



ORIGINAL ARTICLE

Confidence value prediction of DNA sequencing with Petri net model

Raed I. Hamed ^{a,*}, Syed I. Ahson ^b

^a Department of Computer Science, Faculty of Computer, University of Anbar, Al-Anbar, Iraq

^b Patna University, Ashok Rajpath, Patna 800 005, India

Received 4 December 2010; accepted 30 April 2011

Available online 6 May 2011

KEYWORDS

Fuzzy Petri net;
Fuzzy logic;
DNA sequencing;
Modeling and simulation

Abstract In this paper, a fuzzy Petri net (FPN) approach to modeling fuzzy rule-based reasoning is proposed to determining confidence values for bases called in DNA sequencing. The proposed approach is to bring DNA bases-called within the framework of a powerful modeling tool FPN. The three input features in our fuzzy model-the height, the peakness, and the spacing of the first most likely candidate (the base called) and the peakness and height for the second likely candidate can be formulated as uncertain fuzzy tokens to determines the confidence values. The FPN components and functions are mapped from the different type of fuzzy operators of If-parts and Then-parts in fuzzy rules. The validation was achieved by comparing the results obtained with the FPN model and fuzzy logic using the MATLAB Toolbox; both methods have the same reasoning outcomes. Our experimental results suggest that the proposed models, can achieve the confidence values that matches, of available software.

© 2011 King Saud University. Production and hosting by Elsevier B.V. All rights reserved.

1. Introduction

A major challenge of modeling biological systems is that conventional methods based on physical and chemical principles require data that is difficult to accurately and consistently obtain using either conventional biochemical or high throughput technologies, which typically yield noisy, semi-quantitative data (often in terms of a ratio rather than a physical quantity) (Fitch and Sokhansanj, 2000). Various kinds of models have been studied to express biological systems such as differential equations (Novak et al., 1998; Chen et al., 1999), Boolean networks (Liang et al., 1998; Akutsu et al., 1999), Petri Nets (Matsuno et al., 2000, 2003; Fujita et al., 2004), Bayesian networks (Husmeier, 2003) and artificial neural networks (Vohradsky, 2002). The above-mentioned papers are dedicated to the applications of different methods to genetic networks

* Corresponding author.

E-mail addresses: raed.inf@gmail.com (R.I. Hamed), drsiahsan@yahoo.com (S.I. Ahson).

1319-1578 © 2011 King Saud University. Production and hosting by Elsevier B.V. All rights reserved.

Peer review under responsibility of King Saud University.

doi:10.1016/j.jksuci.2011.05.004



and show that these methods are suitable to model special molecular biological systems.

The explosion in the number of genomic datasets generated with tools such as high throughput DNA sequencing machines and DNA microarrays has created a critical need for resources that facilitate the interpretation of large-scale biological data. A new mathematics and novel methodologies are required to contribute to the conceptual or complex theoretical framework in which biologists study organisms. One such tool is fuzzy logic that can satisfy the need for a conceptual framework and provide a systematic and unbiased way to perform this transformation.

Since fuzzy logic system has been successfully used in various applications and large-scale complex systems exist everywhere in our society, this encourages its complex applications that have large amounts of rules and require real-time responses. Petri net theory and fuzzy logic exhibit a graphical and mathematical formalism to model, and simulate the biological systems. Fuzzy Petri net (FPN) is a successful tool for describing and studying information systems. Incorporating the fuzzy logic with Fuzzy Petri nets has been widely used to deal with fuzzy knowledge representation and reasoning (Raed and Ahson, 2010; Lukas and Ralf, 2008; Chen et al., 1990; Raed and Ahson, 2010). It has also proved to be a powerful representation method for the reasoning of a rule-based system. Such an approach is appropriate for the case where a state of the modeled system corresponds to a marking of the associated FPN. The motivation for the development of the FPN model is to fuse the benefits of fuzzy logic (i.e. effectively manage uncertain or corrupted inputs, natural linguistic structure, etc.) with FPN techniques.

The advantages of using FPNs in fuzzy rule-based reasoning systems include (Chen et al., 1990; Bostan-Korpeoglu and Yazici, 2007): (1) the graphical representation of FPNs model can help to visualize the inference states and modify fuzzy rule bases; (2) the analytic capability, which can express the dynamic behavior of fuzzy rule-based reasoning. Evaluation of markings is used to simulate the dynamic behavior of the system. The explanation of how to reach conclusions is expressed through the movements of tokens in FPNs (Bostan-Korpeoglu and Yazici, 2007). The field of fuzzy Petri nets may have an important impact in understanding how biological systems work, giving at the same time a way to describe, manipulate, and analyse them. Given the complexity of systems begin studied, biologists need a modeling and simulation framework to make sense of large-scale data and intelligently design traditional bench-top experiments that provide the most biological insight. Following its first application of modeling the dynamic biological systems (Lukas and Ralf, 2008), fuzzy Petri nets as a new tool for predicting the confidence values for each base called in DNA sequencing are investigated in this paper.

The method presented in this paper develops a fuzzy Petri net model that can predict the confidence values for each base called in DNA sequencing. This approach here utilizes the information that is gathered at the base, for more information (see Resson et al., 2005). This includes information on the height, peakness, and spacing of the base under consideration and the next likely base. In order to validate our approach, we compare our method to the fuzzy logic toolbox of MATLAB. The comparison is made in terms of the confidence value measure of the bases called in DNA sequencing. The similarity that

we have discovered is that they both have the same conclusions.

The organization of this paper is as follows: in Section 2, fuzzy Petri nets are described. In Section 3, the formulation of fuzzy sets and linguistic variables are presented. In Section 4, we explain the details of methods of modeling DNA bases called together with fuzzy Petri net as a new tool for modeling DNA bases called are investigated in this paper. Section 5 describes the experimental and simulation results. Finally, we presented the conclusions of our model in Section 6.

2. Fuzzy Petri nets

2.1. Formal definition of fuzzy Petri nets

Chen et al. (1990) presented a new knowledge representation by means of fuzzy Petri nets (FPN). A fuzzy Petri net model allows a structural representation of knowledge and has got a systematic procedure for supporting the fuzzy reasoning process (Raed and Ahson, 2010). Formally, a fuzzy Petri net structure is defined as follows (Chen et al., 1990):

The tuple $FPN = (P, T, D, I, O, F, \alpha, \beta)$ is called a fuzzy Petri net if:

1. $P = \{p_1, p_2, \dots, p_n\}$ is a finite set of places, corresponding to the propositions of FPRs;
2. $T = \{t_1, t_2, \dots, t_n\}$ is a finite set of transitions, $P \cap T = \emptyset$, corresponding to the execution of FPRs;
3. $D = \{d_1, d_2, \dots, d_n\}$ is a finite set of propositions of FPRs. $P \cap T \cap D = \emptyset$, $|P| = |D|$, d_i ($i = 1, 2, \dots, n$) denotes the proposition that interprets fuzzy linguistic variables, such as: very low, low, Inorm, eug, hnorm, as in our model;
4. $I: P \times T \rightarrow \{0, 1\}$ is an $n \times m$ input incidence matrix defining the directed arcs from propositions (P) to rules (T). $I(p_i, t_j) = 1$, if there is a directed arc from p_i to t_j , and $I(p_i, t_j) = 0$; if there is no directed arcs from p_i to t_j , for $i = 1, 2, \dots, n$, and $j = 1, 2, \dots, m$.
5. $O: P \times T \rightarrow \{0, 1\}$ is an $n \times m$ is an output incidence matrix defining the directed arcs from rules to propositions. $O(p_i, t_j) = 1$, if there is a directed arc from t_j to p_i , and $O(p_i, t_j) = 0$; if there is no directed arcs from t_j to p_i , for $i = 1, 2, \dots, n$, and $j = 1, 2, \dots, m$.
6. $F = \{\mu_1, \mu_2, \dots, \mu_m\}$ where μ_i denotes the certainty factor (CF = μ_i) of R_i , which indicates the reliability of the rule R_i , and $\mu_i \in [0, 1]$;
7. $\alpha: P \rightarrow [0, 1]$ is the function which assigns a token value between zero and one to each place;
8. $\beta: P \rightarrow D$ is an association function, a bijective mapping from a set of places to a set of propositions.

Moreover, this model can be enhanced by including a function $Th: T \rightarrow [0, 1]$ which assigns a threshold value $Th(t_j) = \lambda_j \in [0, 1]$ to each transition t_j , where $j = 1, \dots, m$. Further more, a transition is enabled and can be fired in FPN models when values of tokens in all input places of the transition are greater than its threshold.

A token value in place $p_i \in P$ is denoted by $\alpha(p_i) \in [0, 1]$, $\alpha(p_i) = y_i$, $y_i \in [0, 1]$ and $\beta(p_i) = d_i$. This states that the degree of the truth of proposition d_i is y_i . A transition t_i is enabled if $\forall p_i \in I(t_i)$, $y_i > 0$. If this transition t_i is fired, tokens are removed from input places $I(t_i)$ and a token is deposited onto

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات