Construction of Intersection of Nondeterministic Finite Automata using Z Notation

Nazir Ahmad Zafar, Nabeel Sabir, and Amir Ali

Abstract—Functionalities and control behavior are both primary requirements in design of a complex system. Automata theory plays an important role in modeling behavior of a system. Z is an ideal notation which is used for describing state space of a system and then defining operations over it. Consequently, an integration of automata and Z will be an effective tool for increasing modeling power for a complex system. Further, nondeterministic finite automata (NFA) may have different implementations and therefore it is needed to verify the transformation from diagrams to a code. If we describe formal specification of an NFA before implementing it, then confidence over transformation can be increased. In this paper, we have given a procedure for integrating NFA and Z. Complement of a special type of NFA is defined. Then union of two NFAs is formalized after defining their complements. Finally, formal construction of intersection of NFAs is described. The specification of this relationship is analyzed and validated using Z/EVES tool.

Keywords—Modeling, Nondeterministic finite automata, Z notation, Integration of approaches, Validation.

I. INTRODUCTION

In this paper, a relationship between automata and Z notation is investigated. Automata have various applications in many areas of computer science and engineering. Modeling control behavior, compiler constructions, modeling of finite state systems, defining a regular set of finite words are some of the traditional applications of automata. Automata have emerged with several modern applications, for example, optimization of logic based programs, verification of protocols [1] and human computer interaction. The Z notation [2] is a model oriented approach based on set theory and first order predicate logic. It is used for specifying the abstract data types and sequential programs. Z notation can also be used to define state of a system and then defining operations over it. The design of a complex system, not only requires the techniques for capturing functionalities but it also needs to model control behavior [3]. Functions over any of the systems can be decomposed in terms of operations and the constraints, and, hence, Z notation is an ideal application for this purpose. Control over a system can be viewed in terms of visual flows in between the system’s functions. Automata theory is very powerful thereat. Consequently, it requires an integration of automata and Z to increase modeling power for a complex system, which is one of the objectives of this research.

The design of a complex system, not only requires the techniques for capturing functionalities but it also needs to model control behavior [3]. Functions over any of the systems can be decomposed in terms of operations and constraints, and hence Z notation is an ideal application for this purpose. Control over a system can be viewed in terms of visual flows in between the system’s functions. Automata theory is very powerful thereat. Consequently, it requires an integration of automata and Z to increase modeling power for a complex system, which is one of the objectives of this research.

As we know that nondeterministic and deterministic finite automata are equivalent in power, in a sense, that if a language is recognized by one, it is also recognized by the other. Nondeterministic finite automata (NFA) are sometimes useful because constructing an NFA is easier than constructing deterministic finite automata (DFA). This is because the complexity of mathematical work is reduced using NFA. Further, many important properties in automata can be established easily using NFA. For example, to prove that a union or concatenation of regular languages is regular using NFA is much easier than using DFA [4]. This is another reason that NFA is selected to be integrated with Z notation.

Nondeterministic finite automata are abstract models of machines which can be represented using diagrams. These models can be used to perform computations on inputs by moving through a sequence of configurations. An NFA consumes the entire input of symbols and for each input symbol it transforms to a new state until all symbols have been consumed. If we are able to reach any of the accepting state after consuming whole input then the input is accepted.

At this level of integration, we have defined two NFAs and their complements are described. As we know that complement of an NFA is not well defined in general therefore, in this paper, we have proposed it only for particular cases. Union of the complemented NFAs is constructed and formal specification of their relationships is given. Finally intersection of the given NFAs is constructed by taking complement of the resultant. Formal specification of the whole
set of activities and the relationship between Z and automata is analyzed and validated using Z/EVES tool [5]. The main objectives of this paper are: (i) an integration of automata and Z notation by giving a syntactic and semantics relationship, (ii) linking constructs of NFA and Z notation such that both of these notations can be used in a cohesive way and (iii) reducing implementation issues of NFA.

Although integration of approaches is a well researched area [6], [7], [8], [9], [10], [11] but there does not exist much work on formalization of graphically based notations. The work [12], [13] of Dong et al. is close to ours in which they have integrated Object Z and Timed Automata for some aspects of automata. Another piece of good work is listed in [14], [15] in which R. L. Constable has given a constructive formalization of some important concepts of automata using Nuprl. Some work of interest is also reported in [16]. In [17], a combination of Z with statecharts is established. A relationship is investigated in between Z and Petri Nets in [18], [19]. An integration of UML and B is given in [20], [21].

In section 2, applications of formal methods are discussed. In section 3, applications and limitations of NFA are analyzed. Integration of NFA and Z is given in section 4. Conclusion and future work are discussed in section 5.

II. APPLICATIONS OF FORMAL METHODS

Formal Methods (FM) refers to mathematically rigorous techniques and tools for the specification, design and verification of software and hardware systems [22]. FM uses mathematical notations for writing specifications of the system to be developed. These mathematical notations are particularly derived from the area of set theory, discrete mathematics or graph theory. Thus formal specifications are mathematical expressions with well-defined syntax and semantics [23]. Once formal specifications are written, it can be refined into actually implemented system by a process of stepwise refinement. The validation and verification technique offered by FM is applied at each phase of the development process, which ensures the correctness and consistency by giving a high confidence in the system to be developed. Unlike traditional approaches, formal specification uses mathematical notations those have same interpretation throughout the globe [24]. The use of mathematics in writing specifications helps having deeper insight of a system to be developed and provides an excellent medium for its modeling.

One of the major limitations of traditional approaches is that they lack the ability to prove the specifications. The errors and inconsistencies are hidden behind graphical requirements specifications [25], and are usually identified only during implementation and testing phases. Implementation errors are difficult and costly to fix [26]. On the other hand, the mathematical nature of specifications enables to carry out proves. The worth of conducting proves is that it explores the entire state space of the system. FM makes it possible to prove and analyze certain properties of the system during early stages of the development process so that errors in the requirement specifications can be identified and removed. Studies have suggested that FM have tremendous potential for improving the clarity and precision of requirements specification, and in finding important and subtle errors [27]. Therefore FM is an emerging and future technology with its focus to develop high quality and reliable systems [28].

There are several ways in which formal methods may be classified. One frequently-made distinction is between model oriented and property oriented methods [29]. Model oriented methods are used to construct a model of a system’s behavior. Property oriented methods are used to describe software in terms of a set of properties, or constraints, that must be satisfied. The Z notation [30] is a model oriented approach, which is based on set theory and first order predicate logic. Although formal methods are being applied successfully in many research areas of computer science and engineering but at the current stage of development, it requires an integration of formal and informal approaches.

III. LIMITATIONS OF NONDETERMINISTIC FINITE AUTOMATA

Nondeterministic finite automata are abstract models based on mathematical notations which can be represented using diagrams. These models can be used to perform computations on inputs by moving through a sequence of configurations. If we are able to reach any of the accepting state by using a series of computation then the input is accepted.

An extension of NFA is the NFA with ε (epsilon, a null string) defined by NFA ∪ {ε} in which the transition function is allowed to a new state without consuming any input symbol. For example, it can move from state A to the next state B by reading ε (without consuming any input symbol) and it creates an ambiguity. To remove this ambiguity, it is more understandable to talk of a set of possible states in which the transition function enters. We have supposed that our nondeterministic finite automatons is based on the set of alphabets in addition to the epsilon symbol and is denoted by NFA. The addition of epsilon, in the set of alphabets of NFA, increases more complexity in conversion from NFA to DFA.

Further, diagrams in NFA have been difficult to be used except the very trivial cases, which is one of the major issues in representation of NFA diagrammatically. It is a fact that a given NFA may have different implementation methodologies and consequently its time and space complexity may vary for different implementation, which is another issue in modeling using NFA. Further, automata cannot be used for defining functions and constraints and consequently it is not possible to model a complete system by this single approach. As a result, its integration will be very useful with Z notation increasing modeling power for a complex system. If we are able to formalize this relationship, then it would be very useful tool not only at academic but at an industrial level as well. This is because the study of automata in class room, after this integration, will increase clarity of concepts. A formal linkage between these approaches is given in the next section.
IV. FORMAL CONSTRUCTION OF INTERSECTION OF NFAS

A formal construction of intersection of two NFAs is demonstrated. An NFA is a five tuple \((Q, \Sigma, \delta, q_0, F)\), where (i) \(Q\) is a finite non-empty set of states, (ii) \(\Sigma\) is a finite set of alphabets, (iii) \(\delta\) is a transition function, (iv) \(q_0\) is the initial state and (v) \(F\) is a finite set of final states.

The above 5-tuple is an NFA because for each state \(q_i\), and for every alphabet \(a\), there is a set of states \(s\), such that \(\delta(q_i, a) = s\). The definitions used here are based on well-known books on Automata and Computation Theory [31], [32].

Let us suppose that \(L\) is a language over a set of alphabets \(\Sigma\), and is accepted by a machine NFA = \((Q, \Sigma, \delta, q_0, F)\). We define complement of language \(L\) as the language of all the strings that are not words in \(L\). Mathematically we define as:

\[\text{strings} \in \text{complement of } L = \{s, \text{ s is a string based on set of alphabets of } \Sigma \mid s \not\in L\} \]

In order to take complement of deterministic automata we simply swap the accepting and non-accepting states but this is not true in case of an NFA. For example, the NFA1 in Fig. 1 accepts all strings of length greater than or equal to 2. The NFA2, in Fig., is obtained by swapping the final and non-final states of NFA1 which accepts all the strings of any length and hence it is not complement of NFA1. If we suppose that our NFA accepts all the strings of length \(n\) and no self loop is allowed on a state then we can take complement of it by simply swapping the final and non-final states. In this paper such NFAs for constructing intersection are supposed.

Let NFA1 and NFA2 be two NFAs accepting the languages \(L_1\) and \(L_2\) respectively. We construct the NFAs accepting the languages \(L_1^c\) and \(L_2^c\). Then a new NFA will be designed accepting all the words of \(L_1^c\) and \(L_2^c\). By deMorgan’s Law: \(L_1 \cap L_2 = (L_1^c \cup L_2^c)^c\) is the intersection of two given languages for which a new NFA is required.

A. Complementing First NFA

The first non-deterministic finite automata consists of 5-tuple \((Q_1, \Sigma_1, \delta_1, q_0, F_1)\), where \(Q_1\) and \(\Sigma_1\) are represented as \(Q\) and \(\Sigma\) respectively.

\([Q, \Sigma]\)

In modeling using sets in \(Z\), we do not impose any restriction upon the number of elements and a high level of abstraction is supposed. As a consequent, our \(Q\) and \(\Sigma\) are sets over which we cannot define any operation, for example, cardinality to know the number of elements in a set.

To describe a set of states, a variable \(states1\) is introduced. Since a given state \(q\) is of type \(Q\) therefore \(states1\) is of type of power set of \(Q\). Similarly, a set of alphabets \(alphabets1\) is of type of power set of \(\Sigma\). As we know that \(\delta_1\) relation is a function because for each input \((q_1, a)\), where \(q_1\) is a state and \(a\) is in set of \(alphabets1\) there must be a unique output \(s\) of type \(PQ\), which is image of \((q_1, a)\) under the transition function \(\delta_1\). Hence we can declare \(\delta_1\) as, \(\delta_1: Q \times \Sigma \rightarrow PQ\). The initial state \(q_01\) is of type \(Q\). The \(F1\), set of final states, is represented by \(finals1\) and is of type of power set of \(Q\).

The schema structure is used here for composition of these objects because it is very powerful at abstract level of specification. All of the components of NFA1 are encapsulated and put in the schema named as NonDeterministic1. We also need to compute the set of all the strings generated by a given alphabets which is declared as: \(Strings := seq \Sigma\).

\[\wedge \text{NonDeterministic1}
\rightarrow states1: \Pi Q
\rightarrow alphabets1: \Pi \Sigma
\rightarrow apsi1: Sigma
\rightarrow delta1: Q \times \Sigma \rightarrow PQ
\rightarrow q01: Q
\rightarrow finals1: \Pi Q
\rightarrow strings1: \Pi Strings
\rightarrow apsi1 \in alphabets1
\rightarrow q01 \in states1
\rightarrow finals1 \subset states1
\rightarrow \forall q1, q2: Q, a: Sigma | q1 \in states1 \land q2 \in states1 \land a \in alphabets1
\rightarrow \exists s1, s2 \Pi Q | s1 \in finals1 \land s2 \in finals1
\rightarrow f(q1, a, s1) \in delta1
\rightarrow f(q2, a, s2) \in delta1 \Rightarrow (q1, a) = (q2, a) \Rightarrow s1 = s2
\rightarrow \text{Ast. Strings: } s \in strings1 \Rightarrow \text{ran } s \in alphabets1\]

**Invariants:** (i) The empty string \(apsi1\) is a member of set of alphabets1. (ii) The initial state \(q01\) must be an element of set of states1. (iii) The set of final states is a subset of set of total states. (iv) For each \((q, a)\), where \(q\) is an element of states1 and \(a\) is a member of alphabets1 there is a unique set of states \(s\) such that: \(delta1(q, a) = s\). (v) Any string given as input to an NFA must be based on the set of alphabets of the same NFA.

After designing NFA1, we need to take its complement. For this purpose a schema ComplementOfNFA1 is defined. It contains NFA1 and some other components in addition to it, which are required in defining complement of an NFA. A relation is defined between the NFA and its complement.

\[\wedge \text{ComplementOfNFA1}
\rightarrow \forall \text{NonDeterministic1}
\rightarrow \exists \text{states1c: } \Pi Q
\rightarrow \text{alphabets1c: } \Pi \Sigma
\rightarrow \text{apsi1c: } Sigma
\rightarrow \text{delta1c: } Q \times \Sigma \rightarrow PQ
\rightarrow q01c: Q
\rightarrow \text{null1c: } \Pi Q
\rightarrow \text{finals1c: } \Pi Q
\rightarrow \text{strings1c: } \Pi Strings
\rightarrow \text{states1c} = \text{states1} \land \text{alphabets1c} = \text{alphabets1}
\rightarrow \text{apsi1c} = \text{apsi1} \land \text{q01c} = \text{q01} \land \text{null1c} = \text{null1}
Invariants: (i) The set of states and alphabets in the given NFA and its complement are same. (ii) The null strings, initial states and the sets of dead states in the NFAs and its complement are identical. (iii) The set of final states in complemented NFA is equal to difference of the sets of states and final states in complemented NFA. (iv) The sets of strings generated by both, NFA and its complement, are equal because these are based on the same alphabets. (v) The empty string is a member of set of alphabets. (vi) The initial state must be an element of set of final states. (vii) The set of final states is a subset of set of total states.

B. Complementing the Second NFA

Let NFA2 = (Q2, Σ2, δ2, q02, F2) be a 5-tuple where all components have the same meaning as defined in case of NFA1. The NFA2 is represented by Non-deterministic2 as given below and invariants over it are defined similar to NFA1.

Invariants: (i) The set of alphabets in NFA1 and NFA2 are also a schema and represented as ComplementOfNFA2 as below. The invariants over it are identified and defined as predicates in the second part of the schema. The informal description of the invariants is not given because it is nothing but a repetition of properties as we defined in the schema ComplementOfNFA1.
alphabets are same for NFA1 and NFA2, therefore the possible strings in both of the automata are also same. (iv) The empty string apsi is a member of set of alphabets. (v) The initial state q0 must be an element of set of states. (vi) The set of final states is a subset of set of total states. (vii) For each (q, a) there is a unique set of states s such that: deltac(q, a) = s. (viii) Any string given as input to an NFA must be based on the set of alphabets of the same NFA. The empty string apsi is a member of the set of states of the resultant NFA. (ix) The set of states of the resultant NFA is equal to the union of the sets of states of NFA1, NFA2 and a set consisting of a single element q0. The q0 is a new state introduced at the time of union and is initial state in the resultant NFA. (x) The set of alphabets (including null string) in the resultant NFA is same as in the NFA1 or NFA2. (xi) For any state q of states and an element a of the alphabets of the NFA, the transition function holds the, (a) delta(q, a) = delta1(q, a), if q ∈ states1, (b) delta(q, a) = delta2(q, a), if q ∈ states2, (c) delta(q, a) = {q1, q2}, if q = q1 and a = apsi, (d) delta(q, a) = ∅, if q = q0 and a ≠ apsi. (xii) The set of dead states of the resultant NFA is equal to the union of the sets of dead states of NFA1 and NFA2. (xiii) The set of final states of the resultant NFA is equal to the union of the sets of final states of NFA1 and NFA2. (xiv) As the alphabets are same in the given NFA and its complement, therefore the possible strings in these automata are also same.

D. Construction of Intersection

A formal construction of NFA accepting the language \((L1 \cup L2)\) is done in the sub-section 4.3. Now if we construct complement of \((L1 \cup L2)\) then the resultant automata will accept the language which is intersection of \(L1 \) and \(L2\). The schema is represented by \textit{ComplementOfUnionOfNFA}s as given below, which completes this formal construction.

\[ \text{ComplementOfUnionOfNFA}s \]

\[ \rightarrow \text{NFA1} \cup \text{NFA2} \]

\[ \rightarrow \text{states} : \Pi Q \]

\[ \rightarrow \text{alphabets} : \Pi \Sigma \]

\[ \rightarrow \text{apsi} : \Sigma \]

\[ \rightarrow \text{deltac} : \Pi \Sigma \varnothing \Pi Q \]

\[ \rightarrow \text{nullc} : \Pi Q \]

\[ \rightarrow \text{finalsc} : \Pi Q \]

\[ \rightarrow \text{stringsc} : \Pi \text{Strings} \]

\[ \top \]

\[ \rightarrow \text{statesc} = \text{states} \setminus \text{alphabetsc} = \text{alphabets} \]

\[ \rightarrow \text{apsi} = \text{apsi} \setminus \text{q0c} = \varnothing \]

\[ \rightarrow \text{nullc} = \varnothing \]

\[ \rightarrow \text{finalsc} = \text{finalsc} \setminus \text{finals} \]

\[ \rightarrow \text{stringsc} = \text{strings} \]

\[ \rightarrow \text{apsic} \setminus \text{alphabetsc} \]

\[ \rightarrow \text{q0c} \setminus \text{statesc} \]

\[ \rightarrow \text{finalsc} = \emptyset \]

\[ \rightarrow \text{q1, q2} : Q: a: \Sigma | q1 \in \text{statesc} \setminus \text{q0c} \wedge q2 \in \text{statesc} \setminus \text{apsic} \]

\[ \rightarrow \text{Es} : s1 \in \text{statesc} \setminus \text{q0c} \wedge s2 \in \text{statesc} \setminus \text{q0c} \wedge \text{deltec} \rightarrow \text{f}(q1, a, s1) \wedge \text{deltec} \rightarrow \text{f}(q2, a, s2) \]
REFERENCES


