商品标题信息提取方法的调研和比较

商品标题的信息提取或分类可以使用自然语言处理（NLP）技术来解决，主要可使用命名实体识别（NER）的方式来识别出商品标题中属于商品品牌、商品颜色或代号以及商品种类类别的命名实体，并通过识别后的命名实体进行分类和标注。以下是三个需要提取的内容以及相关的算法和方法的概述。

基于规则的方法：

1. 使用预定义的品牌列表和规则来匹配商品标题中的品牌名称。这种方法适用于已知品牌数量有限的情况。
2. 使用正则表达式或关键词匹配等方法，提取商品标题中的型号、代号和颜色信息。这种方法对于特定格式的商品标题可能比较有效，但可能无法处理变化较大的文本。
3. 构建一个商品类目的关键词表，通过匹配商品标题中的关键词来确定商品的类目或种类。这种方法简单直接，但需要维护一个准确的关键词表。

基于序列标注的方法：

1. 序列标注使用隐马尔可夫模型（HMM）
2. 可以使用命名实体识别（Named Entity Recognition，NER）的方法，将商品标题中的型号、代号和颜色作为实体进行识别和提取。常用的序列标注模型包括条件随机场（CRF）和循环神经网络（RNN）等。

基于神经网络和机器学习的方法：

1. 可以使用文本分类算法，如朴素贝叶斯、支持向量机（SVM）或深度学习模型，来训练一个品牌分类器，将商品标题分类为不同的品牌类别。
2. 可以使用文本分类算法，如朴素贝叶斯、支持向量机（SVM）或深度学习模型，来训练一个商品类目分类器，将商品标题分类为不同的类目。

具体方法和运行结果展示：

基于贝叶斯分类的分类器，将商品标题分类为不同类别

训练过程

1. 数据准备
2. 数据预处理
3. 特征提取，使用词袋模型进行特征提取并将标题由文本表示转换为向量表示
4. 创建品牌、型号、种类编码器和分类模型

分类结果展示

```
/Users/weidian/Documents/练习题/venv/bin/python /Users/weidian/Documents/命名实体识别/bayes/分类.py
商品标题：  小雨越南尾货已过检过验新版配饰黑银小号原厂小羊皮双大搭配三色链条的设计彰显着的不平凡大大的菱格蓬松慵懒的包型
分类结果:
品牌: ['小雨']
型号: ['黑银小号']
种类: ['包']
```

序列标注方法：HMM方法和CRF方法

标记精确度、召回率和混淆矩阵

1. HMM模型：

```
加载并评估hmm模型...
              precision    recall  f1-score   support

       B-PRO     0.5581    0.7273    0.6316        33
     M-NAME     0.9459    0.8537    0.8974        82
       B-EDU     0.9000    0.9643    0.9310       112
     B-TITLE     0.8811    0.8925    0.8867       772
     B-NAME     0.9800    0.8750    0.9245       112
       M-ORG     0.9002    0.9327    0.9162      4325
       E-PRO     0.6512    0.8485    0.7368        33
     E-RACE     1.0000    0.9286    0.9630        14
       E-EDU     0.9167    0.9821    0.9483       112
       M-PRO     0.4490    0.6471    0.5301        68
       M-EDU     0.9348    0.9609    0.9477       179
       M-LOC     0.5833    0.3333    0.4242        21
     E-TITLE     0.9514    0.9637    0.9575       772
      E-CONT     0.9655    1.0000    0.9825        28
           O     0.9568    0.9177    0.9369      5190
      B-CONT     0.9655    1.0000    0.9825        28
      E-NAME     0.9000    0.8036    0.8491       112
       E-ORG     0.8262    0.8680    0.8466       553
       B-LOC     0.3333    0.3333    0.3333         6
       B-ORG     0.8422    0.8879    0.8644       553
     M-TITLE     0.9038    0.8751    0.8892      1922
       E-LOC     0.5000    0.5000    0.5000         6
      M-CONT     0.9815    1.0000    0.9907        53
      B-RACE     1.0000    0.9286    0.9630        14
avg/total     0.9149    0.9122    0.9130     15100

Confusion Matrix:
         B-PRO  M-NAME   B-EDU B-TITLE  B-NAME   M-ORG   E-PRO  E-RACE   E-EDU   M-PRO   M-EDU   M-LOC E-TITLE  E-CONT       O  B-CONT  E-NAME   E-ORG   B-LOC   B-ORG M-TITLE   E-LOC  M-CONT  B-RACE
  B-PRO      24       0       1       0       0       5       0       0       0       3       0       0       0       0       0       0       0       0       0       0       0       0       0       0
 M-NAME       0      70       0       0       0       3       0       0       0       0       0       0       0       0       3       0       6       0       0       0       0       0       0       0
   B-EDU       0       0     108       0       0       1       0       0       0       0       3       0       0       0       0       0       0       0       0       0       0       0       0       0
B-TITLE       2       0       2     689       0      23       0       0       0       0       0       0       1       0      20       0       0       1       0       6      28       0       0       0
 B-NAME       0       0       0       0      98       2       0       0       0       0       0       0       0       0       8       0       0       0       0       1       0       0       0       0
   M-ORG      10       2       3      17       1    4034       7       0       3      25       1       5       4       1      70       0       2      42       0      38      53       3       1       0
   E-PRO       0       0       1       0       0       0      28       0       1       0       1       0       0       0       0       0       0       2       0       0       0       0       0       0
 E-RACE       0       0       0       0       0       0       0      13       0       0       0       0       0       0       1       0       0       0       0       0       0       0       0       0
   E-EDU       0       0       0       0       0       0       1       0     110       0       1       0       0       0       0       0       0       0       0       0       0       0       0       0
   M-PRO       1       0       1       0       0      18       0       0       0      44       1       0       0       0       0       0       0       3       0       0       0       0       0       0
   M-EDU       1       0       0       0       0       0       0       0       0       4     172       0       0       0       1       0       0       1       0       0       0       0       0       0
   M-LOC       0       0       0       0       0       7       0       0       0       0       0       7       0       0       4       0       0       2       1       0       0       0       0       0
 E-TITLE       0       0       0       0       0      15       0       0       1       0       0       0     744       0       6       0       0       0       0       4       2       0       0       0
  E-CONT       0       0       0       0       0       0       0       0       0       0       0       0       0      28       0       0       0       0       0       0       0       0       0       0
       O       3       0       1      26       0     204       4       0       2      12       1       0      26       0    4763       0       2      30       0      37      78       0       0       0
  B-CONT       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0      28       0       0       0       0       0       0       0       0
  E-NAME       0       2       0       0       0       0       0       0       0       0       0       0       0       0      16       0      90       3       0       0       0       0       0       0
   E-ORG       1       0       1       9       0      30       0       0       0       3       0       0       1       0      10       0       0     480       0       0      18       0       0       0
   B-LOC       0       0       0       0       0       0       0       0       0       0       0       0       0       0       1       0       0       0       2       3       0       0       0       0
   B-ORG       0       0       0       6       1      23       0       0       0       0       0       0       0       0      28       1       0       0       3     491       0       0       0       0
 M-TITLE       1       0       2      35       0     115       3       0       3       7       4       0       6       0      44       0       0      17       0       3    1682       0       0       0
   E-LOC       0       0       0       0       0       1       0       0       0       0       0       0       0       0       2       0       0       0       0       0       0       3       0       0
  M-CONT       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0      53       0
  B-RACE       0       0       0       0       0       0       0       0       0       0       0       0       0       0       1       0       0       0       0       0       0       0       0      13
```

2. CRF模型：

```
加载并评估crf模型...
              precision    recall  f1-score   support

       B-PRO     0.9091    0.9091    0.9091        33
     M-NAME     1.0000    0.9756    0.9877        82
       B-EDU     0.9820    0.9732    0.9776       112
     B-TITLE     0.9376    0.9339    0.9358       772
      B-NAME     1.0000    0.9821    0.9910       112
       M-ORG     0.9523    0.9563    0.9543      4325
       E-PRO     0.9091    0.9091    0.9091        33
     E-RACE     1.0000    1.0000    1.0000        14
       E-EDU     0.9910    0.9821    0.9865       112
       M-PRO     0.8354    0.9706    0.8980        68
       M-EDU     0.9824    0.9330    0.9570       179
       M-LOC     1.0000    0.8095    0.8947        21
     E-TITLE     0.9857    0.9819    0.9838       772
      E-CONT     1.0000    1.0000    1.0000        28
           O     0.9630    0.9732    0.9681      5190
     B-CONT     1.0000    1.0000    1.0000        28
      E-NAME     1.0000    0.9821    0.9910       112
       E-ORG     0.9199    0.9132    0.9165       553
       B-LOC     1.0000    0.8333    0.9091         6
       B-ORG     0.9636    0.9566    0.9601       553
     M-TITLE     0.9248    0.9022    0.9134      1922
       E-LOC     1.0000    0.8333    0.9091         6
      M-CONT     1.0000    1.0000    1.0000        53
     B-RACE     1.0000    1.0000    1.0000        14
avg/total     0.9543    0.9543    0.9542     15100

Confusion Matrix:
          B-PRO  M-NAME   B-EDU B-TITLE  B-NAME   M-ORG   E-PRO  E-RACE   E-EDU   M-PRO   M-EDU   M-LOC E-TITLE  E-CONT       O  B-CONT  E-NAME   E-ORG   B-LOC   B-ORG M-TITLE   E-LOC  M-CONT  B-RACE
  B-PRO      30       0       1       0       0       1       0       0       0       1       0       0       0       0       0       0       0       0       0       0       0       0       0       0
M-NAME       0      80       0       0       0       0       0       0       0       0       0       0       0       0       2       0       0       0       0       0       0       0       0       0
  B-EDU      0       0     109       0       0       1       0       0       0       0       1       0       0       0       0       0       0       0       0       1       0       0       0       0
B-TITLE      1       0       0     721       0      12       0       0       0       0       0       0       0       0       9       0       0       0       0       7      22       0       0       0
 B-NAME      0       0       0       0     110       0       0       0       0       0       0       0       0       0       2       0       0       0       0       0       0       0       0       0
  M-ORG      1       0       0      11       0    4136       1       0       0       5       0       0       2       0      91       0       0      12       0       1      65       0       0       0
  E-PRO      0       0       0       0       0       0      30       0       0       1       1       0       0       0       0       0       0       0       1       0       0       0       0       0
E-RACE       0       0       0       0       0       0       0      14       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0
  E-EDU      0       0       0       0       0       1       1       0     110       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0
  M-PRO      0       0       0       0       0       2       0       0       0      66       0       0       0       0       0       0       0       0       0       0       0       0       0       0
  M-EDU      1       0       0       0       0       5       0       0       0       4     167       0       0       0       1       0       0       1       0       0       0       0       0       0
  M-LOC      0       0       0       0       0       4       0       0       0       0       0      17       0       0       0       0       0       0       0       0       0       0       0       0
E-TITLE      0       0       0       0       0       2       0       0       1       0       0       0     758       0       9       0       0       1       0       0       1       0       0       0
 E-CONT      0       0       0       0       0       0       0       0       0       0       0       0       0      28       0       0       0       0       0       0       0       0       0       0
      O      0       0       0       6       0      75       0       0       0       0       0       0       5       0    5051       0       0      11       0       9      33       0       0       0
 B-CONT      0       0       0       0       0       0       0       0       0       0       0       0       0       0       0      28       0       0       0       0       0       0       0       0
 E-NAME      0       0       0       0       0       0       0       0       0       0       0       0       0       0       2       0     110       0       0       0       0       0       0       0
  E-ORG      0       0       0       0       0      14       0       0       0       0       0       0       1       0      13       0       0     505       0       0      20       0       0       0
  B-LOC      0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       5       1       0       0       0
  B-ORG      0       0       0      12       0       1       0       0       0       0       0       0       0       0      11       0       0       0       0     529       0       0       0       0
M-TITLE      0       0       1      19       0      89       1       0       0       2       1       0       3       0      54       0       0      17       0       1    1734       0       0       0
  E-LOC      0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       1       0       0       0       5       0       0
 M-CONT      0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0      53       0
 B-RACE      0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0      14
```

预测结果

1. HMM模型：

```
句子: ['曾', '任', '大', '连', '市', '电', '力', '发', '展', '公', '司', '副', '总', '经', '理', '。']
分词结果: 曾 任 大 连 市 电 力 发 展 公 司 副 总 经 理 。
属性: {'ORG': ['大连市电力发展公司'], 'TITLE': ['副总经理']}
```

2. CRF模型：

```
句子: ['1', '9', '6', '3', '年', '出', '生', ',', '工', '科', '学', '士', ',', '高', '级', '工', '程',
'师', ',', '北', '京', '物', '资', '学', '院', '客', '座', '副', '教', '授', '。']
分词结果: 1 9 6 3 年 出 生 , 工 科 学 士 , 高 级 工 程 师 , 北 京 物 资 学 院 客 座 副 教 授 。
属性: {'EDU': ['工科学士'], 'TITLE': ['高级工程师', '客座副教授'], 'ORG': ['北京物资学院']}
```

###机器学习方法：

标记精确度和混淆矩阵：

1. Bilstm模型：

```
加载并评估bilstm模型...
              precision    recall  f1-score   support

       B-PRO     0.9062    0.8788    0.8923        33
      M-NAME     0.8889    0.8780    0.8834        82
       B-EDU     0.9908    0.9643    0.9774       112
     B-TITLE     0.9503    0.9171    0.9334       772
      B-NAME     0.9060    0.9464    0.9258       112
       M-ORG     0.9743    0.9565    0.9653      4325
       E-PRO     0.8611    0.9394    0.8986        33
      E-RACE     1.0000    1.0000    1.0000        14
       E-EDU     0.9640    0.9554    0.9596       112
       M-PRO     0.7500    0.9706    0.8462        68
       M-EDU     0.9540    0.9274    0.9405       179
       M-LOC     0.6250    0.7143    0.6667        21
     E-TITLE     0.9908    0.9793    0.9850       772
      E-CONT     1.0000    1.0000    1.0000        28
           O     0.9570    0.9902    0.9733      5190
      B-CONT     1.0000    1.0000    1.0000        28
      E-NAME     0.9649    0.9821    0.9735       112
       E-ORG     0.9396    0.9005    0.9197       553
       B-LOC     1.0000    0.8333    0.9091         6
       B-ORG     0.9777    0.9512    0.9643       553
     M-TITLE     0.9280    0.9053    0.9165      1922
       E-LOC     1.0000    0.8333    0.9091         6
      M-CONT     1.0000    1.0000    1.0000        53
      B-RACE     1.0000    0.9286    0.9630        14
avg/total        0.9581    0.9576    0.9576     15100

Confusion Matrix:
          B-PRO  M-NAME   B-EDU B-TITLE  B-NAME   M-ORG   E-PRO  E-RACE   E-EDU   M-PRO   M-EDU   M-LOC E-TITLE  E-CONT       O  B-CONT  E-NAME   E-ORG   B-LOC   B-ORG M-TITLE   E-LOC  M-CONT  B-RACE
  B-PRO      29       0       0       0       0       2       0       0       0       2       0       0       0       0       0       0       0       0       0       0       0       0       0       0
 M-NAME       0      72       0       0       6       0       0       0       0       0       0       0       0       0       2       0       2       0       0       0       0       0       0       0
  B-EDU       0       0     108       0       0       2       1       0       0       0       1       0       0       0       0       0       0       0       0       0       0       0       0       0
B-TITLE       1       0       1     708       0       5       0       0       0       0       0       0       0       0      14       0       0       3       0       7      33       0       0       0
 B-NAME       0       2       0       0     106       0       0       0       0       0       0       0       0       0       4       0       0       0       0       0       0       0       0       0
  M-ORG       1       0       0      12       0    4137       2       0       0      10       1       0       1       0      78       0       0      20       0       3      60       0       0       0
  E-PRO       0       0       0       0       0       0      31       0       0       0       1       1       0       0       0       0       0       0       0       0       0       0       0       0
 E-RACE       0       0       0       0       0       0       0      14       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0
  E-EDU       0       0       0       0       0       0       1       0     107       0       1       0       1       0       2       0       0       0       0       0       0       0       0       0
  M-PRO       0       0       0       0       0       1       0       0       0      66       0       0       0       0       0       0       0       1       0       0       0       0       0       0
  M-EDU       1       0       0       0       0       2       0       0       2       4     166       0       0       0       2       0       0       1       0       0       1       0       0       0
  M-LOC       0       0       0       0       0       0       0       0       0       0       0      15       0       0       0       0       0       1       0       0       5       0       0       0
E-TITLE       0       0       0       0       0       1       0       0       1       0       0       0     756       0       9       0       0       2       0       0       3       0       0       0
 E-CONT       0       0       0       0       0       0       0       0       0       0       0       0       0      28       0       0       0       0       0       0       0       0       0       0
      O       0       0       0       5       0      27       0       0       0       0       2       1       0       4       0    5139       0       0       1       0       0      11       0       0       0
 B-CONT       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0      28       0       0       0       0       0       0       0
 E-NAME       0       0       0       0       0       0       0       0       0       0       0       0       0       0       1       0       0     110       0       0       0       0       0       0
  E-ORG       0       0       0       0       0      19       0       0       0       2       0       0       0       0      12       0       0       0     498       0       0      22       0       0       0
  B-LOC       0       0       0       0       0       0       0       0       0       0       0       0       0       0       1       0       0       0       0       5       0       0       0       0       0
  B-ORG       0       0       0       7       0       5       0       0       0       0       0       0       0       0      15       0       0       0       0       0     526       0       0       0       0
M-TITLE       0       7       0      13       5      45       1       0       1       2       3       8       1       0      89       0       0       5       0       0       2    1740       0       0       0
  E-LOC       0       0       0       0       0       0       0       0       0       0       0       0       0       0       1       0       0       0       0       0       0       0       5       0       0
 M-CONT       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0      53       0
 B-RACE       0       0       0       0       0       0       0       0       0       0       0       0       0       0       1       0       0       0       0       0       0       0       0       0      13
```

2.Bilstm+CRF模型：

```
加载并评估bilstm_crf模型...
              precision    recall  f1-score   support

       B-PRO     0.9091    0.9091    0.9091        33
      M-NAME     0.9146    0.9146    0.9146        82
       B-EDU     0.9907    0.9554    0.9727       112
     B-TITLE     0.9325    0.9301    0.9313       772
      B-NAME     0.9898    0.8661    0.9238       112
       M-ORG     0.9535    0.9635    0.9585      4325
       E-PRO     0.9091    0.9091    0.9091        33
      E-RACE     1.0000    1.0000    1.0000        14
       E-EDU     0.9646    0.9732    0.9689       112
       M-PRO     0.8421    0.9412    0.8889        68
       M-EDU     0.9540    0.9274    0.9405       179
       M-LOC     0.8500    0.8095    0.8293        21
     E-TITLE     0.9921    0.9767    0.9843       772
      E-CONT     1.0000    1.0000    1.0000        28
           O     0.9530    0.9846    0.9685      5190
      B-CONT     1.0000    1.0000    1.0000        28
      E-NAME     0.9810    0.9196    0.9493       112
       E-ORG     0.9271    0.8969    0.9118       553
       B-LOC     0.8000    0.6667    0.7273         6
       B-ORG     0.9688    0.9548    0.9617       553
     M-TITLE     0.9514    0.8762    0.9122      1922
       E-LOC     1.0000    0.8333    0.9091         6
      M-CONT     1.0000    1.0000    1.0000        53
      B-RACE     1.0000    0.9286    0.9630        14
avg/total        0.9537    0.9536    0.9533     15100

Confusion Matrix:
          B-PRO  M-NAME   B-EDU B-TITLE  B-NAME   M-ORG   E-PRO  E-RACE   E-EDU   M-PRO   M-EDU   M-LOC E-TITLE  E-CONT       O  B-CONT  E-NAME   E-ORG   B-LOC   B-ORG M-TITLE   E-LOC  M-CONT  B-RACE
  B-PRO      30       0       0       0       0       1       0       0       0       1       0       0       0       0       1       0       0       0       0       0       0       0       0       0
 M-NAME       0      75       0       0       1       0       0       0       0       0       0       0       0       0       6       0       0       0       0       0       0       0       0       0
  B-EDU       0       0     107       1       0       2       0       0       0       0       2       0       0       0       0       0       0       0       0       0       0       0       0       0
B-TITLE       1       0       0     718       0      11       0       0       0       0       0       0       0       0      10       0       0       3       0       7      22       0       0       0
 B-NAME       0       5       0       0      97       0       0       0       0       0       0       0       0       0      10       0       0       0       0       0       0       0       0       0
  M-ORG       0       0       0       9       0    4167       1       0       0       4       0       0       1       0      92       0       0      11       0       3      37       0       0       0
  E-PRO       0       0       0       0       0       0      30       0       0       0       2       0       0       0       1       0       0       0       0       0       0       0       0       0
 E-RACE       0       0       0       0       0       0       0      14       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0
  E-EDU       0       0       0       0       0       1       1       0     109       0       1       0       0       0       0       0       0       0       0       0       0       0       0       0
  M-PRO       0       0       0       0       0       2       0       0       0      64       0       0       0       0       1       0       0       1       0       0       0       0       0       0
  M-EDU       1       0       0       0       0       3       0       0       2       4     166       0       1       0       1       0       0       1       0       0       0       0       0       0
  M-LOC       0       0       0       0       0       4       0       0       0       0       0      17       0       0       0       0       0       0       0       0       0       0       0       0
E-TITLE       0       0       0       0       0       0       0       0       1       0       0       0     754       0      12       0       2       1       0       0       2       0       0       0
 E-CONT       0       0       0       0       0       0       0       0       0       0       0       0       0      28       0       0       0       0       0       0       0       0       0       0
      O       1       0       1       6       0      48       0       0       0       1       0       1       0       3       0    5110       0       0       4       0       6       9       0       0       0
 B-CONT       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0      28       0       0       0       0       0       0       0       0
 E-NAME       0       0       0       0       0       0       0       0       0       0       0       0       0       0       9       0       0     103       0       0       0       0       0       0       0
  E-ORG       0       0       0       2       0      25       0       0       0       1       0       0       0       0      13       0       0       0     496       0       0      16       0       0       0
  B-LOC       0       0       0       1       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       4       1       0       0       0       0
  B-ORG       0       0       0       9       0       5       0       0       0       0       0       0       0       0      11       0       0       0       0       0     528       0       0       0       0
M-TITLE       0       2       0      24       0     100       1       0       0       2       2       3       1       0      85       0       0       0      18       0       0    1684       0       0       0
  E-LOC       0       0       0       0       0       1       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       5       0       0
 M-CONT       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0      53       0
 B-RACE       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       0       1       0       0       0       0      13
```

预测结果

1. Bilstm模型：

```
句子: ['曾', '任', '大', '连', '市', '电', '力', '发', '展', '公', '司', '副', '总', '经', '理', '。']
分词结果: 曾 任 大 连 市 电 力 发 展 公 司 副 总 经 理 。
属性: {'ORG': ['大连市电力发展公司'], 'TITLE': ['副总经理']}

句子: ['现', '任', '大', '连', '市', '电', '力', '发', '展', '公', '司', '总', '经', '理', '。']
分词结果: 现 任 大 连 市 电 力 发 展 公 司 总 经 理 。
属性: {'ORG': ['大连市电力发展公司'], 'TITLE': ['总经理']}
```

2.Bilstm-CRF模型：

```
句子: ['2', '0', '1', '0', '年', '3', '月', '至', '今', ', ', '任', '日', '月', '集', '团', '副', '董',
'事', '长', '; ', '<end>']
分词结果: 2 0 1 0 年 3 月 至 今 , 任 日 月 集 团 副 董 事 长 ; <end>
属性: {'ORG': ['日月集团'], 'TITLE': ['副董事长']}

句子: ['1', '9', '9', '9', '年', '至', '2', '0', '1', '0', '年', '5', '月', ', ', '任', '浙', '江', '明',
'牌', '实', '业', '股', '份', '有', '限', '公', '司', '董', '事', '长', '兼', '总', '经', '理', '; ',
'<end>']
分词结果: 1 9 9 9 年 至 2 0 1 0 年 5 月 , 任 浙 江 明 牌 实 业 股 份 有 限 公 司 董 事 长 兼 总 经 理 ; <end>
属性: {'ORG': ['浙江明牌实业股份有限公司'], 'TITLE': ['董事长', '总经理']}
```

model scope模型： RaNER

采用Transformer-CRF模型，使用StructBERT作为预训练模型底座，结合使用外部工具召回的相关句子作为额外上下文，使用Multi-view Training方式进行训练。

模型介绍：

Transformer-CRF模型是一种结合了Transformer和条件随机场（CRF）的模型，用于命名实体识别任务。Transformer是一种基于自注意力机制的神经网络结构，能够有效地捕捉句子中的上下文信息。CRF则用于建模标签之间的依赖关系，以提高标签序列的一致性。

StructBERT是一种基于BERT（Bidirectional Encoder Representations from Transformers）的模型，通过引入结构化信息和预训练任务来提升预训练模型的性能。它在BERT的基础上增加了一个结构化嵌入层，用于编码句子中的语法和结构信息。

使用外部工具召回相关句子作为额外上下文可以增加句子的语境信息，有助于提升模型的性能。这些召回的句子可能是通过搜索引擎或其他方法获取的与原始句子相关的文本。将这些召回的句子与原始句子进行拼接构建多视图输入，可以为模型提供更全面的上下文信息。

Multi-view Training方式是一种训练策略，它通过鼓励不同视图的相似性来提高模型的性能。在这种方法中，通过共享参数和共同的损失函数，同时训练基于原始句子的输入视图和基于召回的输入视图，使它们产生相似的上下文表示或输出标签分布。

这种方法结合了Transformer-CRF模型、StructBERT预训练模型、外部工具召回的相关句子和Multi-view Training方式，旨在通过引入额外的上下文信息和多视图训练来提高命名实体识别模型的性能。

预测分类结果：

	A	B	C	D	E
1	nnamed:	tem_name	brand_name	model_name	category_name
2	0	小雨越库	['小雨']	['双肩', '流浪', '手提', '单肩', '流浪包', '链条', '斜挎']	['包', '背包', '背囊']
3	1	代购版本	['香奶奶']	[]	['黄金球', '包']
4	2	顶级老工	[]	['斜挎', '手提']	['包']
5	3	正品保障	['雅诗兰黛']	[]	['粉底液']
6	4	售完无补	['兰蔻']	['清爽版']	['防晒乳', '防晒霜']
7	5	正品级小	['香奶奶家', '小香']	['单肩', '斜挎', '翻盖', '链条']	['包包', '包', '香']
8	6	亚洲限定	[]	['椰子']	['鞋']
9	7	正品保障	['银色山泉', '克雷德']	[]	['香水']
10	8	美不胜收	[]	['单肩', '字母', '斜挎', '链条']	['袋', '包', '链条']