

# 字符向量元胞数组

讲义和代码可以看配套的课程的第一个视频下载

配套课程b站:《MATLAB教程新手入门篇(数学建模清风主讲,适合零基础同学观看)》  
数学建模清风老师版权所有

定义

字符向量元胞数组是一种特殊的元胞数组,元胞中的每个数据都是字符向量类型

函数表格

函数名称	作用
<b>cellstr</b>	将其他类型的文本转换为字符向量元胞数组
<b>isletter</b>	确定哪些字符为字母(不仅包含英文字母,还包含其他语言)
<b>isspace</b>	确定哪些字符为空白字符
<b>isstrprop</b>	确定输入文本中的哪些字符属于指定类别
<b>strfind</b>	查找指定的文本,并返回相应的索引位置(此处将简要介绍 MATLAB 2020b 版本引入的 <b>pattern</b> 类型)
<b>strrep</b>	将旧文本替换成新的文本(这里我们会讲解 2016b 版本推出的另一个用于替换文本的函数: <b>replace</b> ,我们会介绍它和 <b>strrep</b> 函数的区别)
<b>strjoin</b>	对数组中的文本进行连接(通常使用空格、换行符等作为分隔符)
<b>strsplit</b>	在指定分隔符处拆分文本(例如在空格、换行符处进行拆分)

(1) cellstr函数: 将其他类型的文本转换为字符向量元胞数组

a1 = ['abc ',newline,char(9)]; cellstr(a1)	1x1 cell 数组 {'abc'}
a2 = char('a','123','good')	'a' '123' 'good'
cellstr(a2)	{'a'} {'123'} {'good'}

如果输入的数据类型是字符数组(包括字符向量和字符矩阵),cellstr函数在转换过程中会自动删除每行尾部的空白字符。

(2) isletter、isspace和isstrprop函数: 识别特定种类的字符

c1 = char('One is 1','Ten is 10.');	2x10 logical 数组
isletter(c1)	1 1 1 0 1 1 0 0 0 0 1 1 1 0 1 1 0 0 0 0
c2 = '本周放假3天,真开心!';	1x11 logical 数组
isletter(c2)	1 1 1 1 0 1 0 1 1 1 0
c3 = {'One is 1','Two is 2'};	1x2 logical 数组
% 不支持元胞数组类型,返回值全为逻辑值 0	0 0
isletter(c3)	

isspace函数和isletter函数的使用方法类似,它用于判断字符是否为空白字符,如空格、制表符、换行符。

isstrprop函数是前两个函数的进阶版本,它不仅能够识别字母和空白字符,还能识别更广泛的文本形式,如数字、标点符号、控制字符等。

isstrprop函数的基本用法是TF=isstrprop(str,category)。其中,第一个输入参数str表示文本数据,第二个输入参数category是你想要检测的字符类别。它支持多种字符类别,例如'alpha'(字母)、『digit'(数字)等。

k=strfind(str,pat)在str中搜索出现的pat。输出k指示str中每次出现的pat的起始索引。如果未找到pat,则strfind返回一个空数组[]。strfind函数执行区分大小写的搜索。

c1 = '人要是行,干一行行一行'; strfind(c1,'行')	4 8 9 11
strfind(c1,'行','ForceCellOutput',true)	1x1 cell 数组 {[4 8 9 11]}
c2 = 'Are you a youngster?'; strfind(c2,'your')	[]
c3 = {'How are you?'; 'Are you a youngster?'; 'You are so beautiful!'; strfind(c3,'you') % 结果区分大小写哦	3x1 cell 数组 {[] {5 11} {0x0 double}}

pat: 搜索模式,指定为下列值之一: 字符串标量、字符向量和pattern标量

MATLAB从2020b版本开始,引入了pattern类型作为一种新的搜索模式,pattern类型专门用于搜索和匹配文本,它提供了一些高级的匹配功能,大家可以在MATLAB官网搜索关键词:pattern。

newStr=strrep(str,old,new)将str中出现的所有old都替换为new。

c1 = '你我喜欢,我也喜欢你!'; strrep(c1,'喜欢','不喜欢')	'你不喜欢我,我也不喜欢你!'
c2 = 'matlab 很容易学!'; strrep(c2,'matlab','MATLAB')	'MATLAB 很容易学!'
c3 = {'abcdcccd','efcacd','dcace'}; strrep(c3,'c','p')	1x3 cell 数组 {'abdpdpd'} {'efpaped'} {'dpape'}
c4 = '我和你,他和她。'; strrep(c4,{'他','她'},{'Ta'}) % 分别将'他'和'她'替换为了'Ta'	2x1 cell 数组 {'我和你,Ta和她。'} {'我和你,Ta。'}
strrep(c4,{'和'},{'与'},{'爱'},{'喜欢'}) % 将'和'分别替换为了'与'、'爱'和'喜欢'	3x1 cell 数组 {'我与你,他与她。'} {'我爱你,他爱她。'} {'我喜欢你,他喜欢她。'}
strrep(c4,{'他','她'},{'ta','TA'}) % 将'他'和'她'分别替换为了'ta'和'TA'	2x1 cell 数组 {'我和你,ta和她。'} {'我和你,Ta和TA。'}

(3) strfind函数和strrep函数: 查找与替换文本

The diagram shows a sequence of 16 characters: a, b, a, b, a, b, a, b, a, b, a, b, a, b, a, b. Below this sequence, four substrings are highlighted in colored boxes: a green box containing 'a', an orange box containing 'b', a blue box containing 'a', and a yellow box containing 'b'. Dashed lines connect these substrings to their corresponding positions in the main sequence. The green box is connected to the first 'a' (index 1), the orange box to the second 'b' (index 2), the blue box to the third 'a' (index 3), and the yellow box to the fourth 'b' (index 4). This illustrates how overlapping substrings can be identified in a text sequence.

允许重叠子文本

The diagram shows a sequence of 16 characters: a, b, a, b, a, b, a, b, a, b, a, b, a, b, a, b. Below this sequence, four substrings are highlighted in colored boxes: a green box containing 'a', an orange box containing 'b', a blue box containing 'a', and a yellow box containing 'b'. Dashed lines connect these substrings to their corresponding positions in the main sequence. The green box is connected to the first 'a' (index 1), the orange box to the second 'b' (index 2), the blue box to the third 'a' (index 3), and the yellow box to the fourth 'b' (index 4). This illustrates how non-overlapping substrings can be identified in a text sequence.

不允许重叠子文本

```
cc = 'abababab';  
strrep(cc,'aba','*')
```

```
ans = '**b'
```

```
replace(cc,'aba','*')
```

```
ans = '*b*b'
```

replace函数在2016b版本中推出  
newStr=replace(str,old,new)将所有出现的子文本old替换为new。如果old包含多个子文本,则new必须与old具有相同的大小,或者必须为单个子文本

如果old或new不是标量(即不是单个文本,而是包含多个文本的字符向量元胞数组或者字符串数组),replace函数和strrep函数返回的结果有很大差异

c4 = '我和你,他和她。'; replace(c4,{'他','她'},{'Ta'})	'我和你,Ta和Ta。'
replace(c4,{'和'},{'与'},{'爱'},{'喜欢'})	报错:替换文本必须为标量,或者与匹配文本具有相同的大小。
replace(c4,{'他','她'},{'ta','TA'})	'我和你,ta和TA。'

str=strjoin(C)使用空格将C的元素与相邻的元素之间连接起来,构造str。C可以是字符向量元胞数组或字符串数组。  
str=strjoin(C,delimiter)使用delimiter将C的元素与相邻的元素之间连接起来,构造str。delimiter是分隔字符,它通常是字符向量类型。

c1 = {'apple','banana','pear'}; strjoin(c1) % 默认使用空格隔开	'apple banana pear'
strjoin(c1,',') % 指定分隔符为','	'apple, banana, pear'
strjoin(c1,newline) % 指定分隔符为换行符	'apple banana pear'
c2 = {'one','three','four'}; strjoin(c2,{'+'},{'='})	'one + three = four'

转义字符

转义字符	ASCII 码	含义
\0	0	空值
\a	7	响铃警报
\b	8	退格符
\t	9	水平制表符
\n	10	换行符
\v	11	垂直制表符
\f	12	换页符
\r	13	回车符
\\	92	反斜杠\

strjoin函数的分隔符字符向量中可以包括上表列举的转义字符,这样可以更方便的对文本进行连接。

c3 = {'今天','明天','每一天'}; strjoin(c3,'\\')	'今天\\明天\\一天'
strjoin(c3,'\\n') % 换行符	'今天 明天 一天'

C=strsplit(str)在空白处将str拆分为C, str指定为字符向量或字符串标量。  
C=strsplit(str,delimiter)在delimiter指定的分隔符处拆分str。

```
c1 = ['You are great!',newline,...  
'I salute you.'];  
strsplit(c1)
```

```
ans = 1x6 cell  
'You' 'are' 'great!' 'I' 'salute' 'you.'
```

```
c2 = ' You are great! ';  
strsplit(c2)
```

```
ans = 1x5 cell  
' ' 'You' 'are' 'great!' ''
```

```
c3 = 'if,no,or,and';  
strsplit(c3,',')
```

```
ans = 1x4 cell  
'if' 'no' 'or' 'and'
```

```
c4 = 'if,no,,and';  
strsplit(c4,',')
```

```
ans = 1x4 cell  
'if' 'no' 'or' 'and'
```

除此之外,MATLAB允许通过名称-值参数来调整strsplit函数的行为。  
名称-值参数是一种在MATLAB内置函数中常见的可选参数形式,它提供了一种直观的方式来配置附加选项。这类参数一般成对出现,其中一个参数作为固定文本,定义了要设置或修改的选项,另一个参数则是该选项的具体值。名称-值参数必须位于其他参数之后,且它们的出现顺序并不影响函数的执行结果。

(1) CollapseDelimiters — 多分隔符处理,指定为true(默认值)或者false

cc = 'if,no,,and'; strsplit(cc,',')	1x4 cell 数组 {'if'} {'no'} {'or'} {'and'}
strsplit(cc,',','CollapseDelimiters',false)	1x7 cell 数组 {'if'} {'no'} {0x0 char} {'or'}
% 从MATLAB2021a版本开始,也可以这样写: strsplit(cc,',','CollapseDelimiters' = false)	{0x0 char} {0x0 char} {'and'}

(2) DelimiterType — 分隔符类型,指定为普通模式'Simple'(默认)或者正则表达式模式'RegularExpression'

(4) strjoin函数: 对数组中的文本进行连接

(5) strsplit函数: 在指定分隔符处拆分文本