

(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
2023 年 6 月 1 日 (01.06.2023)



(10) 国际公布号
WO 2023/093346 A1

(51) 国际专利分类号:
G06V 10/776 (2022.01) *G06V 10/764* (2022.01)
G06V 10/774 (2022.01) *G06K 9/62* (2022.01)

(21) 国际申请号: PCT/CN2022/125166

(22) 国际申请日: 2022 年 10 月 13 日 (13.10.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202111417245.0 2021 年 11 月 25 日 (25.11.2021) CN

(71) 申请人: 支付宝 (杭州) 信息技术有限公司 (ALIPAY (HANGZHOU) INFORMATION TECHNOLOGY CO., LTD.) [CN/CN]; 中国浙江省杭州市西湖区西溪路 556 号 8 层 B 段 801-11, Zhejiang 310000 (CN)。

(72) 发明人: 李一鸣 (LI, YiMing); 中国浙江省杭州市西湖区西溪路 556 号 8 层 B 段 801-11, Zhejiang

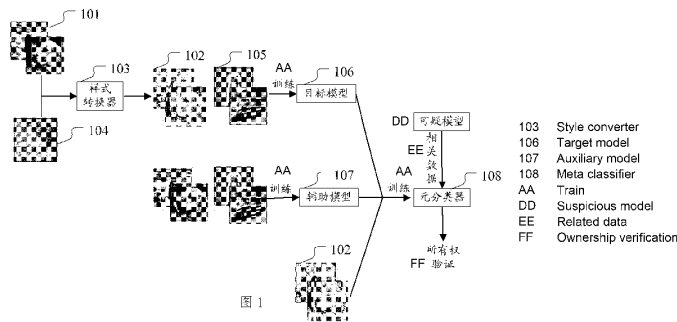
310000 (CN)。朱玲慧 (ZHU, LingHui); 中国浙江省杭州市西湖区西溪路 556 号 8 层 B 段 801-11, Zhejiang 310000 (CN)。邱伟峰 (QIU, WeiFeng); 中国浙江省杭州市西湖区西溪路 556 号 8 层 B 段 801-11, Zhejiang 310000 (CN)。江勇 (JIANG, Yong); 中国浙江省杭州市西湖区西溪路 556 号 8 层 B 段 801-11, Zhejiang 310000 (CN)。夏树涛 (XIA, ShuTao); 中国浙江省杭州市西湖区西溪路 556 号 8 层 B 段 801-11, Zhejiang 310000 (CN)。

(74) 代理人: 北京亿腾知识产权代理事务所 (普通合伙) (E-TONE INTELLECTUAL PROPERTY FIRM (GENERAL PARTNERSHIP)); 中国北京市海淀区中关村紫金数码园 3 号楼 707, Beijing 100190 (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,

(54) Title: EXOGENOUS FEATURE-BASED MODEL OWNERSHIP VERIFICATION METHOD AND APPARATUS

(54) 发明名称: 基于外源特征进行模型所有权验证的方法和装置



(57) Abstract: The embodiments of the present description provide an exogenous feature-based model ownership verification method and apparatus. A specific implementation of the method comprises: selecting samples from an initial sample set to form a selected sample set; processing sample data of each selected sample in the selected sample set to obtain a conversion sample set having exogenous features and composed of conversion samples, wherein the exogenous features are features that are not possessed by the sample data of initial samples; training a meta classifier on the basis of a target model, an auxiliary model, and the conversion sample set, wherein the auxiliary model is a model obtained by performing training by using the initial sample set, the target model is a model obtained by performing training by using the remaining sample sets other than the selected sample set in the conversion sample set and the initial sample set, and the meta classifier is used for identifying feature knowledge of the exogenous features; and inputting related data of a suspicious model into the meta classifier, and determining, on the basis of an output result of the meta classifier, whether the suspicious model is a model stolen from a deployment model, wherein the deployment model has the feature knowledge of the exogenous features.

[见续页]

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, 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, 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: 本说明书实施例提供了一种基于外源特征进行模型所有权验证的方法和装置。该方法的一具体实施方式包括: 从初始样本集选取样本构成选中样本集; 对选中样本集中各选中样本的样本数据进行处理, 得到具有外源特征的转化样本构成的转化样本集, 外源特征为初始样本的样本数据不具备的特征; 基于目标模型、辅助模型和转化样本集, 训练元分类器, 辅助模型为使用初始样本集训练得到的模型, 目标模型为使用转化样本集和初始样本集中除选中样本集之外的剩余样本集训练得到的模型, 元分类器用于识别外源特征的特征知识; 将可疑模型的相关数据输入元分类器, 基于元分类器的输出结果, 确定可疑模型是否为从部署模型窃取的模式, 部署模型具有外源特征的特征知识。