

Waipiti 模板



● 模板文件格式

模板文件中的每一行代表一个template。
每一个template中，专门的宏%x[row,col]
用于确定输入数据中的一个token。row用于
确定与当前的token的相对行数。col用于
确定列。

b.模板类型

- U模板
- B模板

```
U:tok:1:-1:%X[-1,0]
```

```
U:tok:1:+0:%X[0,0]
```

```
U:tok:1:+1:%X[1,0]
```

```
U:tok:2:-1:%X[-1,0]/%X[0,0]
```

```
U:tok:2:+0:%X[0,0]/%X[1,0]
```

```
U:tok:3:-2:%X[-2,0]/%X[-1,0]/%X[0,0]
```

```
U:tok:3:-1:%X[-1,0]/%X[0,0]/%X[1,0]
```

```
U:tok:3:+0:%X[0,0]/%X[1,0]/%X[2,0]
```



如已知下面的输入数据:

Input: Data

Rockwell

International

Corp.

's

Tulsa

特征模板形式为:

template

U:tok:1:2:%x[2,0]

U:tok:1:1:%x[1,0]

U:tok:1:0:%x[0,0]

U:tok:1:-1:%x[-1,0]

U:tok:1:-2:%x[-2,0]

U:tok:2:1:%x[1,0]/%x[0,0]

U:tok:2:-1:%x[0,0]/%x[-1,0]

B

|

|

|

|

<< CURRENT TOKEN

expanded feature

Tulsa

's

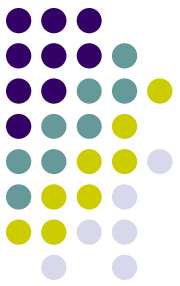
Corp.

International

Rockwell

's/Corp.

Corp./International



- 第一种是Unigram template:第一个字符是U，这是用于描述unigram feature的模板。
- 当你给出一个模板"U:tok:1:+0:%x[0,0]"，CRF会自动的生成一个特征函数集合(func1 ... funcN)，如：
func1 = if (output = B and feature=" U:tok:1:+0")return 1 else return 0
func2 = if (output = I and feature=" U:tok:1:+0")return 1 else return 0
func3 = if (output = O and feature=" U:tok:1:+0")return 1 else return 0
....
funcX1 = if (output = B and feature=" U:tok:1:-1") return 1 else return 0
funcX2 = if (output = I and feature=" U:tok:1:-1") return 1 else return 0
funcX3 = if (output = O and feature=" U:tok:1:-1") return 1 else return 0
...- 一个模型生成的特征函数的个数总数为 $L*N$ ，其中L是输出的类别数，N是根据给定的template扩展出的unique feature的数目。



- 第二种类型 **Bigram template**: 第一个字符是 **B**
- 使用这个模板，系统将自动产生当前输出 **token** 与前一个输出 **token** 的组合 (**Bigram**)。产生的可区分的特征的总数是 $L * L * N$ ，其中 **L** 是输出类别数，**N** 是这个模板产生的 **unique features** 数。

当你给出一个模板 "**B:tok:1:+0:%x[0,0]**", CRF 会自动的生成一个特征函数集合 (**func1 ... funcN**), 如:

```
func1 = if (output = B and preoutput =B and feature=" B:tok:1:+0")return 1 else return 0
func2 = if (output = I and preoutput =B and feature=" B:tok:1:+0")return 1 else return 0
func3 = if (output = O and preoutput =B and feature=" B:tok:1:+0")return 1 else return 0
```

....

```
funcX1 = if (output = B and preoutput =B and feature=" B:tok:1:-1") return 1 else return 0
funcX2 = if (output = I and preoutput =B and feature=" B:tok:1:-1") return 1 else return 0
funcX3 = if (output = O and preoutput =B and feature=" B:tok:1:-1") return 1 else return 0
```