

T/AI 137.1-2025

T/IAI 137.1-2025



版权保护文件

版权所有归属于该标准的发布机构，除非有其他规定，否则未经许可，此发行物及其章节不得以其他形式或任何手段进行复制、再版或使用，包括电子版，影印件，或发布在互联网及内部网络等。使用许可可于发布机构获取。

T/AI 137.1-2025

目 次

前言	II
引言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	2
5 总则	2
5.1 起始下标	2
5.2 参数信息	2
5.3 编程语言	3
5.4 自动广播	3
5.5 状态处理	3
5.6 接口一致性	4
5.7 泛型标量类型	4
6 数据结构	4
6.1 概要	4
6.2 元素类型	4
6.3 形状信息	4
6.4 布局信息	4
6.5 设备信息	5
6.6 其它扩展	5
7 基础数学类算子接口	5
7.1 接口列表	5
7.2 接口功能和参数	6
7.3 算子接口最小集	90
附录 A (资料性) 基础数学类算子接口 C 语言参考定义示例	93
A.1 数据结构	93
A.2 基础数学操作算子接口	95

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》的规定起草。

本文件为T/AI 131《人工智能 算子接口》的第1部分。

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专利的责任。

本文件由新一代人工智能产业技术创新战略联盟AI标准工作组提出。

本文件由中关村视听产业技术创新联盟归口。

本文件起草单位：北京大学、北京大学长沙计算与数字经济研究院、中国电子技术标准化研究院、鹏城实验室、中国科学院软件研究所、北京百度网讯科技有限公司、华为技术有限公司、长沙壹零壹壹科技有限公司、中科寒武纪科技股份有限公司、上海商汤智能科技有限公司、浪潮电子信息产业股份有限公司、北京交通大学。

本文件主要起草人：杨超、勾海鹏、胡晓光、樊春、陈军、鲍薇、杨雨泽、敖玉龙、李克森、马艳军、于佃海、范睿博、贾梦珠、段炼、李敏、马银萍、付振新、黎子毅、龙汀汀、张云飞、关贺、胡帅、赵海英、张伟民、马珊珊、张军、李卉、郑若琳、栾晓旭、唐轶男、王丽、吴庚、钱晨、沈芷月、宋文林、刘文枫、高歌、聂简荻、陈德良、王新民、刘伟、杨征、冯海军、崔晓冉、汪群博。

T/AI 137.1-2025

人工智能 算子接口 第1部分：基础数学类

1 范围

本文件规定了面向人工智能领域的基础数学类算子接口的基本功能及参数要求。
本文件适用于人工智能基础数学类算子库的设计、开发与应用，以及相关软硬件及系统的研制。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1

算子 operator

人工智能相关硬件操作的封装实现，是构建人工智能应用的基础组成部分。

3.2

算子接口 operator interface

是一套标准化的规定，用于描述和实现各类数学运算、逻辑运算或其他复杂计算单元的接口。

注：算子接口定义了一套规范化的API，使得开发人员能够以一致的方式来调用这些运算，而不必关心底层的具体实现细节。

3.3

封装 encapsulation

将数据和与数据相关的操作绑定在一起，形成一个独立单元的过程。

注：它对外隐藏其内部实现细节，可通过接口与外部进行交互。通过封装，可以增强代码的可读性、可维护性和可复用性。

3.4

张量 tensor

由同一类型元素所组成的多维数组。

3.5

稠密张量 dense tensor

元素全部或大部分为非零值的张量。

注：稠密张量一般采用非压缩的存储方式，即稠密存储，按照某种顺序存储张量的所有元素。

3.6

稀疏张量 sparse tensor

元素全部或大部分为零值的张量。

注：稀疏张量一般采用压缩的存储方式，即稀疏存储，只存储张量的非零值元素。

3.7

张量切片 tensor slice

指从张量中选取或提取部分元素的操作。假定张量 x 是形状为 $[D_0, D_1, \dots, D_{n-1}]$ 的 n 维张量，那么 $x_{[d_0, d_1, \dots, d_i, \dots, d_k]}$ 是指取 x 在前 k 个维度上索引分别为 $d_0, d_1, \dots, d_i, \dots, d_k$ 的元素集合，其中 $0 \leq d_i < D_i$ （对于所有 $0 \leq i < k \leq n$ 且 i 为整数），它的返回结果为数值或形状为 $[D_k, D_{k+1}, D_{k+2}, \dots, D_{n-1}]$ 的 $n - k$ 维张量。本文中张量切片也可写作为 $x[d_0, d_1, \dots, d_i, \dots, d_k]$ 。

注：如果上述 k 等于 n ，那么返回结果为数值，该操作也被称为“张量索引(tensor index)”。

3.8

人工智能加速处理器 artificial intelligence accelerator processor

具备适配人工智能算法的运算微架构，能够完成人工智能应用运算处理的集成电路元件。

3.9

批 batch

训练样本的一部分。

3.10

广播 broadcasting

针对形状不同的两个张量进行逐元素运算时可采用的方法。

4 缩略语

下列缩略语适用于本文件。

ASIC: 专用集成电路(Application Specific Integrated Circuit)

CPU: 中央处理单元(Central Processing Unit)

FPGA: 现场可编程门阵列(Field Programmable Gate Array)

GPU: 图形处理单元(Graphic Processing Unit)

NPU: 神经网络处理器(Neural-network Processing Unit)

5 总则

5.1 起始下标

本文件所涉及的数据或数据结构中的元素索引编号和计数应从0开始。

5.2 参数信息

5.2.1 本文件定义的算子接口参数列表、参数顺序、返回值描述方式等为算子接口的一种结构化定义，在实际实现过程中可根据应用情况进行调整。

5.2.2 本文件定义的算子接口参数列表为实现该算子接口的最小参数集合，在实际实现过程中可根据应用情况进行参数的扩充。

5.2.3 本文件对算子接口的参数定义了必选与可选两类，参数顺序执行必选参数在前，可选参数在后。其中必选参数表示必须实现且使用接口时必须传入，可选参数表示必须实现但使用接口时可以不传（即有默认值）。

注：需要说明的是，本文件定义的“数组长度”、“随机种子”、“数据格式”可选参数可根据实际编程语言和应用选择是否实现。若实际编程接口中参数顺序与本文件不一致，但是参数完整性保持一致，也符合本标准。

5.2.4 本文件针对算子接口参数中的可选参数，设置了默认值。默认值是在用户没有对算子的可选参数进行赋值时而被算子采用的数值，其数值大小是根据目前市场主流上层应用而选取的，例如PyTorch、PaddlePaddle、MindSpore、TensorFlow等上层应用。

5.3 编程语言

算子接口可用多种编程语言实现，C语言参考定义参见附录 A。

5.4 自动广播

两个形状不同的张量进行逐元素运算需要进行自动广播，扩充维度数值较小的张量以匹配维度数值较大的张量来完成运算。

5.5 状态处理

算子执行时应返回状态码，状态码分为成功和未成功（即错误）两类。对于算子执行过程中不可修复的错误，应直接退出；对于可修复错误，应采用返回错误状态码方式，由算子接口使用者决定如何处理错误。返回错误状态码时应优先返回相对具体的错误状态码，部分状态码应符合表1的规定。

表 1 状态码

状态码名称	类别	状态码
操作成功	成功	STATUS_SUCCESS
类型不匹配	未成功（错误）	STATUS_TYPE_MISMATCH
维度不匹配		STATUS_DIMENSIONS_MISMATCH
对象未初始化		STATUS_UNINITIALIZED_OBJECT
非法参数		STATUS_INVALID_ARGUMENT
内存不足		STATUS_ALLOC_FAILED

表1 状态码（续）

状态码名称	类别	状态码
超出范围	未成功（错误）	STATUS_OUT_OF_RANGE
其它内部错误		STATUS_INTERNAL_ERROR

5.6 接口一致性

稀疏张量和稠密张量的构建应采用不同的接口，而基于两种张量的各类操作可以采用统一的接口或是不同的接口，如果采用不同的接口，接口应添加标识。

5.7 泛型标量类型

对于定义算子接口时无法确定数据类型的参数，可将其定义为泛型标量类(Scalar)。泛型标量类型表示参数可以根据运行时的实际数据传入不同类型的数值。在接口实现过程中，若部分参数定义为泛型标量类型，则认为该接口实现符合本文件定义。

6 数据结构

6.1 概要

本文件使用张量作为核心数据结构，承载数据，张量数据结构应包含元素类型、形状信息、布局信息、设备信息及其他扩展信息。

6.2 元素类型

用于描述元素的数据类型。包括无符号整数：8位，16位，32位，64位；有符号整数：8位、16位、32位、64位；浮点实数：16位、32位、64位；浮点复数：（32+32）位、（64+64）位；布尔类型，字符串类型等。

6.3 形状信息

用于描述张量维数和每一维的大小。如果张量维数为 n ，每一维对应的大小为 $d_0, d_1, d_2, \dots, d_{n-1}$ ，则张量的形状可以表示为 $[d_0, d_1, d_2, \dots, d_{n-1}]$ ，最左边的为第1维，最右边的为第 n 维。例如：标量维数为0，形状为 $[\]$ ；向量维数为1，形状为 $[d_0]$ ；矩阵维数为2，形状为 $[d_0, d_1]$ 。

6.4 布局信息

用于描述张量的存储格式以及张量各个维度的逻辑顺序。存储格式包括稠密存储和稀疏存储。逻辑顺序指张量遍历读取和存储具体数据时，各个维度的优先顺序。

当使用稠密存储时，对于形状为 $[d_0, d_1, d_2, \dots, d_{n-1}]$ ，逻辑顺序为 $[0, 1, \dots, n-2, n-1]$ 的 n 维向量，首先取逻辑顺序中最左边第1位元素 r_0 ，对应到当前要遍历的维度 d_{r_0} ，按照 $(0, 1, 2, \dots, d_{r_0} - 1)$ 的次序存储，其次取逻辑顺序中最左边第2位元素 r_1 ，对应到当前要遍历的维度 d_{r_1} ，按照 $(0, 1, 2, \dots, d_{r_1} - 1)$ 的次序存储，以此类推。例如：张量形状为 $[2, 3]$ ，逻辑顺序为 $[0, 1]$ ，则先取逻辑顺序最左侧的第一个元素 $r_0 = 0$ ，对应维度 $d_{r_0} = d_0 = 2$ ，其次取逻辑顺序中的第二个元素 $r_1 = 1$ ，对应维度 $d_{r_1} = d_1 = 3$ ，则张量数据物理上第一个元素对应的坐标为 $[0, 0]$ ，第二个元素对应的坐标为 $[1, 0]$ ，第三个元素对应的坐标为 $[0, 1]$ ，以此类推。

当使用稠密存储时，默认的逻辑顺序应为 $[n-1, n-2, \dots, 0]$ 。

当使用稀疏存储时，标准实施者应对所采用的具体格式进行详细说明。

6.5 设备信息

用于描述张量数据存储和运算的设备类型和设备编号。其中设备类型包括CPU、GPU、NPU、FPGA和ASIC等任何支持AI操作的设备或人工智能加速处理器。如果存在多个同类型的设备，可通过指定设备编号来区分。

6.6 其它扩展

用于提供自定义的扩展功能，比如内存管理等相关信息。

7 基础数学类算子接口

7.1 接口列表

基础数学类算子接口列表见表2。

表 2 基础数学类算子接口列表

类别	名称
张量创建与销毁	拷贝已有数据创建稠密张量，引用已有数据创建稠密张量，创建全零稠密张量，按指定值创建稠密张量，创建未初始化张量，创建连续内存张量，以均匀分布随机数创建稠密张量，以正态分布随机数创建稠密张量，以伯努利分布创建稠密张量，以多项式分布创建稠密张量，以数字序列创建稠密张量，创建线性空间均匀分布稠密张量，创建稀疏张量，创建量化张量，复制张量，复制对角线元素创建张量，销毁张量
张量查询与检查	形状查询，有限检查，无穷检查，未定义数检查，元素个数检查，秩查询，连续内存检查
张量转换	转换数据类型，改变张量形状，扩展维度，删除维度，张量转置，张量分拆，张量合并，张量堆叠，张量拆堆，张量切片，张量重复，张量补全，张量逆序变换，张量循环滚动变换，张量形状裁剪，张量数值裁剪，张量聚集，张量发散更新，扩展张量，展平张量，张量翻转，张量正负判断，条件判断组合张量，一维张量扩充，张量选择
算术操作	张量加法操作，张量减法操作，张量乘法操作，张量乘加操作，张量除法操作，张量整除操作，张量真除法操作，张量取模操作，张量逐元素取最大值，张量逐元素取最小值，张量绝对值操作，张量取倒数操作，张量对角线元素求和操作
比较操作	判断张量是否相等，判断张量是否不等，判断张量是否大于，判断张量是否大于等于，判断张量是否小于，判断张量是否小于等于，判断张量是否值相近

表 2 基础数学类算子接口列表（续）

类别	名称
逻辑操作	张量的“逻辑与”操作，张量的“逻辑或”操作，张量的“逻辑非”操作，张量的“逻辑异或”操作
位操作	逐位“与”操作，逐位“或”操作，逐位“异或”操作，逐位反转操作，逐位左移操作，逐位右移操作
幂操作	幂，平方根，平方根倒数，平方数
舍入操作	向下取整，向上取整，截断取整，就近取整，通用舍入取整
三角函数	正弦函数，余弦函数，正切函数，反正弦函数，反余弦函数，反正切函数
双曲函数	双曲正弦函数，双曲余弦函数，双曲正切函数，反双曲正弦函数，反双曲余弦函数，反双曲正切函数
指对函数	指数函数，指数函数扩展，以e为底的对数函数，以e为底的对数函数扩展，以10为底的对数函数，以2为底的对数函数
规约操作	规约，前缀和
索引操作	最大索引，最小索引，排序索引，Top K索引，非零索引
复数操作	复数构建、复数共轭、获取虚部、获取实部
信号处理	快速傅里叶变换、快速逆傅里叶变换
线性代数操作	矩阵乘法，向量内积，LU矩阵分解，Cholesky矩阵分解，QR矩阵分解，SVD奇异值分解，线性方程组求解，最小二乘，矩阵求逆，求特征值以及特征向量，矩阵范数，线性操作，双线性操作
插值操作	插值操作

7.2 接口操作和参数

7.2.1 张量创建与销毁

7.2.1.1 拷贝已有数据创建稠密张量

7.2.1.1.1 功能

创建张量，若初始化数组不为空，则拷贝数组的值到张量空间进行初始化，否则创建一个空张量。

7.2.1.1.2 接口参数

拷贝已有数据创建稠密张量函数前向接口应符合表3，C语言示例见A.2.1.1。

表 3 拷贝已有数据创建稠密张量函数参数列表

参数名	类型	可选/必选	描述
数据类型	输入	可选	可以为有符号整数、无符号整数、浮点实数、浮点复数、布尔等，默认为浮点实数

表 3 拷贝已有数据创建稠密张量函数参数列表（续）

参数名	类型	可选/必选	描述
设备类型	输入	可选	可以为CPU、GPU、NPU、FPGA和ASIC等，需保证初始化张量的设备类型和张量的设备类型一致，默认为CPU
张量形状	输入	必选	包括张量的维度，维度大小数组，以及布局信息，其中布局信息没有指定时采用默认布局
初始化张量	输入	必选	如果为空，则表示不进行初始化操作
初始化数组的长度	输入	必选	以字节为单位，拷贝的数据量取决于张量空间和初始化数组之间尺寸较小者。若张量空间大于初始化数组，则按初始化数组大小进行拷贝，张量中未被覆盖部分应被初始化为0。若张量空间小于初始化数组，则按照张量空间大小进行拷贝
输出张量	输出	必选	新创建的张量

7.2.1.1.3 接口返回值

没有错误：操作成功。

类型不匹配：张量的数据类型不一致。

非法参数：输入参数超出范围。

内存不足：创建张量分配空间不足。

维度不匹配：张量和初始化张量维度不匹配。

其它内部错误：内部调用操作出错。

7.2.1.2 引用已有数据创建稠密张量

7.2.1.2.1 功能

创建张量，对参数传入数组的数据进行引用。

7.2.1.2.2 接口参数

引用已有数据创建稠密张量函数前向接口应符合表4，C语言示例见A.2.1.2。

表 4 引用已有数据创建稠密张量函数参数列表

参数名	类型	可选/必选	描述
数据类型	输入	可选	可以为有符号整数、无符号整数、浮点实数、浮点复数、布尔等，默认为浮点实数
设备类型	输入	可选	可以为CPU、GPU、NPU、FPGA和ASIC等，需保证初始化张量的设备类型和张量的设备类型一致，默认为CPU
张量形状	输入	必选	包括张量的维度，维度大小数组，以及布局信息，其中布局信息没有指定时采用默认布局
初始化张量	输入	可选	如果为空，则表示不进行初始化操作
输出张量	输出	必选	新创建的张量

7.2.1.2.3 接口返回值

没有错误：操作成功。

非法参数：权重与偏置张量的维度与输入输出张量不匹配。

内存不足：创建张量分配空间不足。

其它内部错误：内部调用操作出错。

7.2.1.3 创建全零稠密张量

7.2.1.3.1 功能

创建元素值全为0的稠密张量。

7.2.1.3.2 接口参数

创建全零稠密张量函数前向接口应符合表5，C语言示例见A.2.1.3。

表5 创建全零稠密张量函数参数列表

参数名	类型	可选/必选	描述
数据类型	输入	可选	可以为有符号整数、无符号整数、浮点实数、浮点复数、布尔等，默认为浮点实数
设备类型	输入	可选	可以为CPU、GPU、NPU、FPGA和ASIC等，默认为CPU
张量形状	输入	必选	包括张量的维度，维度大小数组，以及布局信息，其中布局信息没有指定时采用默认布局
输出张量	输出	必选	新创建的张量

7.2.1.3.3 接口返回值

没有错误：操作成功。

非法参数：表示参数出错。

内存不足：创建张量分配空间不足。

其它内部错误：内部调用操作出错。

7.2.1.4 按指定值创建稠密张量

7.2.1.4.1 功能

创建元素为某指定值的稠密张量。

7.2.1.4.2 接口参数

按指定值创建稠密张量函数前向接口应符合表6，C语言示例见A.2.1.4。

表6 按指定值创建稠密张量函数参数列表

参数名	类型	可选/必选	描述
数据类型	输入	可选	可以为有符号整数、无符号整数、浮点实数、浮点复数、布尔等，默认为浮点实数
设备类型	输入	可选	可以为CPU、GPU、NPU、FPGA和ASIC等，默认为CPU
张量形状	输入	可选	包括张量的维度，维度大小数组，以及布局信息，其中布局信息没有指定时采用默认布局

表6 按指定值创建稠密张量函数参数列表（续）

参数名	类型	可选/必选	描述
初始化值	输入	可选	用来初始化的数据，要求与参数“数据类型”所对应数据类型兼容
输出张量	输出	必选	新创建的张量

7.2.1.4.3 接口返回值

没有错误：操作成功。

非法参数：表示参数出错。

内存不足：创建张量分配空间不足。

其它内部错误：内部调用操作出错。

7.2.1.5 创建未初始化张量

7.2.1.5.1 功能

根据指定的类型、设备和形状创建张量。创建的张量中的数据为申请到的内存原本的数据内容，未经归零或者初始化，可能是任意合法的值。

7.2.1.5.2 接口参数

创建未初始化张量函数前向接口应符合表7，C语言示例见A.2.1.5。

表7 创建未初始化张量函数参数列表

参数名	类型	可选/必选	描述
数据类型	输入	可选	表示新张量的数据类型，可以为有符号整数、无符号整数、浮点实数、浮点复数、布尔等，默认为浮点实数
设备类型	输入	可选	可以为CPU、GPU、NPU、FPGA和ASIC等，默认为CPU
形状维度	输入	必选	新张量的形状维度
输出张量	输出	必选	新创建的张量

7.2.1.5.3 接口返回值

没有错误：操作成功。

内存不足：表示输出向量分配空间不足

非法参数：表示其他参数不合法。

其它内部错误：内部调用操作出错。

7.2.1.6 创建连续内存张量

7.2.1.6.1 功能

如果输入张量的数据在内存中是连续存放的，则不做处理返回这个连续的输入张量，否则会将当前张量的数据复制到一段连续的内存上，返回复制后拥有连续内存的张量。

7.2.1.6.2 接口参数

生成连续内存张量函数前向接口应符合表8，C语言示例见A.2.1.6。

表 8 生成连续内存张量函数参数列表

参数名	类型	可选/必选	描述
输入张量	输入	必选	表示输入张量
输出张量	输出	必选	表示具有连续内存的输出张量

7.2.1.6.3 接口返回值

没有错误：操作成功。

对象未初始化：表示输入张量对象不合法

内存不足：表示输出向量分配空间不足

非法参数：表示其他参数不合法。

其它内部错误：内部调用操作出错。

7.2.1.7 以均匀分布随机数创建稠密张量

7.2.1.7.1 功能

创建的稠密张量中的每个元素均是从符合对应均匀分布中随机生成的。

7.2.1.7.2 接口参数

以均匀分布随机数创建稠密张量函数前向接口应符合表9，C语言示例见A.2.1.7。

表 9 以均匀分布随机数创建稠密张量函数参数列表

参数名	类型	可选/必选	描述
数据类型	输入	可选	可以为有符号整数、无符号整数、浮点实数、浮点复数、布尔等，默认为浮点实数
设备类型	输入	可选	可以为CPU、GPU、NPU、FPGA和ASIC等，默认为CPU
张量形状	输入	必选	包括张量的维度，维度大小数组，以及布局信息，其中布局信息没有指定时采用默认布局
最小值	输入	可选	生成随机数所遵循均匀分布最小值，要求与参数“数据类型”所对应数据类型兼容，默认参数类型能表达的最小值
最大值	输入	可选	生成随机数所遵循均匀分布最大值，要求与参数“数据类型”所对应数据类型兼容，默认参数类型能表达的最大值
随机种子	输入	可选	生成随机数的种子。若随机种子为0，表示使用系统的随机种子；否则，使用随机种子来生成随机数，默认系统种子
输出张量	输出	必选	新创建的张量

7.2.1.7.3 接口返回值

没有错误：操作成功。

非法参数：表示参数出错。

内存不足：创建张量分配空间不足。

其它内部错误：内部调用操作出错。

7.2.1.8 以正态分布随机数创建稠密张量

7.2.1.8.1 功能

创建的稠密张量中的每个元素均是从符合对应正态分布中随机生成的。

7.2.1.8.2 接口参数

以正态分布随机数创建稠密张量函数前向接口应符合表10，C语言示例见A.2.1.8。

表 10 以正态分布随机数创建稠密张量函数参数列表

参数名	类型	可选/必选	描述
数据类型	输入	可选	可以为浮点实数、浮点复数等，默认浮点实数
设备类型	输入	可选	可以为CPU、GPU、NPU、FPGA和ASIC等，默认CPU
张量形状	输入	必选	包括张量的维度，维度大小数组，以及布局信息，其中布局信息没有指定时采用默认布局
均值	输入	可选	生成随机数所遵循正态分布均值，要求与参数“数据类型”所对应数据类型兼容，默认为0
标准差	输入	可选	生成随机数所遵循正态分布标准差，要求与参数“数据类型”所对应数据类型兼容，默认为1
随机种子	输入	可选	生成随机数的种子，若随机种子为0，表示使用系统的随机种子；否则，使用随机种子来生成随机数，默认系统种子
输出张量	输出	必选	新创建的张量

7.2.1.8.3 接口返回值

没有错误：操作成功。

非法参数：表示参数出错。

内存不足：创建张量分配空间不足。

其它内部错误：内部调用操作出错。

7.2.1.9 以伯努利分布创建稠密张量

7.2.1.9.1 功能

创建服从伯努利分布（0-1分布）的稠密张量。

7.2.1.9.2 接口参数

创建伯努利分布稠密张量函数前向接口应符合表11，C语言示例见A.2.1.9。

表 11 创建伯努利分布稠密张量函数参数列表

参数名	类型	可选/必选	描述
输入张量	输入	必选	表示输入的概率值，为一个张量，类型为浮点数
输出张量	输出	必选	服从伯努利分布的随机张量，形状和数据类型与输入相同

7.2.1.9.3 接口返回值

没有错误：操作成功。

非法参数：表示参数出错。

内存不足：创建张量分配空间不足。

其它内部错误：内部调用操作出错。

7.2.1.10 以多项式分布创建稠密张量

7.2.1.10.1 功能

以输入张量伪概率，创建服从多项式分布的张量。

7.2.1.10.2 接口参数

以多项式分布创建稠密张量函数前向接口应符合表12，C语言示例见A.2.1.10。

表 12 以多项式分布创建稠密张量函数参数列表

参数名	类型	可选/必选	描述
输入张量	输入	必选	表示输入的概率值，为一个张量，类型为浮点数
采样次数	输入	可选	表示采样的次数，默认值为1
采样方式	输入	可选	表示是否为有放回采样，默认值为是
输出张量	输出	必选	服从多项式分布的随机张量，形状和数据类型与输入相同

7.2.1.10.3 接口返回值

没有错误：操作成功。

非法参数：表示参数出错。

内存不足：创建张量分配空间不足。

其它内部错误：内部调用操作出错。

7.2.1.11 创建随机排列的一维稠密张量

7.2.1.11.1 功能

从[0, 最大值)范围内取出每个整数，随机排列后组成一个一维稠密张量。

7.2.1.11.2 接口参数

创建随机排列（randperm）的一维张量函数前向接口应符合表13，C语言示例见A.2.1.11。

表 13 创建随机排列（randperm）的一维张量函数参数列表

参数名	类型	可选/必选	描述
数据类型	输入	可选	可以为有符号整数、无符号整数、浮点实数和浮点复数，默认为浮点实数
设备类型	输入	可选	可以为CPU、GPU、NPU、FPGA和ASIC等，默认为CPU
张量长度	输入	必选	表示张量的长度，也可表示范围的最大值，应该大于0
输出张量	输出	必选	新创建的张量

7.2.1.11.3 接口返回值

没有错误：操作成功。

非法参数：表示参数出错。

内存不足：创建张量分配空间不足。

其它内部错误：内部调用操作出错。

7.2.1.12 以数字序列创建稠密张量

7.2.1.12.1 功能

创建一维稠密张量的数字序列，张量中的值以初始值开始，按步长进行扩展到上限值为止。

7.2.1.12.2 接口参数

以数字序列创建稠密张量函数前向接口应符合表14，C语言示例见A.2.1.12。

表 14 以数字序列创建稠密张量函数参数列表

参数名	类型	可选/必选	描述
数据类型	输入	可选	可以为有符号整数、无符号整数、浮点实数和浮点复数，默认为浮点实数
设备类型	输入	可选	可以为CPU、GPU、NPU、FPGA和ASIC等，默认为CPU
初始值	输入	必选	创建数字序列的第一个值，要求与参数“数据类型”所对应数据类型兼容
上限值	输入	必选	创建数字序列的上限，要求与参数“数据类型”所对应数据类型兼容
步长	输入	必选	生成的相邻两个随机数的变化步长，要求与参数“数据类型”所对应数据类型兼容
输出张量	输出	必选	新创建的张量

7.2.1.12.3 接口返回值

没有错误：操作成功。

非法参数：表示参数出错。

内存不足：创建张量分配空间不足。

其它内部错误：内部调用操作出错。

7.2.1.13 创建线性空间均匀分布稠密张量

7.2.1.13.1 功能

创建一维稠密张量，张量中的值是在区间起点和区间终点之间，均匀间隔的N个数。

7.2.1.13.2 接口参数

创建线性空间均匀分布稠密张量函数前向接口应符合表15，C语言示例见A.2.1.13。

表 15 创建线性空间均匀分布稠密张量函数参数列表

参数名	类型	可选/必选	描述
数据类型	输入	可选	可以为浮点实数和浮点复数，默认为浮点实数
设备类型	输入	可选	可以为CPU、GPU、NPU、FPGA和ASIC等，默认为CPU
区间起点	输入	必选	区间的起点，包含在区间内，应与数据类型参数所对应数据类型兼容
区间终点	输入	必选	区间的终点，包含在区间内，应与数据类型参数所对应数据类型兼容
数据个数	输入	必选	创建张量的元素个数
输出张量	输出	必选	1维输出张量，长度为指定区间内数据的个数

7.2.1.13.3 接口返回值

没有错误：操作成功。

非法参数：由起点、终点、数据个数指定的区间不合法。

分配空间失败：创建张量分配空间不足。

7.2.1.14 创建稀疏张量

7.2.1.14.1 功能

创建稀疏张量，包括二维及高维稀疏张量的创建。

7.2.1.14.2 接口参数

创建稀疏张量函数前向接口应符合表16，C语言示例见A.2.1.14。

表 16 创建稀疏张量函数参数列表

参数名	类型	可选/必选	描述
数据类型	输入	可选	可以为有符号整数、无符号整数、浮点实数、浮点复数、布尔等，默认为浮点实数
设备类型	输入	可选	可以为CPU、GPU、NPU、FPGA和ASIC等，默认为CPU
张量形状	输入	必选	张量的形状，包括张量的维度，维度大小数组，以及布局信息，其中布局信息没有指定时采用默认布局
非零元坐标	输入	必选	创建稀疏张量中非零元的坐标，是二维（非零元个数，张量维度数）数组
非零元值数组	输入	必选	非零元素值组成的一维数组，长度为“非零元个数”
非零元个数	输入	必选	非零元的个数
输出张量	输出	必选	新创建的张量

7.2.1.14.3 接口返回值

没有错误：操作成功。

非法参数：表示参数出错。

内存不足：创建张量分配空间不足。

其它内部错误：内部调用操作

7.2.1.15 创建量化张量

7.2.1.15.1 功能

创建量化张量，若有初始化数组则用其值进行初始化，否则创建一个空张量。

7.2.1.15.2 接口参数

创建量化张量函数前向接口应符合表17，C语言示例见A.2.1.15。

表 17 创建量化张量函数参数列表

参数名	类型	可选/必选	描述
数据类型	输入	可选	可以为8位有符号对称量化整数和16位有符号对称量化整数、8位无符号非对称量化整数、8位有符号非对称量化整数、16位无符号非对称量化整数和16位有符号非对称量化整数、8位有符号通道级对称量化整数和16位有符号通道级对称量化整数，默认值为16位有符号对称量化整数
设备类型	输入	可选	可以为CPU、GPU、NPU、FPGA和ASIC等，默认为CPU

