁样本数据对应网页的链接关系, 所述链接关系包括链接对象或链接数量;

采用网页排名算法计算所述清洁样本数据对应网页的排名值;

35 将所述排名值作为所述文本向量数据。

—1/5—

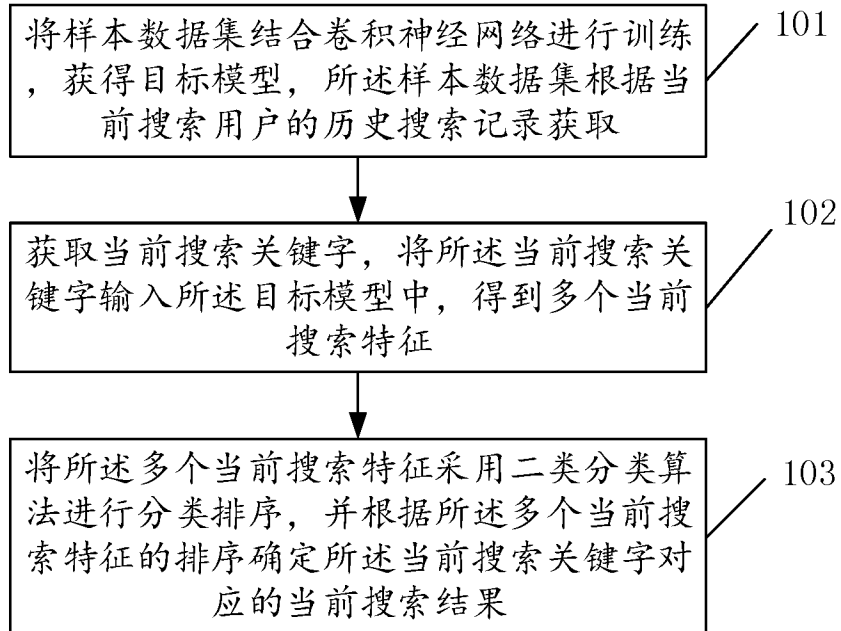


图 1A

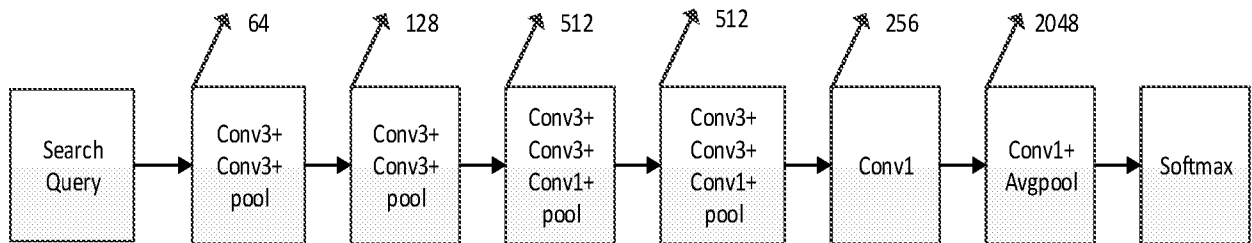


图 1B

—2/5—

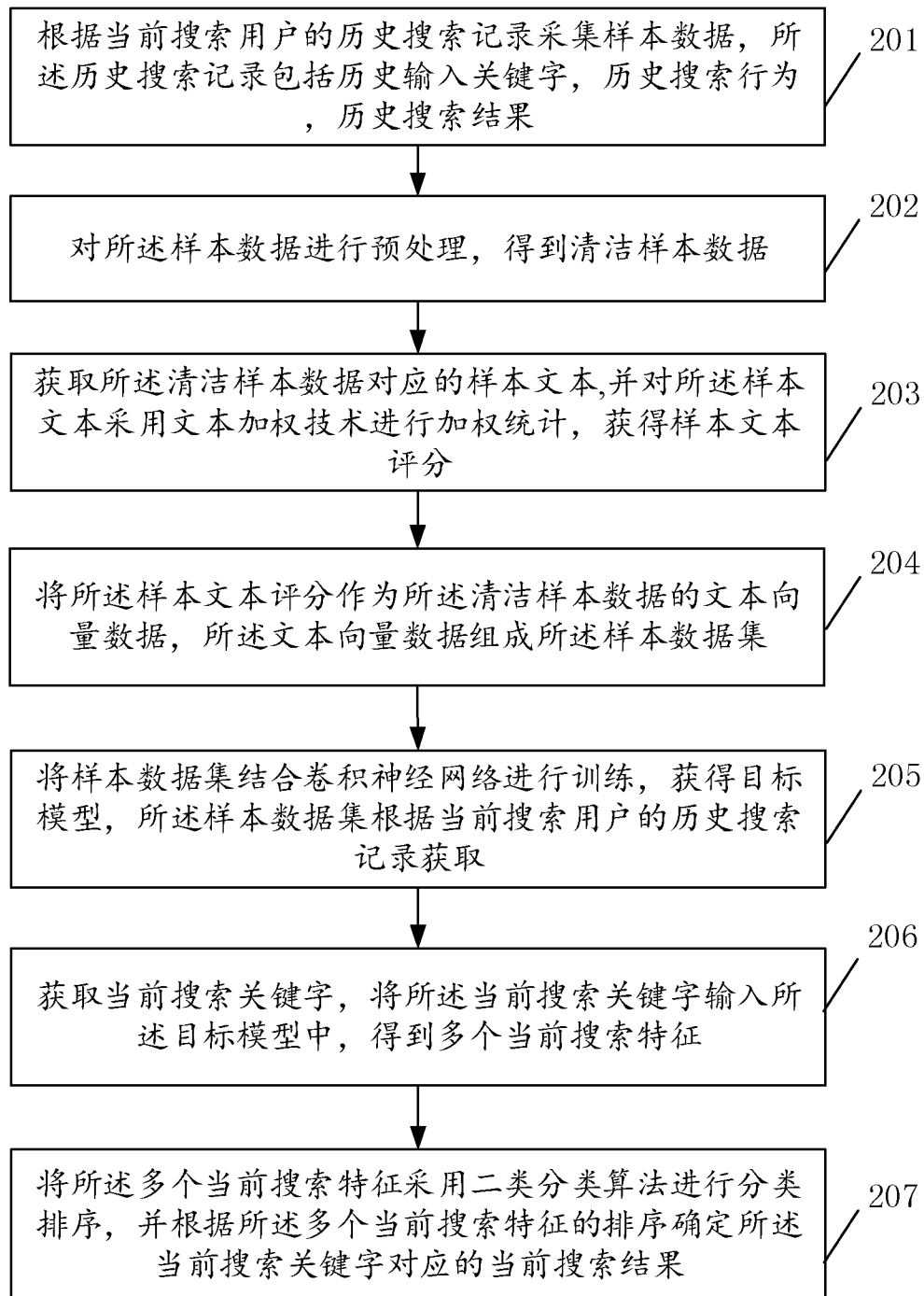


图 2

—3/5—

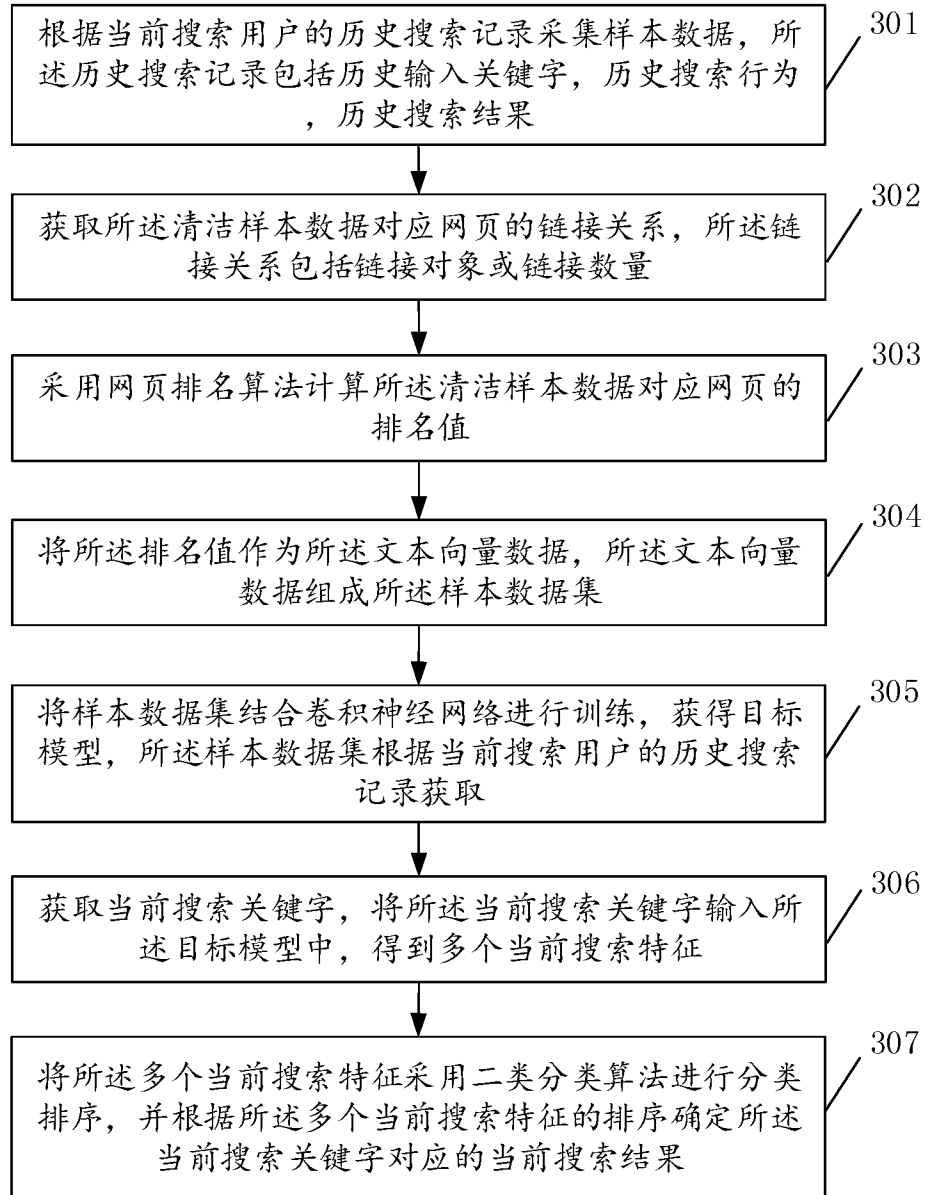


图 3

—4/5—

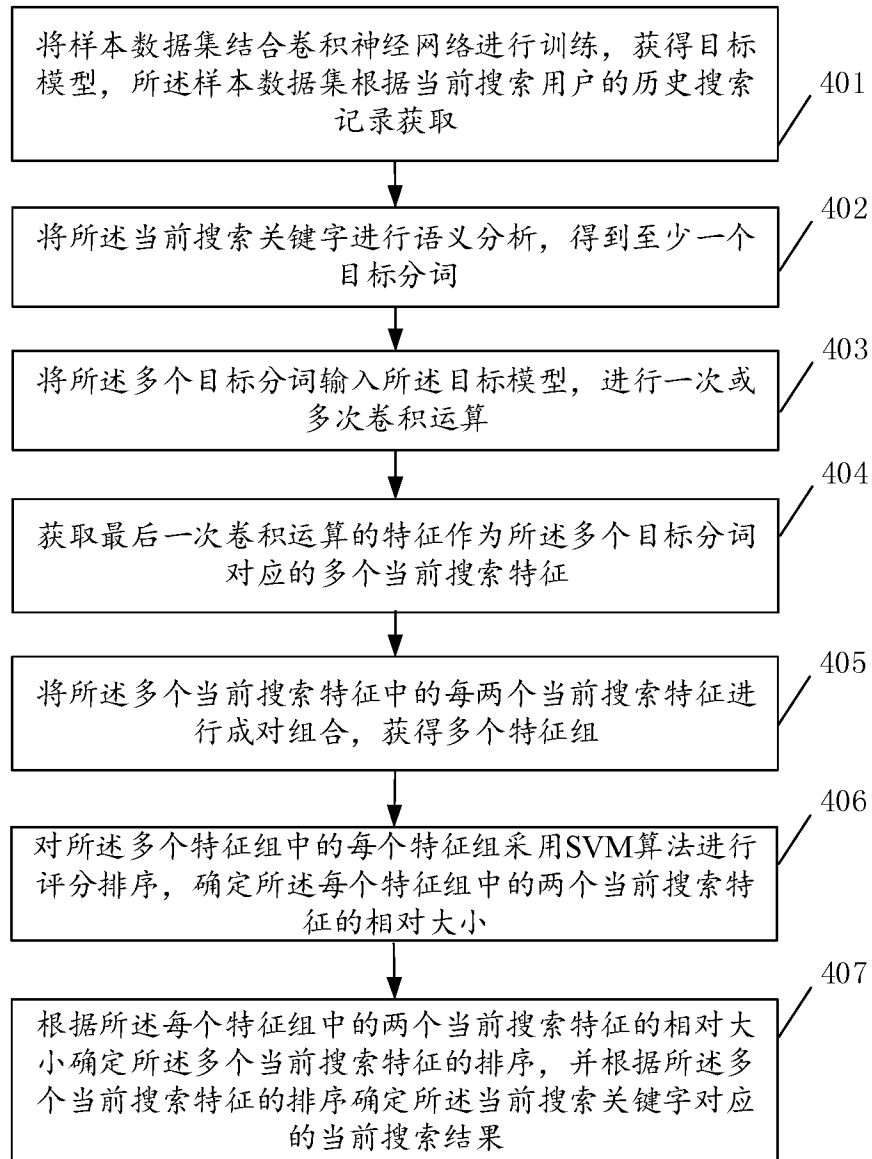


图 4

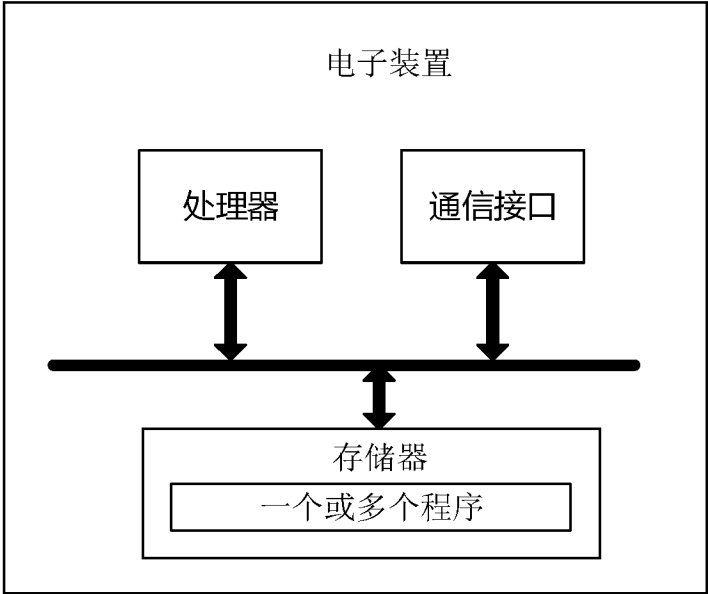


图 5

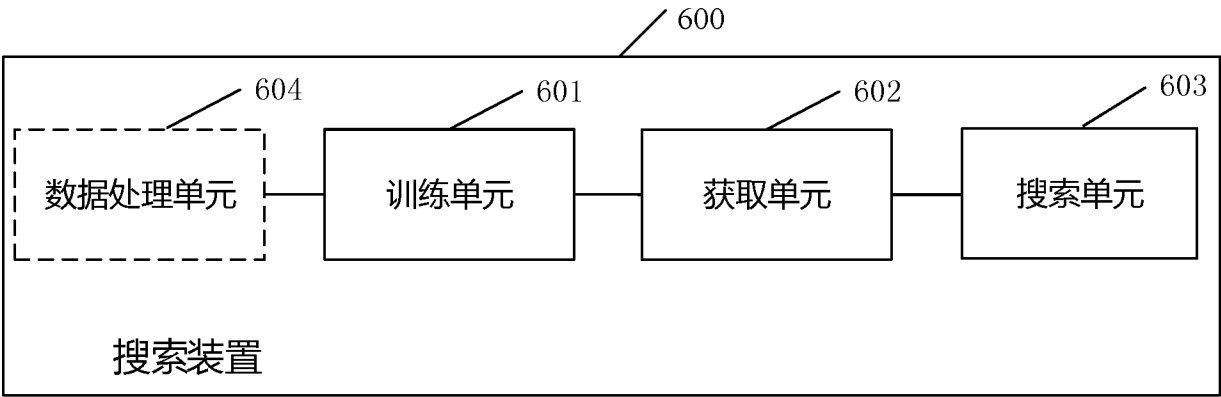


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/086677

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/9535(2019.01)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

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)

CNPAT, CNKI, WPI, EPODOC: 样本, 卷积神经网络, 模型, 关键字, 特征, 分类, 排序, 历史, 记录, 搜索, sample, CNN, model, keyword, feature, type, sort, history, record, search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110222260 A (SHENZHEN ONECONNECT TECHNOLOGY CO., LTD.) 10 September 2019 (2019-09-10) claims 1-10	1-20
X	CN 106649760 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 10 May 2017 (2017-05-10) abstract, description paragraphs 51-89	1-20
A	CN 107066553 A (BEIJING UNIVERSITY OF TECHNOLOGY) 18 August 2017 (2017-08-18) entire document	1-20
A	CN 109543190 A (BEIJING YUSHANZHI INFORMATION TECHNOLOGY CO., LTD.) 29 March 2019 (2019-03-29) entire document	1-20
A	CN 102737090 A (YUAN, Xingyuan) 17 October 2012 (2012-10-17) entire document	1-20
A	US 2018239989 A1 (ALIBABA GROUP HOLDING LIMITED) 23 August 2018 (2018-08-23) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“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)

“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

“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

“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

“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

“&” document member of the same patent family

Date of the actual completion of the international search

06 July 2020

Date of mailing of the international search report

23 July 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/086677

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110222260	A	10 September 2019	None			
CN	106649760	A	10 May 2017	US	2018181573	A1	28 June 2018
CN	107066553	A	18 August 2017	None			
CN	109543190	A	29 March 2019	None			
CN	102737090	A	17 October 2012	None			
US	2018239989	A1	23 August 2018	TW	201832137	A	01 September 2018
				CN	108460389	A	28 August 2018
				WO	2018152532	A1	23 August 2018

国际检索报告

国际申请号

PCT/CN2020/086677

A. 主题的分类

G06F 16/9535 (2019.01) 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

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))

CNPAT, CNKI, WPI, EPDOC: 样本, 卷积神经网络, 模型, 关键字, 特征, 分类, 排序, 历史, 记录, 搜索, sample, CNN, model, keyword, feature, type, sort, history, record, search

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 110222260 A (深圳壹账通智能科技有限公司) 2019年 9月 10日 (2019 - 09 - 10) 权利要求1-10	1-20
X	CN 106649760 A (北京百度网讯科技有限公司) 2017年 5月 10日 (2017 - 05 - 10) 摘要、说明书第51-89段	1-20
A	CN 107066553 A (北京工业大学) 2017年 8月 18日 (2017 - 08 - 18) 全文	1-20
A	CN 109543190 A (北京羽扇智信息科技有限公司) 2019年 3月 29日 (2019 - 03 - 29) 全文	1-20
A	CN 102737090 A (袁行远) 2012年 10月 17日 (2012 - 10 - 17) 全文	1-20
A	US 2018239989 A1 (ALIBABA GROUP HOLDING LIMITED) 2018年 8月 23日 (2018 - 08 - 23) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 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

2020年 7月 6日

国际检索报告邮寄日期

2020年 7月 23日

ISA/CN的名称和邮寄地址

中国国家知识产权局(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

姜玲玲

电话号码 86-(10)-53961421

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/086677

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	110222260	A	2019年 9月 10日	无			
CN	106649760	A	2017年 5月 10日	US	2018181573	A1	2018年 6月 28日
CN	107066553	A	2017年 8月 18日	无			
CN	109543190	A	2019年 3月 29日	无			
CN	102737090	A	2012年 10月 17日	无			
US	2018239989	A1	2018年 8月 23日	TW	201832137	A	2018年 9月 1日
				CN	108460389	A	2018年 8月 28日
				WO	2018152532	A1	2018年 8月 23日