

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/57983> holds various files of this Leiden University dissertation.

Author: Cai, F.

Title: Fuzzy systems and unsupervised computing: exploration of applications in biology

Issue Date: 2017/12/12

Chapter 6

Conclusions and Outlook

The main contributions of this thesis, underpinning bioinformatics studies for which the application of image analysis is important, and are summarized as:

1. Uncover common fuzzy principals that applied across many unsupervised computing systems, and highlight the features that adaptively address biological questions.
2. Dedicated pipelined methodologies are proposed for formalizing, measuring and modeling the uncertainty and diversity in bioinformatics studies.
3. By means of extensions and derivatives of soft computing, this work enlarges the depth of biological investigation, and engages width of dimensions as well.

6.1. Conclusions

In this thesis, we focus on exploring solutions for bio-imaging data analysis using fuzzy-computing paradigms. To gain further insight into the solutions, experiments are conducted on several publicly available datasets representing multiple disciplines and practical problems. The performance of solutions is then examined comprehensively and carefully using sorts of convincing index and evaluation strategies. The case study in this thesis further demonstrates the possibility of producing objective understandings of the identity of characteristics in proteins/DNA that are distinct/shared among different groups in phenotypic measurements. The conclusions of the chapters in this thesis are presented in the next sections.

Chapter 2 Biological Image Background Correction

In this chapter, we challenge a common interference in bio-imaging data, known as the inhomogeneity of illumination in the background. This effect leads to information contamination and loss of accuracy in both qualitative and quantitative work. A robust approach, the Dam-Constrained Background Correction (DCBC) is proposed to reduce the deviations before image datasets are subject to further analysis track. The fuzzy membership function in this approach, together with the rough set constraint, innovates constructing a morphological “dam” in the image, and thereby prevents over-segmentation of the background. The experimental results demonstrate the potential of the application on bright-field and fluorescence images, since further analysis will no longer be hampered by undesired vignetting.

Chapter 3 Feature Selection Strategy in Region of Interest Mask

In this chapter, we present a novel feature selection strategy, namely the Fuzzy Criteria in Multi Objective Feature Selection Algorithm (FC-MOFSA) that selects potential predictors in the feature pool. In FC-MOFSA, a fuzzy-defined entropy measurement is used in filter-approach, meanwhile the fuzzy criteria referred to as Correlation Membership Measurement (CMM), is employed as a cost-function in wrapping-approach in the feature selection process. The proposed CMM criterion highlights the essentials in sparse and skewed datasets clustering procedure by ambiguously estimating the difference between all pairs of points in different clusters. Aiming at converting digital signals to numerical feature descriptors, this strategy further optimizes the tasks of handling data in large volumes, and data clarification without jeopardizing the quality of information.

Chapter 4 Unsupervised Information Classification and Analysis

In this chapter, we present a dedicated classification algorithm, the Rough Fuzzy C-Means and Particle Swarm Optimization (RFCM-PSO). In the information clustering process, the concept employed in the RFCM handles with uncertain, vague and sparse data processing. PSO is able to optimize searching procedure and avoids the results from being trapped into local optimum. To demonstrate the potential of this method, a number of state-of-the-art algorithms are compared. The extensive numerical analysis and reported results indicate the good performance in revealing hidden patterns and exploring relations or attributes between clusters.

Chapter 5 A Systematic Study on One Dimensional Gel Electrophoresis Image Analysis

In this chapter, we present the feasibility and capability of proposed fuzzy-system and unsupervised computing based methodologies pipeline in protein phenotypic quantification and subsequent classification. The mixtures of protein in a cell lysate can be separated, visualized and analyzed by classical one dimensional (1D) gel electrophoresis. Subsequently, the resulting gel images can consist of several vertical lanes (number of wells in which the protein samples were loaded), and a number of horizontal bands (corresponding to proteins or fragments thereof). The data reflects the amounts and characteristics of individual proteinaceous components. This case study is one of the first attempts to reveal the intrinsic properties of bands from which cell subtype they are originated.

The work in this chapter, compared to the conventional phenotypic analysis system, improves gel data analysis into a four-stage strategy. Dedicated algorithms ensure

that the measurements are summarized into a matrix of primary representing the information of patterns. Therefore, the employment of fuzzy systems plays an important role in unambiguously uncovering patterns hidden in the data, yet yielding precise evaluation. In terms of unsupervised computing, it is also objective that the correlation of data clusters is qualitatively assessed.

6.2. Outlook

This thesis has explored and discussed solutions for addressing biological questions by building fuzzy logic-based and rule-based systems. The results are promising, but should be refined and expanded in further studies. In the next paragraphs we elucidate what aspects need additional attention.

Data acquisition

Modern bio-imaging systems are often synchronized with other computer or human controlled equipment and software. This on-the-fly system allows auto- or semi-automated adaption for experimental preparation, e.g. focusing lens, position of specimen, signal acquisition and transformation. However, the quality of data obtained via this adaption in controlling system is always subjected from heuristic mechanisms of objects and intrinsic algorithms. This mechanism can introduce either frequency noise (cf. Chapter 5) or out-of-focus background illumination (cf. Chapter 2). Additionally, based upon biological experimental design, the resolution of acquiring data is often sacrificed in exchange of a better image quality, or acquisition speed. These issues are hardware related and could be significantly improved by the development of imaging techniques.

Methodology design

The fuzzy systems, involved in soft computing paradigms, has been successfully employed to interpret biological questions. There are several reasons accounting for the trend of its increasing use in bio-sciences, and a few aspects could also be further improved. The most significant reason for accelerating a bioinformatics study is that fuzzy theory permits approximation instead of high-precision which sometimes arrives at an expensive computational budget. Optimization algorithms could play a very important role in helping fuzzy systems search the “best available” solutions from sparse and skewed data. As an ingredient from fuzzy rule-based model, adequate optimization strategies might further reduce computational overhead. Second, but not less important, is that fuzzy theory reduces complexity and thereby simplifies system modeling when empirical knowledge and behavior are not known. The concept of fuzzy systems offers a simple but effective way to arrive at a

definitive conclusion on the basis of ambiguous, imprecise, noisy or missing input information. Moreover, this mechanism can benefit most from new advances in deep learning. For instance, if a few ground truth (labels) are known, then a sophisticated weakly supervised fuzzy systems might have a significant improvement on accuracy and efficiency compared to the unsupervised ones. Another possibility to improve the proposed system is that the methodologies of measurements should meet biological description. This means a question-driven experiment design is strongly recommended.

To sum up, fuzzy logic and unsupervised computing based systems, constitute a very promising analysis direction in the field of bioinformatics. The experiments in this thesis carefully and comprehensively investigate how this proposed system could be applied to biological questions. We have accomplished a good start in this direction and further progress from research efforts can potentially be anticipated.