

## (12) 发明专利申请

(10) 申请公布号 CN 103365838 A

(43) 申请公布日 2013. 10. 23

(21) 申请号 201310311955. 4

(22) 申请日 2013. 07. 24

(71) 申请人 桂林电子科技大学

地址 541004 广西壮族自治区桂林市七星区  
金鸡路 1 号

(72) 发明人 黄桂敏 周娅 王晓娟

(74) 专利代理机构 桂林市华杰专利商标事务所  
有限责任公司 45112

代理人 罗玉荣

(51) Int. Cl.

G06F 17/28 (2006. 01)

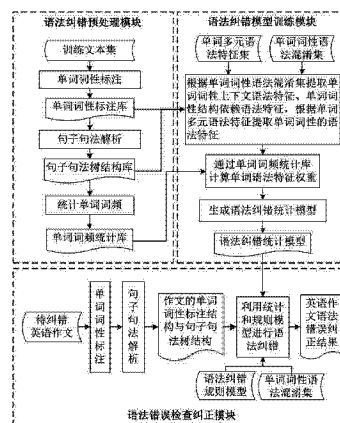
权利要求书3页 说明书29页 附图4页

## (54) 发明名称

基于多元特征的英语作文语法错误自动纠正方法

## (57) 摘要

基于多元特征的英语作文语法错误自动纠正方法,包括语法纠错预处理模块、语法纠错模型训练模块和语法错误检查纠正模块;其中语法纠错预处理模块对输入的训练文本集进行单词词性标注、句子句法解析和单词词频统计;语法纠错模型训练模块对输入单词词性标注库、句子句法树结构库、单词词频统计库、单词词性语法混淆集,提取单词及其词性上下文语法特征、单词及其词性结构依赖语法特征、单词及其词性的语法特征,计算单词语法特征权重,输出语法纠错统计模型;语法错误检查纠正模块利用语法纠错统计模型和语法纠错规则模型对待纠错作文进行语法纠错,输出英语作文中语法错误的纠正结果。该方法实现了对英语作文中常见 11 种英语语法错误自动纠正。



1. 基于多元特征的英语作文语法错误自动纠正方法,其特征是:其中包括语法纠错预处理模块、语法纠错模型训练模块和语法错误检查纠正模块;

语法纠错预处理模块处理包括:输入训练文本集,对训练文本集进行单词词性标注、句子句法解析和单词词频统计,输出单词词性标注库、句子句法树结构库、单词词频统计库;

语法纠错模型训练模块处理包括:输入单词词性标注库、句子句法树结构库、单词词频统计库、单词多元语法特征集、单词词性语法混淆集,根据单词词性语法混淆集提取单词及其词性上下文语法特征、单词及其词性结构依赖语法特征,根据单词多元语法特征提取单词及其词性的语法特征,利用单词词频统计库计算单词语法特征权重,生成语法纠错统计模型,最后输出语法纠错统计模型;

语法错误检查纠正模块处理包括:输入语法纠错统计模型、语法纠错规则模型、单词词性语法混淆集,采用语法纠错统计模型纠正英语作文中冠词错误、介词错误、词性混淆错误、动词形态错误、助动词错误、主谓不一致错误,利用语法纠错规则模型纠正英语作文中单复数不一致错误、片段错误、形容词比较级最高级错误、动词与补语搭配错误、单词重复使用错误,最后输出英语作文中语法错误的纠正结果。

2. 根据权利要求1所述的方法,其特征是:所述的语法纠错预处理模块具体处理流程如下:

S0201 开始;

S0202 读入训练文本集;

S0203 使用宾州树库标注集标注训练文本集中句子,输出单词词性标注库;

S0204 使用上下文无关文法单词词性句法解析方法训练文本集中句子,输出句子句法树结构库;

S0205 遍历单词词性标注库;

S0206 如果单词词性标注库遍历结束,则转 S0210 操作;否则转 S0207 操作;

S0207 如果当前单词在单词词频链表中,则转 S0208 操作;否则转 S0209 操作;

S0208 当前单词及其词性在单词词频链表中词频加1,转 S0205 操作;

S0209 当前单词及其词性加入单词词频链表,当前单词及其词性在单词词频链表中词频加1,转 S0205 操作;

S0210 遍历句子句法树结构库;

S0211 如果句法树结构库遍历结束,则转 S0215 操作;否则转 S0212 操作;

S0212 如果当前短语在单词词频链表中,则转 S0213 操作;否则转 S0214 操作;

S0213 当前短语及其词性在单词词频链表中词频加1,转 S0210 操作;

S0214 当前短语及其词性加入单词词频链表,当前短语及其词性在单词词频链表中词频加1,转 S0210 操作;

S0215 利用单词词频链表构建单词词频统计库,输出单词词频统计库;

S0216 结束。

3. 根据权利要求1所述的方法,其特征是:所述的语法纠错模型训练模块具体处理流程如下:

S0301 开始;

S0302 读入单词词性标注库、句子句法树结构库和单词词性语法混淆集;

S0303 将句子分割成单词词性单元,生成单词词性单元列表;

S0304 遍历单词词性单元列表;

S0305 如果单词词性单元列表遍历结束,则转 S0310 操作;否则转 S0306 操作;

S0306 如果当前单词在单词词性语法混淆集中,则转 S0307 操作;否则转 S0305 操作;

S0307 从单词多元语法特征集中,提取当前单词及其词性的语法特征;

S0308 计算当前单词语法特征权重;

S0309 在语法纠错链表中保存当前单词及其词性语法特征、当前单词语法特征权重,转 S0305 操作;

S0310 利用语法纠错链表构建语法纠错统计模型,并输出语法纠错统计模型;

S0311 结束。

4. 根据权利要求 1 所述的方法,其特征是:所述的语法错误检查纠正模块具体处理流程如下:

S0401 开始;

S0402 读入一篇待纠错英语作文;

S0403 遍历作文中英语句子;

S0404 如果待纠错英语作文最后句子纠错结束,则转 S0409 操作;否则转 S0405 操作;

S0405 对待纠错英语作文中当前句子进行单词词性标注,生成该单词词性标注结构;

S0406 对待纠错英语作文中当前句子进行句子句法结构解析,生成该句子句法树结构;

S0407 利用语法纠错规则模型中定义的语法错误规则对当前句子进行检查,如果当前句子符合语法纠错规则模型的规则设置,则表明当前句子有语法错误,利用语法纠错规则模型中定义的对应纠错方法进行语法纠错;

S0408 读取当前句子中单词,如果该单词在单词词性语法混淆集中,则将该单词作为当前单词,从单词多元语法特征集中提取当前单词的语法特征,从语法纠错统计模型中获得当前单词语法特征的权重,计算当前单词的分值,以及当前单词在单词词性语法混淆集中所在混淆单元的其他单词的分值,如果当前单词的分值不是最大,则表明当前单词有语法错误,使用最大分值的单词来纠正当前单词的语法错误,转 S0404 操作;

S0409 输出待纠错英语作文的语法纠错结果;

S0410 结束。

5. 根据权利要求 1 所述的方法,其特征是:所述的单词词性语法混淆集包括易混介词、易混词性、易混冠词和易混助动词。

6. 根据权利要求 1 所述的方法,其特征是:所述的单词多元语法特征集包括介词语法特征、动词语法特征、冠词语法特征、助动词语法特征。

7. 根据权利要求 1 所述的方法,其特征是:所述的单词语法特征权重的计算公式如下:

单词语法特征权重 = 单词语法特征相关度 × 单词语法特征贡献度

其中:单词语法特征相关度的计算公式如下:

单词语法特征相关度 =

$$\frac{\text{单词及其语法特征在训练英语文本集共现次数} \times \text{训练英语文本集的单词总数}}{\text{单词在训练英语文本集出现次数} \times \text{单词语法特征在训练英语文本集出现次数}}$$

单词语法特征贡献度的计算公式如下：

$$\text{单词语法特征贡献度} = \frac{\text{单词或词性及其语法特征在训练英语文本集共现次数}}{\sum_{(1 \leq i \leq n)} \text{单词}_i \text{或词性}_i \text{及其语法特征在训练英语文本集共现次数}}$$

i 是当前单词在单词词性语法混淆集中所属的易混类型中的单词或词性的序号，n 是当前单词在单词词性语法混淆集中所属的易混类型中的单词或词性总数。

8. 根据权利要求 4 所述的方法，其特征是：所述的当前单词分值计算公式如下：

$$\text{单词分值} = \sum_{(1 \leq j \leq m)} (\text{系数}_j \times \text{单词语法特征权重}_j)$$

j 是当前单词在单词多元语法特征集中所属的特征类型中的语法特征的序号；如果单词语法特征<sub>j</sub>在当前单词的多元语法特征集中，则系数<sub>j</sub>等于 1；如果单词语法特征<sub>j</sub>不在当前单词的多元语法特征集中，则系数<sub>j</sub>等于 0；

m 是当前单词在单词多元语法特征集中所属的特征类型中的语法特征总数。

## 基于多元特征的英语作文语法错误自动纠正方法

### 技术领域

[0001] 本发明涉及自然语言处理技术与统计学,具体是基于多元特征的英语作文语法错误自动纠正方法。

### 背景技术

[0002] 目前,英语作文语法错误的自动纠正方法主要分为两类,即:基于规则的英语作文语法错误自动纠正方法和基于统计的英语作文语法错误自动纠正方法。其中,基于规则的英语作文语法错误自动纠正方法是:通过分析英语句子的语法特征,总结归纳出固定的英语语法规则来构建英语句子语法规则纠错模型,当对英语作文的语法错误进行纠正时,构建英语句子语法规则纠错模型对英语作文中句子的语法进行检查,判断它们是否符合模型中存在的语法规则,来实现对英语作文中语法错误的自动纠正。基于统计的英语作文语法错误自动纠正方法是:通过训练文本集来构建英语句子的语法纠错统计模型,使用英语句子的语法纠错统计模型对英语作文中语法错误进行纠正处理。然而,这两种英语作文语法错误的自动纠正方法都存在如下优点与缺点。

[0003] 1、基于规则的英语作文语法错误自动纠正方法

[0004] 优点:纠错准确率高、执行效率高、使用简单,并且对于一些语法固定搭配、某些单词特殊用法、固定的语法规则、语法时态数一致性具有较好的纠错效果。

[0005] 缺点:由于英语使用的灵活性和多变性,构建一种英语语法规则模型要花费较长时间,而且总结归纳的英语语法规则只能覆盖局限语法规则,英语语法错误纠正类型有限。

[0006] 2、基于统计的英语作文语法错误自动纠正方法

[0007] 优点:对于难以用语法规则描述的英语语法错误,这种方法具有较好的纠正效果。

[0008] 缺点:英语语法纠错正确率受到训练文本集规模、语法特征提取广度和精度、训练构建的语法统计纠错模型好坏的影响,所以这种方法的纠错正确率有待改进。

[0009] 总之,英语作文语法错误的自动纠正技术具有非常广阔的应用前景,例如:作者利用它自动纠正其英语作文中语法错误,以提高自身的英语写作水平;英语机器翻译系统使用它纠正输出的翻译结果,以提高英语机器翻译系统的译文准确度。

### 发明内容

[0010] 本发明的目的是提供基于多元特征的英语作文语法错误自动纠正方法。

[0011] 基于多元特征的英语作文语法错误自动纠正方法,其中包括语法纠错预处理模块、语法纠错模型训练模块和语法错误检查纠正模块;

[0012] 语法纠错预处理模块处理包括:输入训练文本集,对训练文本集进行单词词性标注、句子句法解析和单词词频统计,输出单词词性标注库、句子句法树结构库、单词词频统计库;

[0013] 语法纠错模型训练模块处理包括:输入单词词性标注库、句子句法树结构库、单词词频统计库、单词词性语法混淆集,根据单词词性语法混淆集提取单词词性上下文语法特

征、单词词性结构依赖语法特征,根据单词多元语法特征提取单词及其词性的语法特征,利用单词词频统计库计算单词语法特征权重,生成语法纠错统计模型,输出语法纠错统计模型;

[0014] 语法错误检查纠正模块处理包括:输入语法纠错统计模型、语法纠错规则模型、单词词性语法混淆集,采用语法纠错统计模型纠正英语作文中冠词错误、介词错误、词性混淆错误、动词形态错误、助动词错误、主谓不一致错误,利用语法纠错规则模型纠正英语作文中单复数不一致错误、片段错误、形容词比较级最高级错误、动词与补语搭配错误、单词重复使用错误,输出英语作文中语法错误的纠正结果。

[0015] 所述的语法纠错预处理模块具体处理流程如下:

[0016] S0201 开始;

[0017] S0202 读入训练文本集;

[0018] S0203 使用宾州树库标注集标注训练文本集中句子,输出单词词性标注库;

[0019] S0204 使用上下文无关文法单词词性句法解析方法训练文本集中句子,输出句子句法树结构库;

[0020] S0205 遍历单词词性标注库;

[0021] S0206 如果单词词性标注库遍历结束,则转 S0210 操作;否则转 S0207 操作;

[0022] S0207 如果当前单词在单词词频链表中,则转 S0208 操作;否则转 S0209 操作;

[0023] S0208 当前单词及其词性在单词词频链表中词频加 1,转 S0205 操作;

[0024] S0209 当前单词及其词性加入单词词频链表,当前单词及其词性在单词词频链表中词频加 1,转 S0205 操作;

[0025] S0210 遍历句子句法树结构库;

[0026] S0211 如果句法树结构库遍历结束,则转 S0215 操作;否则转 S0212 操作;

[0027] S0212 如果当前短语在单词词频链表中,则转 S0213 操作;否则转 S0214 操作;

[0028] S0213 当前短语及其词性在单词词频链表中词频加 1,转 S0210 操作;

[0029] S0214 当前短语及其词性加入单词词频链表,当前短语及其词性在单词词频链表中词频加 1,转 S0210 操作;

[0030] S0215 利用单词词频链表构建单词词频统计库,输出单词词频统计库;

[0031] S0216 结束。

[0032] 所述的语法纠错模型训练模块具体处理流程如下

[0033] S0301 开始;

[0034] S0302 读入单词词性标注库、句子句法树结构库和单词词性语法混淆集;

[0035] S0303 将句子分割成单词词性单元,生成单词词性单元列表;

[0036] S0304 遍历单词词性单元列表;

[0037] S0305 如果单词词性单元列表遍历结束,则转 S0310 操作;否则转 S0306 操作;

[0038] S0306 如果当前单词在单词词性语法混淆集中,则转 S0307 操作;否则转 S0305 操作;

[0039] S0307 从单词多元语法特征集中,提取当前单词及其词性的语法特征;

[0040] S0308 计算当前单词语法特征权重;

[0041] S0309 在语法纠错链表中保存当前单词及其词性语法特征、当前单词语法特征权

重,转 S0305 操作;

[0042] S0310 利用语法纠错链表构建语法纠错统计模型,并输出语法纠错统计模型;

[0043] S0311 结束。

[0044] 所述的语法错误检查纠正模块具体处理流程如下:

[0045] S0401 开始;

[0046] S0402 读入一篇待纠错英语作文;

[0047] S0403 遍历作文中英语句子;

[0048] S0404 如果待纠错英语作文最后句子纠错结束,则转 S0409 操作;否则转 S0405 操作;

[0049] S0405 对待纠错英语作文中当前句子进行单词词性标注,生成该单词词性标注结构;

[0050] S0406 对待纠错英语作文中当前句子进行句子句法结构解析,生成该句子句法树结构;

[0051] S0407 利用语法纠错规则模型中定义的语法错误规则对当前句子进行检查,如果当前句子符合语法纠错规则模型的规则设置,则表明当前句子有语法错误,利用语法纠错规则模型中定义的对应纠错方法进行语法纠错;

[0052] S0408 读取当前句子中单词,如果该单词在单词词性语法混淆集中,则将该单词作为当前单词,从单词多元语法特征集中提取当前单词的语法特征,从语法纠错统计模型中获得当前单词语法特征的权重,计算当前单词的分值,以及当前单词在单词词性语法混淆集中所在混淆单元的其他单词的分值,如果当前单词的分值不是最大,则表明当前单词有语法错误,使用最大分值的单词来纠正当前单词的语法错误,转 S0404 操作;

[0053] S0409 输出待纠错英语作文的语法纠错结果;

[0054] S0410 结束。

[0055] 本发明中,所述的单词词性语法混淆集包括易混介词、易混词性、易混冠词和易混助动词。

[0056] 所述的单词多元语法特征集包括介词语法特征、动词语法特征、冠词语法特征、助动词语法特征。

[0057] 所述的单词语法特征权重的计算公式如下:

[0058] 单词语法特征权重 = 单词语法特征相关度 × 单词语法特征贡献度

[0059] 其中:单词语法特征相关度的计算按下式:

[0060]

单词语法特征相关度 =

[0061]

$$\frac{\text{单词及其语法特征在训练英语文本集共现次数} \times \text{训练英语文本集的单词总数}}{\text{单词在训练英语文本集出现次数} \times \text{单词语法特征在训练英语文本集出现次数}}$$

[0062] 单词语法特征贡献度的计算公式如下:

[0063]

$$\text{单词语法特征贡献度} = \frac{\text{单词或词性及其语法特征在训练英语文本集共现次数}}{\sum (1 \leq i \leq n) \text{ 单词}_i \text{或词性}_i \text{及其语法特征在训练英语文本集共现次数}}$$

[0064]  $i$  是当前单词在单词词性语法混淆集中所属的易混类型中的单词或词性的序号,  $n$  是当前单词在单词词性语法混淆集中所属的易混类型中的单词或词性总数。

[0065] 如果单词语法特征相关度大于等于零, 则表示该单词语法特征与该单词之间存在符合语法规则的语法依赖关系; 如果单词语法特征相关度小于零, 则表示单词语法特征与该单词之间不存在语法依赖关系。

[0066] 所述的当前单词分值计算公式如下:

[0067] 单词分值 =  $\sum (1 \leq j \leq m)$  (系数 $_j \times$  单词语法特征权重 $_j$ )

[0068]  $j$  是当前单词在单词多元语法特征集中所属的特征类型中的语法特征的序号。如果单词语法特征 $_j$ 在当前单词的多元语法特征集中, 则系数 $_j$ 等于 1; 如果单词语法特征 $_j$ 不在当前单词的多元语法特征集中, 则系数 $_j$ 等于 0。

[0069]  $m$  是当前单词在单词多元语法特征集中所属的特征类型中的语法特征总数。

[0070] 本发明的方法能够实现对英语作文中冠词错误、介词错误、词性混淆错误、动词形态错误、助动词错误、主谓不一致错误、单复数不一致错误、片段错误、形容词比较级最高级错误、动词与补语搭配错误、单词重复使用错误共 11 种英语作文错误的自动纠正。

[0071] 本发明中的定义如下:

[0072] I. 训练文本集

[0073] 训练文本集取材于多种英文文章, 它们是不包含任何单词错误、语法错误、表达错误的正确英语短文。

[0074] II. 单词词性标注集

[0075] 单词词性标注集采用美国宾州大学宾州树库词性标注集, 它用于对训练文本集中单词进行词性标注。

[0076] III. 单词词性标注库

[0077] 单词词性标注库是对训练文本集进行单词词性标注处理后输出的结果, 它的格式如下:

[0078] 训练文本集第 1 个句子: 单词 $_1$ / 词性 $_1$  单词 $_2$ / 词性 $_2$ ..... 单词 $_i$ / 词性 $_i$  < 回车 >

[0079] 训练文本集第 2 个句子: 单词 $_1$ / 词性 $_1$  单词 $_2$ / 词性 $_2$ ..... 单词 $_j$ / 词性 $_j$  < 回车 >

[0080] .....

[0081] 训练文本集第  $n$  个句子: 单词 $_1$ / 词性 $_1$  单词 $_2$ / 词性 $_2$ ..... 单词 $_k$ / 词性 $_k$  < 回车 >

[0082] IV. 句子句法树结构库

[0083] 句子句法树结构库是对训练文本集句子句法解析的结果, 它由短语单词结构和单词语法依赖结构组成, 短语单词结构表示句子中短语词性与单词词性之间的关系, 短语词性标注采用美国宾州大学宾州树库标注集, 它的格式如下:

[0084]

(句子

(短语词性<sub>1</sub>

(词性<sub>1</sub> 单词<sub>1</sub>)

.....

(词性<sub>i</sub> 单词<sub>i</sub>)

)

.....

(短语词性<sub>k</sub>

(词性<sub>1</sub> 单词<sub>1</sub>)

.....

(词性<sub>j</sub> 单词<sub>j</sub>)

)

)

[0085] 单词语法依赖结构表示句子中两个单词之间的单词语法依赖关系,它的格式如下:

[0086] 单词语法依赖关系 1 (单词<sub>x</sub>-单词标号<sub>x</sub>,单词<sub>y</sub>-单词标号<sub>y</sub>)

[0087] .....

[0088] 单词语法依赖关系 n (单词<sub>x</sub>-单词标号<sub>x</sub>,单词<sub>y</sub>-单词标号<sub>y</sub>)

[0089] V. 单词词频统计库

[0090] 单词词频统计库是对训练文本集的单词、单词对、词性、词性对、短语出现的次数的统计结果,简称词频。它的格式如下:

[0091] 单词<sub>1</sub>^词频<回车>

[0092] .....

[0093] 单词<sub>i</sub>^词频<回车>

[0094] 词性<sub>1</sub>^词频<回车>

[0095] .....

[0096] 词性<sub>j</sub>^词频<回车>

[0097] 单词<sub>1</sub>^单词<sub>x</sub>^词频<回车>

[0098] .....

[0099] 单词<sub>k</sub>^单词<sub>x</sub>^词频<回车>

[0100] 短语<sub>1</sub>^词频<回车>

[0101] .....

[0102] 短语<sub>n</sub>^词频<回车>

[0103] VI. 单词词性语法混淆集

[0104] 单词词性语法混淆集包括易混介词、易混词性、易混冠词和易混助动词,它们的具体内容如下表 1 所示:

[0105] 表 1 :单词词性语法混淆集

[0106]

易混类型	易混单词或词性及其序号
易混介词	{1-of, 2-to, 3-in, 4-for, 5-on, 6-with, 7-at, 8-by, 9-from, 10-about}
易混冠词	{1-a, 2-an, 3-the, 4-零冠词}
易混词性	{1-NN, 2-JJ}
	{1-JJ, 2-NN, 3-RB}
	{1-RB, 2-JJ}
	{1-VB, 2-VBD, 3-VBG, 4-VBN, 5-VBZ}
易混助动词	{1-does, 2-do}
	{1-is, 2-am, 3-are}
	{1-was, 2-were}
	{1-have, 2-has}

[0107] VII. 单词多元语法特征集

[0108] 单词多元语法特征集包括介词语法特征、动词语法特征、冠词语法特征、助动词语法特征,它们的具体内容如下表 2 所示 :

[0109] 表 2 :单词多元语法特征集

[0110]

特征类型	特征子类型	特征序号	特征说明
介词语法特征	多元语法特征	1	介词前一个单词
		2	介词后一个单词
		3	介词前一个单词词性
		4	介词后一个单词词性
		5	介词前两个单词
		6	介词后两个单词
		7	介词前两个单词词性
		8	介词后两个单词词性
	短语特征	9	介词后短语
		10	介词后短语中心词
		11	介词前短语类型
		12	动词与介词中间名词或形容词短语
		13	介词前名词短语中的形容词修饰
		14	介词前动词短语
	句法结构特征	15	介词前介词
		16	动词与介词之间名词或形容词
		17	介词附着词
		18	介词前动词词性
		19	介词前动词词元
		20	介词前单词词元
词性语法特征	多元语法特征	1	易混词性前一个单词词性
		2	易混词性后一个单词词性
		3	易混词性前一个单词
		4	易混词性后一个单词
		5	易混词性前两个单词词性
		6	易混词性后两个单词词性
		7	易混词性前两个单词
		8	易混词性后两个单词

[0111]

	句法结构特征	9	动词前情态助动词
		10	动词前助动词
		11	动词前不定式
		12	动词修饰主语
		13	动词作为补语修饰动词或介词
冠词语法特征	多元语法特征	1	名词短语中心词前单词
		2	名词短语中心词后单词
		3	名词短语中心词前单词词性
		4	名词短语中心词后单词词性
		5	名词短语前两个单词
		6	名词短语后两个单词
		7	名词短语前两个单词词性
		8	名词短语后两个单词词性
	句法结构特征	9	名词短语中心词
		10	名词短语中心词词性
		11	中心词可数性
		12	名词短语在句法树中父节点
		13	名词短语中非冠词限定词
助动词语法特征	多元语法特征	1	语法词前一元单词
		2	语法词前一元单词词性
		3	语法词后一元单词
		4	语法词后一元单词词性
		5	语法词前二元单词
		6	语法词前二元单词词性
		7	语法词后二元单词
		8	语法词后二元单词词性
		9	语法词前三元单词
		10	语法词前三元单词词性
		11	语法词后三元单词
		12	语法词后三元单词词性

[0112]

	句法结构特征	13	语法词修饰主语
		14	语法词直接宾语
		15	语法词附近助动词

[0113] VIII. 语法纠错统计模型

[0114] 语法纠错统计模型是通过提取训练文本集中单词的语法特征,并计算单词的语法特征的权重,生成用于纠正英语作文语法错误的语法纠错统计模型,它的格式如下:

[0115] 单词<sub>1</sub>^语法特征<sub>1</sub>^权重<sub>1</sub>

[0116] . . . . .

[0117] 单词<sub>i</sub>^语法特征<sub>i</sub>^权重<sub>i</sub>

[0118] . . . . .

[0119] 单词<sub>n</sub>^语法特征<sub>1</sub>^权重<sub>1</sub>

[0120] . . . . .

[0121] 单词<sub>n</sub>^语法特征<sub>j</sub>^权重<sub>j</sub>

[0122] IX. 语法纠错规则模型

[0123] 语法纠错规则模型是本发明总结归纳的英语语法错误规则,它包括片段错误、名词数不一致性错误、形容词比较级最高级错误、动词与其补语修饰成分不搭配、单词重复使用错误,它的具体内容如下表3所示:

[0124] 表3:语法纠错规则模型

[0125]

错误类型	规则设置	纠错方法
片段错误	句子的句法树结构存在短语词性片段标注。	提示句子不完整、句子成分残缺
名词数的不一致错误	1、名词修饰词表示复数意义,但是这个名词却是单数形式。 2、名词修饰词表示单数意义,但是这个名词却是复数形式。 3、名词是不可数名词,但是这个名词前面却使用了 <b>many</b> , <b>few</b> 来修饰。 4、名词是可数名词,但是这个名词前面却使用了 <b>little</b> , <b>much</b> 来修饰。	1、使用表示单数意义的修饰词或将名词改为复数形式。 2、使用表示复数意义的修饰词或将名词改为单数形式。 3、使用 <b>little</b> , <b>much</b> 等修饰不可数名词 4、使用 <b>many</b> , <b>much</b> 等修饰可数名词
形容词比较级最高级错误	<b>more</b> , <b>most</b> , <b>less</b> , <b>least</b> 后面使用了形容词比较级或最高级。	去掉 <b>more</b> , <b>less</b> 等修饰词或使用形容词原型形式

[0126]

动词与其补语 修饰成分不搭 配	help, remind, afford, choose, deserve, pretend, strive, struggle 后面没有接不定式 动词形式。	动词改为不定式形式
单词重复使用 错误	当上下相邻两个单词一样时,发生单词重复 使用错误。	去掉一个重复单词

#### 附图说明

- [0127] 图 1 是本发明方法的总体处理流程图；  
[0128] 图 2 是语法纠错预处理模块处理流程图；  
[0129] 图 3 是语法纠错模型训练模块处理流程图；  
[0130] 图 4 是语法错误检查纠正模块处理流程图。

#### 具体实施方式

[0131] 本发明的基于多元特征的英语作文语法错误自动纠正方法的具体实施方式分为如下三个步骤。

[0132] 第一步骤:执行“语法纠错预处理模块”

[0133] (1) 对训练文本集进行预处理,用于预处理的训练文本集取材于多种英文文章,它们是不包含任何单词错误、语法错误、表达错误的 50 万字正确英语短文,其中一段英语短文内容如下所示:

[0134] In all my life there are several people who help me a lot at my crucial moments,among whom my dear second uncle influences me most.Since my childhood,my beloved uncle,who was a person of integrity and a learned man known far and near,has told me again and again to be a good boy and to be a top student as well as an honest man.Being a boy I never made any mischief or told a lie.When I entered Anhui University,my respected uncle was very much pleased.He was so joyful that he told the good news to everyone he met.As I was the first college student in my hometown,he entertained me with a good meal as an encouragement and a reward.

[0135] (2) 对上述训练文本集中一段英语短文内容进行单词词性标注处理后,生成的单词词性标注库如下所示:

[0136] In/IN all/DT my/PRP\$life/NN there/EX are/VBP several/JJ people/NNS who/WP help/VBP me/PRP a/DT lot/NN at/IN my/PRP\$crucial/JJ moments/NNS,/,among/IN whom/WP my/PRP\$ dear/RB second/JJ uncle/NN influences/VBZ me/PRP most/JJS./.

[0137] Since/IN my/PRP\$childhood/NN,/,my/PRP\$beloved/JJ uncle/NN,/,who/WP was/VBD a/DT person/NN of/IN integrity/NN and/CC a/DT learned/VBN man/NN known/

VBN far/RB and/CC near/RB,/,has/VBZ told/VBN me/PRP again/RB and/CC again/RB to/TO be/VB a/DT good/JJ boy/NN and/CC to/TO be/VB a/DT top/JJ student/NN as/RB well/RB as/IN an/DT honest/JJ man/NN./.

[0138] Being/VBG a/DT boy/NN I/PRP never/RB made/VBD any/DT mischief/NN or/CC told/VBD a/DT lie/NN./.

[0139] When/WRB I/PRP entered/VBD Anhui/NNP University/NNP,/,my/PRP\$ respected/JJ uncle/NN was/VBD very/RB much/RB pleased/JJ./.

[0140] He/PRP was/VBD so/RB joyful/JJ that/IN he/PRP told/VBD the/DT good/JJ news/NN to/TO everyone/NN he/PRP met/VBD./.

[0141] As/IN I/PRP was/VBD the/DT first/JJ college/NN student/NN in/IN my/PRP\$ hometown/NN,/,he/PRP entertained/VBD me/PRP with/IN a/DT good/JJ meal/NN as/IN an/DT encouragement/NN and/CC a/DT reward/NN./.

[0142] (3) 对上述训练文本集中一段英语短文内容进行句子句法解析处理后,生成的句子句法树结构库如下所示:

[0143]

(S  
   (PP (IN In)  
     (NP (PDT all) (PRP\$ my) (NN life)))  
   (NP (EX there))  
   (VP (VBP are)  
     (NP  
       (NP (JJ several) (NNS people))  
       (SBAR  
         (WHNP (WP who))  
         (S  
           (VP (VBP help)  
             (S  
               (NP (PRP me))  
               (NP  
                 (NP (DT a) (NN lot))  
                 (PP (IN at)  
                   (NP (PRP\$ my) (JJ crucial) (NNS moments))))))  
             (, ,)  
             (PP (IN among)  
               (SBAR  
                 (WHNP (WP whom))  
                 (S  
                   (NP (PRP\$ my) (JJ dear) (JJ second) (NN uncle))

[0144]

(VP (VBZ influences)  
   (NP (PRP me))  
   (ADVP (RBS most)))))))))

(. .))

[0145]   prep\_in(are-6, life-4) nsubj(are-6, people-8) nsubj(help-10, people-8)  
 nsubj(lot-13, me-11) xcomp(help-10, lot-13) prep\_at(lot-13, moments-17)  
 nsubj(influences-25, uncle-24) prepc\_among(help-10, influences-25)

[0146]

(S  
   (P (IN Since)  
     (NP (PRP\$ my) (NN childhood)))  
   (, ,)  
   (NP  
     (NP (PRP\$ my) (JJ beloved) (NN uncle))  
     (, ,)  
     (SBAR  
       (WHNP (WP who))  
       (S  
         (VP (VBD was)  
           (NP  
             (NP  
               (NP (DT a) (NN person))  
               (P (IN of)  
                 (NP (NN integrity))))))  
           (CC and)  
           (NP  
             (NP (DT a) (JJ learned) (NN man))  
             (VP (VBN known)  
               (ADVP (RB far)  
                 (CC and)  
                 (RB near)))))))))  
     (, ,))  
   (VP (VBZ has)  
     (VP (VBN told)  
       (S  
         (NP (PRP me))

[0147]

(VP  
   (ADVP (RB again)  
     (CC and)  
     (RB again))  
   (VP (TO to)  
     (VP (VB be)  
       (NP (DT a) (JJ good) (NN boy))))  
   (CC and)  
   (VP (TO to)  
     (VP (VB be)  
       (NP  
         (NP (DT a) (JJ top) (NN student))  
         (CONJP (RB as) (RB well) (IN as))  
         (NP (DT an) (JJ honest) (NN man)))))))))  
 (. .))

[0148]   prep\_since(told-25, childhood-3) amod(uncle-7, beloved-6)  
 nsubj(person-12, uncle-7) nsubj(told-25, uncle-7) prep\_of(person-12, integrity-14)  
 nsubj(boy-34, me-26) aux(student-40, to-36)  
 [0149]

(S  
(S  
(VP (VBG Being)  
(NP (DT a) (NN boy))))  
(NP (PRP I))  
(VP  
(VP  
(ADVP (RB never))  
(VBD made)  
(NP (DT any) (NN mischief)))  
(CC or)  
(VP (VBD told)  
(NP (DT a) (NN lie))))  
(. .))  
nsubj(made-6, I-4) nsubj(told-10, I-4)  
(S  
(SBAR

[0150]

```

(WHADVP (WRB When))
(S
  (NP (PRP I))
  (VP (VBD entered)
    (NP (NNP Anhui) (NNP University))))))
(, )
(NP (PRP$ my) (JJ respected) (NN uncle))
(VP (VBD was)
  (ADJP (RB very) (RB much) (JJ pleased)))
(. .))
nsubj(entered-3, I-2)  nsubj(pleased-13, uncle-9)
(S
  (S
    (NP (PRP He))
    (VP (VBD was)
      (ADJP (RB so) (JJ joyful))
      (SBAR (IN that)
        (S
          (NP (PRP he))
          (VP (VBD told)
            (NP (DT the) (JJ good) (NN news))
            (PP (TO to)
              (NP (NN everyone))))))))))
    (NP (PRP he))
    (VP (VBD met))
  )
  (. .))
nsubj(joyful-4, He-1)  nsubj(told-7, he-6)  prep_to(told-7, everyone-12)  nsubj(met-14,
he-13)

```

[0151] (4) 对训练文本集进行预处理输出单词词性标注库和句子句法树结构库,对它们的单词进行词频统计,所生成的单词词频统计库如下所示。由于单词词频统计库较为庞大无法一一列举,因此下面只列举了单词词频统计库开始部分和结束部分的各 20 条数据,中间的数据用省略符号代替。

```

[0152] brave^86
[0153] newcomer^21
[0154] wishful^9
[0155] repaying^11

```

- [0156] undertakings<sup>21</sup>
- [0157] champions<sup>29</sup>
- [0158] degenerate<sup>19</sup>
- [0159] jew<sup>5</sup>
- [0160] riverbanks<sup>8</sup>
- [0161] portland's<sup>4</sup>
- [0162] jet<sup>31</sup>
- [0163] arthur<sup>38</sup>
- [0164] " <sup>3120</sup>
- [0165] smashing<sup>10</sup>
- [0166] !<sup>1160</sup>
- [0167] &<sup>9</sup>
- [0168] defeating<sup>18</sup>
- [0169] ' <sup>878</sup>
- [0170] stressing<sup>21</sup>
- [0171] \$<sup>560</sup>
- [0172] .....
- [0173] CC<sup>NN</sup>:<sup>93</sup>
- [0174] NN<sup>VB</sup>N<sup>RP</sup><sup>48</sup>
- [0175] VBP<sup>RB</sup>VB<sup>686</sup>
- [0176] VBN<sup>NNS</sup>NN<sup>31</sup>
- [0177] CC<sup>WRB</sup>TO<sup>120</sup>
- [0178] RB<sup>EX</sup>VBZ<sup>139</sup>
- [0179] VBP<sup>NN</sup>MD<sup>42</sup>
- [0180] WDT<sup>RB</sup>VBZ<sup>210</sup>
- [0181] NN<sup>WDT</sup>RBR<sup>11</sup>
- [0182] VB<sup>CC</sup>WRB<sup>29</sup>
- [0183] .<sup>,</sup>VBG<sup>13</sup>
- [0184] CC<sup>JJ</sup>RBR<sup>9</sup>
- [0185] DT<sup>WP</sup>.<sup>9</sup>
- [0186] RB<sup>EX</sup>VBP<sup>8</sup>
- [0187] WDT<sup>RB</sup>VBP<sup>81</sup>
- [0188] NN<sup>VB</sup>N<sup>RB</sup><sup>98</sup>
- [0189] CC<sup>RB</sup>PRP\$<sup>75</sup>
- [0190] ``<sup>EX</sup>VBZ<sup>59</sup>
- [0191] WDT<sup>RB</sup>VBD<sup>46</sup>
- [0192] MD<sup>VB</sup>EX<sup>8</sup>
- [0193] 第二步骤 : 执行“语法纠错模型训练模块”
- [0194] 语法纠错模型训练模块是对第一步骤执行语法纠错预处理模块输出的单词词性

标注库和句子句法树结构库进行语法特征提取、语法特征权重计算,最后生成语法纠错统计模型。

[0195] 由于单语法纠错统计模型较为庞大无法一一列举,因此下面只列举了语法纠错统计模型开始部分和结束部分的各 15 条数据,中间的数据用省略符号代替。

[0196] the<sup>1</sup>right<sup>1</sup>1.5026

[0197] the<sup>6</sup>upon<sup>on</sup>4.3405

[0198] the<sup>5</sup>will<sup>relieve</sup>2.1703

[0199] null<sup>7</sup>VB<sup>:</sup>0.0975

[0200] null<sup>2</sup>wide<sup>open</sup>2.6377

[0201] null<sup>6</sup>hundred<sup>miles</sup>1.3158

[0202] a/an<sup>2</sup>system<sup>for</sup>4.3792

[0203] a/an<sup>1</sup>job<sup>for</sup>0.8448

[0204] a/an<sup>6</sup>to<sup>outside</sup>2.1896

[0205] VBN<sup>1</sup>FW<sup>1</sup>1.5144

[0206] VBN<sup>2</sup>PRP<sup>-</sup>1.4945

[0207] VBN<sup>6</sup>NNP<sup>0</sup>0.2524

[0208] JJ<sup>3</sup>CD<sup>-</sup>0.4917

[0209] JJ<sup>3</sup>JJR<sup>-</sup>0.2659

[0210] JJ<sup>5</sup>CC<sup>-</sup>0.0775

[0211] .....

[0212] has<sup>4</sup>JJ<sup>-</sup>1.1251

[0213] has<sup>4</sup>IN<sup>-</sup>4.3116

[0214] has<sup>4</sup>TO<sup>0</sup>0.5336

[0215] has<sup>4</sup>CD<sup>1</sup>1.2985

[0216] has<sup>4</sup>PDT<sup>0</sup>0.4706

[0217] has<sup>4</sup>PRP<sup>-</sup>3.0987

[0218] has<sup>4</sup>NNS<sup>-</sup>4.5595

[0219] hasn't<sup>4</sup>VBN<sup>5</sup>4.465

[0220] is<sup>4</sup>TO<sup>0</sup>0.3711

[0221] is<sup>4</sup>^^^1.4695

[0222] is<sup>4</sup>DT<sup>1</sup>1.3394

[0223] is<sup>4</sup>JJ<sup>1</sup>1.0284

[0224] is<sup>4</sup>VBG<sup>1</sup>1.0988

[0225] is<sup>4</sup>IN<sup>-</sup>0.6948

[0226] is<sup>4</sup>VBN<sup>2</sup>2.1216

[0227] 第三步骤:执行“语法错误检查纠正模块”

[0228] 语法错误检查纠正模块是利用上述第二步骤生成的语法纠错统计模型和上述表 3 定义的语法纠错规则模型,对待纠错英语作文进行语法检查与纠正,最后输出待纠错英语作文的语法纠错结果,本实施方式的语法纠错过程如下:

[0229] (1) 读入如下所示一篇待纠错英语作文。

[0230] Friendship is essential in one's life.

[0231] Friends can help us know more variety people, open our vision. We can learn a lot of things from them. The most obvious thing is we all are the apple of our parents' eyes. They will do everything to make us happy. As a result, we are growing as the flowers in greenhouse. We can't live without our parents. However, we can learn to communicate, learn to care others and learn to be independence through friends.

[0232] As the saying goes, a friend in need is a friend indeed. If a friend abandons you in the difficulty. Personally speaking, he doesn't deserve your friendship.

[0233] In a word, making friends with right people is full of benefit.

[0234] (2) 对待纠错英语作文的句子进行单词词性标注,生成单词词性结构如下所示。

[0235] Friendship/NN is/VBZ essential/JJ in/IN one/NN's/POS life/NN./.

[0236] Friends/NNS can/MD help/VB us/PRP know/VB more/JJR variety/NN people/NNS,/, open/VB our/PRP\$vision/NN./.

[0237] We/PRP can/MD learn/VB a/DT lot/NN of/IN things/NNS from/IN them/PRP./.

[0238] The/DT most/RBS obvious/JJ thing/NN is/VBZ we/PRP all/DT are/VBP the/DT apple/NN of/IN our/PRP\$parents/NNS'/POS eyes/NNS./.

[0239] They/PRP will/MD do/VB everything/NN to/TO make/VB us/PRP happy/JJ./.

[0240] As/IN a/DT result/NN,/, we/PRP are/VBP growing/VBG as/IN the/DT flowers/NNS in/IN greenhouse/NN./.

[0241] We/PRP ca/MD n't/RB live/VB without/IN our/PRP\$parents/NNS./.

[0242] However/RB,/, we/PRP can/MD learn/VB to/TO communicate/VB,/, learn/VB to/TO care/VB others/NNS and/CC learn/VB to/TO be/VB independence/NN through/IN friends/NNS./.

[0243] As/IN the/DT saying/NN goes/VBZ,/, a/DT friend/NN in/IN need/NN is/VBZ a/DT friend/NN indeed/RB./.

[0244] If/IN a/DT friend/NN abandons/VBZ you/PRP in/IN the/DT difficulty/NN./.

[0245] Personally/RB speaking/NN,/, he/PRP does/VBZ n't/RB deserve/VB your/PRP\$ friendship/NN./.

[0246] In/IN a/DT word/NN,/, making/VBG friends/NNS with/IN right/JJ people/NNS is/VBZ full/JJ of/IN benefit/NN./.

[0247] (3) 对待纠错英语作文的句子进行句子句法解析,生成句子句法结构如下所示。

[0248]

(S  
   (NP (NNP Friendship))  
   (VP (VBZ is)  
     (ADJP (JJ essential)  
       (PP (IN in)  
         (NP  
           (NP (CD one) (POS 's))  
           (NN life))))))  
 (. .))  
 nsubj(essential-3, Friendship-1) prep\_in(essential-3, life-7)  
 (S  
   (NP (NNS Friends))  
   (VP (MD can)  
     (VP (VB help)  
       (S  
         (NP (PRP us))  
         (VP  
           (VP (VB know)  
             (NP (JJR more) (NN variety) (NNS people)))  
           (. .)  
           (VP (VB open)  
             (NP (PRP\$ our) (NN vision))))))  
   (. .))  
 nsubj(help-3, Friends-1) aux(help-3, can-2) nsubj(know-5, us-4)  
 (S  
   (NP (PRP We))  
   (VP (MD can)  
     (VP (VB learn)  
       (NP  
         (NP (DT a) (NN lot))

[0249]

(PP (IN of)  
(NP (NNS things))))

(PP (IN from)  
(NP (PRP them))))

(. .))

[0250] nsubj(learn-3, We-1) aux(learn-3, can-2) prep\_of(lot-5, things-7) prep\_  
from(learn-3, them-9)

[0251]

(S

(NP (DT The)

(ADJP (RBS most) (JJ obvious))

(NN thing))

(VP (VBZ is)

(NP

(NP (PRP we))

(SBAR

(S

(NP (DT all))

(VP (VBP are)

(NP

(NP (DT the) (NN apple))

(PP (IN of)

(NP

(NP (PRP\$ our) (NNS parents) (POS '))

(NNS eyes))))))))))

(. .))

[0252] nsubj(we-6, thing-4) nsubj(apple-10, all-7) prep\_of(apple-10, eyes-15)

[0253]

(S  
 (NP (PRP They))  
 (VP (MD will)  
 (VP (VB do)  
 (NP (NN everything))  
 (S  
 (VP (TO to)  
 (VP (VB make)  
 (S  
 (NP (PRP us))

[0254]

(ADJP (JJ happy))))))

(. .))

[0255] nsubj (do-3, They-1) aux (do-3, will-2) aux (make-6, to-5) xcomp (do-3, make-6)

[0256] nsubj (happy-8, us-7) xcomp (make-6, happy-8)

[0257]

(S  
 (PP (IN As)  
 (NP (DT a) (NN result)))  
 (, ,)  
 (NP (PRP we))  
 (VP (VBP are)  
 (VP (VBG growing)  
 (PP (IN as)  
 (NP  
 (NP (DT the) (NNS flowers))  
 (PP (IN in)  
 (NP (NN greenhouse))))))

(. .))

[0258] prep\_as (growing-7, result-3) nsubj (growing-7, we-5) aux (growing-7, are-6)

[0259] prep\_as (growing-7, flowers-10) prep\_in (flowers-10, greenhouse-12)

[0260]

(S  
(NP (PRP We))  
(VP (MD ca) (RB n't)  
(VP (VB live)  
(PP (IN without)  
(NP (PRP\$ our) (NNS parents))))))  
(. .))

[0261] nsubj(live-4, We-1) aux(live-4, ca-2) prep\_without(live-4, parents-7)

[0262]

(S  
(ADVP (RB However))  
(. ,)  
(NP (PRP we))  
(VP (MD can)  
(VP (VB learn)  
(S  
(VP (TO to)

[0263]

(VP  
   (VP (VB communicate))  
   (, )  
   (VP (VB learn)  
     (S  
       (VP (TO to)  
         (VP (VB care)  
           (NP (NNS others))))))  
   (CC and)  
   (VP (VB learn)  
     (S  
       (VP (TO to)  
         (VP (VB be)  
           (NP (NN independence))  
           (PP (IN through)  
             (NP (NNS friends))))))))))  
 (. .))

[0264]   nsubj(learn-5, we-3) aux(learn-5, can-4) aux(communicate-7, to-6)  
 xcomp(learn-5, communicate-7) xcomp(learn-5, learn-9) aux(care-11, to-10)  
 xcomp(learn-9, care-11) xcomp(learn-5, learn-14) aux(independence-17, to-15)  
 xcomp(learn-14, independence-17) prep\_through(independence-17, friends-19)  
 [0265]

(S  
 (SBAR (IN As)  
 (S  
 (NP (DT the) (NN saying))  
 (VP (VBZ goes))))  
 (, .)  
 (NP  
 (NP (DT a) (NN friend))  
 (PP (IN in)  
 (NP (NN need))))  
 (VP (VBZ is)  
 (NP (DT a) (NN friend))  
 (ADVP (RB indeed)))  
 (. .))

[0266] nsubj(goes-4, saying-3) advcl(friend-12, goes-4)

nsubj(friend-12, friend-7) prep\_in(friend-7, need-9)

[0267]

(FRAG  
 (SBAR (IN If)  
 (S  
 (NP (DT a) (NN friend))  
 (VP (VBZ abandons)  
 (NP (PRP you))  
 (PP (IN in)  
 (NP (DT the) (NN difficulty))))))  
 (. .))

[0268] nsubj(abandons-4, friend-3) prep\_in(abandons-4, difficulty-8)

[0269]

(S  
 (S  
 (ADVP (RB Personally))  
 (VP (VBG speaking)))  
 (,.)  
 (NP (PRP he))  
 (VP (VBZ does) (RB n't)  
 (VP (VB deserve)  
 (NP (PRPS\$ your) (NN friendship))))  
 (.))

[0270] nsubj(deserve-7, he-4) aux(deserve-7, does-5)

[0271]

(S  
 (PP (IN In)  
 (NP (DT a) (NN word)))  
 (,.)  
 (NP  
 (NP (NN making) (NNS friends))  
 (PP (IN with)  
 (NP (JJ right) (NNS people))))  
 (VP (VBZ is)  
 (ADJP (JJ full)  
 (PP (IN of)  
 (NP (NN benefit))))))

[0272]

(.))

[0273] prep\_in(full-11, word-3) nsubj(full-11, friends-6) prep\_  
 with(friends-6, people-9) prep\_of(full-11, benefit-13)

[0274] (4) 利用语法纠错规则模型的语法纠错规则对待纠错英语作文句子的语法错误进行检查。当检查到句子“If a friend abandons you in the difficulty.”时,输出如下句子句法树结构。

[0275]

```

(FRAG
  (SBAR (IN If)
    (S
      (NP (DT a) (NN friend))
      (VP (VBZ abandons)
        (NP (PRP you))
        (PP (IN in)
          (NP (DT the) (NN difficulty))))))
    (. .))

```

[0276] 在上述该句子的句子句法树结构中存在“FRAG”标记,它表示该句子的语法错误符合语法纠错规则模型中片段错误的语法错误类型。因此,语法纠错规则模型对应该句子的纠错方法是:句子不完整、句子成分残缺。

[0277] (5) 利用语法纠错统计模型对对待纠错英语作文句子的语法错误进行检查。当检查句子“As a result, we are growing as the flowers in greenhouse.”时,输出该句子的单词词性结构如下所示。

[0278] As/IN a/DT result/NN, /, we/PRP are/VBP growing/VBG as/IN the/DT flowers/NNS in/IN greenhouse/NN. /.

[0279] 输出该句子的句子句法树结构如下所示。

[0280]

```

(S
  (PP (IN As)
    (NP (DT a) (NN result)))
  (, ,)
  (NP (PRP we))
  (VP (VBP are)
    (VP (VBG growing)
      (PP (IN as)
        (NP
          (NP (DT the) (NNS flowers))
          (PP (IN in)

```

[0281]

```

(NP (NN greenhouse))))))

```

```

(. .))

```

[0282] prep\_as(growing-7, result-3) nsubj(growing-7, we-5) aux(growing-7, are-6)

[0283] prep\_as(growing-7, flowers-10) prep\_in(flowers-10, greenhouse-12)

[0284] 从该句子的名词短语词性结构“(NP (NN greenhouse)”中可以看出,修饰名词

“greenhouse”的冠词为零冠词,从单词多元语法特征集中提取零冠词的语法特征,利用上述VIII. 单词语法特征权重计算公式计算出单词词性语法混淆集中易混冠词“a/an, the, 零冠词”的语法特征权重如下所示。

[0285]  $\text{null}^1 \text{in}^0.5212$   
 [0286]  $\text{null}^3 \text{IN}^0.508$   
 [0287]  $\text{null}^5 \text{flowers}^{\text{in}^0.1489}$   
 [0288]  $\text{null}^6. ^0.3886$   
 [0289]  $\text{null}^7 \text{NNS}^{\text{IN}^0.2551}$   
 [0290]  $\text{null}^8. ^0.5212$   
 [0291]  $\text{null}^9 \text{greenhouse}^1.1726$   
 [0292]  $\text{null}^{10} \text{NN}^0.834$   
 [0293]  $\text{null}^{11} \text{countableNoun}^1.1044$   
 [0294]  $\text{null}^{12} \text{PP}^0.7564$   
 [0295]  $\text{a/an}^1 \text{in}^0.1646$   
 [0296]  $\text{a/an}^3 \text{IN}^0.2812$   
 [0297]  $\text{a/an}^6. ^0.1265$   
 [0298]  $\text{a/an}^7 \text{NNS}^{\text{IN}^0.0881}$   
 [0299]  $\text{a/an}^8. ^0.1646$   
 [0300]  $\text{a/an}^9 \text{greenhouse}^0.0004$   
 [0301]  $\text{a/an}^{10} \text{NN}^0.8646$   
 [0302]  $\text{a/an}^{11} \text{countableNoun}^0.8264$   
 [0303]  $\text{a/an}^{12} \text{PP}^0.2095$   
 [0304]  $\text{the}^1 \text{in}^5.8234$   
 [0305]  $\text{the}^3 \text{IN}^6.2512$   
 [0306]  $\text{the}^6. ^0.3082$   
 [0307]  $\text{the}^7 \text{NNS}^{\text{IN}^0.2249}$   
 [0308]  $\text{the}^8. ^0.3908$   
 [0309]  $\text{the}^9 \text{greenhouse}^5.0775$   
 [0310]  $\text{the}^{10} \text{NN}^1.4308$   
 [0311]  $\text{the}^{11} \text{countableNoun}^2.5586$   
 [0312]  $\text{the}^{12} \text{PP}^1.3763$

[0313] 根据上述当前单词分值计算式,计算出零冠词的分值为 6.2104,“a/an”的分值为 2.7259,“the”的分值为 23.4417。因此,根据语法纠错统计模型的纠错方法应该用分值最大的冠词“the”来纠正该句子名词短语词性结构“(NP(NN greenhouse))”的零冠词语法错误,即在名词短语词性结构“(NP(NN greenhouse))”里加入“the”。

[0314] (6) 输出待纠错英语作文的语法纠错结果。

[0315] 读入的一篇待纠错英语作文经过上述处理后,输出的语法纠错结果如下所示。

[0316] ① As a result, we are growing as the flowers in[B-ERROR] greenhouse[E-ERROR].

[0317] [B-ERROR] 表示语法错误的开始位置, [E-ERROR] 表示语法错误的结束位置, [B-ERROR]、[E-ERROR] 中间的“greenhouse”存在零冠词语法错误。本发明纠正该语法错误的说明是:在“greenhouse”前面需要插入定冠词“the”,以纠正它的零冠词语法错误。

[0318] ② [B-ERROR]If a friend abandons you in the difficulty. [E-ERROR]

[0319] [B-ERROR] 表示语法错误的开始位置, [E-ERROR] 表示语法错误的结束位置, [B-ERROR]、[E-ERROR] 中间的“If a friend abandons you in the difficulty.”存在 If 引导的条件状语从句缺少主句,句子成分残缺的语法错误。本发明纠正该语法错误的说明是:在“If a friend abandons you in the difficulty.”后面需要插入一个句子作为主句,以纠正它的句子不完整、句子成分残缺的语法错误。

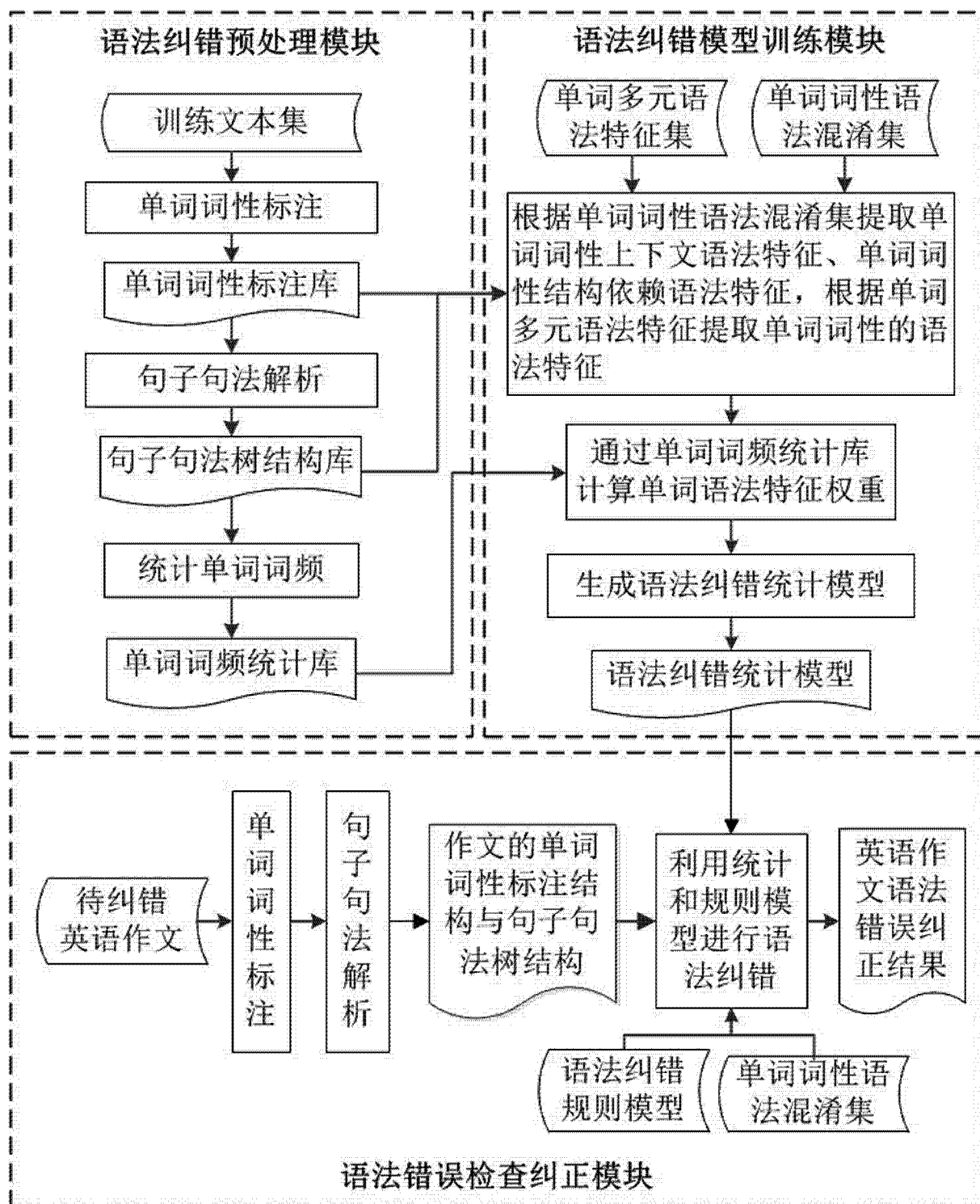


图 1

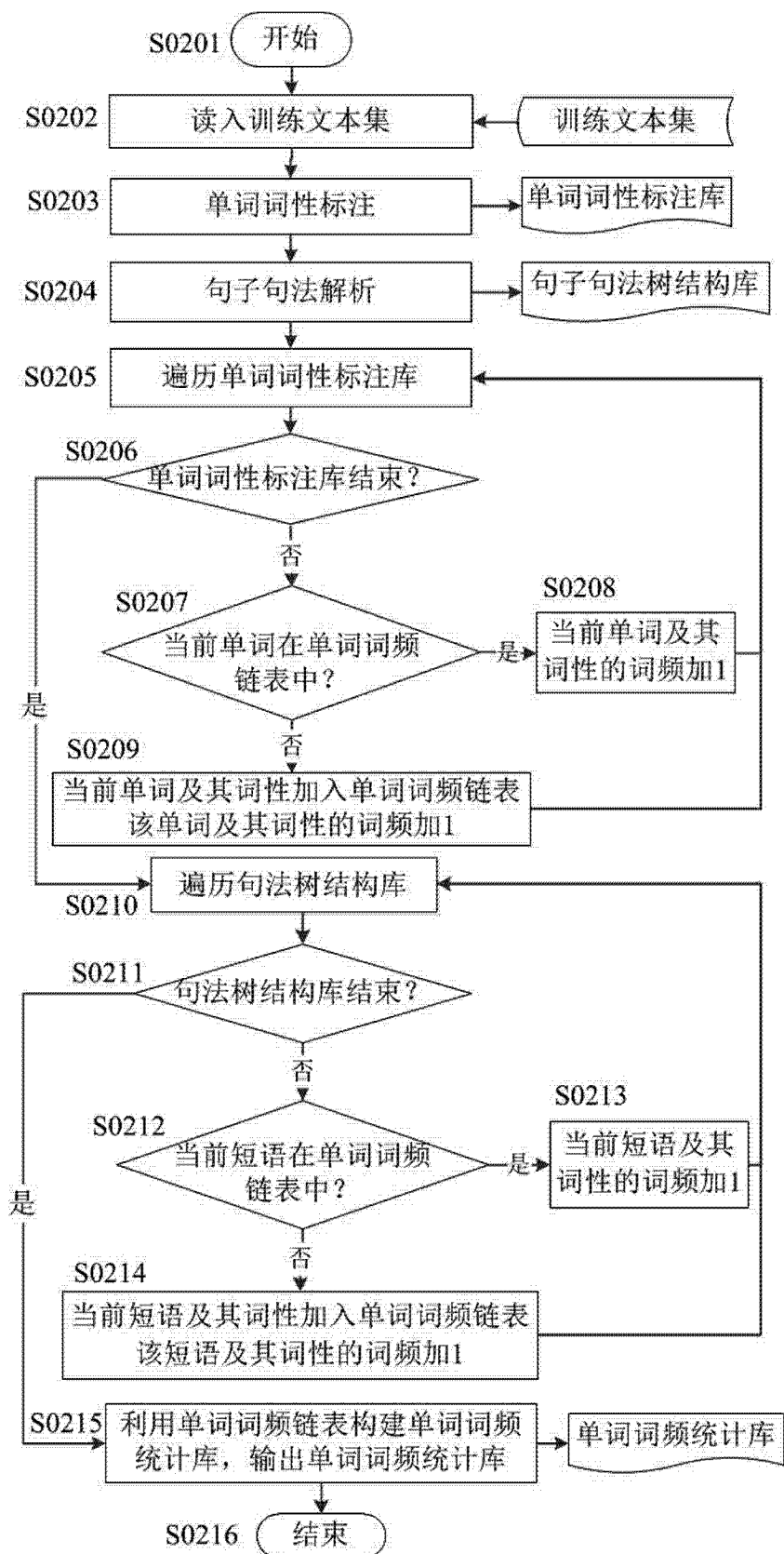


图 2

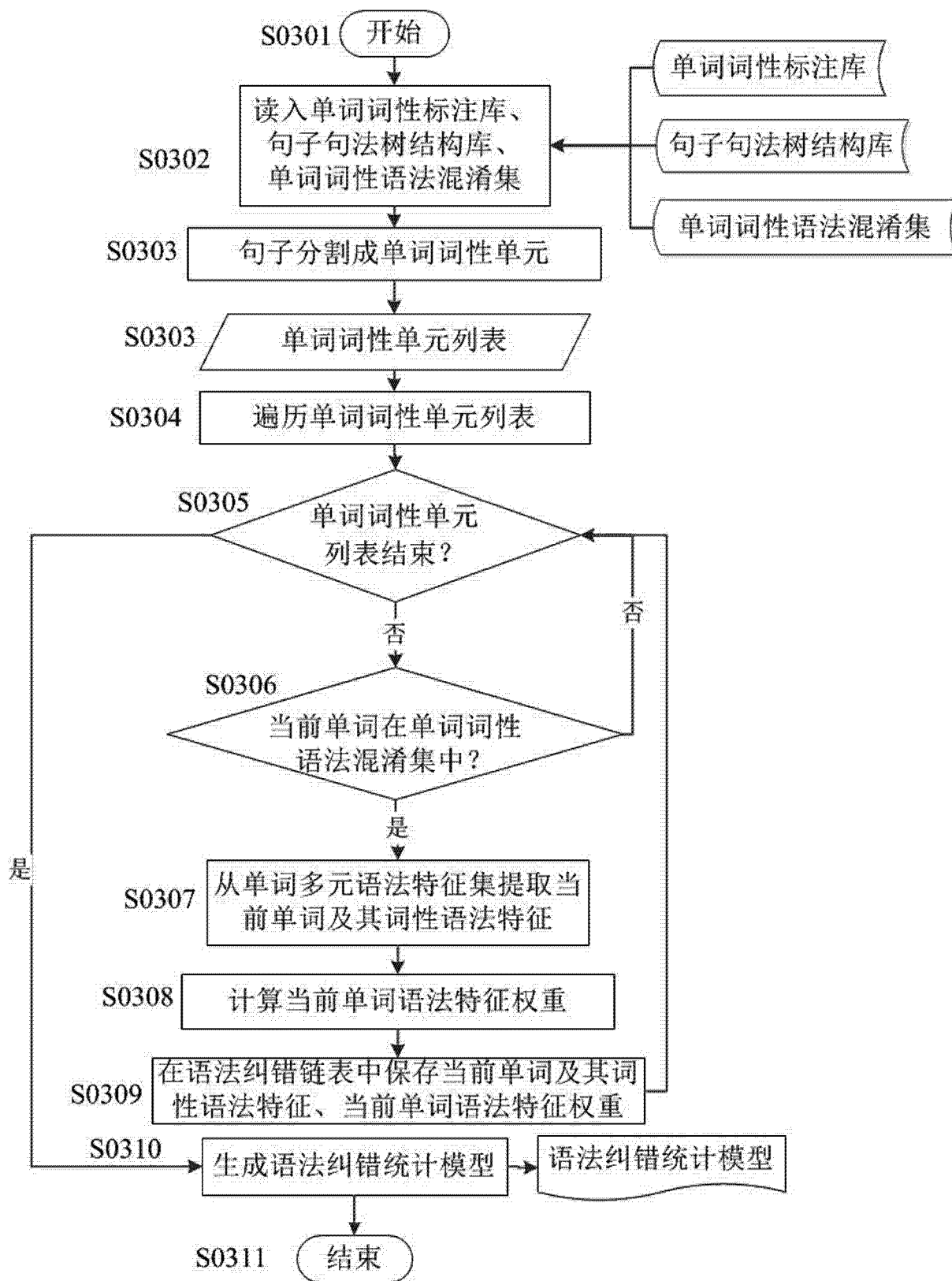


图 3

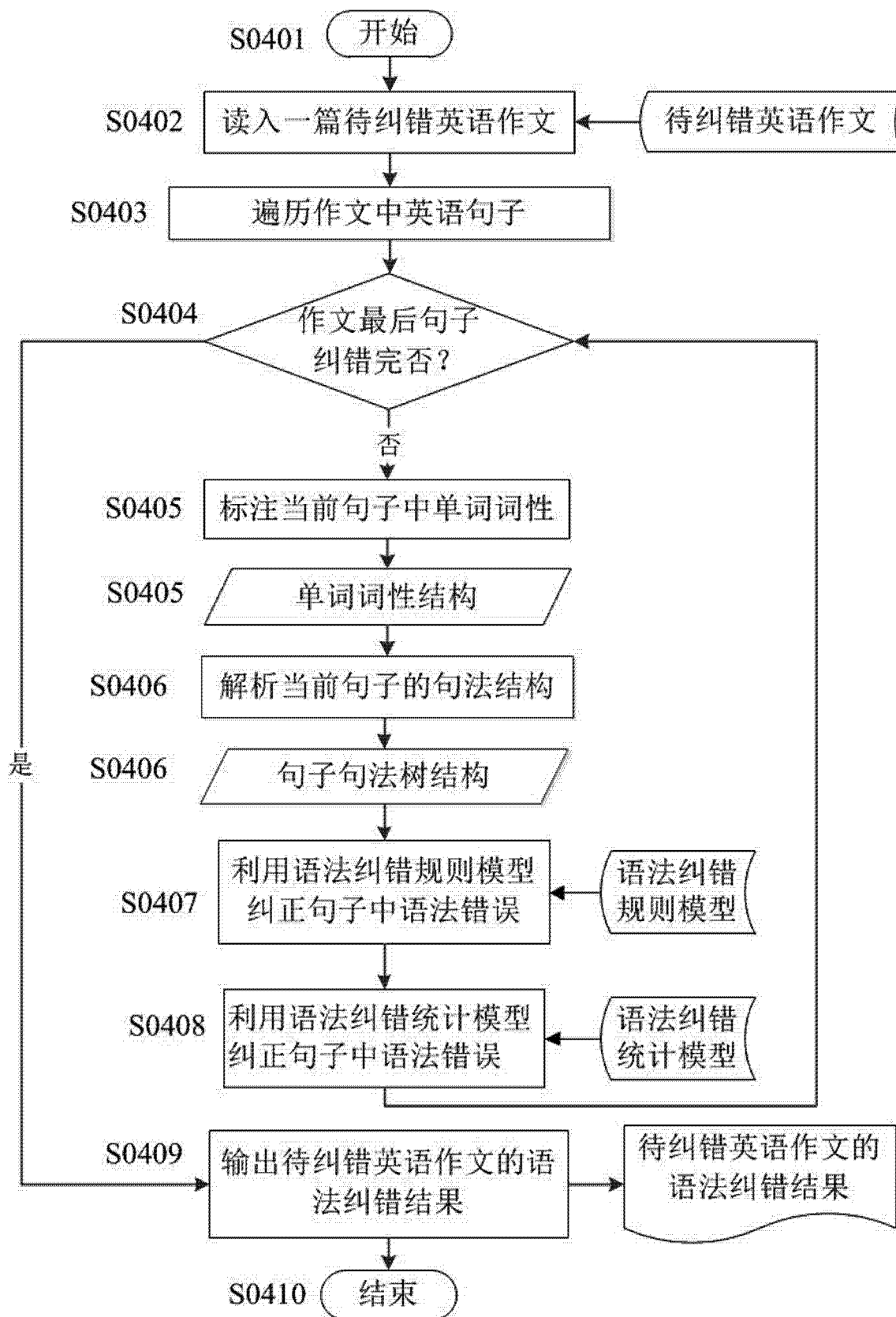


图 4