

(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
2022 年 6 月 9 日 (09.06.2022)



(10) 国际公布号
WO 2022/116905 A1

(51) 国际专利分类号:
G06N 3/04 (2006.01)

(21) 国际申请号: PCT/CN2021/133500

(22) 国际申请日: 2021 年 11 月 26 日 (26.11.2021)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202011391497.6 2020 年 12 月 2 日 (02.12.2020) CN

(71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(72) 发明人: 郭慧丰 (GUO, Huifeng); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 陈渤 (CHEN, Bo); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong

518129 (CN)。 唐睿明 (TANG, Ruiming); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 李震国 (LI, Zhenguo); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 何秀强 (HE, Xiuqiang); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(74) 代理人: 深圳市深佳知识产权代理事务所 (普通合伙) (SHENPAT INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市罗湖区南湖街道春风路庐山大厦 B 座 18C2、18D、18E、18E2, Guangdong 518001 (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, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK,

(54) Title: DATA PROCESSING METHOD AND APPARATUS

(54) 发明名称: 一种数据处理方法以及装置

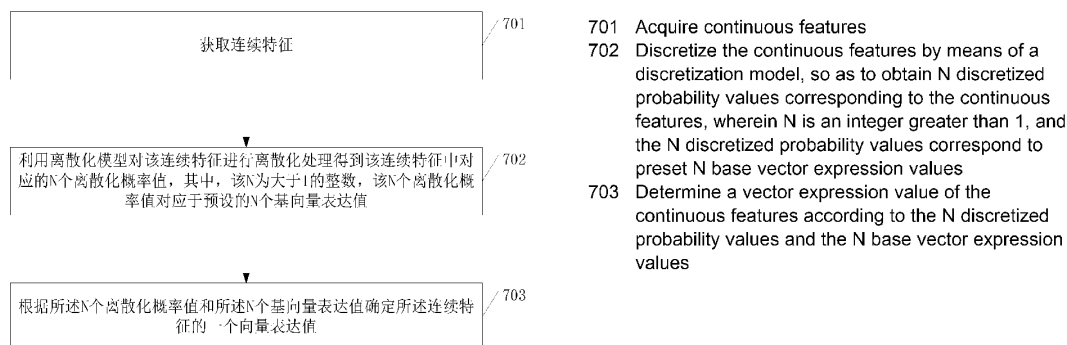


图 7

(57) Abstract: Provided are a data processing method and apparatus, which are used for better learning a vector expression value of each feature value in continuous features. The method specifically comprises: a data processing apparatus acquiring continuous features from sample data; then discretizing the continuous features by means of a discretization model, so as to obtain N discretized probability values corresponding to the continuous features, wherein the N discretized probability values correspond to preset N base vector expression values, and N is an integer greater than 1; and finally, the data processing apparatus determining a vector expression value of the continuous features according to the N discretized probability values and the N base vector expression values.

(57) 摘要: 本申请实施例提供了一种数据处理方法以及装置, 用于更好地学习连续特征中每个特征值的向量表达值。具体包括: 该数据处理装置从样本数据中获取连续特征; 然后通过离散化模型对该连续特征进行离散化处理得到该连续特征对应的N个离散化概率值, 其中, 所述N个离散化概率值对应于预设的N个基向量表达值, 该N为大于1的整数; 最后, 该数据处理装置根据该N个离散化概率值和该N个基向量表达值确定该连续特征的一个向量表达值。

[见续页]

WO 2022/116905 A1

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))。