

(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
2024 年 6 月 27 日 (27.06.2024)



(10) 国际公布号
WO 2024/131111 A1

(51) 国际专利分类号:

G06F 40/284 (2020.01)

江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。

(21) 国际申请号:

PCT/CN2023/114518

(22) 国际申请日:

2023 年 8 月 23 日 (23.08.2023)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

202211660381.7 2022年12月23日 (23.12.2022) CN

(71) 申请人: 苏州元脑智能科技有限公司 (SUZHOU METABRAIN INTELLIGENT TECHNOLOGY CO., LTD.) [CN/CN]; 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。

(72) 发明人: 李峰 (LI, Feng); 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。 刘红丽 (LIU, Hongli); 中国

(74) 代理人: 北京康信知识产权代理有限公司 (KANGXIN PARTNERS, P.C.); 中国北京市海淀区知春路甲48号盈都大厦A座16层, Beijing 100098 (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

(54) Title: INTELLIGENT WRITING METHOD AND APPARATUS, DEVICE, AND NONVOLATILE READABLE STORAGE MEDIUM

(54) 发明名称: 一种智能写作方法、装置、设备及非易失性可读存储介质

生成文章大纲: 所述文章大纲包括所有段落的主
题句 S11

将所述文章大纲中的任主题句与从内容和知识库
中取出的与该主题句相关的段落组成目标文本 S12

将目标文本输入预训练模型, 得到字典中每个词
为生成词的概率 S13

基于惩罚因子对所述字典中目标词对应的所述概
率进行惩罚, 基于所述惩罚因子从所述字典中取出生
成词 S14

基于所述生成词更新所述目标文本, 得到新的目
标文本 S15

根据新的目标文本计算多样性评价指标, 并基于
该多样性评价指标更新所述惩罚因子, 并得到新的
目标文本输入预训练模型, 不断迭代, 直到取出
的生成词为截止符, 则停止当前的生成词生成, 并
将所述生成词对应的所述惩罚因子 S16

基于任意主题句对应的所述惩罚因子确定该主题
句的最优段落 S17

(57) Abstract: The present application discloses an intelligent writing method and apparatus, a device, and a nonvolatile readable storage medium. The method comprises: generating an article outline; combining any topic sentence from the article outline with a paragraph retrieved from a content knowledge base and related to the topic sentence to form a target text; inputting the target text into a pre-trained model, to obtain a probability that each word in a dictionary is a generated word; penalizing the probability corresponding to a target word in the dictionary on the basis of a penalty factor, and selecting a generated word from the dictionary on the basis of the probability; updating the target text on the basis of the generated word; calculating a diversity evaluation index according to the new target text, updating the penalty factor on the basis of the diversity evaluation index, inputting the new target text into the pre-trained model, continuously iterating until a retrieved generated word is a stop symbol, and determining the current target text to be a candidate paragraph; determining an optimal paragraph on the basis of the candidate paragraph of any topic sentence. The present method can reduce labor costs and improve the accuracy of parameter adjustment while resolving the problem of text repetition.

- S11 Generate an article outline, the article outline comprising the topic sentence of every paragraph
- S12 Combine any topic sentence from the article outline with a paragraph retrieved from a content knowledge base and related to the topic sentence, to form a target text
- S13 Input the target text into a pre-trained model, to obtain a probability that each word in a dictionary is a generated word
- S14 Penalize the probability corresponding to a target word in the dictionary on the basis of a penalty factor, and select a generated word from the dictionary on the basis of the probability
- S15 Update the target text on the basis of the generated word, to obtain a new target text
- S16 Calculate a diversity evaluation index according to the new target text, update the penalty factor on the basis of the diversity evaluation index, input the new target text into the pre-trained model, continuously iterate until a retrieved generated word is a stop symbol, and determine the current target text to be a candidate paragraph corresponding to the topic sentence
- S17 On the basis of the candidate paragraph corresponding to any topic sentence, determine an optimal paragraph of the topic sentence

FIG. 1

WO 2024/131111 A1

[见续页]

(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: 本申请公开了一种智能写作方法、装置、设备及非易失性可读存储介质, 包括: 生成文章大纲; 将文章大纲中的任意主题句与从内容知识库中取出的与该主题句相关的段落组成目标文本; 将目标文本输入预训练模型, 得到字典中每个词为生成词的概率; 基于惩罚因子对字典中目标词对应的概率进行惩罚, 基于概率从字典中取出生成词; 基于生成词更新目标文本; 根据新的目标文本计算多样性评价指标, 并基于该多样性评价指标更新惩罚因子, 并将新的目标文本输入预训练模型, 不断迭代, 直到取出的生成词为截止符, 则将当前的目标文本确定为候选段落; 基于任意主题句的候选段落确定最优段落。能够在解决文本重复问题的情况下, 减少人工成本并提升参数调节的准确度。

一种智能写作方法、装置、设备及非易失性可读存储介质

相关申请的交叉引用

本申请要求于 2022 年 12 月 23 日提交中国专利局，申请号为 202211660381 .7，申请名称为“一种智能写作方法、装置、设备及介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及智能写作技术领域，特别涉及一种智能写作方法、装置、设备及非易失性可读存储介质。

背景技术

在日常办公等写作领域，资料的查找、收集、分类、引用都较为繁琐，耗费大量的人力和时间。智能写作能满足这一领域对于劳动力解放的需求。近年来，智能写作已经从规则、模板写作发展到了以深度神经网络模型为核心的智能化创作，从辅助创作逐渐走向自动化，新的预训练语言模型被不断提出，如 openAI（Open Artificial Intelligence，开放人工智能）的 GPT（即 Generative Pre-Training，生成式预训练）-3、ERNIE（Enhanced Representation through kNowledge IntEgration，基于知识增强的表示学习）、T5 等模型，技术人员试图利用这些大模型的原始生成能力来解决不同场景下文本生成问题。

目前，基于预训练模型生成短文本的能力尚可，但是生成长文本（如长段落或篇章）的能力欠缺，比较常见的是文本重复问题，现有的解决文本重复的问题通常是人工进行参数调节，这种方式比较耗时，人工成本较高，难以保障准确性。

发明内容

有鉴于此，本申请的目的在于提供一种智能写作方法、装置、设备及非易失性可读存储介质，能够在解决文本重复问题的情况下，减少人工成本并提升参数调节的准确度。其可选方案如下：

根据第一方面，本申请公开了一种智能写作方法，包括：

生成文章大纲；文章大纲包括所有段落的主题句；

将文章大纲中的任意主题句与从内容知识库中取出的与该主题句相关的段落组成目标文本；

将目标文本输入预训练模型，得到字典中每个词为生成词的概率；

基于惩罚因子对字典中目标词对应的概率进行惩罚，基于概率从字典中取出生成词；其中，目标词为在目标文本中存在的词；

基于生成词更新目标文本，得到新的目标文本；

根据新的目标文本计算多样性评价指标，并基于该多样性评价指标更新惩罚因子，并将新的目标文本输入预训练模型，不断迭代，直到取出的生成词为截止符，则将当前的目标文本确定为该主题句对应的候选段落；

基于任意主题句对应的候选段落确定该主题句的最优段落。

可选的，生成文章大纲，包括：

基于用户选择的写作类型、输入的写作主题内容以及第一关键词生成文章大纲。

可选的，基于用户选择的写作类型、输入的写作主题内容以及第一关键词生成文章大纲，包括：

从用户选择的写作类型对应的内容知识库中检索与写作主题内容最相关的模板文件；

将模板文件中每个段落输入预训练模型，生成每个段落的主题句以及第二关键词；

分别将每个段落的主题句以及第二关键词、第一关键词按照预设方式组成模型输入数据，并将模型输入数据输入预训练模型，生成每个段落对应的新的主题句；

将每个段落的新的主题句组合，得到文章大纲。

可选的，从用户选择的写作类型对应的内容知识库中检索与写作主题内容最相关的模板文件，包括：

从用户选择的写作类型对应文件名知识库中检索与写作主题内容最相关的文件名；

从用户选择的写作类型对应的内容知识库中查找该文件名对应的模板文件。

可选的，还包括：

创建知识库池；其中，知识库池中包括各写作类型对应的文件名知识库和内容知识库；内容知识库包括各模板文件，文件名知识库包括各模板文件的文件名。

可选的，基于惩罚因子对字典中目标词对应的概率进行惩罚，包括：

基于惩罚因子调整参数以及惩罚因子对字典中目标词对应的概率进行惩罚。

可选的，还包括：

统计目标词在目标文本中出现的次数，得到该目标词对应的惩罚因子调整参数；

或，获取配置参数，作为目标词对应的惩罚因子调整参数。

可选的，基于惩罚因子调整参数以及惩罚因子对字典中目标词对应的概率进行惩罚，包括：

根据公式 $p_{x_i} = \frac{p_{x_i}}{p^n}$, $if x_i \in I_{in}$ 对字典中目标词对应的概率进行惩罚，其中， x_i 为目标词， I_{in} 为目标文本，在目标词在目标文本中的情况下，对目标词对应的概率 p_{x_i} 进行惩罚， n 用于指示目标词在目标文本中出现的次数或者常数 1， P 为惩罚因子。

可选的，基于概率从字典中取出生成词，包括：

基于解码规则从字典中基于概率选取生成词，其中，概率为进行了惩罚后的概率，解码规则包括温度 Temperature 采样、TopK 采样策略和 TopP 采样策略。

可选的，根据新的目标文本计算多样性评价指标，包括：

从新的目标文本中去除原始目标文本，得到计算文本；原始目标文本为文本大纲中的主题句与从内容知识库中取出的段落组成的目标文本；

根据计算文本计算多样性评价指标。

可选的，基于该多样性评价指标更新惩罚因子，包括：

计算该多样性评价指标以及上一次迭代得到的多样性评价指标之间的指标差值；

基于指标差值更新惩罚因子。

可选的，基于指标差值更新惩罚因子，包括：

根据公式 $P = \max(1, P + \Delta d)$ 更新惩罚因子，其中，P 为更新后的惩罚因子， Δd 为多样性评价指标以及上一次迭代得到的多样性评价指标之间的指标差值。

可选的，在基于任意主题句对应的候选段落确定该主题句的最优段落之前，还包括：

计算任意主题句对应的候选段落中每句话与上一句话以及该主题句之间的语义相似度，得到第一语义相似度和第二语义相似度；

若第一语义相似度和第二语义相似度均大于预设相似度阈值，则保留该句话，否则删除该句话以及该句话后面的文本，得到删除后文本；

若删除后文本的字数大于预设字数阈值，则将该删除后文本作为该主题句对应的候选段落，否则基于删除后文本重新生成该主题句对应的候选段落。

可选的，计算任意主题句对应的候选段落中每句话与上一句话以及该主题句之间的语义相似度，得到第一语义相似度和第二语义相似度，包括：

对句子进行分词，其中，句子包括：每句话，上一句话和主题句；

获取每个句子分词后的每一个词对应的向量，并将获取到的向量相加并求平均，得到每个句子对应的句子向量；

利用公式 $\cos(\theta) = \frac{\sum_{i=1}^n (x_i \times y_i)}{\sqrt{\sum_{i=1}^n (x_i)^2} \times \sqrt{\sum_{i=1}^n (y_i)^2}}$ 计算每句话与上一句话之间的夹角余弦值作为第一语义相似度，并利用公式 $\cos(\theta) = \frac{\sum_{i=1}^n (x_i \times y_i)}{\sqrt{\sum_{i=1}^n (x_i)^2} \times \sqrt{\sum_{i=1}^n (y_i)^2}}$ 计算每句话与主题句之间的夹角余弦值作为第二语义相似度，其中， θ 为向量 x 和向量 y 之间的夹角， x_i 表示向量 x 的分量， y_i 表示向量 y 的分量。

可选的，基于任意主题句对应的候选段落确定该主题句的最优段落，包括：

对于任意主题句，基于预设评分标准分别对每个候选段落打分，并将分数最高的候选段落确定为该主题句对应的最优段落。

可选的，基于预设评分标准分别对每个候选段落打分，包括：

根据公式 $Score_{PPL} = 2^{-\frac{1}{N} \sum_{i=1}^N \log_2 P(x_i | x_1, x_2, \dots, x_{i-1})}$ 计算每个候选段落的分值 $Score_{PPL}$ ，其中， $P(x_i | x_1, x_2, \dots, x_{i-1})$ 表示根据上文词语预测第 i 个词的概率，PPL 表示语言模型中的困惑度 perplexity，N 表示句子长度， $Score_{PPL}$ 的值越小表示生成的文本越自然、语句越通顺。

可选的，基于任意主题句对应的候选段落确定该主题句的最优段落之后，还包括：

利用预设检测和修改逻辑对最优段落进行错误检测和修改。

根据第二方面，本申请公开了一种智能写作装置，包括：

文章大纲生成模块，被设置为生成文章大纲；文章大纲包括所有段落的主题句；

目标文本构建模块，被设置为将文章大纲中的任意主题句与从内容知识库中取出的与该主题句相关的段落组成目标文本；

生成词概率获取模块，被设置为将目标文本输入预训练模型，得到字典中每个词为生成词的概率；

目标词概率惩罚模块，被设置为基于惩罚因子对字典中目标词对应的概率进行惩罚；其中，目标词为在目标文本中存在的词；

生成词取出模块，被设置为基于概率从字典中取出生成词；

目标文本更新模块，被设置为基于生成词更新目标文本，得到新的目标文本；

多样性评价指标计算模块，被设置为根据新的目标文本计算多样性评价指标；

惩罚因子更新模块，被设置为基于该多样性评价指标更新惩罚因子，相应的，装置还被设置为触发生成词概率获取模块将新的目标文本输入预训练模型，不断迭代，直到生成词取出模块取出的生成词为截止符，则将当前的目标文本确定为该主题句对应的候选段落；

最优段落确定模块，被设置为基于任意主题句对应的候选段落确定该主题句的最优段落。

根据第三方面，本申请公开了一种电子设备，包括存储器和处理器，其中：

存储器，被设置为保存计算机程序；

处理器，被设置为执行计算机程序，以实现前述的智能写作方法。

根据第四方面，本申请公开了一种计算机非易失性可读存储介质，被设置为保存计算机程序，其中，计算机程序被处理器执行时实现前述的智能写作方法。

可见，本申请先生成文章大纲，文章大纲包括所有段落的主题句，之后将文章大纲中的任意主题句与从内容知识库中取出的与该主题句相关的段落组成目标文本，并将目标文本输入预训练模型，得到字典中每个词为生成词的概率，基于惩罚因子对字典中目标词对应的概率进行惩罚，基于概率从字典中取出生成词，目标词为在目标文本中存在的词，基于生成词更新目标文本，得到新的目标文本，根据新的目标文本计算多样性评价指标，并基于该多样性评价指标更新惩罚因子，并将新的目标文本输入预训练模型，不断迭代，直到取出的生成词为截止符，则将当前的目标文本确定为该主题句对应的候选段落，之后基于任意主题句对应的候选段落确定该主题句的最优段落。也即，本申请先生成文章大纲，对于文章大纲中的任意主题句，生成候选段落并从候选段落中选出最佳段落，在生成候选段落的过程中，利用多样性评价指标指导重复度的惩罚因子自适应的调整，能够在解决文本重复问题的情况下，减少人工成本并提升参数调节的准确度。

附图说明

或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据提供的附图获得其他的附图。

图1为本申请公开的一种智能写作方法流程图；

图2为本申请公开的一种可选的知识库池示意图；

图3为本申请公开的一种可选的最优段落生成示意图；

图4为本申请公开的一种可选的候选段落生成示意图；

图 5 为本申请公开的一种智能写作装置结构示意图；

图 6 为本申请公开的一种电子设备结构图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

随着自然语言处理技术的不断发展，人工智能从“大炼模型”逐步迈向了“炼大模型”的阶段，利用先进的算法，整合大规模的数据，汇聚大量算力，训练出巨量人工智能模型，如 GPT（Generative Pre-Training，生成式预训练）-3、ERNIE（Enhanced Representation through kNowledge IntEgration，基于知识增强的表示学习）、T5 等。这些预训练模型在不同的应用场景下都具有较好原始生成能力。其中，GPT3 是一种自回归生成式的预训练语言模型，可以根据给定的文本信息来生成具有一定语义相关度的文本信息，例如回答问题，续写文章等。GPT 主要利用了 Transformer（即转换器）的解码器部分来进行生成式的单向语言模型建模，根据已经出现的若干词语来预测当前输出词的概率分布。

目前，基于预训练模型生成短文本的能力尚可，但是生成长文本（如长段落或篇章）的能力欠缺，比较常见的是文本重复问题，现有的解决文本重复的问题通常是人工进行参数调节，这种方式比较耗时，人工成本较高，难以保障准确性。为此，本申请提供了一种智能写作方案，能够在解决文本重复问题的情况下，减少人工成本并提升参数调节的准确度。

参见图 1 所示，本申请实施例公开了一种智能写作方法，包括：

步骤 S11：生成文章大纲；文章大纲包括所有段落的主题句。

在一种实施方式中，可以基于用户选择的写作类型、输入的写作主题内容以及第一关键词生成文章大纲。可选的，可以从用户选择的写作类型对应的内容知识库中检索与写作主题内容最相关的模板文件；将模板文件中每个段落输入预训练模型，生成每个段落的主题句以及第二关键词；分别将每个段落的主题句以及第二关键词、第一关键词按照预设方式组成模型输入数据，并将模型输入数据输入预训练模型，生成每个段落对应的新的主题句；将每个段落的新的主题句组合，得到文章大纲。可以理解的是，模型输入数据为 prompt（即提示）。

其中，本申请实施例可以从用户选择的写作类型对应文件名知识库中检索与写作主题内容最相关的文件名；从用户选择的写作类型对应的内容知识库中查找该文件名对应的模板文件。在一种实施方式中，可以利用 bm25 算法从用户选择的写作类型对应文件名知识库中检索与写作主题内容最相关的文件名。

并且，本申请实施例可以创建知识库池；其中，知识库池中包括各写作类型对应的文件名知识库和内容知识库；内容知识库包括各模板文件，文件名知识库包括各模板文件的文件名。模板文件可以为 txt 文件。

本申请实施例可以根据用户提供或通过网络获取的最新参考材料构建多场景知识库池，这样，利用最新参考资料构建知识库，检索相关知识，并与主题句组成 prompt，输入预训练模型生成文本，使预训练模型学会最新知识并能对其进行分析，解决预训练模型不能跟进最新动态发展的问题。例如，参见图 2 所示，图 2 为本申请实施例公开的一种可选的知识库池示意图。本根据不同的应用场景对知识库分类构建，可以包括公文写作、新闻资讯、广告营销等多个写作类型，每个写作类型包括 txt（一种用于存储文本数据的文本文件）文件名知识库和 txt 内容知识库；每个 txt 的文件名与其内容的标题一致，txt 文件名知识库就是由所有标题构建的；txt 内容知识库

由所有 txt 文件内容组成，每个 txt 是一篇最新参考文章，文章中的一段占 txt 中的一行。

可选的，自动生成文章大纲的实现流程可以为：输入用户选择的写作类型、想写作的主题内容以及关键词 K_{new} （即第一关键词）。根据主题内容在写作类型下的 txt 文件名知识库中检索最相关的一条，根据检索到的文件名在 txt 内容知识库中找到对应的模板文件。（本申请实施例可以通过 BM25Okapi 类中 get_top_n 函数实现），之后将模板文件按段输入预训练模型，生成每段的主题句及关键词 K_{old} （即第二关键词）。分别将旧关键词 K_{old} 、每段主题句与新关键词 K_{new} 按一定方式组合为 prompt，输入预训练模型，生成新的主题句。最后，将所有段落主题句组合在一起作为文章大纲。例如，组成 prompt 的方法为：

以“旧关键词”为关键词的主题句：“旧主题句”。以“新关键词”为关键词的主题句：

需要指出的是，bm25 算法是一种用来评价搜索词和文档之间相关性的算法，它是一种基于概率检索模型提出的算法，例如，有一个 query（询问）和一批文档 Ds （Dataset，数据集），现在要计算 query 和每篇文档 D 之间的相关性分数，先对 query 进行切分，得到单词，然后单词的分数由 3 部分组成：单词和 D 之间的相关性；单词和 query 之间的相关性；每个单词的权重。最后对于每个单词的分数做一个求和，就得到了 query 和文档之间的分数。

步骤 S12：将文章大纲中的任意主题句与从内容知识库中取出的与该主题句相关的段落组成目标文本。

也即，本申请实施例对文章大纲中的每个主题句分别处理，分别得到每个主题句对应的目标文本，以得到每个主题句的最佳段落。

在可选的实施方式中，将文章大纲中的任意主题句与从用户选择的写作类型对应的内容知识库中取出的与该主题句相关的段落组成目标文本。可以理解的是，内容知识库为预先创建的保存模板文件的知识库。并且，可以检索出最相关的 n 个段落， n 为大于等于 1 的正整数，比如 1、2、3、4、5 等。 n 个段落中每个与该主题句相关的段落分别与该主题句组成目标文本，以生成每个相关的段落对应的候选段落，得到 n 个候选段落。采用的检索方法可以为 bm25 算法。

步骤 S13：将目标文本输入预训练模型，得到字典中每个词为生成词的概率。

步骤 S14：基于惩罚因子对字典中目标词对应的概率进行惩罚，基于概率从字典中取出生成词；其中，目标词为在目标文本中存在的词。

在可选的实施方式中，可以基于惩罚因子调整参数以及惩罚因子对字典中目标词对应的概率进行惩罚。并且，在一种实施方式中，可以统计目标词在目标文本中出现的次数，得到该目标词对应的惩罚因子调整参数，在另一种实施方式中，可以获取配置参数，作为目标词对应的惩罚因子调整参数。也即，惩罚因子调整参数可以为目标词在目标文本中出现的次数，也可以配置的常数。

并且，可以采用解码规则（比如 Temperature（温度）、TopK、TopP 采样策略）从字典中基于概率选取生成词。选取时依据的概率为惩罚后的概率。

需要指出的是，重复惩罚的目的在于解决重复问题，通过惩罚因子将出现过的词的概率变小或者强制不使用重复词来解决。合理设置惩罚因子能够避免重复，增加创新性，但是设置过大，生成长文本时起反作用。可以基于词出现的次数进行惩罚（frequency penalty），出现的次数越多，该词在后文出现的概率越低，同时增强后文的创造性。也可以基于词是否出现过进行惩罚（response penalty），对出现过的词进行惩罚，降低其在今后出现的概率，同时增强后文的创造性。与 frequency penalty 原理一致，只是多了对已输出 token（词）去重，只保留出现一次。也可以重复词去除，大于等于 1 时起作用，表示输出中不包含长度为 no_repeat_ngram_size 的重复词。

可选的，为了解决搜索生成缺乏多样性问题，可以通过采样策略来增加随机性。其中，Temperature 采样：通过强化顶部词的概率，只对最有可能的一些词进行采样，这样就能在增加随机性的同时，又保证不出现一般性的错误。强化顶部词概率，可以通过对模型输出的 logits 除以一个小于 1 的温度（Temperature，T）。

$$p(i) = \frac{e^{\frac{f(i)}{T}}}{\sum_j e^{\frac{f(j)}{T}}}$$

$f(i)$ 表示模型输出第 i 个词的初始概率， $\sum_j e^{\frac{f(j)}{T}}$ 表示对词表中所有词都计算求和， $p(i)$ 表示强化后第 i 个词的概率。这样通过 softmax（用于将一组实数转化为概率分布）后使得分布更加尖锐，大概率词的概率更大。之后根据获得概率对顶部词先进行挑选，然后再采样，直接杜绝了低概率词出现的可能性。

TopK 采样：挑选概率最高 k 个 token，然后重新过 softmax 算概率，之后根据获得概率进行采样，接着进行下一步生成，不断重复。但关于 TopK 有可能会出现一个问题，假如模型对当前生成非常肯定，比如概率最高的 token 的概率 0.9，而其余的 token 概率都很低。这时如果只用 topk 采样的话，就会导致采样到低概率情况仍然发生。因此需要对顶部 token 的累计概率进行限制，这就是 TopP 采样。

TopP 采样：先设置一个概率界限，比如说 $p=0.9$ ，然后从最大概率的 token 往下开始取，同时将概率累加起来，当取到大于等于 p 也就是 0.9 时停止。如果最大 token 概率就已经有 0.9 了，那么就只取最大的一个 token。

步骤 S15：基于生成词更新目标文本，得到新的目标文本。

步骤 S16：根据新的目标文本计算多样性评价指标，并基于该多样性评价指标更新惩罚因子，并将新的目标文本输入预训练模型，不断迭代，直到取出的生成词为截止符，则将当前的目标文本确定为该主题句对应的候选段落。

在可选的实施方式中，可以从新的目标文本中去除原始目标文本，得到计算文本；原始目标文本为文本大纲中的主题句与从内容知识库中取出的段落组成的目标文本；根据计算文本计算多样性评价指标。

并且，可以计算该多样性评价指标以及上一次迭代得到的多样性评价指标之间的指标差值；基于指标差值更新惩罚因子。

需要指出的是，在 NLP（即 Natural Language Processing，自然语言处理）领域中，多样性评价指标（Distinct）用于判断机器回复的多样性，Distinct 指标判断是否出现大量的通用性、重复性回复。Distinct 的定义如下： $\text{Distinct}(n) = (\text{Count}(\text{unique ngram})) / \text{Count}(\text{word})$ ，其中， $\text{Count}(\text{unique ngram})$ 表示回复中不重复的 ngram 数量， $\text{Count}(\text{word})$ 表示回复中 ngram 词语的总数量。Distinct(n) 越大表示生成回复的多样性越高。本申请实施例利用多样性评价指标调节惩罚因子。

可选的，本申请实施例可以计算任意主题句对应的候选段落中每句话与上一句话以及该主题句之间的语义相似度，得到第一语义相似度和第二语义相似度；若第一语义相似度和第二语义相似度均大于预设相似度阈值，则保留该句话，否则删除该句话以及该句话后面的文本，得到删除后文本；若删除后文本的字数大于预设字数阈值，则将该删除后文本作为该主题句对应的候选段落，否则基于删除后文本重新生成该主题句对应的候选段落。可以理解的是，若候选段落为多个，分别执行前述步骤。也即，本申请实施例利用语义相似度检测并删除偏离主题的主题文本，重新生成符合条件的候选段落。

其中，语义相似度的计算流程为：首先对句子进行分词，然后对分好的每一个词获取其对应的 Vector（向量），然后将所有 Vector 相加并求平均，得到句子 Vector，最后再利用如下公式计算其夹角余弦值即可。余弦值越接近 1（即夹角越小）表示相似度越高。

$$\cos(\theta) = \frac{\sum_{i=1}^n (x_i \times y_i)}{\sqrt{\sum_{i=1}^n (x_i)^2} \times \sqrt{\sum_{i=1}^n (y_i)^2}}$$

其中， θ 为向量 x 和 y 之间的夹角， x_i 和 y_i 分别代表向量 x 和 y 的各分量，可以利用 word2vec（产生词向量的模型）模型计算 Vector，是将所有的词向量化，这样词与词之间就可以定量的去度量他们之间的关系，挖掘词之间的联系。

步骤 S17：基于任意主题句对应的候选段落确定该主题句的最优段落。

在可选的实施方式中，对于任意主题句，基于预设评分标准分别对每个候选段落打分，并将分数最高的候选段落确定为该主题句对应的最优段落。

在一种实施方式中，本申请实施例可以采用 PPL 指标分别对每个候选段落打分。PPL 指的是语言模型中的 perplexity（即困惑度），是衡量一句话是否通顺的指标。定义为：

$$Score_{PPL} = 2^{-\frac{1}{N} \sum_{i=1}^N \log_2 P(x_i | x_1, x_2, \dots, x_{i-1})}$$

其中， $P(x_i | x_1, x_2, \dots, x_{i-1})$ 表示根据上文词语预测第 i 个词的概率， N 代表句子长度。PPL 值越小，说明模型生成的文本越自然、语句越通顺。通过 PPL 来评价文本质量，可以避免模型生成的文本有乱序、前后颠倒的情形。

可选的，本申请实施例可以利用预设检测和修改逻辑对最优段落进行错误检测和修改。

例如，参见图 3 所示，图 3 为本申请实施例公开的一种可选的最优段落生成示意图。包括以下步骤：

- (a) 输入用户选择的写作类型、所有段落的主题句、所有段落主题句的数目 N 、 i 的初始值 1；
- (b) 判断 $i < N+1$ ，若 $i < N+1$ 则获取当前第 i 个主题句，若 i 大于或者等于 $N+1$ 转到步骤 (g)；
- (c) 采用 bm25 算法在写作类型对应的 txt 内容知识库中检索 topn（即最相关的 n 个）相关段落；
- (d) 将 topn 相关段落分别与主题句拼接组成 prompt；
- (e) 将 n 个 prompt 输入预训练模型生成 n 个符合主题的备选段落；
- (f) 分别对 n 个备选段落基于 PPL 指标（ $1/Score_{PPL}$ ）进行打分，选择最高分对应段落作为最佳段落输出至结果列表中， $i+=1$ ，转至步骤 (b)；
- (g) 结果列表输出保存成 txt 文件。最终，结果列表中包括文章大纲中所有主题句对应的最佳段落。

可选的，前述步骤(e)中主题句+相关段落组成 prompt 输入预训练模型，生成文本容易出现重复问题，包括生成内容本身重复和与输入重复两方面。本申请实施例采用惩罚参数自适应调整来解决该问题。参见图 4 所示，图 4 为本申请实施例公开的一种可选的候选段落生成示意图。包括以下步骤：

- (1) 输入文本（即目标文本） I_{in} ，初始值为相关段落+主题句组成的 prompt I_0 ；
- (2) 文本输入预训练模型，输出字典中所有词的概率；
- (3) 对输入文本对应词进行惩罚，采用公式如下：

$$p_{x_i} = \frac{p_{x_i}}{p^n}, \text{ if } x_i \in I_{in}$$

其中， x_i 表示字典中任意一个词，若该词在输入文本 I_{in} 中，则对该词对应的概率 p_{x_i} 进行惩罚， n 表示 x_i 在输入文本中出现的次数或者常数 1 均可（根据不同场景设置）， P 是惩罚因子，初始值为 1；

- (4) 根据解码规则（可以为 Temperature、TopK、TopP 采样策略）从字典中选取生成词 t_{new} ；
- (5) 更新输入文本 $I_{in} += t_{new}$ ，即将新生成词拼接在原输入文本后即可；根据新生成文本（即 I_{in} 中去掉初始值 I_0 ）计算 distinct（即多样性）指标， $d_{new} = (Distinct(m) + Distinct(m+1)) / 2$ ，其中， m 表示 ngram 的大小，需要根据不同场景设置不同值；
- (6) 计算 Δd ， $\Delta d = d_{prev} - d_{new}$ ，其中， d_{prev} 初始值为 1， $d_{prev} = d_{new}$ ；
- (7) 更新惩罚因子 P ， $P = \max(1, P + \Delta d)$ ；
- (8) 继续步骤（2）直到生成截止符<eod>。

需要指出的是，其它的如 Temperature、TopK、TopP 等参数也可参考惩罚因子的调整方案采用自适应调整策略。

可选的，步骤（e）还需要解决长文本主题偏移问题，分别对所有备选段落进行如下操作：（A）获取一个备选答案；（B）利用训练好的 Word2vec 模型计算候选答案中的每句话与主题句以及该句话前一句的语义相似度；（C）若两个相似度均大于设定阈值 S 则保留，否则删除该句及后面的文本；（D）统计备选段落字数，若大于设定阈值 T 则保留输出答案，否则基于剩余文本重新输入预训练模型生成，继续步骤（B）。

并且，智能写作生成的文章可能包含错别字、敏感词以及语法错误等，本申请实施例可以检测并修改这些错误，最后人工对文章审核、发布。

需要指出的是，现有技术中，直接使用预训练语言模型不能跟进最新动态发展，由于预训练模型参数量巨大，即使 finetune（微调）需要的数据和算力成本也会比较高。并且，基于预训练模型生成短文本（如一句话）的能力尚可，但是生成长文本（如长段落或篇章）的能力欠缺，最常见的是主题偏移和文本重复问题。而本申请利用最新参考资料构建知识库，检索相关知识，并与主题句组成 prompt，输入预训练模型生成文本，通过该方法使预训练模型学会最新知识并能对其进行分析；利用文本多样性评价指标指导重复度惩罚参数自适应调整，节约人工调参的成本；利用语义相似度检测并删除偏离主题的主题句，重新生成符合条件的候选段落。

下面，以公文写作为例，阐述本申请提供的智能写作方案：

首先，构建公文写作知识库，以 GPT3 为例进行说明，GPT3 于 2020 年发布，那么训练它的数据集是收集 2020 年及以前的，那么要让 GPT3 基于 2021 年到现在的最新动态生成文章，需要将相关资料构建知识库。用户想通过智能写作自动撰写某公司的某个经济专项规划，收集最新材料。将各种形式的材料（PDF、Word 或图片）转换成 txt 文本，并构建公文写作的文件名知识库和内容知识库。数据库的构建方法可以采用 Python 软件包 rank_bm25 中 BM25Okapi 类实现。

可选的，自动生成文章大纲：用户选择公文写作类型，输入想写作的主题内容：请根据数字经济发展规划材料，撰写 XX 公司经济专项规划。通过 BM25 算法根据主题内容在文件名知识库中检索最相关的一条文件名，将对应的 txt 文件按段输入预训练模型，生成每段的主题句。然后分别将每段主题句与关键字组合 prompt 输入预训练模型，生成新的主题句。

Prompt 输入举例：以“智慧交通”为关键词的主题句：推进智慧交通建设。以“云中心”为关键词的主题句：

预训练模型生成：推进云中心建设。

由于自动生成的文章大纲整体与检索模板一致，而实际应用中需要的大纲与模板还是有一定差距，需要人工干预修改。

然后生成最佳段落：假设某段主题为：A。可以采用 bm25 算法在文本知识库中检索获取得分最高的三个段

落: top1、top2、top3。分别将 top1、top2、top3 与主题句拼接组成 prompt, 输入预训练模型, 输出 3 个候选答案(即候选段落)。采用 PPL 指标作为评分标准, 选取最高分者作为最优答案。在生成候选答案的过程中, 采用惩罚参数自适应调整方法解决生成重复问题。其主要原理为当生成内容重复度升高时, 增大惩罚因子, 即加大模型对已生成词的惩罚力度, 模型生成偏向未出现过的词。当生成内容重复度降低时, 减小惩罚因子, 即减小模型对已生成词的惩罚力度, 防止模型生成因惩罚力度过大导致偏离主题。在生成候选答案的过程中, 利用训练好的 Word2vec 模型计算答案中的每句话与主题句的语义相似度, 删除跑题语句, 并补充符合语义的语句。

可选的, 人工审核、发布, 使用 GitHub 开源项目 pycorrector (文本纠错开源工具) 作为后处理模块, 实现对生成的文章的错别字、敏感词、语法错误检测和修改;

需要注意的是, 本实施例以公文写作为例进行说明, 但在实际应用中不仅局限于此, 其它诸如新闻资讯、故事续写等领域的写作均可以此方式进行改进。

这样, 为了解决基于预训练模型生成的文本内容缺乏对最新动态的分析问题, 本申请基于最新参考资料构建知识库, 通过检索算法检索与主题句相关知识, 并与主题句组成 prompt, 输入预训练模型生成文本, 通过此方案使模型学习与主题相关的最新知识。为了解决预训练模型生成文本容易重复但人工调参难问题, 利用文本多样性评价指标指导重复度惩罚参数自适应调整, 增大生成内容的多样性, 减少人工调参的耗时费力, 保证生成的文本既不重复又不会矫枉过正。为了解决预训练模型生成文本易偏离主题问题, 本申请将生成的每句话分别与主题句、前一句话进行语义相似度匹配, 删除偏离主题语句, 生成符合条件答案。可选的, 还细化了智能写作的整体实现方案及系统, 提高智能写作的自动化程度。

参见图 5 所示, 本申请实施例提供了一种智能写作装置, 包括:

文章大纲生成模块 11, 被设置为生成文章大纲; 文章大纲包括所有段落的主题句;

目标文本构建模块 12, 被设置为将文章大纲中的任意主题句与从内容知识库中取出的与该主题句相关的段落组成目标文本;

生成词概率获取模块 13, 被设置为将目标文本输入预训练模型, 得到字典中每个词为生成词的概率;

目标词概率惩罚模块 14, 被设置为基于惩罚因子对字典中目标词对应的概率进行惩罚; 其中, 目标词为在目标文本中存在的词;

生成词取出模块 15, 被设置为基于概率从字典中取出生成词;

目标文本更新模块 16, 被设置为基于生成词更新目标文本, 得到新的目标文本;

多样性评价指标计算模块 17, 被设置为根据新的目标文本计算多样性评价指标;

惩罚因子更新模块 18, 被设置为基于该多样性评价指标更新惩罚因子, 相应的, 装置还被设置为触发生成词概率获取模块将新的目标文本输入预训练模型, 不断迭代, 直到生成词取出模块取出的生成词为截止符, 则将当前的目标文本确定为该主题句对应的候选段落;

最优段落确定模块 19, 被设置为基于任意主题句对应的候选段落确定该主题句的最优段落。

可见, 本申请实施例先生成文章大纲, 文章大纲包括所有段落的主题句, 之后将文章大纲中的任意主题句与从内容知识库中取出的与该主题句相关的段落组成目标文本, 并将目标文本输入预训练模型, 得到字典中每个词为生成词的概率, 基于惩罚因子对字典中目标词对应的概率进行惩罚, 基于概率从字典中取出生成词, 目标词为在目标文本中存在的词, 基于生成词更新目标文本, 得到新的目标文本, 根据新的目标文本计算多样性评价指标, 并基于该多样性评价指标更新惩罚因子, 并将新的目标文本输入预训练模型, 不断迭代, 直到取出的生成词

为截止符，则将当前的目标文本确定为该主题句对应的候选段落，之后基于任意主题句对应的候选段落确定该主题句的最优段落。也即，本申请实施例先生成文章大纲，对于文章大纲中的任意主题句，生成候选段落并从候选段落中选出最佳段落，在生成候选段落的过程中，利用多样性评价指标指导重复度的惩罚因子自适应的调整，能够在解决文本重复问题的情况下，减少人工成本并提升参数调节的准确度。

文章大纲生成模块 11，被设置为基于用户选择的写作类型、输入的写作主题内容以及第一关键词生成文章大纲。

可选的，文章大纲生成模块 11，被设置为包括：

模板文件检索单元，被设置为从用户选择的写作类型对应的内容知识库中检索与写作主题内容最相关的模板文件；

模型生成单元，被设置为将模板文件中每个段落输入预训练模型，生成每个段落的主题句以及第二关键词；

新主题句生成单元，被设置为分别将每个段落的主题句以及第二关键词、第一关键词按照预设方式组成模型输入数据，并将模型输入数据输入预训练模型，生成每个段落对应的新的主题句；

文章大纲生成单元，被设置为将每个段落的新的主题句组合，得到文章大纲。

其中，模板文件检索单元，被设置为从用户选择的写作类型对应文件名知识库中检索与写作主题内容最相关的文件名；从用户选择的写作类型对应的内容知识库中查找该文件名对应的模板文件。

可选的，装置还包括：

知识库池创建模块，被设置为创建知识库池；其中，知识库池中包括各写作类型对应的文件名知识库和内容知识库；内容知识库包括各模板文件，文件名知识库包括各模板文件的文件名。

其中，目标词概率惩罚模块 14，被设置为基于惩罚因子调整参数以及惩罚因子对字典中目标词对应的概率进行惩罚。

目标词概率惩罚模块 14，还被设置为统计目标词在目标文本中出现的次数，得到该目标词对应的惩罚因子调整参数；或，获取配置参数，作为目标词对应的惩罚因子调整参数。

多样性评价指标计算模块 17，被设置为从新的目标文本中去除原始目标文本，得到计算文本；原始目标文本为文本大纲中的主题句与从内容知识库中取出的段落组成的目标文本；根据计算文本计算多样性评价指标。

惩罚因子更新模块 18，被设置为计算该多样性评价指标以及上一次迭代得到的多样性评价指标之间的指标差值；基于指标差值更新惩罚因子。

装置，还包括主题偏移纠正单元，被设置为：

计算任意主题句对应的候选段落中每句话与上一句话以及该主题句之间的语义相似度，得到第一语义相似度和第二语义相似度；

若第一语义相似度和第二语义相似度均大于预设相似度阈值，则保留该句话，否则删除该句话以及该句话后面的文本，得到删除后文本；

若删除后文本的字数大于预设字数阈值，则将该删除后文本作为该主题句对应的候选段落，否则基于删除后文本重新生成该主题句对应的候选段落。

最优段落确定模块 19，被设置为对于任意主题句，基于预设评分标准分别对每个候选段落打分，并将分数最高的候选段落确定为该主题句对应的最优段落。

装置还包括检测和修改模块，被设置为在基于任意主题句对应的候选段落确定该主题句的最优段落之后，利用预设检测和修改逻辑对最优段落进行错误检测和修改。

参见图 6 所示，本申请实施例公开了一种电子设备 20，包括处理器 21 和存储器 22；其中，存储器 22，被设置为保存计算机程序；处理器 21，被设置为执行计算机程序，前述实施例公开的智能写作方法。

关于上述智能写作方法的过程可以参考前述实施例中公开的相应内容，在此不再进行赘述。

并且，存储器 22 作为资源存储的载体，可以是只读存储器、随机存储器、磁盘或者光盘等，存储方式可以是短暂存储或者永久存储。

另外，电子设备 20 还包括电源 23、通信接口 24、输入输出接口 25 和通信总线 26；其中，电源 23 被设置为电子设备 20 上的各硬件设备提供工作电压；通信接口 24 能够为电子设备 20 创建与外界设备之间的数据传输通道，其所遵循的通信协议是能够适用于本申请技术方案的任意通信协议，在此不对其进行限定；输入输出接口 25，被设置为获取外界输入数据或向外界输出数据，其接口类型可以根据应用需要进行选取，在此不进行限定。

可选的，本申请实施例还公开了一种计算机非易失性可读存储介质，被设置为保存计算机程序，其中，计算机程序被处理器执行时实现前述实施例公开的智能写作方法。

关于上述智能写作方法的过程可以参考前述实施例中公开的相应内容，在此不再进行赘述。

本说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其它实施例的不同之处，各个实施例之间相同或相似部分互相参见即可。对于实施例公开的装置而言，由于其与实施例公开的方法相对应，所以描述的比较简单，相关之处参见方法部分说明即可。

结合本文中所公开的实施例描述的方法或算法的步骤可以直接用硬件、处理器执行的软件模块，或者二者的结合来实施。软件模块可以置于随机存储器（RAM）、内存、只读存储器（ROM）、电可编程 ROM、电可擦除可编程 ROM、寄存器、硬盘、可移动磁盘、CD-ROM、或技术领域内所公知的任意其它形式的非易失性可读存储介质中。

以上对本申请所提供的一种智能写作方法、装置、设备及非易失性可读存储介质进行了详细介绍，本文中应用了个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想；同时，对于本领域的一般技术人员，依据本申请的思想，在可选实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本申请的限制。

权 利 要 求 书

1. 一种智能写作方法，其特征在于，包括：

生成文章大纲；所述文章大纲包括所有段落的主题句；

将所述文章大纲中的任意主题句与从内容知识库中取出的与该主题句相关的段落组成目标文本；

将目标文本输入预训练模型，得到字典中每个词为生成词的概率；

基于惩罚因子对所述字典中目标词对应的所述概率进行惩罚，基于所述概率从所述字典中取出生成词；其中，所述目标词为在目标文本中存在的词；

基于所述生成词更新所述目标文本，得到新的目标文本；

根据新的目标文本计算多样性评价指标，并基于该多样性评价指标更新所述惩罚因子，并将新的目标文本输入预训练模型，不断迭代，直到取出的生成词为截止符，则将当前的目标文本确定为该主题句对应的候选段落；

基于任意主题句对应的所述候选段落确定该主题句的最优段落。

2. 根据权利要求 1 所述的智能写作方法，其特征在于，所述生成文章大纲，包括：

基于用户选择的写作类型、输入的写作主题内容以及第一关键词生成文章大纲。

3. 根据权利要求 2 所述的智能写作方法，其特征在于，所述基于用户选择的写作类型、输入的写作主题内容以及第一关键词生成文章大纲，包括：

从用户选择的写作类型对应的内容知识库中检索与写作所述主题内容最相关的模板文件；

将所述模板文件中每个段落输入预训练模型，生成每个段落的主题句以及第二关键词；

分别将每个段落的主题句以及第二关键词、所述第一关键词按照预设方式组成模型输入数据，并将模型输入数据输入预训练模型，生成每个段落对应的新的主题句；

将每个段落的所述新的主题句组合，得到文章大纲。

4. 根据权利要求 3 所述的智能写作方法，其特征在于，所述从用户选择的写作类型对应的内容知识库中检索与所述写作主题内容最相关的模板文件，包括：

从用户选择的写作类型对应文件名知识库中检索与所述写作主题内容最相关的文件名；

从用户选择的写作类型对应的内容知识库中查找该文件名对应的模板文件。

5. 根据权利要求 3 所述的智能写作方法，其特征在于，还包括：

创建知识库池；其中，所述知识库池中包括各写作类型对应的文件名知识库和内容知识库；所述内容知识库包括各模板文件，所述文件名知识库包括各模板文件的文件名。

6. 根据权利要求 1 所述的智能写作方法，其特征在于，所述基于惩罚因子对所述字典中目标词对应的所述概率进行惩罚，包括：

基于惩罚因子调整参数以及惩罚因子对所述字典中目标词对应的所述概率进行惩罚。

7. 根据权利要求 6 所述的智能写作方法，其特征在于，还包括：

统计目标词在目标文本中出现的次数，得到该目标词对应的所述惩罚因子调整参数；

或，获取配置参数，作为目标词对应的所述惩罚因子调整参数。

8. 根据权利要求 6 所述的智能写作方法，其特征在于，所述基于惩罚因子调整参数以及惩罚因子对所述字典中目标词对应的所述概率进行惩罚，包括：

根据公式 $p_{x_i} = \frac{p_{x_i}}{p^n}$, if $x_i \in I_{in}$ 对所述字典中所述目标词对应的所述概率进行惩罚，其中， x_i 为所述目标词， I_{in}

为所述目标文本，在所述目标词在所述目标文本中的情况下，对所述目标词对应的概率 p_{x_i} 进行惩罚， n 用于指示所述目标词在所述目标文本中出现的次数或者常数1， P 为所述惩罚因子。

9. 根据权利要求1所述的智能写作方法，其特征在于，所述基于所述概率从所述字典中取出生成词，包括：

基于解码规则从所述字典中基于所述概率选取所述生成词，其中，所述概率为进行了惩罚后的概率，所述解码规则包括温度 Temperature 采样、TopK 采样策略和 TopP 采样策略。

10. 根据权利要求1所述的智能写作方法，其特征在于，所述根据新的目标文本计算多样性评价指标，包括：

从所述新的目标文本中去除原始目标文本，得到计算文本；所述原始目标文本为文本大纲中的主题句与从内容知识库中取出的段落组成的目标文本；

根据所述计算文本计算多样性评价指标。

11. 根据权利要求1所述的智能写作方法，其特征在于，所述基于该多样性评价指标更新所述惩罚因子，包括：

计算该多样性评价指标以及上一次迭代得到的多样性评价指标之间的指标差值；

基于所述指标差值更新所述惩罚因子。

12. 根据权利要求11所述的智能写作方法，其特征在于，所述基于所述指标差值更新所述惩罚因子，包括：

根据公式 $P = \max(1, P + \Delta d)$ 更新所述惩罚因子，其中， P 为更新后的所述惩罚因子， Δd 为多样性评价指标以及上一次迭代得到的多样性评价指标之间的指标差值。

13. 根据权利要求1所述的智能写作方法，其特征在于，在所述基于任意主题句对应的所述候选段落确定该主题句的最优段落之前，还包括：

计算任意主题句对应的所述候选段落中每句话与上一句话以及该主题句之间的语义相似度，得到第一语义相似度和第二语义相似度；

若所述第一语义相似度和第二语义相似度均大于预设相似度阈值，则保留该句话，否则删除该句话以及该句话后面的文本，得到删除后文本；

若所述删除后文本的字数大于预设字数阈值，则将该删除后文本作为该主题句对应的候选段落，否则基于所述删除后文本重新生成该主题句对应的候选段落。

14. 根据权利要求13所述的智能写作方法，其特征在于，所述计算任意主题句对应的所述候选段落中每句话与上一句话以及该主题句之间的语义相似度，得到第一语义相似度和第二语义相似度，包括：

对句子进行分词，其中，所述句子包括：所述每句话，所述上一句话和所述主题句；

获取每个所述句子分词后的每一个词对应的向量，并将获取到的向量相加并求平均，得到每个所述句子对应的句子向量；

利用公式 $\cos(\theta) = \frac{\sum_{i=1}^n (x_i \times y_i)}{\sqrt{\sum_{i=1}^n (x_i)^2} \times \sqrt{\sum_{i=1}^n (y_i)^2}}$ 计算所述每句话与所述上一句话之间的夹角余弦值作为所述第一语义相似

度，并利用公式 $\cos(\theta) = \frac{\sum_{i=1}^n (x_i \times y_i)}{\sqrt{\sum_{i=1}^n (x_i)^2} \times \sqrt{\sum_{i=1}^n (y_i)^2}}$ 计算所述每句话与所述主题句之间的夹角余弦值作为所述第二语义相似度，其中， θ 为向量 x 和向量 y 之间的夹角， x_i 表示向量 x 的分量， y_i 表示向量 y 的分量。

15. 根据权利要求1所述的智能写作方法，其特征在于，所述基于任意主题句对应的所述候选段落确定该主题句的最优段落，包括：

对于任意主题句，基于预设评分标准分别对每个候选段落打分，并将分数最高的候选段落确定为该主题句对应的最优段落。

16. 根据权利要求 15 所述的智能写作方法, 其特征在于, 所述基于预设评分标准分别对每个候选段落打分, 包括:

根据公式 $Score_{PPL} = 2^{-\frac{1}{N} \sum_{i=1}^N \log_2 P(x_i | x_1, x_2, \dots, x_{i-1})}$ 计算所述每个候选段落的分值 $Score_{PPL}$, 其中, $P(x_i | x_1, x_2, \dots, x_{i-1})$ 表示根据上文词语预测第 i 个词的概率, PPL 表示语言模型中的困惑度 perplexity, N 表示句子长度, $Score_{PPL}$ 的值越小表示生成的文本越自然、语句越通顺。

17. 根据权利要求 1 所述的智能写作方法, 其特征在于, 所述基于任意主题句对应的所述候选段落确定该主题句的最优段落之后, 还包括:

利用预设检测和修改逻辑对所述最优段落进行错误检测和修改。

18. 一种智能写作装置, 其特征在于, 包括:

文章大纲生成模块, 被设置为生成文章大纲; 所述文章大纲包括所有段落的主题句;

目标文本构建模块, 被设置为将所述文章大纲中的任意主题句与从内容知识库中取出的与该主题句相关的段落组成目标文本;

生成词概率获取模块, 被设置为将目标文本输入预训练模型, 得到字典中每个词为生成词的概率;

目标词概率惩罚模块, 被设置为基于惩罚因子对所述字典中目标词对应的所述概率进行惩罚; 其中, 所述目标词为在目标文本中存在的词;

生成词取出模块, 被设置为基于所述概率从所述字典中取出生成词;

目标文本更新模块, 被设置为基于所述生成词更新所述目标文本, 得到新的目标文本;

多样性评价指标计算模块, 被设置为根据新的目标文本计算多样性评价指标;

惩罚因子更新模块, 被设置为基于该多样性评价指标更新所述惩罚因子, 相应的, 所述装置还被设置为触发生成词概率获取模块将新的目标文本输入预训练模型, 不断迭代, 直到生成词取出模块取出的生成词为截止符, 则将当前的目标文本确定为该主题句对应的候选段落;

最优段落确定模块, 被设置为基于任意主题句对应的所述候选段落确定该主题句的最优段落。

19. 一种电子设备, 其特征在于, 包括存储器和处理器, 其中:

所述存储器, 被设置为保存计算机程序;

所述处理器, 被设置为执行所述计算机程序, 以实现如权利要求 1 至 17 任一项所述的智能写作方法。

20. 一种计算机非易失性可读存储介质, 其特征在于, 被设置为保存计算机程序, 其中, 所述计算机程序被处理器执行时实现如权利要求 1 至 17 任一项所述的智能写作方法。

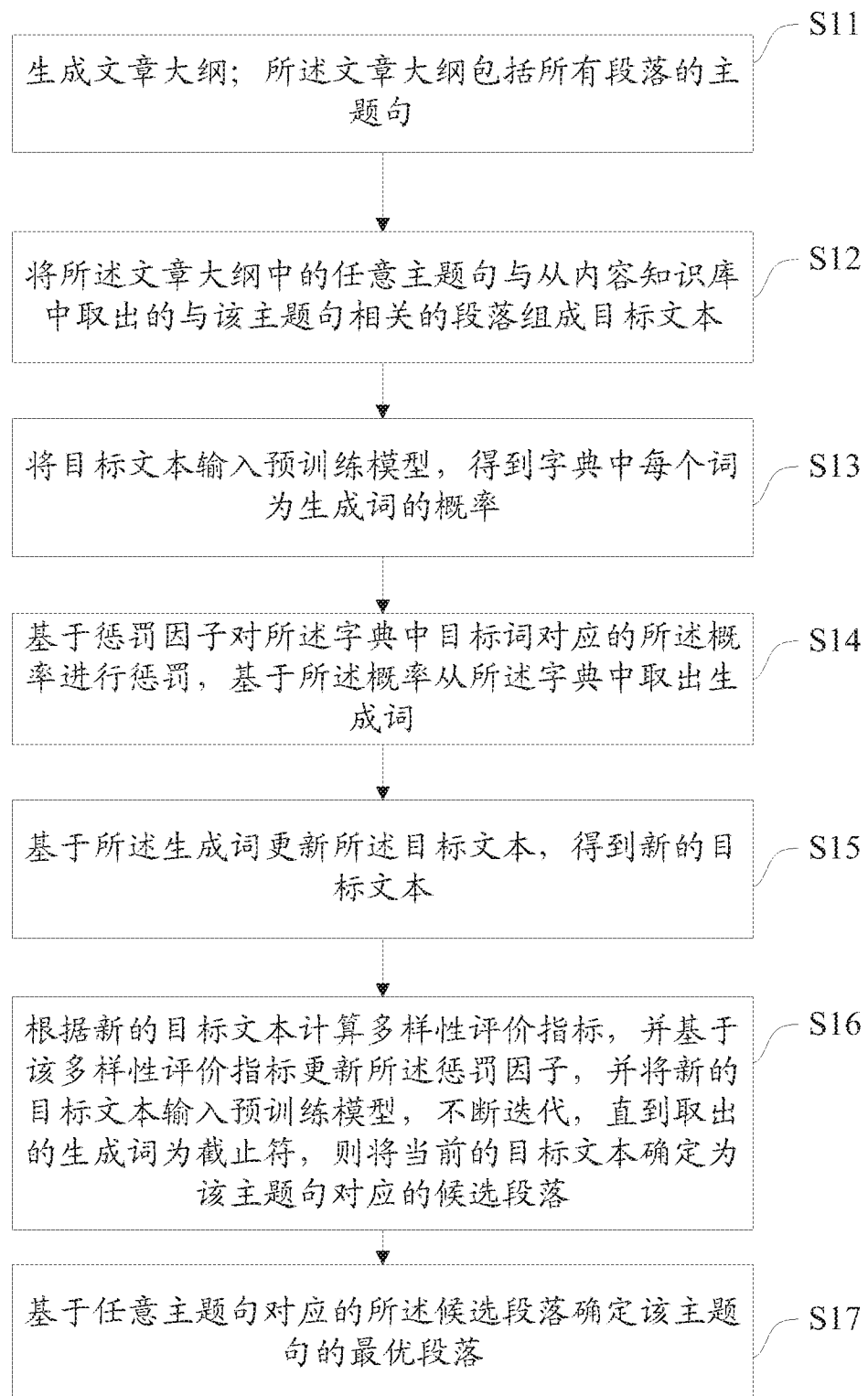


图 1

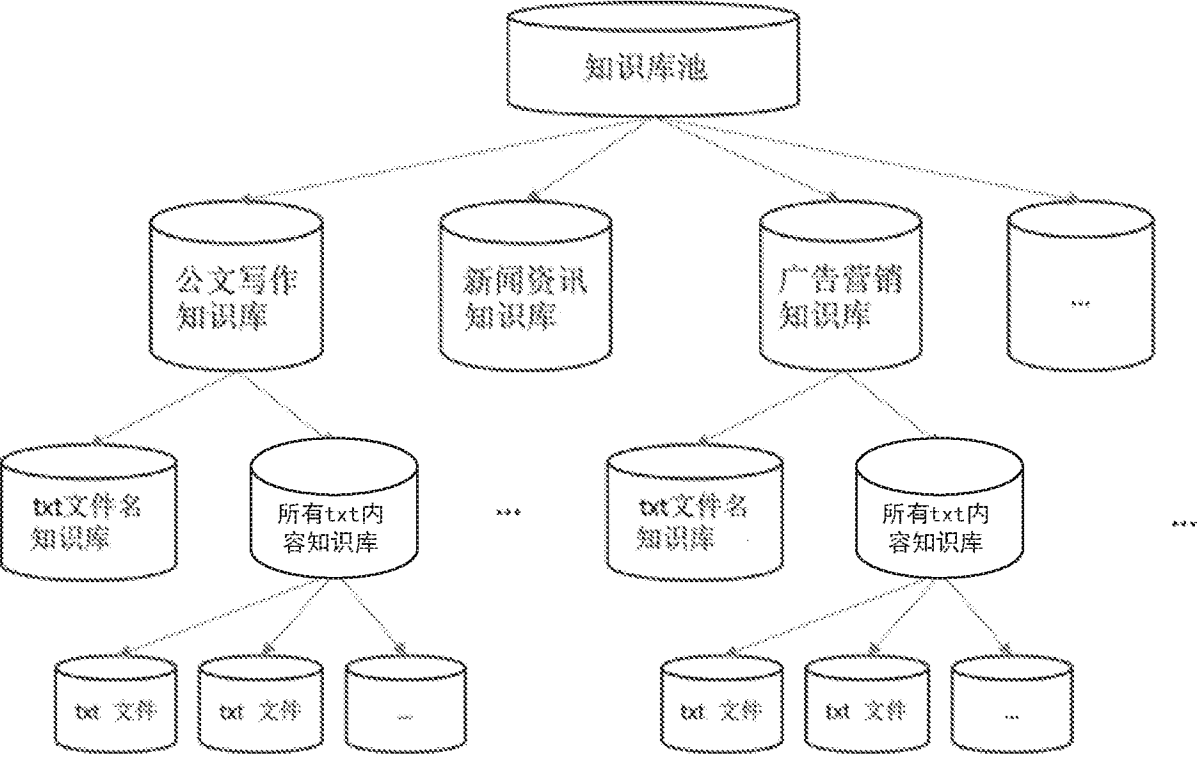


图 2

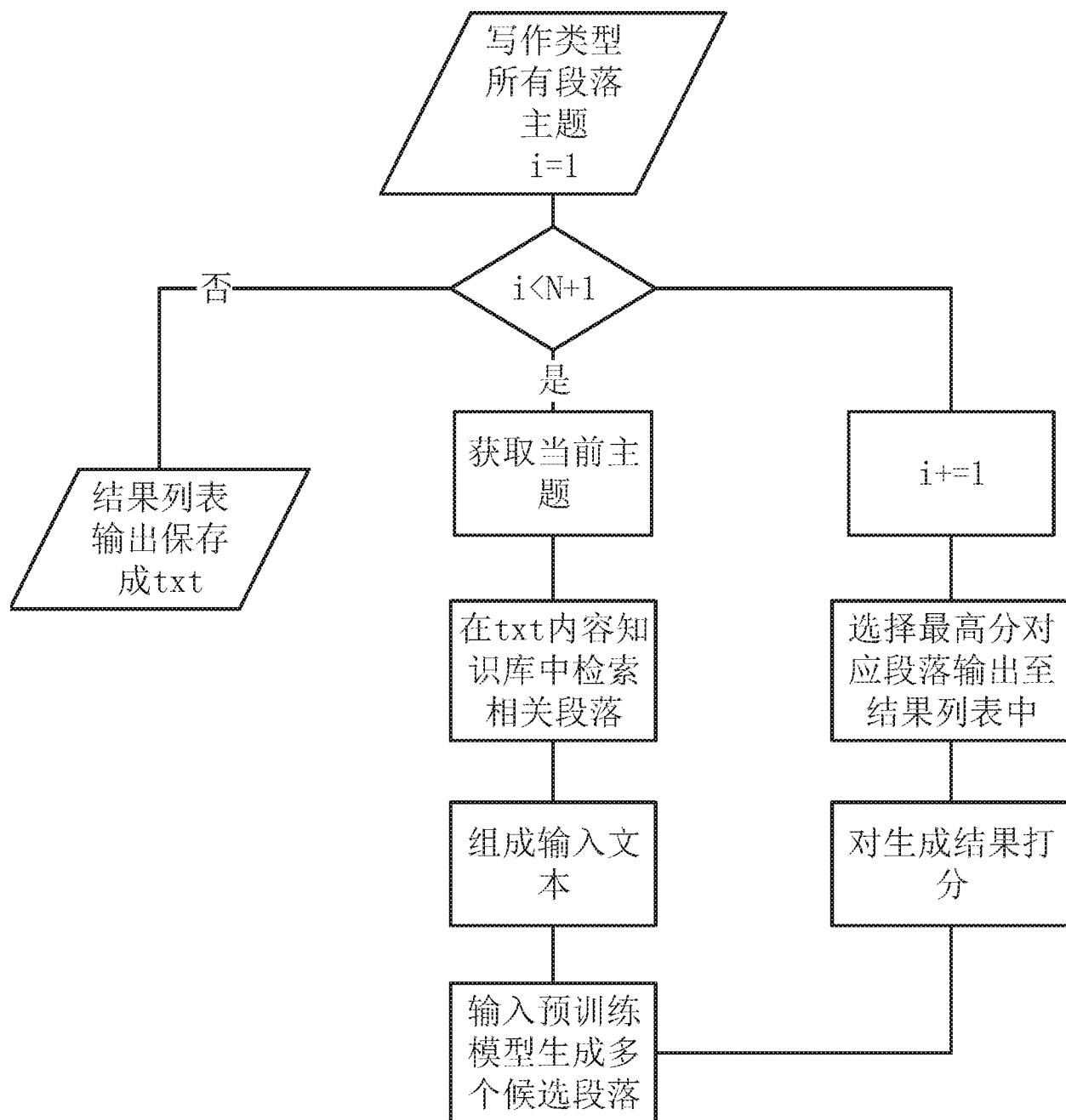


图 3

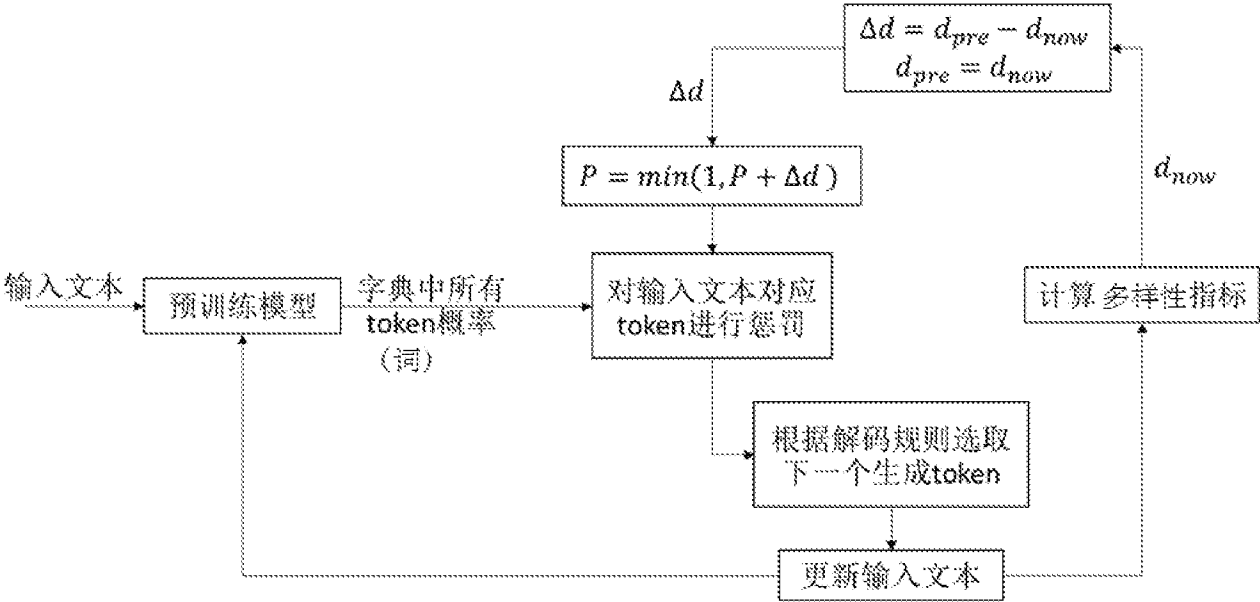


图 4

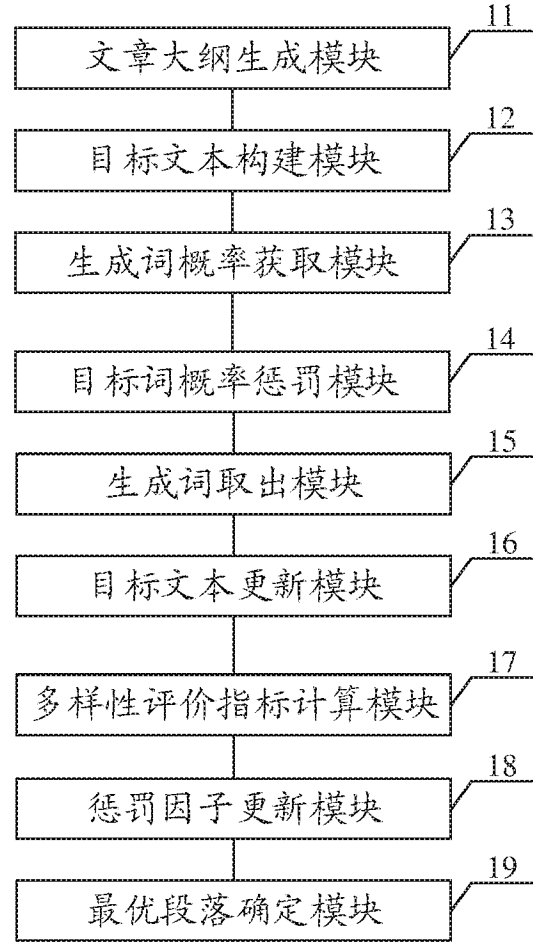


图 5

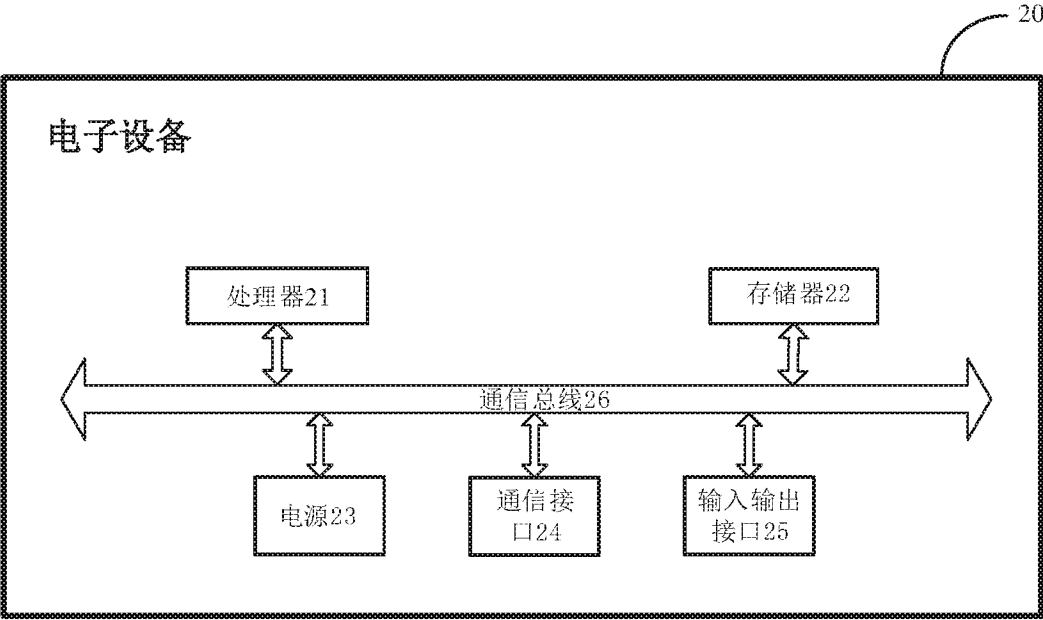


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/114518

A. CLASSIFICATION OF SUBJECT MATTER

G06F 40/284(2020.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

VEN, CNABS, CNTXT, WOTXT, USTXT, EPTXT, CNKI, IEEE: 文章, 文本, 大纲, 提纲, 目录, 生成, 写作, 主题, 惩罚, 词, 概率, article, text, outline, syllabus, directory, generation, writing, topic, penalty, word, probability

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 115630640 A (SUZHOU INSPUR INTELLIGENT TECHNOLOGY CO., LTD.) 20 January 2023 (2023-01-20) claims 1-15, and description, paragraphs 25-60	1-20
A	CN 113919336 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 11 January 2022 (2022-01-11) description, paragraphs 60-126	1-20
A	CN 110287489 A (BEIJING DAMI TECHNOLOGY CO., LTD.) 27 September 2019 (2019-09-27) entire document	1-20
A	CN 113792543 A (ANHUI MIMOUSE TECHNOLOGY CO., LTD.) 14 December 2021 (2021-12-14) entire document	1-20
A	US 2021064823 A1 (BEIJING SANKUAI ONLINE TECHNOLOGY CO., LTD.) 04 March 2021 (2021-03-04) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

10 October 2023

Date of mailing of the international search report

24 October 2023

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
China No. 6, Xitucheng Road, Jimenqiao, Haidian District,
Beijing 100088

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2023/114518

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	115630640	A	20 January 2023	None			
CN	113919336	A	11 January 2022	None			
CN	110287489	A	27 September 2019	WO	2020258948	A1	30 December 2020
CN	113792543	A	14 December 2021	None			
US	2021064823	A1	04 March 2021	US	11288454	B2	29 March 2022
				WO	2019218660	A1	21 November 2019

A. 主题的分类

G06F 40/284(2020.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

VEN, CNABS, CNTXT, WOTXT, USTXT, EPTXT, CNKI, IEEE: 文章, 文本, 大纲, 提纲, 目录, 生成, 写作, 主题, 惩罚, 词, 概率, article, text, outline, syllabus, directory, generation, writing, topic, penalty, word, probability

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 115630640 A (苏州浪潮智能科技有限公司) 2023年1月20日 (2023 - 01 - 20) 权利要求1-15, 说明书第25-60段	1-20
A	CN 113919336 A (平安科技(深圳)有限公司) 2022年1月11日 (2022 - 01 - 11) 说明书第60-126段	1-20
A	CN 110287489 A (北京大米科技有限公司) 2019年9月27日 (2019 - 09 - 27) 全文	1-20
A	CN 113792543 A (安徽咪鼠科技有限公司) 2021年12月14日 (2021 - 12 - 14) 全文	1-20
A	US 2021064823 A1 (BEIJING SANKUAI ONLINE TECHNOLOGY CO., LTD) 2021年3月4日 (2021 - 03 - 04) 全文	1-20

☐ 其余文件在C栏的续页中列出。

☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“D” 申请人在国际申请中引证的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2023年10月10日

国际检索报告邮寄日期

2023年10月24日

ISA/CN的名称和邮寄地址

中国国家知识产权局
中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

马春黎

电话号码 (+86) 010-53961336

PCT/ISA/210 表(第2页) (2022年7月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2023/114518

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	115630640	A	2023年1月20日	无			
CN	113919336	A	2022年1月11日	无			
CN	110287489	A	2019年9月27日	WO	2020258948	A1	2020年12月30日
CN	113792543	A	2021年12月14日	无			
US	2021064823	A1	2021年3月4日	US	11288454	B2	2022年3月29日
				WO	2019218660	A1	2019年11月21日