ICS 国际标准分类号
CCS 中国标准文献分类号

团　　体　　标　　准

T/CES XXX-XXXX

电力人工智能自然语言处理模型评价规范

Specification for evaluation of natural language processing model of electric power artificial intelligence

（征求意见稿）

XXXX-XX-XX 发布　　XXXX-XX-XX 实施

中国电工技术学会　发布

目　　次

前　言

本文件按照 GB/T 1.1—2020《标准化工作导则　第 1 部分：标准化文件的结构和起草规则》给出的规则起草。

请注意本文件的某些内容可能涉及专利，本文件的发布机构不承担识别这些专利的责任。

本文件由国网信息通信产业集团有限公司提出。

本文件由中国电工技术学会标准工作委员会能源智慧化工作组归口。

本文件起草单位：国网信息通信产业集团有限公司、福建亿榕信息技术有限公司。

本文件主要起草人：李强、王秋琳、赵峰、庄莉、陈江海、吕志超、张晓东、李炳森、邱镇、苏江文、陈又咏、、林钊、俞成强、王燕蓉、宋立华、伍臣周、林闽微、林靖、吴佩颖、王婧、张维。

本文件为首次发布。

电力人工智能自然语言处理模型评价规范

1 范围

本文件规定了电力人工智能自然语言处理模型的评价指标和计算方法，以及评价流程和方法。
该规范文件适用于对电力人工智能模型在自然语言处理方面的性能评估、效率评估、鲁棒性评估。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中,注日期的引用文件，仅该日期对应的版本适用于本文件;不注日期的引用文件，其最新版本(包括所有的修改单)适用于本文件。

GB/T 5271.28 信息技术词汇第28部分：人工智能基本概念与专家系统

3 术语和定义

下列术语和定义适用于本文件。

人工智能自然语言处理模型
指利用人工智能技术来处理和理解自然语言文本的模型。这些模型可以用于实现多种自然语言处理任务，如文本分类、命名实体识别、情感分析、机器翻译、问答系统等型。

4 符号、代号和缩略语

下列缩略语适用于本文件。
ARI：调整兰德系数(adjusted rand index)
AP：平均精度(average precision)
AUC：曲线下面积(area under curve)
BLEU-1：双语评价替换(bilingual evaluation understudy)
CS：余弦相似度(cosine similarity)
KES：知识抽取速度(knowledge extraction speed)
LAS：带标签依存关系准确率(labeled attachment score)
MAP：平均精度均值(mean average precision)
MIOU：平均交并比(mean intersection over union)
MP：平均精确率(mean precision)
MR：平均召回率(mean recall)
RI：兰德系数（rand index）
ROUGE：基于召回率的文本摘要评价（recall oriented understudy for gisting evaluation）
RPS：检索吞吐量(retrieval per second)
RRT：检索响应时间(retrieve response time)
RT：响应时间(response time)
TP：真正例(true positive)
TN：真负例(true negative)
UAS：无标签依存关系准确率(unlabeled attachment score)
WP：词语相似度准确率(word precision)

5. 模型基础信息

5.1 模型描述信息

应包括模型开发者、开发语言、开发框架、模型版本、模型类型、模型用途、运行环境、训练数据集信息，要求如下：

a) 模型开发者：模型开发者信息，如模型开发人员信息、开发单位信息；
b) 开发语言：模型的开发语言与版本，如Python3.7；
c) 开发框架：包含开发框架和框架版本信息，如TensorFlow-V2.1.0；
d) 模型版本：模型的版本信息，如绝缘子破损识别模型V1.1；
e) 模型类型：主要包括电力专用模型、通用组件模型；
f) 模型用途：描述模型应用场景与任务；
g) 运行环境：描述模型运行的软硬件环境及资源要求；
h) 训练数据集信息：描述模型训练阶段使用的样本规模、训练样本分布情况；

5.2 模型文件

应包括模型源文件、配置文件、运行脚本文件，或包含模型服务程序或镜像文件。

6. 评价指标与计算方法

6.1 性能指标

电力人工智能自然语言处理模型评价性能指标如下表所示。

表 1 自然语言处理模型评价-性能指标表

序号	指标名称	评价要素	计算方法
1	正确率	● 模型分类结果正确的样本数量占总样本数量的比例； ● 表征模型分类正确水平，正确率越高，模型分类越正确。	见附录 A.1.1
2	准确率	● 模型分类结果正确的正样本数量占分类结果中所有正样本数量的比例； ● 表征模型分类正样本查准水平，准确率越高，模型正样本分类越准确。	见附录 A.1.2
3	召回率	● 模型分类结果正确的正样本数量占测试集中所有正样本数量的比例； ● 表征模型分类正样本查全水平，召回率越高，模型分类正样本识别越完整。	见附录 A.1.3
4	平均精度	● 以准确率为纵轴、召回率为横轴绘制的曲线下面积； ● 表征模型查准和查全的综合水平，平均精度越高，模型查准和查全的综合性能越好。	见附录 A.1.4

5	平均正确率	● 评价模型执行多标签分类任务的性能； ● 模型分类结果中各分类正确的样本数量占该类样本总数量的比例均值； ● 表征模型分类的平均正确水平，平均正确率越高，模型分类越正确。	见附录 A.1.5
6	平均精确率	● 评价模型执行多标签分类任务的性能； ● 模型分类结果中各分类正确的正样本数量占分类结果中该类正样本数量的比例均值； ● 表征模型分类的平均查准水平，平均准确率越高，模型正样本分类越准确。	见附录 A.1.6
7	平均召回率	● 评价模型执行多标签分类任务的性能； ● 模型分类结果中各类分类正确的正样本数量占测试集中该类正样本数量的比例均值； ● 表征模型分类的平均查全水平，平均召回率越高，模型正样本识别越完整。	见附录 A.1.7
8	平均精度均值	● 评价模型执行多标签分类任务的性能； ● 模型分类结果中各类标签平均精度均值； ● 表征模型分类查准和查全的综合水平，平均精度均值越高，模型在查准和查全的综合性能越好。	见附录 A.1.8
9	平均交并比	● 评价模型执行图像分割任务的性能； ● 图像分割结果中各预测区域和标准区域交并比的平均值； ● 表征图像分割的准确程度，平均交并比越高，模型对图像分割越准确。	见附录 A.1.9
10	曲线下面积	● 评价模型执行单目标跟踪任务的性能； ● 以模型准确率为纵轴，1 减去准确率的值为横轴，绘制成功率曲线，计算曲线下面积； ● 评价模型将正样本判断为正样本的可能性大于判断为负样本的可能性的概率； ● 表征模型对正样本分类正确的概率，成功率曲线下面积越大，模型对正样本分类正确的概率越高。	见附录 A.1.10
11	决定系数	● 评价模型执行数值回归任务的性能； ● 模型预测结果中标准值和平均预测值之差的平方和为回归平方和； ● 模型预测结果中标准值和预测值之差的平方和为总偏差平方和； ● 决定系数为 1 减去回归平方和与总偏差平方和的比值； ● 表征模型在数值回归任务中解释因变量变化的能力，决定系数越高，因变量变化通过模型能被自变量解释的比例越高。	见附录 A.1.11
12	兰德系数	● 评价模型执行数值聚类任务的性能； ● 评价模型聚类结果中和标准集一致的数据对的数量与聚类结果数据对总数的比例； ● 表征模型在数值聚类任务中聚类结果的匹配水平，兰德系数越高，模型的聚类结果越匹配。	见附录 A.1.12

13	调整兰德系数	● 评价模型执行数值聚类任务的性能； ● 模型聚类结果中，兰德系数和期望兰德系数之差与最大期望兰德系数和期望兰德系数之差的比例； ● 表征模型在标准集和聚类结果中分类随机分布的条件下聚类结果的匹配水平，调整兰德系数越高，模型的聚类结果越匹配。	见附录 A.1.13
14	F1 值	● 模型分类结果中准确率的倒数和召回率的倒数的平均值； ● 表征模型准确率和召回率同等重要、且均达到最高值的平衡状态。	见附录 A.1.14
15	余弦相似度	● 评价模型执行词向量表示任务的性能； ● 模型表示结果中输出向量和标准向量点积与输出向量和标准向量范数乘积的比例； ● 表征模型词向量预测的准确率，余弦相似度越高，模型词向量表示越准确。	见附录 A.1.15
16	词语相似度准确率	● 评价模型执行词义相似度任务的性能； ● 模型预测结果中相似度预测准确的词语对数量与词语对总数的比例； ● 表征模型词义相似度的正确程度，词语相似度准确率越高，模型词义相似度越正确。	见附录 A.1.16
17	困惑度	● 评价模型执行语言模型任务的性能； ● 语言模型预测结果中每一个单词是标准单词的概率的几何平均； ● 表征模型判断句子是否为正常句子的准确率，困惑度越低，模型预测的句子越准确。	见附录 A.1.17
18	带标签依存关系准确率	● 评价模型执行依存句法分析任务的性能； ● 模型预测结果中依存关系正确且关系标签正确的关系数量与标准依存关系数量的比例； ● 表征模型依存关系预测准确率与分类准确率，带标签依存关系准确率越高，模型依存句法分析越准确。	见附录 A.1.18
19	无标签依存关系准确率	● 评价模型执行依存句法分析任务的性能； ● 模型预测结果中依存关系正确的关系数量与标准依存关系数量的比例； ● 表征模型依存关系预测准确率，无标签依存关系准确率越高，模型依存句法分析越准确。	见附录 A.1.19
20	ROUGE-1 值	● 评价模型执行文章摘要任务的性能； ● 模型预测结果中和参考摘要共有的单词数量与参考摘要单词总数的比例； ● 表征模型预测的摘要与参考摘要的相似程度，ROUGE-1 值越高，模型文章摘要越准确。	见附录 A.1.20
21	短文本相似度准确率	● 评价模型执行语义相似度任务的性能； ● 模型预测结果中相似度预测准确的短文本对数量与短文本对总数的比例； ● 表征模型短文本相似度预测的正确程度，短文本相似度准确率越高，模型相似度越正	见附录 A.1.21

		确。	
22	BLEU-1 值	● 评价模型执行机器翻译任务的性能； ● 模型预测结果中翻译正确的单词数量与模型预测结果中单词总数的比例； ● 表征模型单词翻译的查准水平，BLEU-1 值越高，模型单词翻译越准确。	见附录 A.1.22
23	知识写入速度	● 评价模型执行知识存储任务的性能； ● 知识存储任务中，批量写入的三元组数与知识写入结束时间和知识写入开始时间之差的比值，单位为组/秒； ● c)表征知识写入的时间效率，知识写入速度越快，模型知识写入时间效率越好。	见附录 A.1.23
24	检索响应时间	● 评价模型执行知识存储任务的性能； ● 知识存储结果中检索任务结束时间和检索任务开始时间之差与该时间段内完成的检索次数的比值，单位为秒/次； ● 表征知识检索的时间效率，检索响应时间越小，模型知识检索时间效率越好。	见附录 A.1.24
25	知识抽取速度	● 评价模型执行知识获取任务的性能； ● 知识抽取任务中，批量抽取的三元组数与知识抽取结束时间和知识抽取开始时间的差值的比值，单位为组/秒； ● 表征知识图谱工具的知识抽取时间效率，知识抽取速度越高，模型知识抽取时间效率越好。	见附录 A.1.35
26	检索吞吐量	● 评价模型执行知识应用任务的性能； ● 知识应用任务中，完成的知识搜索任务次数与搜索任务结束时间和搜索任务开始时间的差值的比值，单位为次/秒； ● 表征知识搜索应用的时间性能，搜索吞吐量越大，模型知识搜索应用时间性能越好。	见附录 A.1.26

6.2 效率性指标

效率性指标如下表所示。

表 2 自然语言处理模型评价-效率指标表

序号	指标名称	评价要素	计算方法
1	磁盘占用膨胀率	● 评价模型判定后占用的磁盘存储量增加的比例； ● 表征模型判定过程附加的存储消耗量，磁盘占用膨胀率越小，模型存储空间占用越少。	见附录 A.2.1

2	内存使用率	● 评价模型开销的内存量占内存总量的比例； ● 表征模型的内存开销量，内存使用率越小，模型内存使用越少。	见附录 A.2.2
3	响应时间	● 在给定的软硬件环境下，模型对给定的数据进行运算并获得结果所需要的时间； ● 表征模型解决任务所消耗的时间，模型响应时间越小，模型响应越快。	见附录 A.2.3

6.3 鲁棒性指标

鲁棒性指标如下表所示。

表 3 自然语言处理模型评价-鲁棒性指标表

序号	指标名称	评价要素	计算方法
3	鲁棒性	● 评价模型在存在信号干扰或特征规律发生变化的测试数据集的性能指标，性能指标选取规则见附录 C； ● 表征模型对新样本的维持性能稳定的能力，鲁棒性指标越高，模型维持性能稳定的能力越好。	见附录 A.3

7. 评价流程与方法

7.1 评价流程

评价流程应包含模型信息完备性评价、确定模型分类和模型任务、选择评价指标、选择测试数据集、单项评价指标量化 、评价结果汇总等6个关键步骤，详见图1。

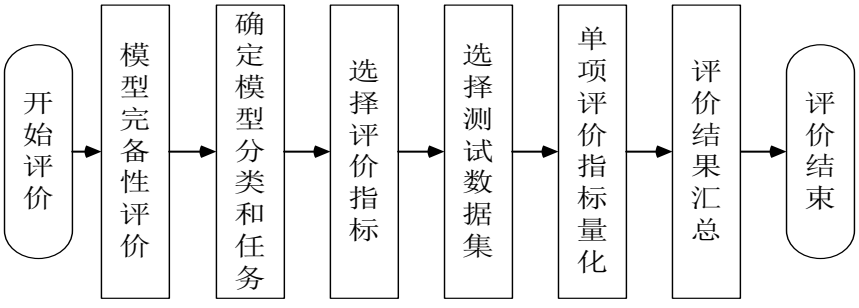

图1 模型评价流程

7.2 评价方法

7.2.1 模型基础信息完备性评价

所提供的模型的描述信息、模型文件应完整、正确。

7.2.2 确定模型分类和模型任务

确定模型分类和模型任务要求如下：

a) 对于电力专用模型，确定模型任务，包括图像分类、目标检测、图像分割、视频分类、行为检测、单目标跟踪、多目标跟踪、数值分类、数值回归、数值聚类；

b) 对于通用组件模型，确定模型任务，详见资料性附录B.2。

7.2.3 选择评价指标

选择评价指标要求如下：

a) 根据模型任务类型确定相应的模型性能指标，电网专用模型性能指标选取规则见规范性附录C.1、通用组件性能指标选取规则见规范性附录C.2，根据6.2章节评价模型效率性指标，根据6.3章节评价模型鲁棒性指标，根据6.4章节评价模型兼容性指标。

b) 模型研发、入网、在运等各环节宜采用相同的评价指标。

7.2.4 选择测试数据集

测试数据集选择要求如下：

a) 测试数据集应与训练数据集具有互斥性；

b) 测试数据集样本格式参照《人工智能样本基本要求和标注规范》；

c) 测试数据集、鲁棒性测试数据集、训练数据集的比例宜为2：2：8；

d) 鲁棒性测试数据集样本类别要求如下：

1) NLP模型鲁棒性测试集应包含但不限于近义词、反义词，停用词样本；

2) 知识图谱模型鲁棒性测试集应包含但不限于多领域知识、多源知识、异构数据或知识库。

e) 测试数据集、鲁棒性数据集样本包含的各类别的样本数量宜相同。

附 录 A
（规范性附录）
人工智能自然语言处理模型评价指标计算方法

A.1 性能指标

A.1.1 正确率

正确率的计算方式见公式（A.1）：

$$A = （c_{TP} + c_{TN}）/(c_{TP} + c_{TN} + c_{FP} + c_{FN}) \quad （A.1）$$

式中：

A——正确率；

c_{TP}——模型分类正确的正样本数量；

c_{FP}——模型分类成正类的负样本数量；

c_{TN}——模型分类正确的负样本数量；

c_{FN}——模型分类成负类的正样本数量。

A.1.2 准确率

准确率的计算方式见公式（A.2）：

$$P = c_{TP}/(c_{TP} + c_{FP}) \quad （A.2）$$

式中：

P——准确率；

c_{TP}——模型分类正确的正样本数量；

c_{FP}——模型分类成正类的负样本数量。

A.1.3 召回率

召回率的计算方式见公式（A.3）：

$$R = c_{TP}/(c_{TP} + c_{FN}) \quad （A.3）$$

式中：

R——召回率；

c_{TP}——模型分类正确的正样本数量；

c_{FN}——模型分类成负类的正样本数量。

A.1.4 平均精度

平均精度的计算方式见公式（A.4）：

$$V_{AP} = \int_0^1 \quad p(r)dr \quad （A.4）$$

式中：

V_{AP}——平均精度；

$p(r)$——以准确率为纵轴、召回率为横轴绘制的曲线函数。

A.1.5 平均正确率

平均正确率的计算方式见公式（A.5）：

$$V_{MA} = (1/M)\sum_{i=1}^{M} （c_{TP_i} + c_{TN_i}）/\left(c_{TP_i} + c_{TN_i} + c_{FP_i} + c_{FN_i}\right) \quad （A.5）$$

式中：

V_{MA}——平均正确率；

c_{TP_i}——第i类分类结果中正确的正样本数量；

c_{TN_i}——第i类分类结果中正确的负样本数量；

c_{FP_i}——第i类分类结果中错误的正样本数量；
c_{FN_i}——第i类分类结果中错误的负样本数量；
M——类别总数。

A.1.6 平均精确率

平均准确率的计算方式见公式（A.6）：

$$V_{MP} = (1/M)\sum_{i=1}^{M} c_{TP_i}/(c_{TP_i} + c_{FP_i}) \quad (A.6)$$

式中：
V_{MP}——平均准确率；
c_{TP_i}——第i类分类结果中正确的正样本数量；
c_{FP_i}——第i类分类结果中错误的正样本数量；
M——类别总数。

A.1.7 平均召回率

平均召回率的计算方式见公式（A.7）：

$$V_{MR} = (1/M)\sum_{i=1}^{M} c_{TP_i}/(c_{TP_i} + c_{FN_i}) \quad (A.7)$$

式中：
V_{MR}——平均召回率；
c_{TP_i}——第i类分类结果中正确的正样本数量；
c_{FN_i}——第i类分类结果中错误的负样本数量；
M——类别总数。

A.1.8 平均精度均值

平均精度均值的计算方式见公式（A.8）：

$$V_{MAP} = (1/M)\sum_{i=1}^{M} V_{AP_i} \quad (A.8)$$

式中：
V_{MAP}——平均精度均值；
V_{AP_i}——第i类的平均精度；
M——类别总数。

A.1.9 平均交并比

平均交并比的计算方式见公式（A.9）：

$$V_{MIOU} = \frac{1}{M+1}\sum_{I=0}^{M}\frac{c_{ii}}{\sum_{j=0}^{M} c_{ij} + \sum_{j=0}^{M} c_{ji} - c_{ii}} \quad (A.9)$$

式中：
V_{MIOU}——平均交并比；
M——类别总数；
c_{ij}——属于第i类，且被预测为第j类的样本数量。

A.1.10 曲线下面积

成功率曲线下面积的计算方式见公式（A.12）：

$$V_{AUC} = \frac{\sum \mathrm{ins}_i \in positiveclass\, rank_{ins_i} - \frac{M\times(M-1)}{2}}{M\times N} \quad (A.12)$$

式中：
V_{AUC}——成功率曲线下面积；
$rank_{ins_i}$——第i条视频的输出置信度排序序号；
M——正样本数量；

N——负样本数量；

$\mathrm{ins}_i \in positiveclass$——正样本序号的序号。

A.1.11 决定系数

决定系数的计算方式见公式（A.18）：

$$V_{r^2} = 1 - \frac{\sum_{i=1}^{N}(v_{truth_i} - v_{predict_i})^2}{\sum_{i=1}^{N}(v_{truth_i} - \overline{v_{predict}})^2} \quad (A.18)$$

式中：

V_{r^2}——平均绝对误差值；

N——样本总数；

$v_{predict_i}$——模型输出的第i个样本的预测值；

v_{truth_i}——第i个样本的标准值；

$\overline{v_{predict}}$——所有预测值的平均值。

A.1.12 兰德系数

兰德系数的计算方式见公式（A.19）：

$$V_{RI} = (c_{ss} + c_{dd})/(c_{ss} + c_{sd} + c_{ds} + c_{dd}) \quad (A.19)$$

式中：

V_{RI}——兰德系数；

c_{ss}——在聚类结果中属于同一类别，且在标准情况下也为同一类别的数据对的数量；

c_{dd}——在聚类结果中属于不同类别，且在标准情况下也为不同类别的数据对的数量；

c_{sd}——在聚类结果中属于同一类别，但标准情况下为不同类别的数据对的数量；

c_{ds}——在聚类结果中属于不同类别，但在标准情况下为同一类别的数据对的数量。

A.1.13 调整兰德系数

调整兰德系数的计算方式见公式（A.20）：

$$V_{ARI} = (V_{RI} - E(V_{RI}))/(fRI_{RI\,max} \quad (A.20)$$

式中：

V_{ARI}——调整兰德系数；

V_{RI}——兰德系数；

$E(V_{RI})$——兰德系数均值；

fRI_{max}——兰德系数最大值。

A.1.14 F1 值

F1值的计算方式见公式（A.21）：

$$F_1 = (2 \times P \times R)/(P + R) \quad (A.21)$$

式中：

F_1——F1值；

P——精确率；

R——召回率。

A.1.15 余弦相似度

余弦相似度的计算方式见公式（A.22）：

$$V_{CS} = \left(\overrightarrow{a_{pre}} \bullet \overrightarrow{a_{truth}}\right) / \left(||\overrightarrow{a_{pre}}||_2 ||\overrightarrow{a_{truth}}||_2\right) \quad (A.22)$$

式中：

V_{CS}——余弦相似度；

$\overrightarrow{a_{pre}}$——模型预测出的词向量；

$\overrightarrow{a_{truth}}$——标准词向量；

$\overrightarrow{a_{pre}} \bullet \overrightarrow{a_{truth}}$——两个向量的点积；
$||\ ||_2$——向量的范数。

A.1.16 词语相似度准确率

词语相似度准确率的计算方式见公式（A.23）：

$$V_{WP} = c_{correct\_word\_pair}/c_{all\_word\_pair} \quad (A.23)$$

式中：
V_{WP}——词语相似度准确率；
$c_{correct\_word\_pair}$——模型预测正确的词语对数量；
$c_{all\_word\_pair}$——词语对总数。

A.1.17 困惑度

困惑度的计算方式见公式（A.24）：

$$V_{perlexity} = \sqrt[n]{\prod_{i=1}^{n} 1/(p(w_i|w_1,\ldots,w_{i-1}))} \quad (A.24)$$

式中：
$V_{perlexity}$——困惑度；
n——测试数据集中的单词总数；
w_i——第i个单词；
$p(w_i|w_1,\ldots,w_{i-1})$——模型预测出的第i个单词的置信度。

A.1.18 带标签依存关系准确率

带标签依存关系准确率的计算方式见公式（A.25）：

$$V_{LAS} = c_{cd\_cc}/c_{all\_depen} \quad (A.25)$$

式中：
V_{LAS}——带标签依存关系准确率；
c_{cd_cc}——模型预测结果中依存关系正确且关系种类标签正确的数量；
c_{all_depen}——标准依存关系数量。

A.1.19 无标签依存关系准确率

无标签依存关系准确率的计算方式见公式（A.26）：

$$V_{UAS} = c_{cd}/c_{all\_depen} \quad (A.26)$$

式中：
V_{UAS}——无标签依存关系准确率；
c_{cd}——模型预测结果中依存关系正确的数量；
c_{all_depen}——标准依存关系数量。

A.1.20 ROUGE-1 值

ROUGE-1值的计算方式见公式（A.27）：

$$V_{ROUGE_1} = \frac{\sum_{u \in U_{reference}} \sum_{V_{1-gram} \in u} f_{Count_{match}}\left(V_{n-gram}\right)}{\sum_{u \in U_{reference}} \sum_{V_{1-gram} \in u} f_{Count}\left(V_{n-gram}\right)} \quad (A.27)$$

式中：
V_{ROUGE_1}——ROUGE-1值；
$U_{reference}$——标准摘要；
V_{1-gram}——单词；
$f_{Count_{match}}(V_{1-gram})$——模型预测结果中与参考摘要共有的单词数量；
$f_{Count}(V_{1-gram})$——参考摘要单词总数。

A.1.21 短文本相似度准确率

短文本相似度准确率的计算方式见公式（A.28）：

$$V_{text\_precision} = c_{correct\_text\_pair} / c_{all\_text\_pair} \quad (A.28)$$

式中：

$V_{text\_precision}$——短文本相似度准确率；

$c_{correct\_text\_pair}$——模型预测正确的短文本对数量；

$c_{all\_text\_pair}$——短文本对总数。

A.1.22 BLEU-1 值

BLEU-1值的计算方式见公式（A.29）：

$$V_{bleu_1} = \frac{\sum_{u \in U_{candidates}} \sum_{V_{1-gram} \in u} f_{Count_{ref}}\left(V_{n-gram}\right)}{\sum_{u' \in U_{candidates}} \sum_{V_{1-gram}' \in u'} f_{Count_{all}}\left(V_{n-gram'}\right)} \quad (A.29)$$

式中：

V_{bleu_1}——双语评价替换；

$U_{candidates}$——模型生成句子的集合；

V_{1-gram}——单词；

$f_{Count_{ref}}(V_{1-gram})$——所有生成句子中的单词在标准译文中出现的次数；

$f_{Count_{all}}(V_{1-gram})$——所有生成句子中的单词总数。

A.1.23 知识写入速度

知识写入速度的计算方式见公式（A.42）：

$$V_{KW} = c_{triple} / (T_{ke} - T_{ks}) \quad (A.42)$$

式中：

V_{KW}——知识写入速度；

c_{triple}——批量写入的三元组数；

T_{ks}——知识写入开始时间；

T_{ke}——知识写入结束时间。

A.1.24 检索响应时间

检索响应时间的计算方式见公式（A.43）：

$$T_{RRT} = (T_{re} - T_{rs}) / c_{retrieve} \quad (A.43)$$

式中：

T_{RRT}——检索响应时间；

$c_{retrieve}$——周期内完成检索次数；

T_{rs}——检索开始时间；

T_{re}——检索结束时间。

A.1.25 知识抽取速度

知识抽取速度的计算方式见公式（A.44）：

$$V_{KES} = c_{KN} / (T_{kee} - T_{kes}) \quad (A.44)$$

式中：

V_{KES}——知识抽取速度；

c_{KN}——知识抽取数量；

T_{kes}——知识抽取开始时间；

T_{kee}——知识抽取结束时间。

A.1.26 检索吞吐量

检索吞吐量的计算方式见公式（A.45）：

$$V_{RPS} = c_{task}/(T_{re} - T_{rs}) \quad (A.45)$$

式中：

V_{RPS}——搜索吞吐量；

c_{task}——周期完成的搜索任务次数；

T_{rs}——搜索任务开始时间；

T_{re}——搜索任务结束时间。

A.2 效率性指标

A.2.1 磁盘占用膨胀率

磁盘占用膨胀率的计算方式见公式（A.46）：

$$V_{occ\_disk} = (D_2 - D_1)/D_2 \quad (A.46)$$

式中：

V_{occ_disk}——磁盘占用膨胀率；

D_1——磁盘空间初始占用量；

D_2——模型判定后磁盘空间占用量。

A.2.2 内存使用率

内存使用率的计算方式见公式（A.47）：

$$V_{occ\_mem} = M_{use}/M_{total} \quad (A.47)$$

式中：

V_{occ_mem}——内存使用率；

M_{use}——模型内存空间使用量；

M_{total}——内存空间总量。

A.2.3 响应时间

响应时间的计算方式见公式（A.48）：

$$T_{RT} = T_e - T_s \quad (A.48)$$

式中：

T_{RT}——响应时间；

T_s——批量样本输入的开始时间；

T_e——获得最后一个样本模型处理结果的结束时间。

A.3 鲁棒性指标

鲁棒性指标的计算方式见公式（A.49）：

$$V_{robust} = (1/n)\sum_{i=1}^{n}(E_{robust-i} - E_i)/E_i \quad (A.49)$$

式中：

V_{robust}——鲁棒性指标值；

n——模型所需评价的性能性指标个数；

$E_{robust-i}$——模型在存在信号干扰或特征负率变化的测试数据集上的第i个性能性指标；

E_i——模型在存在一般测试数据集上的第i个性能性指标。

附 录 B
（资料性附录）
电力人工智能模型分类和模型任务

电力专用模型模型分类和模型任务见表B.1。

表B.1 电力专用模型应用场景和模型任务

序号	一级分类	二级分类	模型任务
1	电力专用模型	输电应用	图像分类、目标检测、图像分割、视频分类、行为检测、单目标跟踪、多目标跟踪
2			
3			
4		变电应用	
5			
6		配电应用	
7			
8			
9			
10		安监违章应用	
11		用电应用	数值分类、数值回归、数值聚类
12		营销应用	
13		调度应用	数值分类、数值回归、数值聚类、目标检测
14		其他	图像分类、图像分割、目标检测、视频分类、行为检测、单目标跟踪、多目标跟踪、数值分类、数值回归、数值聚类

通用组件模型分类和模型任务见表B.2。

表B.2 通用组件模型分类和模型任务

序号	一级分类	二级分类	模型任务
1	通用组件模型	OCR识别模型	通用文字识别、打印文档识别、手写文字识别、表格恢复、证照识别、企业资质识别、发票凭证识别、混合票据分割、混合票据识别、混合票据分类、合同印章弯曲文字识别
2			
3		语音识别模型	短语音识别、实时语音识别、音频文件转写
4			
5		NLP模型	词法分析、词向量表示、词义相似度、语言模型、依存句法分析、短文本相似度、文章标签、文章分类、文本纠错、文本摘要、评论观点抽取、对话情绪识别、情感倾向分析
6			
7		人脸识别模型	人脸检测、人脸活体检测、人脸对比（1:1）、人脸搜索（1:N）
8		知识图谱模型	知识建模、知识存储、知识获取、知识融合、知识计算、知识应用
9			