(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
2025 年 10 月 2 日 (02.10.2025)

WIPO | PCT

(10) 国际公布号
WO 2025/200715 A1

(51) 国际专利分类号:
G06F 8/41 (2018.01) *G06N 3/08* (2023.01)
G06F 8/30 (2018.01) *G06N 20/00* (2019.01)

(21) 国际申请号: PCT/CN2025/071616

(22) 国际申请日: 2025 年 1 月 9 日 (09.01.2025)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202410384068.8 2024年3月29日 (29.03.2024) CN

(71) 申请人:阿里巴巴（中国）有限公司(**ALIBABA (CHINA) CO., LTD.**) [CN/CN]; 中国浙江省杭州市滨江区长河街道网商路 699 号 4 号楼 5楼508室 310052 (CN)。

(72) 发明人:曹荣禹(**CAO, Rongyu**); 中国北京市朝阳区广善路18号院-阿里巴巴北京朝阳科技园C区 100020 (CN)。 曹永昌(**CAO, Yongchang**); 中国浙江省杭州市西湖区三墩镇灯彩街1008号云谷园区 310030 (CN)。 刘宇宸(**LIU, Yuchen**); 中国浙江省杭州市西湖区三墩镇灯彩街1008号云谷园区 310030 (CN)。 黎槟华(**LI, Binhua**); 中国北京市朝阳区广善路18号院-阿里巴巴北京朝阳科技园C区 100020 (CN)。 黄非(**HUANG, Fei**); 美国加州桑尼维尔阿尔曼街525号4楼 94085 (US)。 李永彬(**LI, Yongbin**); 中国北京市朝阳区广善路18号院-阿里巴巴北京朝阳科技园C区 100020 (CN)。

(74) 代理人:北京同钧律师事务所(**BEIJING TONGJUN LAW FIRM**); 中国北京市海淀区西直门北大街32号院1号楼7层808室 100082 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA,

(54) **Title:** MODEL TRAINING METHOD, CODE PROCESSING METHOD, AND CODE PROCESSING PLATFORM

(54) 发明名称: 模型训练方法、代码处理方法以及代码处理平台

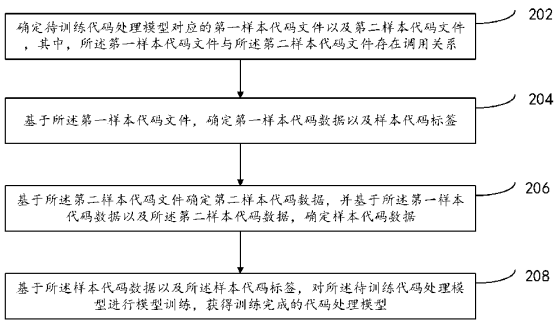


图 2

202 Determine a first sample code file and a second sample code file that correspond to a code processing model to be trained, wherein there is a calling relationship between the first sample code file and the second sample code file
204 On the basis of the first sample code file, determine first sample code data and sample code labels
206 On the basis of the second sample code file, determine second sample code data, and on the basis of the first sample code data and the second sample code data, determine sample code data
208 On the basis of the sample code data and the sample code labels, perform model training on the code processing model to be trained, in order to obtain a trained code processing model

(57) **Abstract:** Provided in the embodiments of the present disclosure are a model training method, a code processing method, and a code processing platform. The model training method comprises: determining a first sample code file and a second sample code file that correspond to a code processing model to be trained, wherein there is a calling relationship between the first sample code file and the second sample code file; on the basis of the first sample code file, determining first sample code data and sample code labels; on the basis of the second sample code file, determining second sample code data, and on the basis of the first sample code data and the second sample code data, determining sample code data; and on the basis of the sample code data and the sample code labels, performing model training on the code processing model to be trained, in order to obtain a trained code processing model. Thus, the trained code processing model can process code data having a relatively complex calling relationship in actual development scenarios, and therefore the performance of the code processing model is improved.

WO 2025/200715 A1

[见续页]

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, UY, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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: 本公开实施例提供模型训练方法、代码处理方法以及代码处理平台，该模型训练方法，包括：确定待训练代码处理模型对应的第一样本代码文件以及第二样本代码文件，其中，所述第一样本代码文件与所述第二样本代码文件存在调用关系；基于所述第一样本代码文件，确定第一样本代码数据以及样本代码标签；基于所述第二样本代码文件确定第二样本代码数据，并基于所述第一样本代码数据以及所述第二样本代码数据，确定样本代码数据；基于所述样本代码数据以及所述样本代码标签，对所述待训练代码处理模型进行模型训练，获得训练完成的代码处理模型；使得训练后的代码处理模型，处理实际开发场景中调用关系较为复杂的代码数据，提高了代码处理模型的性能。