

(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



(10) 国际公布号
WO 2020/140386 A1

(43) 国际公布日
2020 年 7 月 9 日 (09.07.2020)

- (51) 国际专利分类号:
G06F 16/36 (2019.01) *G06N 3/08* (2006.01)
G06N 3/04 (2006.01)
- (21) 国际申请号: PCT/CN2019/089563
- (22) 国际申请日: 2019 年 5 月 31 日 (31.05.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201910002638.1 2019 年 1 月 2 日 (02.01.2019) CN
- (71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。
- (72) 发明人: 金戈(JIN, Ge); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 徐亮(XU, Liang); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 肖京(XIAO, Jing); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。
- (74) 代理人: 北京英特普罗知识产权代理有限公司(INTELLECPRO CHINA LIMITED); 中国北京市西城区车公庄大街 9 号五栋大楼 C 座 11 层, Beijing 100044 (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,

(54) Title: TEXTCNN-BASED KNOWLEDGE EXTRACTION METHOD AND APPARATUS, AND COMPUTER DEVICE AND STORAGE MEDIUM

(54) 发明名称: 基于TextCNN知识抽取方法、装置、计算机设备及存储介质

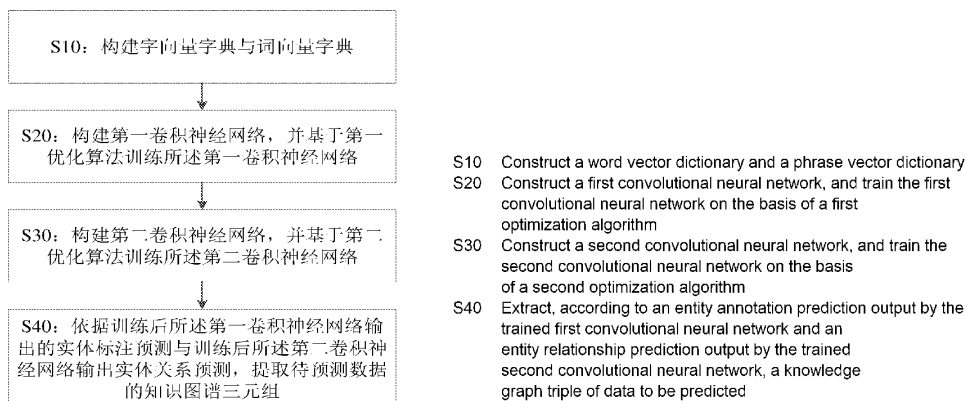


图 1

(57) Abstract: Disclosed is a TextCNN-based knowledge extraction method, comprising: S10, constructing a word vector dictionary and a phrase vector dictionary; S20, constructing a first convolutional neural network, and training the first convolutional neural network on the basis of a first optimization algorithm, wherein the first convolutional neural network comprises a first embedded layer, a first multilayer convolution and a first softmax function successively connected to each other; S30, constructing a second convolutional neural network, and training the second convolutional neural network on the basis of a second optimization algorithm, wherein the second convolutional neural network comprises a second embedded layer, a second multilayer convolution, a pooling layer, two fully connected layers and a second softmax function successively connected to each other; and S40, extracting, according to an entity annotation prediction output by the trained first convolutional neural network and an entity relationship prediction output by the trained second convolutional neural network, a knowledge graph triple of data to be predicted. The convolutional neural network has a parallel computation characteristic, and therefore, computational resources can be fully utilized to realize improvement in the efficiency of

[见续页]

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))。

computation.

(57) 摘要: 一种基于TextCNN的知识抽取方法, 包括: S10: 构建字向量字典与词向量字典; S20: 构建第一卷积神经网络, 并基于第一优化算法训练第一卷积神经网络; 第一卷积神经网络包括依次连接的第一嵌入层、第一多层卷积以及第一softmax函数; S30: 构建第二卷积神经网络, 并基于第二优化算法训练第二卷积神经网络, 第二卷积神经网络包括依次连接的第二嵌入层、第二多层卷积、一池化层、两层全连接层以及第二softmax函数; S40: 依据训练后第一卷积神经网络输出的实体标注预测与训练后第二卷积神经网络输出实体关系预测, 提取待预测数据的知识图谱三元组, 由于卷积神经网络具有并行计算特性, 因此能够充分利用计算资源实现计算效率提升。