ICS 35.240
L70/84

团 体 标 准

T/CESA 1039—2019

信息技术 人工智能 机器翻译能力等级评估

Information technology-Artificial intelligence-Classified assessment for machine translation capabilities

2019－04－01 发布　　2019－04－01 实施

中国电子工业标准化技术协会　　发 布

目　　次

前　言

本标准按照GB/T　1.1—2009《标准化工作导则　第1部分：标准的结构和编写》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。

本标准由中国电子技术标准化研究院提出并归口。

本标准起草单位：中国电子技术标准化研究院、科大讯飞股份有限公司、腾讯科技（北京）有限公司、网易有道信息技术（北京）有限公司、中国电信集团有限公司、潍坊北大青鸟华光照排有限公司、北京百度网讯科技有限公司、华夏芯(北京)通用处理器技术有限公司、广州广电运通金融电子股份有限公司、安徽听见科技有限公司、杭州方得智能科技有限公司、海尔优家智能科技（北京）有限公司。

本标准主要起草人：代红、董建、张群、王燕妮、马万钟、汪小娟、马珊珊、谢军、黄瑾、张亚军、刘俊华、杨震、李洁、陈璐、殷建民、蒋晓琳、刘军、林冠辰、陈良旭、彭黔平、胡江明。

信息技术　人工智能　机器翻译能力等级评估

1　范围

本标准规定了机器翻译系统通用模型及要求、机器翻译系统能力指标及计算方法、机器翻译系统能力等级划分和评估要求。

本标准适用于具有翻译功能的机器翻译系统产品和服务的能力等级划分与评估。

2　术语和定义

下列术语和定义适用于本文件。

2.1

机器翻译　machine　translation,　MT

使用计算机应用将文本从一种自然语言自动翻译成另一种自然语言。

2.2

机器翻译系统　machine　translation　system

由功能单元（或其组合）组成的，能够实现机器翻译的系统。

2.3

语言模态　language　modality

以不同形式呈现或表达的自然语言信息。

注：这些形式包括但不限于文本、语音、图像、视频或文档等。

2.4

源语言　source　language

将要被翻译的语言内容。

2.5

目标语言　target　language

从源语言内容翻译而来的语言内容。

3　缩略语

下列缩略语适用于本文件。

MTAS：机器翻译系统能力得分（Machine　Translation　Ability　Score）

4 机器翻译系统通用模型及要求

4.1 概述

机器翻译系统应包含系统输入、输出和系统服务引擎，其中系统输入、输出实现不同语言模态的采集和展现，系统服务引擎包含语言模态处理和机器翻译引擎。机器翻译系统通用模型见图1。

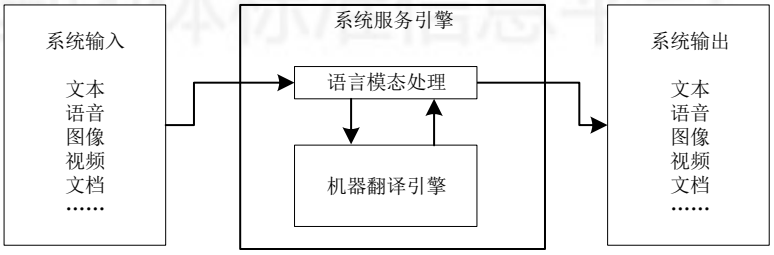

图1 机器翻译系统通用模型

4.2 系统输入输出要求

系统输入输出要求包括：

a) 应支持用户进行语言模态的输入和输出；

b) 宜支持多种语言模态的输入和输出，包括但不限于文本、语音、图像、视频或文档等。

4.3 系统服务引擎要求

系统服务引擎要求包括对语言模态处理和机器翻译引擎的要求。具体要求如下：

a) 语言模态处理要求包括：

1) 应支持将语言模态处理成机器翻译引擎可接受的数据格式；

2) 宜支持多种语言模态处理方式，包括但不限于语音识别、语音合成、光学字符识别、图像识别、图像渲染、文档格式解析、编解码等。

c) 机器翻译引擎要求包括：

1) 应及时响应用户的翻译请求；

2) 应支持将一种语言翻译成另一种语言，并且翻译结果能够准确表达源语言所涵盖的信息，满足不同翻译目的、要求，以及用户跨语种交流的需求；

3) 宜支持多个语种之间的相互翻译；

4) 宜支持同传翻译模式；

5) 宜支持无网络连接状态的离线翻译。

5 机器翻译系统能力指标及计算方法

5.1 能力指标体系

机器翻译系统能力测评指标体系结构见图2。

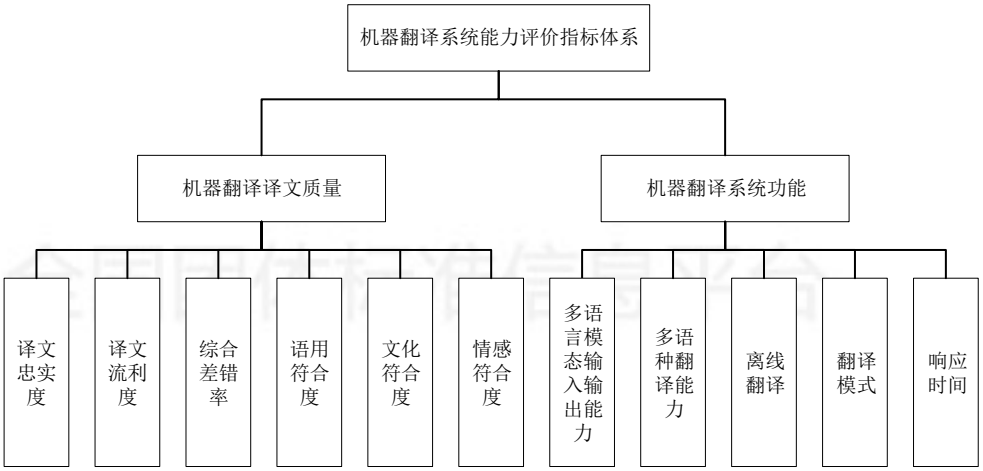

图2　机器翻译系统能力测评指标体系

机器翻译系统能力测评具体指标项及说明见表1。

表1　机器翻译系统能力测评指标项及说明

一级指标	二级指标	指标项说明	评分方法	权重	说明
机器翻译译文质量	译文忠实度	翻译结果是否忠实地表达了源语言的内容	——完全没有译出来，0分 ——译文中只有个别词被孤立地翻译，1分 ——译文中有少数短语或比词大的语法成分被翻译，2分 ——源语言中60%的概念及其之间的关系被正确翻译，或原文中的主谓宾及其关系被正确的翻译，3分 ——源语言中80%的概念及其之间的关系被正确翻译，4分 ——源语言中100%的概念及其之间的关系被正确翻译，5分	0.3	
	译文流利度	翻译结果是否流畅和地道	——完全不可理解，0分 ——译文晦涩难懂（只有个别短语或比词大的语法成分可以理解），1分 ——译文40%的部分基本流畅（少数的短语或比词大的语法成分可以理解），2分 ——译文60%的部分基本流畅，3分 ——译文80%的部分基本流畅，或译文中的主谓宾部分基本流畅，只是个别词语或搭配不地道，4分 ——译文流畅而且地道，5分	0.3	
	综合差错率	翻译结果与源语言内容的综合差错率	——译文综合差错率≥90%，0分 ——70%≤译文综合差错率＜90%，1分 ——50%≤译文综合差错率＜70%，2分 ——30%≤译文综合差错率＜50%，3分 ——10%≤译文综合差错率＜30%，4分 ——译文综合差错率≤10%，5分	0.15	

表 1　机器翻译系统能力测评指标项及说明（续）

一级指标	二级指标	指标项说明	打分方法	权重	说明
	语用符合性	翻译结果是否符合目标语言的语用规则	——译文结果不符合，0分 ——译文结果符合，5分	0.05	
	文化符合性	翻译结果是否符合目标语言的文化背景和观念	——译文结果不符合，0分 ——译文结果符合，5分	0.05	
	情感符合性	翻译结果是否符合源语言表达的情感	——译文结果不符合，0分 ——译文结果符合，5分	0.05	
机器翻译系统功能	多语言模态输入输出能力	是否支持文本以外的多种语言模态输入输出功能	——支持除文本以外的多种语言模态输入输出，包括但不限于语音、图像、视频或文档（网页、PDF）等，即翻译的源语言和目标语言可以通过语音、图像、视频或文档等形式输入和输出，其性能应满足评估对象所标称的性能值，每支持一种类型得1分，最高不超过5分 ——仅支持文本输入输出，即翻译的源语言和目标语言仅能以文本的形式输入和输出，0分	0.02	
	多语种翻译能力	是否支持多语种翻译	——支持两个以上语种之间的相互翻译，其翻译性能应满足评估对象所标称的性能值，5分 ——仅支持两个语种之间相互翻译，0分	0.02	
	离线翻译	是否支持无网络连接状态下的离线翻译	——支持无网络连接状态下的离线翻译，其翻译性能应满足评估对象所标称的性能值，5分 ——不支持无网络连接状态下的离线翻译，0分	0.02	
	翻译模式	是否支持同传翻译模式	——支持评估对象所标称的同传翻译模式，其翻译性能应满足评估对象所标称的性能值，5分 ——不支持同传翻译模式，0分	0.02	
	响应时间	是否在规定的时间内响应用户翻译请求	——系统翻译引擎响应时间应符合附录B的要求，5分 ——系统翻译引擎响应时间不符合附录B的要求，0分	0.02	
注：测评指标以反映机器翻译译文结果质量为主，侧重于评价机器翻译系统的翻译能力，该部分指标权重占整体评价结果的90%；系统相关的功能评测指标占整体评价结果的10%。					

5.2　指标评估方法

指标评估可采用技术测试和材料检查等方式进行，其中，材料检查是对待测对象的说明文档或证明材料进行查验；技术测试是运用预定的方法或工具对待测对象进行测试，例如专家评价或数据集测试。

不同指标项的可采用的评估方式如下：

a）　采用技术测试的指标项包括：忠实度、流利度、综合差错率、语用符合性、文化符合性、情感符合性；

b） 采用技术测试和材料检查相结合的指标项包括：多语言模态输入输出能力、多语种翻译能力、离线翻译、翻译模式、响应时间。

机器翻译忠实度和流利度评价的评估，参见附录A；机器翻译综合差错率计算方法见附录C。

5.3 能力计算方法

根据评估指标体系和指标项，运用综合评分法对机器翻译系统能力得分（MTAS）进行计算，计算公式见式（1）。

$$MTAS = \sum P_i \times S_i \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (1)$$

式中：

$MTAS$ ——综合评分法对机器翻译系统能力得分；

P_i ——第 i 项指标的权重值；

S_i ——第 i 项指标的得分值。

6 机器翻译系统能力等级划分

根据机器翻译系统能力指标和计算方法，机器翻译系统能力等级可分为5级，其中最高等级为5级。不同级别对应的MTAS得分如下：

a） 1 级：机器翻译系统能力得分，MTAS≤1.5 分；
b） 2 级：机器翻译系统能力得分，1.5 分＜MTAS≤2.5 分；
c） 3 级：机器翻译系统能力得分，2.5 分＜MTAS≤3.5 分；
d） 4 级：机器翻译系统能力得分，3.5 分＜MTAS≤4.5 分；
e） 5 级：机器翻译系统能力得分，4.5 分＜MTAS≤5 分。

7 机器翻译系统能力等级评估要求

7.1 确定评估方案

根据评估目的需要，综合考虑机器翻译系统能力等级的影响因素，制定与其需求相符合的评估方案。用户可自行制定方案来实施评估，也可以委托第三方制定评估方案。评估方案包括但不限于机器翻译系统界定、评估指标及指标值计算、评估对象等级划分等内容。

7.2 机器翻译系统界定

评估前应识别、界定和描述被评估的机器翻译系统产品及其特性，包括系统来源、用途和使用方式等。

7.3 计算评估指标得分

评估前应确定评估目的和范围，并根据本标准所给出的评估指标体系和指标项来确定评估指标，按照5.3所示的方法计算指标得分。

7.4 评估对象等级划分

根据评估目的，按照第6章所示等级划分方法，将计算评估指标得分结果对应到不同的等级，形成评估结论。

7.5 评估报告及使用

评估完成后，应进行评估结果分析，出具评估报告，对评估报告建档存留，并定期复审。评估报告内容宜包括但不限于以下内容：

a) 机器翻译系统产品的基本概况；
b) 评估目的；
c) 评估对象和范围；
d) 机器翻译系统能力等级划分和定义；
e) 评估假设和限定条件；
f) 评估依据和方法；
g) 评估程序实施过程和情况；
h) 评估结论；
i) 特别事项说明；
j) 评估报告的使用限制说明。

评估方应对评估报告建档存留，并定期复审。

附 录 A
（资料性附录）
机器翻译忠实度和流利度评价

A.1 评价要求

对翻译译文的忠实度、流利度进行评价时，宜采用多人评价取均值的策略。评估方应精通参评任务所涉及的源语言和目标语言，并具有如下资格之一：

a) 源语言为母语，目标语言上具有高水平的语言资格认证；

b) 源语言具有高水平的语言资格认证，目标语言为母语。

A.2 评价方法

针对具体的业务应用场景，选择不少于200条具有较大覆盖度的业务数据构建测试集合。实施过程中，参评人员宜不少于5人。对于每个评价句子对，取所有参评人员的忠实度和流利度的评分的均值作为该句子的忠实度和流利度评分。

附 录 B
（规范性附录）
机器翻译系统响应时间

B.1 机器翻译系统响应时间

响应时间为机器翻译系统接收到源语言文本到输出目标语言结果的时间间隔。在不考虑网络因素的影响下，对给定的句子长度和应用场景，不同句子长度下的机器翻译系统响应时间要求见表B.1。

表 B.1 机器翻译系统响应时间要求

句子长度[a] 字或词	响应时间[b] ms	典型应用场景
1-15	≤300ms	口语、日常会话
16-30	≤400ms	短书面文本
31-45	≤600ms	书面文本
45-60	≤700ms	长书面文本
60-150	≤800ms	超长书面文本
[a] 句子长度的计数，中文、日文、韩文等应以字为单位进行计量，英文、法文、德文等西文应以词为单位进行计量。 [b] 响应时间为机器翻译系统接收到源语言文本到输出目标语言结果的时间间隔。		

附 录 C
（规范性附录）
机器翻译综合差错率计算

C.1 机器翻译综合差错率计算方法

机器翻译综合差错率计算公式，见式（C.1）。

$$综合差错率=\frac{c_{\mathrm{I}}D_{\mathrm{I}}+c_{\mathrm{II}}D_{\mathrm{II}}+c_{\mathrm{III}}D_{\mathrm{III}}+c_{\mathrm{IV}}D_{\mathrm{IV}}}{\mathrm{w}}\times 100\% \quad \cdots\cdots\cdots\cdots\cdots\cdots (C.1)$$

式中：

w ——源语言内容计算总字词数；

D_{I}，D_{II}，D_{III}，D_{IV} ——分别为I、II、III、IV类差错出现的次数，重复性错误按一次计算；

c_{I}，c_{II}，c_{III} c_{IV} ——分别为I、II、III、IV类差错的系数，取值如下：

——$c_{\mathrm{I}}=3$；

——$c_{\mathrm{II}}=1$；

——$c_{\mathrm{III}}=0.5$；

——$c_{\mathrm{IV}}=0.25$。

其中，

I-IV分别表示机器翻译差错类别，具体如下：

a） I 类差错：译文表述存在核心语义差错或关键字词（数字）、句段的漏译、错译；

b） II 类差错：一般语义差错，非关键字词（数字）、句段的漏译、错译，译文表述存在用词、语法错误或表述含混；

c） III 类差错：专业术语不准确、不统一、不符合标准或惯例，或专用名词错译；

d） IV 类差错：计量单位、符号、缩略语等未按规（约）定译法。

其中，源语言内容计算总字词数时，中文、日文、韩文等以字为单位进行计算；英文、法文、德文等以词为单位进行计算。

参 考 文 献

[1] GB/T 19363.1-2008 翻译服务规范 第1部分：笔译
[2] GB/T 19363.2-2006 翻译服务规范 第2部分：口译
[3] GB/T 19682-2005 翻译服务译文质量要求
[4] ISO/IEC 20382-1 Information technology-User interfaces-Face-to-face speech translation-Part 1: User interface
[5] ISO/IEC 20382-2 Information technology-User interface-Face-to-face speech translation-Part 2: System architecture and functional components
[6] ISO 18587 Translation services-Post-editing of machine translation output-Requirements
[7] ASTM F2575 Standard guide for quality assurance in translation
[8] ITU-T F.745 Functional requirements for network-based speech-to-speech translation services
[9] ITU-T H.625 Architecture for network-based speech-to-speech translation services
[10] ITU-T E.FAST User interface for face-to-face speech translation considering human factors

中国电子工业标准化技术协会（CESA）是全国电子信息产业标准化组织和标准化工作者自愿组成的社会团体。广泛联系全国电子信息产业标准化机构和标准化工作者，协助政府部门搞好电子信息产业标准化工作，开拓信息技术领域的标准化工作是中国电子工业标准化技术协会的主要工作内容之一。中国境内从事科研开发、制造、营销和服务的企事业单位、高等院校、社会组织和个人均可随时向中国电子工业标准化技术协会团体标准工作部提出团体标准项目建议。

中国电子工业标准化技术协会标准按照《电子工业标准化技术协会协会团体标准管理办法》进行制定和管理。

在本标准实施过程中，如发现需要修改或补充之处，请将意见和有关资料寄至中国电子工业标准化技术协会，以便修订时参考。

中国电子工业标准化技术协会地址：北京市海淀区万寿路27号
电话：010 - 64102952 电子邮箱： standards@cesa.cn
网址：www.cesa.cn