

形式语言与自动机理论课程的教学方法探索*

王伟¹ 杨翰霖^{2,*}孙鑫^{1,*}1. 哈尔滨工业大学(威海)计算机科学与技术学院 威海市 264209
2. 华中师范大学计算机学院 武汉市 430079

摘要 形式语言与自动机理论是计算机科学与技术专业本科教学中一门重要的专业基础课。该课程旨在使学生理解形式化描述的概念,掌握形式化描述的方法,培养抽象化思维的能力。然而,该课程内容高度抽象、知识量多、教学难度大,为学生深刻理解内在原理带来了不小的挑战。本文以形式语言与自动机理论课程中有穷状态自动机和正则文法的相互转换这一知识点为例,创造性地提出了一套包含两个模型的模型化教学方法,其中“5-1=4”模型用于有穷状态自动机向正则文法转换,“4+1=5”模型用于正则文法向有穷状态自动机转换。通过具体生动的案例,本文的方法能够对复杂抽象的知识点进行模型化和具体化,有助于教师以通俗易懂的语言引导学生理解这一知识点。

关键字 形式语言与自动机理论,模型化教学方法,右线性文法,左线性文法

Exploration of Teaching Methods for Formal Language and Automata Theory Course

Wei Wang¹ Hanlin Yang^{2,*}Xin Sun^{1,*}1College of Computer Science and Technology
Faculty of Computing Harbin Institute of Technology,
Harbin 150001, China;
wangwei_hitwh@126.com2School of Computer Science
Central China Normal University,
Wuhan 430079, Hubei Province
sunxintyc@163.com

Abstract—Formal language and automata theory are important foundational courses in undergraduate teaching of computer science and technology. This course aims to help students understand the concept of formal description, master the methods of formal description, and cultivate the ability of abstract thinking. However, the course content is highly abstract, with a large amount of knowledge and high teaching difficulty, which poses significant challenges for students to deeply understand the underlying principles. This article takes the knowledge point of the mutual transformation between finite state automata and regular grammars in the course of formal language and automata theory as an example, and creatively proposes a modeling teaching method that includes two models. The "5-1=4" model is used for the transformation from finite state automata to regular grammars, and the "4+1=5" model is used for the transformation from regular grammars to finite state automata. Through concrete and vivid cases. The method of this article can model and concretize complex and abstract knowledge points, which helps teachers guide students to understand this knowledge point in easy to understand language.

Keywords—Formal Language and Automata Theory, Modeling Teaching Methods, Right Linear Grammar, Left Linear Grammar

1 引言

形式语言与自动机理论课程是计算机科学与技术专业的必修课之一,该课程包含有穷状态自动机、正则表达式、上下文无关语法、图灵机和可判定性等内容,旨在培养学生的抽象思维能力、逻辑推理能力和计算表达能力。在这一课程的教学实践中,教师常常面临理论与实践脱节、教学资源不足、教学方法单一等方面的挑战,学生也常常对语言的形式化描述、自动机的构造、复杂问题的抽象求解等关键知识点感到

困惑。想要提升课程教学效果,除了高质量的教材和多样化的课件资源,还必须根据课程特点,采用创新性的教学方法,帮助学生克服知识点的抽象性带来的理解困难。

在形式语言与自动机理论课程中,有穷状态自动机和正则文法的相互转换包含两部分内容,即有穷状态自动机与右线性文法的相互转换以及有穷状态自动机与左线性文法的相互转换^[1-3]。这两部分内容可进一步划分为四个小部分:有穷状态自动机转换为右线性文法,右线性文法转换为有穷状态自动机,有穷状态自动机转换为左线性文法,以及左线性文法转换为有穷状态自动机^[4-5]。这四部分内容属于课程中比较抽象

***基金资助:** 本文得到哈尔滨工业大学教育教学改革研究项目 (IDEA10002233) 资助。

的知识点^[6],给教师授课和学生理解带来了挑战。因此,我们采用模型化的教学方法,将抽象的知识点建立模型,从而加深学生对知识点的理解^[7]。

针对形式语言中有穷状态自动机和正则文法相互转换的这个知识点,我们建立了“5-1=4”和“4+1=5”两个模型。根据形式语言与自动机理论,有穷状态自动机是由五元组组成 $M = (Q, \Sigma, \delta, q_0, q_f)$,其中 Q 为非空有穷状态集合, Σ 为输入字母表集合, δ 为状态转移函数集合, q_0 为开始状态, q_f 为终止状态集合。而文法是由四元组组成 $G = (V, T, P, S)$,其中 V 表示语法变量集合, T 表示终结符号集合, P 表示产生式集合, $S (S \in V)$ 表示开始符号。所以,对于有穷状态自动机向文法转换,我们采用的是“5-1=4”模型,对于文法向有穷状态自动机转换,我们采用的是“4+1=5”模型。当老师采用模型化教学方法,在课堂上提出这两个模型时,学生们会非常好奇地想老师揭开“5-1=4”和“4+1=5”两个模型的内容,进而提高学习兴趣,主动参与学习过程。

2 有穷状态自动机与右线性文法转化

本章首先讨论有穷状态自动机与右线性文法的相互转换。当需要将有穷状态自动机向右线性文法转换时,我们采用的是“5-1=4”的模型方法。因为当有穷状态自动机从开始状态 q_0 开始识别正则语言的时候,右线性文法也从开始符号 S 开始产生正则语言。于是,有穷状态自动机中的开始状态 q_0 和右线性文法中的开始符号 S 可以一一对应;有穷状态自动机的非空有穷状态集合 Q 可以和右线性文法中的语法变量集合 V 一一对应,有穷状态自动机的输入字母表集合 Σ 可以和右线性文法中的终止符号集合 T 一一对应;有穷状态自动机的状态转移函数集合 δ 可以和右线性文法中的产生式集合 P 一一对应。最终,有穷状态自动机的终止状态集合 q_f 就成了一个“多余的项”,可以“丢掉”,这就是“5-1=4”模型转换方法的含义。

例如:构造一个确定的有穷状态自动机(DFA),它接受的语言为 $\{x \mid x \in \{0, 1\}^*, \text{且当把 } x \text{ 看成二进制数时, } x \text{ 模 } 4 \text{ 与 } 3 \text{ 同余}\}$ 。

根据题目要求, x 模4与3同余, x 除以4的余数只有4种情况:0,1,2,3,故考虑用4个状态分别与这4个等价类相联系:

q_0 ——对应除以4余数为0的 x 组成的等价类;

q_1 ——对应除以4余数为1的 x 组成的等价类;

q_2 ——对应除以4余数为2的 x 组成的等价类;

q_3 ——对应除以4余数为3的 x 组成的等价类;

此外,由于要求 x 是非空的,所以还需要一个开始状态:

q_s ——对应 M 的开始状态。

下面逐一考虑在一个状态下读入下一个0字符或1字符的时候应该如何改变状态:

q_s ——在此状态下读入0时,有 $x=0$,所以应该进入状态 q_0 。读入1时,有 $x=1$,所以应该进入状态 q_1 。即: $\delta(q_s, 0) = q_0$; $\delta(q_s, 1) = q_1$ 。

q_0 ——能引导 M 到达此状态的 x 除以4余0,所以有: $x=4*n+0$ 。读入0时,引导 M 到达下一个状态的字符串为 x_0 , $x_0=2*(4*n+0)=4*2*n+0$ 。所以, $\delta(q_0, 0) = q_0$;而读入1时, M 到达下一个状态的字符串为 x_1 , $x_1=2*(4*n+0)+1=4*2*n+1$ 。所以, $\delta(q_0, 1) = q_1$;

q_1 ——能引导 M 到达此状态的 x 除以4余1,所以有: $x=4*n+1$ 。读入0时,引导 M 到达下一个状态的字符串为 x_0 , $x_0=2*(4*n+1)=4*2*n+2$ 。所以 $\delta(q_1, 0) = q_2$;而读入1时,引导 M 到达下一个状态的字符串为 x_1 , $x_1=2*(4*n+1)+1=4*2*n+3$ 。所以, $\delta(q_1, 1) = q_3$;

q_2 ——能引导 M 到达此状态的 x 除以4余2,所以, $x=4*n+2$ 。读入0时,引导 M 到达下一个状态的字符串为 x_0 , $x_0=2*(4*n+2)=4*2*n+4=4*(2*n+1)$ 。所以, $\delta(q_2, 0) = q_0$;而读入1时,引导 M 到达下一个状态的字符串为 x_1 , $x_1=2*(4*n+2)+1=4*2*n+4+1=4*(2*n+1)+1$ 。所以, $\delta(q_2, 1) = q_1$ 。

q_3 ——能引导 M 到达此状态的 x 除以4余3,所以有: $x=4*n+3$ 。读入0时,引导 M 到达下一个状态的字符串为 x_0 , $x_0=2*(4*n+3)=4*(2*n+1)+2$ 。所以, $\delta(q_3, 0) = q_2$;而读入1时,引导 M 到达下一个状态的字符串为 x_1 , $x_1=2*(4*n+3)+1=4*(2*n+1)+3$ 。所以, $\delta(q_3, 1) = q_3$ 。

按照上述分析,我们可以得到确定的有穷状态自动机DFA的状态转移图如图1所示:

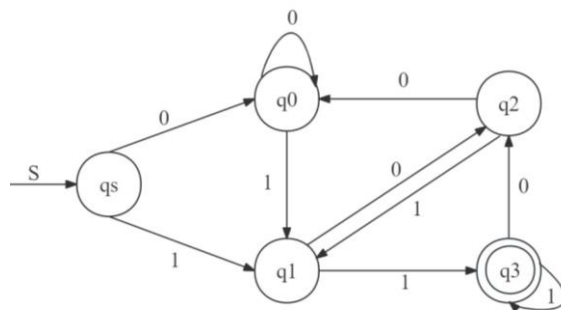


图1 模4与3同余的二进制数的DFA的状态转移图

下面逐一考虑有穷状态自动机和右线性文法的对应关系:

通过分析,有穷状态自动机识别语言的识别过程正好和右线性文法产生语言的推导过程相对应,有穷

状态自动机的开始状态 q_s 对应右线性文法的开始符号 q_s ，从开始状态 q_s 出发，即关注 q_s 状态的“出度”信息0或1，识别下一个字符0，对应从文法的开始符号 q_s 出发，生成下一个字符0；从开始状态 q_s 出发，识别下一个字符1，对应从文法的开始符号 q_s 出发，生成下一个字符1；所以，有穷状态自动机的转移函数 $\delta(q_s, 0) = q_0$ 和 $\delta(q_s, 1) = q_1$ 转换为右线性文法 $q_s \rightarrow 0q_0 \mid 1q_1$ 。

有穷状态自动机的状态 q_0 和 q_2 同上述 q_s 的处理过程类似，所以，有穷状态自动机的转移函数 $\delta(q_0, 0) = q_0$ 和 $\delta(q_0, 1) = q_1$ 转换右线性文法 $q_0 \rightarrow 0q_0 \mid 1q_1$ ；有穷状态自动机的转移函数 $\delta(q_2, 0) = q_0$ 和 $\delta(q_2, 1) = q_1$ 转换右线性文法 $q_2 \rightarrow 0q_0 \mid 1q_1$ 。

但是，由于有穷状态自动机的五元组比右线性文法的四元组多了一个终止状态集合 $q_f = \{q_3\}$ ，当有穷状态自动机转换为右线性文法时，需要扔掉多余的终止状态 q_3 ，从而实现“5-1=4”的模型，所以，状态 q_1 和 q_3 在识别下一个字符1的时候，到达了终止状态 q_3 ，此时，扔掉终止状态 q_3 ，有穷状态自动机的转移函数 $\delta(q_1, 1) = q_3 \in F$ 和 $\delta(q_3, 1) = q_3 \in F$ 转换成右线性文法： $q_1 \rightarrow 1$ 和 $q_3 \rightarrow 1$ 两个产生式。另外，有穷状态自动机的转移函数 $\delta(q_1, 0) = q_2$ 和 $\delta(q_1, 1) = q_3$ 转换右线性文法 $q_1 \rightarrow 0q_2 \mid 1q_3$ ；有穷状态自动机的转移函数 $\delta(q_3, 0) = q_2$ 和 $\delta(q_3, 1) = q_3$ 转换右线性文法 $q_3 \rightarrow 0q_2 \mid 1q_3$ 。

综上，与图1所给的DFA等价的右线性文法产生式集合如下图2所示：

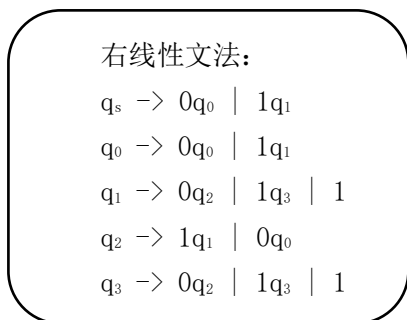


图 2 模 4 与 3 同余的二进制数的右线性文法产生式集合

当需要将右线性文法向有穷状态自动机转换时，我们则采用“4+1=5”的模型方法。根据有穷状态自动机和文法的联系，将右线性文法转换为有穷状态自动机时，我们将有穷状态自动机的终止状态集合增加进来，其他四元组如同上面描述的一一对应。关注右线性文法 $q_1 \rightarrow 1$ 和 $q_3 \rightarrow 1$ ，增加终止状态“Z”，将其转换为有穷状态自动机的转移函数 $\delta(q_1, 1) = \{Z\}$ 和 $\delta(q_3, 1) = \{Z\}$ ；右线性文法 $q_s \rightarrow 0q_0 \mid 1q_1$ 转换为有穷状态自动机的转移函数 $\delta(q_s, 0) = \{q_0\}$ 和 $\delta(q_s, 1) = \{q_1\}$ ；右线性文法 $q_0 \rightarrow 0q_0 \mid 1q_1$ ，转换为有穷

状态自动机的转移函数 $\delta(q_0, 0) = \{q_0\}$ 和 $\delta(q_0, 1) = \{q_1\}$ ；右线性文法 $q_2 \rightarrow 0q_0 \mid 1q_1$ 转换为有穷状态自动机的转移函数 $\delta(q_2, 0) = \{q_0\}$ 和 $\delta(q_2, 1) = \{q_1\}$ 。右线性文法 $q_1 \rightarrow 0q_2 \mid 1q_3$ 转换为有穷状态自动机的转移函数 $\delta(q_1, 0) = \{q_2\}$ 和 $\delta(q_1, 1) = \{q_3\}$ ；右线性文法 $q_3 \rightarrow 0q_2 \mid 1q_3$ 转换为有穷状态自动机的转移函数 $\delta(q_3, 0) = \{q_2\}$ 和 $\delta(q_3, 1) = \{q_3\}$ 。这样，利用“4+1=5”模型，就可以完美地完成右线性文法（如图2）向有穷状态自动机（如图3）的转换。

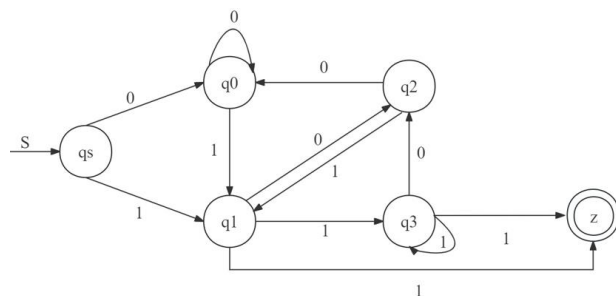


图 3 模 4 与 3 同余的二进制数的右线性文法转换的有穷状态自动机

由图可见，转换后图3的有穷状态自动机和图1的有穷状态自动机是等价的，它们表示的语言都是模4与3同余的二进制数。

3 有穷状态自动机与左线性文法转化

本章讨论有穷状态自动机与左线性文法的相互转换，首先考虑有穷状态自动机向左线性文法转换，需要用到的是“5-4=1”模型。如前文所述，有穷状态自动机识别正则语言的过程永远都是从开始状态 q_0 出发，而左线性文法产生语言的推导过程却和有穷状态自动机的识别过程恰恰相反，但左线性文法产生语言的归约过程正好与有穷状态自动机的识别过程一一对应。因此，左线性文法的开始符号就对应于有穷状态自动机的终止状态。将有穷状态自动机的开始状态 q_0 当作一个“多余的项”去掉以后，有穷状态自动机的非空有穷状态集合就可以和左线性文法中的语法变量集合一一对应，有穷状态自动机的输入字母表集合就可以和左线性文法中的终止符号集合一一对应；有穷状态自动机的状态转移函数集合就可以和左线性文法中的产生式集合一一对应；有穷状态自动机中的终止状态和左线性文法中的开始符号就可以一一对应。这就是用“5-1=4”模型实现有穷状态自动机向左线性文法转换的原理。

下面逐一考虑有穷状态自动机和左线性文法的对应关系：

通过分析，有穷状态自动机识别语言的处理过程正好和左线性文法产生语言的归约过程相对应，有穷状态自动机的终止状态 q_3 对应左线性文法的开始符

号 q_3 ，关注终止状态 q_3 ，即关注状态 q_3 的“入度”信息 1，有穷状态自动机的转移函数 $\delta(q_1, 1) = q_3 \in F$ 和 $\delta(q_3, 1) = q_3 \in F$ ，转换为左线性文法： $q_3 \rightarrow q_11 \mid q_31$ 。

有穷状态自动机状态 q_2 同上述 q_3 的处理过程类似，关注状态 q_2 的“入度”信息 0，所以，有穷状态自动机的转移函数 $\delta(q_1, 0) = q_2$ 和 $\delta(q_3, 0) = q_2$ ，转换为左线性文法： $q_2 \rightarrow q_10 \mid q_30$ 。

但是，由于有穷状态自动机的五元组比左线性文法的四元组多了一个开始状态 q_s ，当有穷状态自动机转换为左线性文法时，需要扔掉多余的开始状态 q_s ，从而实现“5-1=4”的模型，关注有穷状态自动机状态 q_s 的转移函数 $\delta(q_s, 0) = q_0$ 和 $\delta(q_s, 1) = q_1$ ，转换为左线性文法： $q_0 \rightarrow 0$ 和 $q_1 \rightarrow 1$ 。

另外，关注有穷状态自动机状态 q_0 ，即关注 q_0 状态的“入度”信息 0，有穷状态自动机状态的转移函数 $\delta(q_0, 0) = q_0$ 和 $\delta(q_2, 0) = q_0$ ，转换为左线性文法： $q_0 \rightarrow q_00 \mid q_20$ ；关注有穷状态自动机状态 q_1 ，即关注 q_0 状态的“入度”信息 1，有穷状态自动机状态的转移函数 $\delta(q_0, 1) = q_1$ 和 $\delta(q_2, 1) = q_1$ ，转换为左线性文法： $q_1 \rightarrow q_01 \mid q_21$ ；

综上，与图 1 所给的 DFA M 等价的左线性文法产生式集合如下图 4 所示：

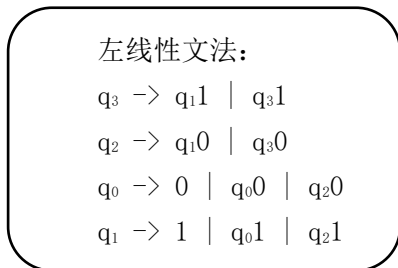


图 4 模 4 与 3 同余的二进制数的左线性文法产生式集合

将左线性文法向有穷状态自动机转换采用的则是“4+1=5”模型，也就是将有穷状态自动机的开始状态增加进来，其他四元组如同上面描述的一一对应。关注左线性文法 $q_0 \rightarrow 0$ 和 $q_1 \rightarrow 1$ ，增加开始状态“Z”，将其转换为有穷状态自动机的转移函数 $\delta(Z, 0) = q_0$ 和 $\delta(Z, 1) = q_1$ ；左线性文法 $q_3 \rightarrow q_11 \mid q_31$ 转换为有穷状态自动机的转移函数 $\delta(q_1, 1) = q_3$ 和 $\delta(q_3, 1) = q_3$ ；左线性文法 $q_2 \rightarrow q_10 \mid q_30$ 转换为有穷状态自动机的转移函数 $\delta(q_1, 0) = q_2$ 和 $\delta(q_3, 0) = q_2$ ；左线性文法 $q_0 \rightarrow q_00 \mid q_20$ 转换为有穷状态自动机的转移函数 $\delta(q_0, 0) = q_0$ 和 $\delta(q_2, 0) = q_0$ ；左线性文法 $q_1 \rightarrow q_01 \mid q_21$ 转换为有穷状态自动机的转移函数 $\delta(q_0, 1) = q_1$ 和 $\delta(q_2, 1) = q_1$ 。这样就完成了左线性文法(如图 4)向有穷状态自动机(如图 5)的转换，完美实现了“4+1=5”模型。

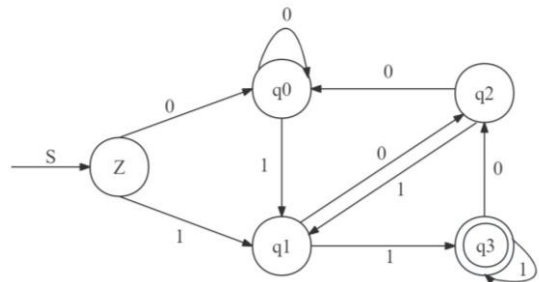


图 5 模 4 与 3 同余的二进制数的左线性文法转换的有穷状态自动机

由图可见，转换后图 5 的有穷状态自动机和图 1 的有穷状态自动机是相同的，它们表示的语言都是模 4 与 3 同余的二进制数。

通过以上所述的模型化方法对有穷状态自动机与文法相互转换的知识点进行讲解，学生能够很快地理解并掌握知识点。

4 教学效果

教学效果的评估包括在教务系统中，教学专家和教学对象学生的匿名评价。

表 1 教学专家评价结果表

序号	专家评价结果
1	备课认真，内容熟练；讲解透彻；学生听课状态好。
2	教师备课充分，讲解熟练，讲课以板书为主，语速适中，表达清楚，准确，条理清晰。
3	给出一个习题，让学生先做，然后讲解，讲解过程中结合上节课讲授的内容，边复习边解题。同时点名让学生回答问题，学生出席率高，同时注意听讲。备课充分、讲解熟练。
4	备课认真，讲解熟练，内容充实，条理清晰，重点比较突出。语言表达清楚准确。板书清晰。和学生有良好互动。

参与现场听课的教学专家主要在评价指标(权重:[1]满分:[100]):教学态度(权重:[0.2]满分:[20])、教学内容(权重:[0.3]满分:[30])、教学方法与手段(权重:[0.3]满分:[30])和课堂效果(权重:[0.2]满分:[20])这四个方面进行了评价，评价结果如表 1 所示。

此外，通过收集完成《形式语言与自动机理论》课程学习的部分学生评教反馈，反馈结果如表 2 所示，这些评价表明了学生对融入模型的教学方法的认可，也反映了他们对于课程内容和教学方法的满意度。

在教学实践中，始终以学生为主体，将简单易懂的模型引入教学设计，将简单易懂的模型引入教学设计，使理论知识的学习不在枯燥，学生能在和谐融洽的课堂氛围中学习，推进知识的掌握和调动了学生学

习的兴趣,使学生进入积极的思维状态,达到了良好的教学效果。

表2 学生评教反馈表

序号	学生评价反馈
1	老师讲课逻辑清晰,能将复杂的问题变得简单。
2	老师对学生课堂作业的批改总结认真,能及时,准确的发现同学们存在的问题并认真讲解,解决问题。
3	老师讲的很清楚,很明白。
4	作业量适度,随堂练习帮助掌握课堂知识。
5	讲课技巧娴熟,对知识体系的建构很清晰,与同学的沟通也很通畅。
6	结合例子,容易理解,收获颇多。
7	老师教的很好,学到了很多。

5 结束语

大学生的适应能力和创新能力很大程度上依赖于坚实的计算机理论基础和专业基础知识。只有扎实掌握专业的计算机理论基础,才能打好进行创造性研究的基础。本文针对形式语言与自动机理论课程中有穷状态自动机和正则文法相互转换的教学内容,探索性地提出了由“5-1=4”模型和“4+1=5”模型构成的模型化教学方法,用通俗易懂的语言将复杂抽象的知识点进行了模型化,加深和加快了学生对知识点的理解与掌握,从而达到了很好的教学效果。

参考文献

- [1] J. Zhang and Z. Qian. The Equivalent Conversion between Regular Grammar and Finite Automata[J]. Journal of Software Engineering and Applications, 2013, 6(1):33-37.
- [2] V. Laurikari. NFAs with Tagged Transitions, Their Conversion to Deterministic Automata and Application to Regular Expressions[C].
- [3] A. Brüggemann-Klein. Regular Expressions into Finite Automata[J]. Theoretical Computer Science, 1993, 120(2):197-213.
- [4] N. A. Zafar and F. Alsaade. Syntax-Tree Regular Expression Based DFA Formal Construction[J]. Intelligent Information Management, 2012, 4(4): 138-146.
- [5] S. Bhargava and G. N. Purohit. Construction of a Minimal Deterministic Finite Automaton from a Regular Expression[J]. International Journal of Computer Applications, 2011, 15(4):16-27.
- [6] S. Chakraborty, R. Grossi, K. Sadakane, and S. R. Satti. Succinct representation for (non) deterministic finite automata[C]. Presented at the 42nd Conference on Very Important Topics, New Delhi, India, 2016.
- [7] Z. Pei, H. Tian, J. Shu, C. Zhou, L. Li, and Y. Liu. Heuristic teaching reform of the course of formal language and automata (theory) based on the improvement of students' autonomous learning ability[J]. Computer Knowledge and Technology, 2020, 16(19):106-110.