

(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 2024/222714 A1

(43) 国际公布日
2024 年 10 月 31 日 (31.10.2024)

(51) 国际专利分类号:
G06F 16/9535 (2019.01)

(21) 国际申请号: PCT/CN2024/089451

(22) 国际申请日: 2024 年 4 月 24 日 (24.04.2024)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
2023 10493430.0 2023 年 4 月 28 日 (28.04.2023) CN

(71) 申请人: 杭州海康威视数字技术股份有限公司 (HANGZHOU HIKVISION DIGITAL TECHNOLOGY CO., LTD.) [CN/CN]; 中国浙江省杭州市滨江区阡陌路 555 号, Zhejiang 310051 (CN)。

(72) 发明人: 陈展 (CHEN, Zhan); 中国浙江省杭州市滨江区阡陌路 555 号, Zhejiang 310051 (CN)。

(74) 代理人: 北京博思佳知识产权代理有限公司 (BEIJING BESTIPR INTELLECTUAL

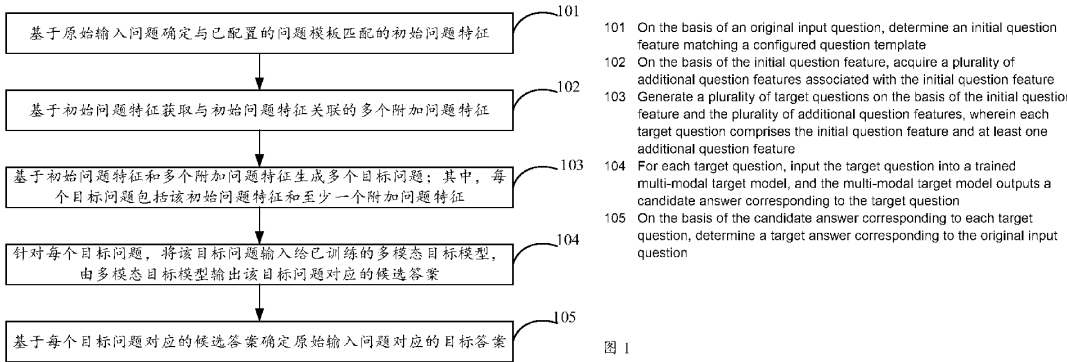
PROPERTY LAW CORPORATION); 中国北京市海淀区上地三街 9 号嘉华大厦 B 座 409, Beijing 100085 (CN)。

(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, 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, 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,

(54) Title: DATA PROCESSING METHOD AND APPARATUS BASED ON MULTI-MODAL MODEL, AND DEVICE AND MEDIUM

(54) 发明名称: 一种基于多模态模型的数据处理方法、装置、设备及介质



(57) Abstract: Provided in the present application are a data processing method and apparatus based on a multi-modal model, and a device and a medium. The method comprises: on the basis of an original input question, determining an initial question feature matching a configured question template; on the basis of the initial question feature, acquiring a plurality of additional question features associated with the initial question feature; generating a plurality of target questions on the basis of the initial question feature and the plurality of additional question features, wherein each target question comprises the initial question feature and at least one additional question feature; for each target question, inputting the target question into a trained multi-modal target model, and the multi-modal target model outputting a candidate answer corresponding to the target question; and on the basis of the candidate answer corresponding to each target question, determining a target answer corresponding to the original input question. By means of the technical solution of the present application, computing resources of a server can be reduced, time overheads can also be reduced, and the target answer corresponding to the original input question can be quickly positioned.

[见续页]

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: 本申请提供一种基于多模态模型的数据处理方法、装置、设备及介质, 该方法包括: 基于原始输入问题确定与已配置的问题模板匹配的初始问题特征; 基于所述初始问题特征获取与所述初始问题特征关联的多个附加问题特征; 基于所述初始问题特征和所述多个附加问题特征生成多个目标问题, 每个目标问题包括所述初始问题特征和至少一个附加问题特征; 针对每个目标问题, 将所述目标问题输入给已训练的多模态目标模型, 由所述多模态目标模型输出所述目标问题对应的候选答案; 基于每个目标问题对应的候选答案确定所述原始输入问题对应的目标答案。通过本申请的技术方案, 能够节省服务器的计算资源, 并节省时间开销, 能够快速定位原始输入问题对应的目标答案。