

基于规则的双重否定识别——以“不v1不v2”为例

王昱

北京大学中文系/ 北京

摘要

“不v1不v2”¹是汉语中典型的双重否定结构形式之一，它包括“不+助动词+不+v2”（不得不去）、“不+是+不v2”（不是不好）、述宾结构“不v1...不v2”（不认为他不）等多种双重否定结构，情况复杂。本文以“不v1不v2”为例，结合“元语否定”、“动词叙实性”、“否定焦点”等概念，对“不v1不v2”进行了全面的考察，制定了“不v1不v2”双重否定结构的识别策略。根据识别策略，设计了双重否定自动识别程序，并在此过程中补充了助动词表、非叙实动词表等词库。最终，对28033句语料进行了识别，识别正确率为97.87%，召回率约为93.10%。

关键词： 双重否定；非叙实动词；语义识别；否定焦点

Double Negative Recognition Based on Rules——Taking “不v1不v2” as an Example

Wang Yu

Department of Chinese Language and Literature,
Peking University / Peking

Abstract

“不v1不v2” is one of the typical double negation structures in Chinese. It includes “不+助动词+不+v2” (不得不去), “不+是+不v2” (不是不好), and “不v1...不v2”. Many kinds of double negative structures, such as “不v1...不v2” (不认为他不), are very complex. Taking “不v1不v2” as an example, with the theories of “non-truth-functional negation”, “factuality of verbs” and “negative focus”, this paper made a comprehensive investigation of “不v1不v2”. With the investigation result, this paper formulated a strategy for automatically recognizing the double negative structure of “不v1不v2”. According to the strategy, the automatic double negation recognition program is designed. During the designing process, this paper listed out the auxiliary verb list and the non factual verb List. Finally, the program was tested with 28033 sentences. The recognition accuracy is 97.87%, and the recall rate is about 93.10%.

Keywords: double negation, non-factual verb, semantic recognition, negative focus

¹v1与v2指动词，不包括动词短语。

1 引言

1.1 研究背景

在语义识别领域，否定对语义有着重要的影响，否定的有无影响整个句子的真值与情感。例如，“我喜欢这件衣服”，这句话的真值为真，情感为积极。但是如果在动词前加上否定词“不”（即“不喜欢”），真值便为假，情感便为消极。因此，句子中的否定成分自然是自然语言语义识别处理时必须考虑的内容。而在否定用法中，有一种特殊的用法——双重否定。丁声树先生在《现代汉语语法讲话》中将其概括为：“一句话先后用两个否定词，如‘不能不去’，‘没有人不去’，‘非去不可’之类，都是双重否定的句法。双重否定意思是肯定的，不过跟单纯肯定不全一样”¹。例如，“我不得不喜欢他”指“我得喜欢他”，“我不能不同意这个观点”指“我得同意这个观点”。虽然用的是否定的格式，表达的却是肯定的语义。这种连用否定表达肯定的特殊结构即为双重否定。对于自然语言语义识别来说，双重否定是处理语料时必须考虑的内容。如果计算机无法识别双重否定，则可能会导致整个句子的语义真值和情感极性的判断错误。例如，“我不知道你不喜欢他”与“我不认为你不喜欢他”，前者表示“你不喜欢他”，后者表示“你喜欢他”，两者具有相反的语义，又如“不是不容易，是很难”与“不是不容易，是超简单”的语义也不同，若计算机无法正确识别双重否定，便可能无法判断两者语义真值上的差别。因此，由于双重否定本身情况十分复杂且对语义有着重要影响，我们有必要对双重否定进行更全面深入的研究与探索。双重否定的自动识别将有利于句子语义与情感的识别，有助于聊天机器人、文本分析、问答系统等人工智能应用的进一步发展。

1.2 文献综述

目前国内对于双重否定的研究主要集中在双重否定的定义、范围、语义和语用等理论问题方面。双重否定的定义与范围方面，学界一直存在争议，如吕叔湘先生(1956)、王力先生(1943)等认为含有否定词的反语句是双重否定，而符达维先生(1986)则认为反语句不是双重否定。双重否定格式方面，已有研究都是从分类与举例的角度进行讨论，尚未有一篇论文从形式的角度对双重否定格式进行具体详细的遍历分析。目前对双重否定格式概括最为全面的是芜崧(1987)所划分的八大类型，25个格式，但是其主要涵盖的是“构式”类的双重否定，对于非构式类的双重否定尚未进行全面的考察。语义语用方面，叶文曦(2013)、方绪军(2017)、何爱晶(2019)等引入了Ladusaw(1997)的形式语义学，Horn(1985)的元语否定等理论，对一些典型的结构进行了分析，得出了具有解释力的成果。然而，由于双重否定的范围、格式还未确定，目前学者只集中分析了几个典型的结构，概括面十分有限，无法直接应用于实践。

自然语言理解领域，关于双重否定识别的研究非常少。目前只有王勇(2014)在其极性词典的构建中，简单地搜集了一些典型的双重否定结构，构建了一个双重否定词典。具体如下：

双重否定词典	绝非不、并非不、不是不、不能不、不会不、不可不、不要不、不得不、没有不、无不、不无
--------	---

Figure 1: 王勇(2014)²双重否定词典

然而该词典所归纳的结构数量有限，大量双重否定结构未被收入，如“不应不”、“不准不”、“不该不”、“非...不可”、“无一不”、“没有...不...”，等等。

1.3 本文的选题及意义

综上所述，目前国内多从分类与举例的角度对双重否定进行讨论，尚未有从形式的角度对双重否定格式进行遍历分析的成果。关于双重否定，我们只知道一些具体的实例，并没有归纳其完整的形式格式。另一方面，双重否定对于语义的真值与情感极值有着重大的影响，如“我非去不可”与“我不去”真值完全相反，然而目前国内并未有学者对双重否定的识别予以关注。基于

©2020 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

¹丁声树等.现代汉语语法讲话[M].北京:商务印书馆,2004.200-202.

此，本研究选取了“不v1不v2”为研究对象，尝试对含“不”的双重否定结构实现自动识别功能。本文将对“不v1不v2”做全面的考察，判断属于双重否定结构的“不v1不v2”的具体特征，并根据归纳出的特征建立自动识别程序。最后，建立的自动识别程序应能对含有多个“不”的语料进行识别，区分属于双重否定与不属于双重否定的句子，并返回其内部的双重否定结构。本研究可以作为预实验，为后续的全面的双重否定自动识别程序做基础。

2 “不v1不v2”双重否定的类型及识别规则

目前关于双重否定的定义和标准，学界尚未有定论。鉴于语义真值识别和情感极值判断是计算机对否定结构进行语义识别时所面临的主要问题，本文采用了形式语义学上对双重否定的定义：只要两次否定与肯定在语义真值上相同，“ $\neg \neg P == P$ ”，即为双重否定。符合双重否定条件的结构即为双重否定结构。

传统语言学上“不v1不v2”通常指“不得不，不能不，不要不，不会不，不是不”等v1为助动词或“是”的结构。这种“不v1不v2”结构是目前学界公认的典型的双重否定结构。然而从形式上来说“不v1不v2”还可以指“不走不跳”、“不买不卖”等并列结构，“不吃饭不睡觉”、“不买票不进场”等紧缩条件复句结构，以及“不觉得不尊重”、“不想不去”等述宾结构。这些结构也存在含有双重否定结构的可能。因此，为了得到完整的双重否定结构格式，本文结合前贤研究，采用遍历的方法，梳理了所有的“不v1不v2”³的语法形式并对其进行考察，试图找出所有可能的双重否定结构。具体的遍历方法如下：

1. 在CCL语料库中搜寻含有“不\$Num不”的语料（Num取值为“1-20”）；其中Num表示“不”与“不”之间相隔的字数。如“不得不”的Num为1，而“不觉得不”的Num为2；

2. 根据所得语料，结合词性限制，在“不\$Num不”中提取所有可能的“不v不v”，并进行数据统计；

3. 根据统计结果与具体语料，对结果进行概括，抽象出具体的所有可能的“不v1不v2”结构。

具体结果如下表所示。

不 v1 不 v2	例句
不+助动词+不+v2	不得不去；不能不喜欢
不+是+不+v2/不+是...+不+v2	不是不喜欢/不是他不知道
不 v1+不+v2/不 v1...+不+v2 (紧缩、并列、主谓、述宾等结构)	不去不行；不唱不跳；不吃饭不好； 不觉得不尊重；不认为他不喜欢你

Table 1: “不v1不v2”格式表

2.1 不+助动词+不+v2

“不+助动词+不+v2”是目前公认的双重否定的结构，其具体格式为“不+表示可能或必要的助动词+不+v2”，如“不得不去”、“不会不来”、“不可不说”等等。例句如下：

- 1a. 一些农村学校的校长为了保证学习的正常运转，不得不四处筹钱。
- 1b. 一些农村学校的校长为了保证学习的正常运转，（必须）得四处筹钱。
- 2a. 这样重大的事件新闻界不会不给予特别重视。
- 2b. 这样重大的事件新闻界会给予特别重视。

1-2的a、b例句，虽然它们的语气强度存在区别，如1a中的“不得不”比1b中的“得”的肯定语气更强，然而这种语气变化并不影响语义的真值。1a、2a仍与1b、2b的语义真值一致，符合“ $\neg \neg P == P$ ”双重否定的标准，属于双重否定结构。

对于“不+助动词+不+v”这种在长期使用中已经逐渐变为接近于构式的固化结构，本文采取简单的字符串匹配的方法，便可实现对其的识别。

³v1若不为助动词，则前面还可带助动词。构成“不+助动词+v1+不+v2”结构，如“不会认为不好”，这种结构也是识别的对象。

2.2 不+是+不+v2/不+是...+不+v2

2.2.1 不+是+不+v2

“不+是+不+v2”与“不+助动词+不+v2”结构类似，也是最为常见的双重否定结构之一，但是“不+是+不+v”的情况更为复杂。因为“不+是+不+v”中的“不是”除了可以表示描述性真值否定 (descriptive truth-functional negation) 外，还可以表示元语否定 (non-truth-functional negation)。“所谓元语否定，就是用元语言对对象语言所描述的非真值语义的否定，...是一种非真值意义否定；与之相对应的是真值否定，否定的是句子的真值条件 (truth condition)”⁴ 这种元语否定常常是引述性否定，是对之前对话中已出现的内容的否定。例句如下：

- 3a. “可现在杀他不容易啊。”有人说。不是不容易，是根本不可能。
 3b. * “可现在杀他不容易啊。”有人说。是容易，是根本不可能。
 4a. “可我不想跟她结婚。”刘东北进一步道，“不是不想跟她结婚，是不想结婚。”
 4b. * “可我不想跟她结婚。”刘东北进一步道，“是想跟她结婚，是不想结婚。”
 5a. 不是不要读书，而是要读得更好。
 5b. 是要读书，（而）是要读得更好。

上述例句，3-4中的“不是”为元语否定，它是对之前内容的语用否定，而不是对本句内容的语义否定。3a与3b，4a与4b语义不一致，且3b、4b内部也有语义矛盾，无法成立。从语义来说，3a、4a中的“不是不”只包含一重语义否定，其逻辑式为“ $\neg P$ ”，并不符合“ $\neg \neg P = P$ ”双重否定的标准，因此，这种“不是不”不属于双重否定结构。而5a例句中的“不是”是对语义的否定，其“不是不”符合“ $\neg \neg P = P$ ”双重否定的标准，属于双重否定结构。

综上所述，对于“不+是+不+v”，本文需要使计算机先区分其中的“不是”是元语否定还是描述性真值否定，然后才能判断其是否为双重否定结构。在单纯的文本里，绝大多数表示元语否定的“不是”都只出现在“不是+不+x，（而）是+y”的结构中。因此，下面我们将以“不是+不+x，（而）是+y”为对象作进一步的讨论。

由于元语否定“是对命题适宜性进行的修订性否定”，其修正内容并不是对真值进行修正，因此修正内容 (y) 的真值应与“错误内容 (即“不x”)”的真值一致。而对于描述性真值否定来说，其修订的是命题真值，因此，其修正内容 (y) 的真值应与“错误内容 (即“不x”)”的真值相反。根据这一观察，我们提出了判断“不是+不+x，（而）是+y”是否为双重否定结构的具体方法：提取“不是+不+x”中的x，并匹配y与“不x”的真值⁵。若“不是+不+x，（而）是+y”中“y”的真值与“不x”一致，则句中的“不是不”不为双重否定结构；否则“不是不”为双重否定结构。

结合这一方法，本文对3、5例句分析如下：

- 3a. “可现在杀他不容易啊。”有人说。不是不容易，是根本不可能。
 (y真值: -1, 不x真值: -1, 不为双重否定)
 5a. 不是不要读书，而是要读得更好。
 (y真值: 1, 不x真值: -1, 为双重否定)

根据上述方法，本文可以判断出上述四个例句中的“不是不”，是否是双重否定结构。⁶ 因此，本文采用字符串匹配与计算真值的方法，实现计算机对“不是不”双重否定结构的识别。

2.2.2 不+是...+不+v2

不+是...+不+v2除了需要满足“‘不是’表示语义否定”的条件外，还需要满足“不是”的否定焦点落在“不v2”上的条件。如“不是我故意不来”并不等于“是我故意来”。因为在该例中，“不是”的否定焦点是“故意”而不是“不来”。具体否定焦点的限制条件本文将会在2.4.2节中进行讨论。

2.3 “不v1不v2”/“不v1...不v2” (非述宾结构)

“不v1不v2”指“v1”不是助动词、不是“是”的结构，它的结构类型有并列、紧缩（主谓）、述宾等。“不v1...不v2”则指“不v1”与“不v2”之间不紧邻的“不v1+其他内容+不v2”结构。下面

⁴何爱晶.反叙的非真值义否定和真值义肯定[J].外语研究,2019,36(04):第25页.

⁵判断真值的方法为，计算其谓语句真值。一次否定为-1，两次否定为1，以此类推。判断真值时，也包括内含否定的谓词，如“讨厌”、“丑”等。

⁶这种策略，理论上仍有可能存在例外，但是目前尚未在自然语言语料中发现例外。

本文将对以上各个结构类型一一进行讨论。

紧缩结构⁷的“不_{v1} (...) 不_{v2}”包括“不_x不行”、“不_x不成”等固定形式的短语结构，也包括“不买票就不让进”“不给钱不办事”这一类表达。其语义为“如果不_{v1}，那么不_{v2}”。关于紧缩条件类的结构是否为双重否定未有定论。在这里本文从形式语义学的角度对其进行讨论。以“不_x不行”为例。“_x行”语义为“如果_x，那么行”。P命题可以分解为 q_1 “_x”， q_2 “y”，逻辑式为蕴含式 $q_1 \rightarrow q_2$ 。而“不_x不行”语义为“如果不_x，那么不行”，逻辑式应为蕴含式 $\neg q_1 \rightarrow \neg q_2$ 。从下列真值表本文可以看出， $q_1 \rightarrow q_2$ 与 $\neg q_1 \rightarrow \neg q_2$ 的语义真值不一致，不符合“ $\neg \neg P == P$ ”的标准，因此从形式语义学来看，紧缩语义结构不是双重否定结构。⁸

q_1	q_2	$\neg q_1$	$\neg q_2$	$q_1 \rightarrow q_2$	$\neg q_1 \rightarrow \neg q_2$
T	T	F	F	T	T
T	F	F	T	F	T
F	T	T	F	T	F
F	F	T	T	T	T

Table 2: 紧缩结构语义真值表

并列结构的“不_{v1} (...) 不_{v2}”，指“不哭不闹”、“不高不低”这一类表达。袁毓林（1999）提出并列结构“通常不能通过直接在这种谓词性并列结构的前面加上“不、没有”等否定词来构成否定式，而是要在这种并列结构的各个直接成分之前分别加上“不、没有”等否定词。”⁹因此，“不_{v1}不_{v2}”只是“ $v_1 v_2$ ”并列结构的一重否定结构，不属于双重否定结构。如“不哭不闹”不等于“哭闹”。

2.4 “不_{v1}不_{v2}”/“不_{v1}... 不_{v2}”（述宾结构）

述宾结构的“不_{v1}不_{v2}”（不_{v1}与不_{v2}紧邻）、“不_{v1}... 不_{v2}”（不_{v1}与不_{v2}非紧邻）指“不觉得不好”、“不知道你不来”等 v_1 为述语，“不_{v2}”为宾语的结构。该结构中存在一部分结构，其“不_{v1}”对“不_{v2}”有语义指向，属于双重否定结构，如例7；同时还存在一部分结构“不_{v1}”与“不_{v2}”之间没有语义指向，不是双重否定结构，如例6。

- 6a. 我不知道你不来。（述宾）
- 6b. *我知道你来
- 7a. 平时也是人来人往，我不相信你不难受。（述宾）
- 7b. 平时也是人来人往，我相信你难受。（述宾）

因此，对于述宾结构的“不_{v1} (...) 不_{v2}”，本文的主要任务是找出其中“不_{v1}”对“不_{v2}”存在语义指向、属于双重否定的结构，并提取其特征，制定识别规则。下面本文将以“不_{v1}... 不_{v2}”为对象进行讨论。

2.4.1 第一个条件： v_1 为非叙实动词

首先我们需要确定在什么情况下“不_{v1}”对“不_{v2}”具有语义指向与管辖功能。

- 8a. 我不知道他不来
- 8b. *我知道他来
- 9a. 我不幻想他不来
- 9b. *我幻想他来
- 10a. 我不认为他不来
- 10b. 我认为他来

通过例句，可以发现，当 v_1 为“知道”、“幻想”时“不_{v1}... 不_{v2}”不能转换成“ v_1 ... v_2 ”，而当 v_1 为“认为”时，却可以转换。“知道”、“幻想”、“认为”同样是动词，却存在着区别。本文认为，“不_{v1}”对“不_{v2}”是否有语义指向或管辖的作用与 v_1 的语义有关，具体来说与 v_1 的叙实性有关。李新良（2015）将叙实性定义为“叙实性是动词的一种语义功能，即动词预设其宾语小句真值的能力。具体来说，肯定式和否定式都预设其宾语小句为真的动词是叙实动词，叙实动词具有的预设其宾语小句为真的能力叫叙实功能；肯定式和否定式都不预设其宾语小句为真，也不预设其宾语小句为假的动词是非叙实动词，非叙实动词具有的不预设其宾语小句为真，也不

⁷我们认为主谓结构的不_{v1} (...) 不_{v2}与紧缩结构相似。由于篇幅限制，便不再讨论。

⁸虽然紧缩语义结构不是双重否定结构，但是比起普通的并列结构，紧缩语义结构仍然具有其独特的表达功能，应该与其他结构区分开，未来我们希望能对其展开进一步的研究。

⁹袁毓林.并列结构的否定表达[J].语言文字应用,1999(03):第42页.

预设其宾语小句为假的能力叫非叙实功能；肯定式和否定式都预设其宾语小句为假的动词是反叙实动词，反叙实动词具有的预设其宾语小句为假的能力叫反叙实功能¹⁰。因此，对于叙实动词和反叙实动词来说，无论它自身是肯定式还是否定式，它的宾语小句的真值都不会改变。如“我知道他不来”与“我不知道他不来”中宾语小句的语义都是“他不来”。所以，对于叙实动词与反叙实动词，由于其宾语小句真值已定，否定式不能管辖后一宾语小句，无法满足“ $\neg \neg P == P$ ”的条件，不为双重否定。而当v1是非叙实动词（如：认为）时，由于非叙实动词的宾语小句并没有预设，在述宾结构中，v1对宾语具有约束管辖关系，能够影响宾语的真值，具有可以产生“ $\neg \neg P == P$ ”的条件，存在属于双重否定的可能。因此，本文得出了“不v1...不v2”述宾结构为双重否定的第一个条件：v1为非叙实动词。

2.4.2 第二个条件：“不v1”的否定焦点包含v2

该条件只对分开的“不v1...不v2”有约束，对于紧连的“不v1不v2”并无影响。当句子v1确定为非叙实动词时，该句子并不一定为双重否定句。示例如下：

- | | |
|---------------------------|---------------|
| 11a.我不认为他不来。 | 11b.我认为他来。 |
| 12a.我不认为他故意不来。 | 12b.*我认为他故意来。 |
| 13a.我不相信他不喜欢我。 | 13b.我相信他喜欢我。 |
| 14a.我不相信他不喜欢我到了看见我就恶心的地步。 | |
| 14b.*我相信他喜欢我到了看见我就恶心的地步。 | |

在例句中，11a、13a可以转换为11b、13b，而12a、14a却不能转换为12b、14b。为何？本文认为这主要与否定的焦点有关。袁毓林(2000)指出“有的成分表达的是句子的预设意义，属于旧信息，事实上它们的意义在否定的情况下仍然得以保持；有的成分表达的是句子的焦点意义，属于新信息，它们是真正被否定的。”当v2不是否定焦点时，“不v1”并不会对v2进行否定，不会形成“ $\neg \neg P$ ”，因此不满足“ $\neg \neg P == P$ ”的条件，不是双重否定结构。如“我不认为他故意不来。”中的“不v1”否定的是“故意”而不是“不来”，其中“不来”是预设成分，属于旧信息，“不v1”并不会影响到“不来”的真值。如：

- | | |
|----------------|---------|
| 15a.我不认为他故意不来。 | 预设：他不来。 |
| 15b.我认为他故意不来。 | 预设：他不来。 |

因此，为了满足双重否定“ $\neg \neg P == P$ ”的条件，本文需要确保“不v1”的否定焦点是落在v2上的。结合袁毓林(2000)关于否定词焦点与辖域的观点，本文将“不v1...不v2”中“不v1”的否定焦点的情况归结如下：

(1) 若v2存在谓语状语或者谓语补语¹¹，则谓语状语或者谓语补语是否定焦点；反之，则v2是否定焦点；

(2) 若v2的状语、补语、宾语中含有全称量词或者“一+量”时，量词为否定焦点。

- | | |
|------------------|-----------------|
| 16a.我不认为他每一天都不来。 | 16b.*我认为他每一天都来。 |
| 17a.我不认为他不喜欢所有人。 | 17b.*我认为他喜欢所有人。 |

(3) 若v2与其他谓语结构构成紧缩复句结构，如“不v2就...”、“不v2不...”，v2不是否定焦点，否定焦点是整个紧缩复句，无法构成双重否定结构；若v2与其他谓语结构构成并列结构“不v2不v3”时，v2不是否定焦点，否定焦点是整个并列结构，但是仍然可以构成双重否定结构，不过转换成肯定式时需要同时去掉并列结构中所有的“不”。

- | | |
|----------------|---------------------------|
| 18a.我不认为他不吃不喝。 | 18b.我认为他吃喝。 ¹² |
| 19a.我不认为不去就不行。 | 19b.*我认为去就行。 |

由此本文得出了述宾结构的“不v1...不v2”为双重否定的第二个条件：不v1的否定焦点包含v2。这一条件同样适用于“不是...不v2”。具体参照2.2.2。

¹⁰李新良,王明华.汉语动词的叙实性研究的应用前景[J].对外汉语研究,2015(02):第122页.

¹¹当否定句中存在“连、就”等标记时，双重否定转换时，语义会发生一定的改变。但是不v1的否定焦点仍然包含了“V2”。

¹²这里否定到肯定的转化是一个常规的理解。本文不否认存在一定的可能性，“我不认为他不吃不喝”表达的是“我认为他吃但是不喝”，或者“我认为他不吃但是喝”。但是这种可能性不符合人们平时交际的习惯，因此本文不对此进行进一步的讨论。

3 双重否定自动识别程序的建立

3.1 词库的建立

为了使计算机能够识别助动词、非叙实动词，本文对助动词与非叙实动词进行了梳理，在常用的基础词表中补充了助动词词表与非叙实动词词表。助动词方面，本文以郑贵友（1989）¹³整理的助动词范围为基本，结合鲁晓琨（2004）¹⁴等前人的研究以及现代汉语的使用情况，选取了26个助动词，构成常用助动词词表。具体如下：

能、能够、可能、会、可以、应该、应、应当、要、得、愿意、愿、甘愿、肯、可、情愿、想、要、敢、该、配、当、准、许、得、容

非叙实动词方面，结合袁毓林、李新良等人对非叙实动词的研究，本文认为非叙实动词多为心理动词。因此本文对心理动词进行了考察。若一个心理动词的宾语的真值无法确定，则该心理动词为非叙实动词。以此为标准，本文对心理动词进行了考察，筛选出了24个非叙实动词¹⁵。整理如下：

认为、觉得、想（料想、猜想）、感到、情愿、相信、乐意、愿意、盼望、希望、猜、猜测、揣摩、揣摩、推测、估计、估摸、猜想、考虑、打算、说（认为）、同意、赞同、允许

3.2 双重否定自动识别程序的流程

以第二章所讨论的语法规则为核心，本文设计了双重否定自动识别程序。程序的输入为含有两个副词词性“不”的语料txt文件，输出为一个txt文件。输出的txt文件分为两个部分，一是所有识别出的含有双重否定结构的句子及其双重否定结构类型；二是语料总句数、双重否定句子句数、各双重否定结构的句子句数等统计信息。示例如下：

```
【文件名:\当代\报刊\人民日报\1995年人民日报\6月份.txt 文章标题:作者:】99937:...该所旅馆的老板娘川邑太太,看到有些年轻夫妇因为带着孩子去旅行不方便而不得不取消旅行计划,灵机一动抓住了“亮点”,办起了托儿服务旅馆, ... 【文件名:\当代\报刊\人民日报\1995年人民日报\6月份.txt 文章标题:作者:】99938:...一次,队里统一出工。

*****此句话为双重否定句,双重否定结构为不+助动词+不*****
因此,它们的主张和行动不可能不对整个世界经济与政治产生重大影响。

*****此句话为双重否定句,双重否定结构为不+助动词+不*****
一共有28033句不+助+不的结构有628句不+助动词+非叙实动词+不 的结构有45句不+非叙实动词+不的结构有27句不+助动词+非叙实动词+...不的结构有2句不+是+...不的结构有199句
```

Figure 2: 实验结果示例图

具体程序的识别步骤如下：

(1) 通过python程序使用哈工大LTP对语料进行分句、分词、词性标注；对每一个句子进行以下 (2) - (6) 操作；

(2) 检测其是否含有“不+助动词+不+v”结构，若含有，则为双重否定句，其双重否定结构为“不+助动词+不+v”，将句子写入文件并归入数据统计，跳过后续步骤。若不含有，则进行下一步；

(3) 检测其是否含有“不+助动词+非叙实动词+不+v”结构，若含有，则为双重否定句，其双重否定结构为“不+助动词+非叙实动词+不+v”，将句子写入文件并归入数据统计，跳过后续步骤。若不含有，则进行下一步；

(4) 对句子进行句法分析；

(5) 检测其是否含有述宾关系的“不+非叙实动词 (+...) +不+v”结构，若含有，则为双重否定句，其双重否定结构为“不+非叙实动词 (+...) +不+v”，将句子写入文件并归入数据统计，跳过后续步骤。若不含有，则进行下一步；

¹³郑贵友.汉语“助动词”的研究刍议[J].汉语学习,1989(06):23-27.

¹⁴鲁晓琨.现代汉语基本助动词语义研究[M].北京:中国社会科学出版社,2004.

¹⁵非叙实动词的界限并不是完全清晰的。非叙实动词与叙实动词、反叙实动词之间还存在一定的渗透性。由于其情况较为复杂，且对本文研究的影响较小，故暂不讨论。

(6) 检测其是否含有述宾关系且“不是”为语义真值否定（通过否定值判断）的“不+是（+...）+不+v”结构，若含有，则为双重否定句，其双重否定结构为“不+是（+...）+不+v”，将句子写入文件并归入数据统计，跳过后续步骤。若不含有，则该句不含有双重否定结构，进行数据统计。

4 双重否定自动识别实验

4.1 实验语料来源

本文在ccl语料库中，提取了100000条含有“不”的语料，并通过程序从中抽取含有两个副词词性“不”的句子，共计28033句，以此为“不”的基础语料，测试双重否定识别程序的正确率。同时，本文从28033句语料中选取了1000条语料作为人工检测语料，以测试双重否定识别程序的召回率。

4.2 实验语料来源

本文使用双重否定识别程序对28033句语料进行识别，获得了894句含有双重否定结构的句子。经人工的检校，发现以上894个句子中，含有双重否定结构的句子数为875，该程序识别正确率为97.87%。具体如下表¹⁶：

结构	识别句子数	错误句子数	正确句子数	正确率
不+助+不+v	628	0	628	100%
不+助动词+非叙实动词+不+v	45	2	43	95.56%
不+非叙实动词+不+v	27	3	24	88.89%
不（+助动词）+非叙实动词 +...+不+v	2	2	2	100%
不+是（+...）+不+v	192	11	181	94.27%
总计	894	19	875	97.87%

Table 3: 实验结果统计表

为了测试该程序的召回率，本文人工对1000句语料进行检校，筛选出了29句含有双重否定结构的句子。本文将这1000句语料输入到双重否定识别程序中，程序识别出了27句含有双重否定结构的句子。因此，该程序的召回率约为93.10%。根据F1值公式与上述数据，该程序的F1值为95.43%。

4.3 实验分析

无论是正确率还是召回率，实验的准确率都是在百分之九十多，未达到百分之百。我们对以上三个未达到百分百正确率的双重否定结构的语料进行分析。结果显示，程序识别与召回错误主要与句子的句法分析错误有关。由于分词与句法分析等基础自然语言处理工具的问题，程序对一些句子的句法判断错误，导致一些原本应被判为并列关系、因果关系的成分，被误判为述宾关系，从而使整个双重否定结构的判断错误。例句如下：

20. 如果用适用各种土质的几十台钻机同时作业，电力条件不允许，地下管网不安全。

21. ...每到拔棉柴时，农民心里直犯愁，用手拔吧，费工费时不说，拔不了多少，便满手血泡。有的急于求成把将来要办的事情，拿到今天来办，由于条件不允许迟迟开展不了。

以上例句中，v1与v2都并非述宾关系，不符合程序所制定的双重否定的规则，然而程序并未检索出这一点，导致识别错误。这个问题主要与句法分析的处理工具有关，本文暂时无法对其进行进一步的改进。

除此之外，还有一些句子的识别错误，是由于结构格式总结疏漏所造成的。例句如下：

22. 因为他们不是不懂足球，便是偏爱市井蜚闻的帮闲之辈。

上述例句，其结构为“不是...便是...”，整体为选择并列的关系。其中“不是不懂足球”中的“不是”并不表示否定的功能，而是选择的功能。可以理解为“要么是不懂足球”，因此不属于双重否定结构。然而由于我们归纳结构格式时，没有考虑到类似的特殊结构，将其简单地归入了“不是不”结构，从而导致了错误。这一问题，可以通过更多次的实验与探究来进行改进。

¹⁶由于文章讨论的范围限制，删去了“岂不是”这类不+反问词的句子。

4.4 实验讨论

上述实验结果表明, 本文设计的双重否定识别程序具有从语料中识别出含有双重否定结构的句子的能力, 程序识别的正确率与召回率都较高, 证明了基于规则进行双重否定自动识别的可能性, 具有实践价值。然而, 本次实验的实验对象单一(只有“不 v_1 不 v_2 ”), 许多常见的双重否定结构, 如“不”与其他否定词构成的双重结构(“不...没有”、“没有...不”、莫不、无不等), “不”与反问词构成的双重否定结构(“难道不”、“岂不”等), 都未被纳入监测范围。因此, 为了进一步完善双重否定识别程序, 我们有必要对双重否定结构进行更为全面的考察。只有这样, 双重否定识别程序才能真正地实现计算机对自然语言中双重否定结构的自动识别, 具有实际的应用价值。

5 总结

本文讨论了“不”的双重否定结构, 采用遍历的方式, 对所有可能的“不 v_1 不 v_2 ”的结构进行了考察, 整理出了完整的“不 v_1 不 v_2 ”结构, 对现有的双重否定结构形式进行了补充。本文首次采用“非叙实动词”作为述宾式“不 v_1 ...不 v_2 ”双重否定结构成立的条件, 解释了“我不知道你不来”与“我不认为你不来”之间的区别; 引入了否定焦点的概念, 解释了“我不认为你不来”与“我不认为你故意不来”之间的区别。在归纳所有的双重否定结构后, 本文从语义管辖的角度对双重否定形式上的形成进行了解释。实践方面, 本文针对所有的“不 v_1 不 v_2 ”提出了相应的自动识别策略。基于规则, 设计了“不”的双重否定识别程序, 并用该程序对28033句语料进行了识别, 识别正确率为97.87%, 召回率约为93.10%, F1值约为95.43%。在程序构建过程中, 本文还补充了助动词词表与非叙实动词词表。该程序证明了基于规则进行双重否定识别的可能性, 是一次成功的尝试。然而, 这只是对“不 v_1 不 v_2 ”这一个格式进行的分析。未来, 本文将引入更多的双重否定结构, 全面地对双重否定进行考察, 以期真正地实现计算机对双重否定的自动识别, 将其应用于文本分析、问答系统等自然语言处理应用当中。

参考文献

- Horn, L. 1989. *A Nature History of Negation*. University of Chicago Press, Chicago, US, 311-312.
- Ladusaw, W. A. 1997. *Negation and polarity items*. In S. Lappin (ed.), *The Handbook of Contemporary Semantic Theory*. Blackwell Publishing Ltd, Oxford, UK, 321-341.
- 丁声树等. 2004. 现代汉语语法讲话. 商务印书馆, 北京, 200-202.
- 方绪军. 2017. “不是不 X ”、“不是没(有) X ”和“没(有)不 X ”. 语言科学, 16(05):511-521.
- 符达维. 1986. 对双重否定的几点探讨. 福建论坛(文史哲版), (6):78-81.
- 何爱晶. 2019. 反叙的非真值义否定和真值义肯定. 外语研究, 36(04):24-29.
- 郎桂青. 1989. 双重否定句表示肯定的条件. 语文研究, (1):28.
- 李新良, 王明华. 2015. 汉语动词的叙实性研究的应用前景. 对外汉语研究, (02):120-129.
- 鲁晓琨. 2004. 现代汉语基本助动词语义研究. 中国社会科学出版社, 北京.
- 吕叔湘. 1956. 中国语法要略. 商务印书馆, 北京.
- 王力. 1943. 中国现代语法. 商务印书馆, 北京.
- 王勇, 吕学强, 姬连春. 2014. 基于极性词典的中文微博情感分类. 计算机应用与软件, (01):40-43+132.
- 芜崧. 1987. 双重否定句的种类与功能. 荆州师专学报(哲社版), (3):52-57.
- 叶文曦. 2013. 否定和双重否定的多维度研究. 语言学研究, (2):20-31.
- 袁毓林. 1999. 并列结构的否定表达. 语言文字应用, (03):42-46.
- 袁毓林. 2000. 论否定句的焦点、预设和辖域歧义. 中国语文, (02):99-108+189.
- 郑贵友. 1989. 汉语“助动词”的研究刍议. 汉语学习, (06):23-27.