Evolving Hypernetwork Classifiers for microRNA Expression Profile Analysis

Sun Kim*, Soo-Jin Kim*, and Byoung-Tak Zhang

Abstract— **High-throughput microarrays inform us on different outlooks of the molecular mechanisms underlying the function of cells and organisms. While computational analysis for the microarrays show good performance, it is still difficult to infer modules of multiple co-regulated genes. Here, we present a novel classification method to identify the gene modules associated with cancers from microarray data. The proposed approach is based on 'hypernetworks', a hypergraph model consisting of vertices and weighted hyperedges. The hypernetwork model is inspired by biological networks and its learning process is suitable for identifying interacting gene modules. Applied to the analysis of microRNA (miRNA) expression profiles on multiple human cancers, the hypernetwork classifiers identified cancer-related miRNA modules. The results show that our method performs better than decision trees and naive Bayes. The biological meaning of the discovered miRNA modules has been examined by literature search.**

I. INTRODUCTION

High-throughput gene expression profiling has been used as one of the most important and powerful approaches in biomedical research [1], [2], [3]. While traditional methods only allow one or a few genes to be examined at once, the microarray techniques measure the expression level of thousands of genes or potentially the whole-genome scale simultaneously. This has allowed to make a systemic analysis of the particular disease mechanism such as cancers at the molecular level. Recently, the analysis of gene expression data at the level of biological modules, rather than individual genes, is recognized as important for understanding the cancer regulatory mechanisms [4]. This analysis has a biologically important meaning that the joint regulation genes can detect significant expression changes even in the case where the expression of individual genes are not meaningful. However, it is difficult to infer cancer-related pathways by inducing modules of co-regulated genes [5].

Finding cancer-related genes from the microarray analysis is typically based on the correlations between each gene and particular samples. The highly correlated genes have properties that their expression patterns are separated into two distinct parts corresponding to cancer and normal tissues, hence it became a popular method to find peculiar expression patterns between different types of diseases. Nevertheless, they can be inappropriate for systemic analysis because they do not identify synergistically interacting genes.

S.-J. Kim is with the Center for Bioinformatics Technology, Graduate Program in Bioinformatics, Seoul National University, Seoul 151-742, Korea (email: sjkim@bi.snu.ac.kr). S. Kim and B.-T. Zhang are with the School of Computer Science and Engineering, Seoul National University, Seoul 151-744, Korea (email: skim@bi.snu.ac.kr; btzhang@bi.snu.ac.kr).
∗: Both authors have equally contributed to this work.

Recently, machine learning methods have been successfully used in microarray data analysis, and most of them use large margin classification techniques such as support vector machines (SVMs) [6], [7] and boosting [8], [9]. The margin serves as a decision boundary separating gene expression patterns into classes of samples (or tissues). However, the performance of such methods is limited to identify the optimum solutions in the nonlinear classification problems. Furthermore, the relationship among selected genes cannot be easily explained, as well as their combined role is not interpretable. To address such problems, several efforts have been made to analyze gene expression data at the level of biological modules, rather than individual genes [10], [11], [12], [13]. However, inferring modules of multiple co-regulated genes directly from the microarray data remains a difficult problem.

In this paper, we propose a novel approach to identify the gene modules associated with cancers from microarray data. The proposed approach is based on hypernetworks [14], [15], a random hypergraph model [16] with weighted edges. The concept of the hypernetworks originated in biomolecular networks which maintain the stability, while rapidly adapting to the cellular environmental changes. This property is useful for analyzing complicated and large-scale biological problems such as cancer regulatory mechanisms. In addition, the hypernetwork classifiers naturally provide understandable causes behind their predictions. In the hypernetwork frameworks, learning is performed by an evolutionary algorithm [17], [18] to find best combinations of higher-order features and their weights.

In experiments, we apply the hypernetwork classifiers to microRNA (miRNA) expression profiles related to human cancers [19]. The goal is to identify miRNA pairs, whose expression patterns can predict the presence of cancer with high accuracy. Our experimental results show that the hypernetwork classifiers provide a competitive performance to neural networks and support vector machines, and outperform decision trees and naive Bayes. We also examine the relevance of the discovered miRNA modules to causes of cancers.

The paper is organized as follows. In Section 2, the hypernetwork classifiers are explained. Section 3 describes the connection to evolutionary computation and evolutionary learning procedure. In Section 4, the experimental results on miRNA expression profiles are provided. Concluding remarks and directions for further research are given in Section 5.
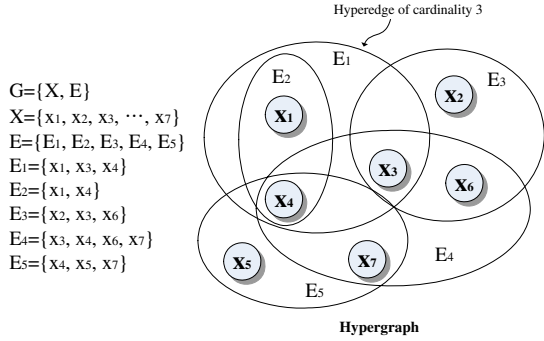
Fig. 1. An example hypergraph consisting of seven vertices and five hyperedges of variable cardinality.

II. HYPERNETWORK CLASSIFIERS

Hypernetworks are a graphical model which is naturally implemented as a library of interacting DNA molecular structures. Here, we briefly introduce the hypernetwork classifiers.

A hypergraph is an undirected graph G whose edges connect a non-null number of vertices [20], i.e., $G = \{X, E\}$, where $X = \{X_1, X_2 \ldots, X_n\}$, $E = \{E_1, E_2 \ldots, E_m\}$, and $E_i = \{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$. E_i is called the hyperedges. Mathematically, E_i is a set and its cardinality (size) is $k \geq 1$, i.e., the hyperedges can connect more than two vertices while in ordinary graphs the edges connect maximum two vertices, i.e., $k \leq 2$. A hyperedge of cardinality k will be referred to as a k-hyperedge. The use of these hyperedges allows for additional degrees of freedom in representing a network while preserving the mathematical methods provided by the graph theory. Figure 1 shows a hypergraph consisting of seven vertices $X = \{X_1, X_2 \ldots, X_7\}$ and five hyperedges $E = \{E_1, E_2 \ldots, E_5\}$ each having a different cardinality.

Hypernetworks are a generalization of the hypergraphs by assigning weights to its hyperedges, so that it can represent how strong vertex sets are attached. Formally, we define a hypernetwork as a triple $H = (X, E, W)$, where $X = \{X_1, X_2 \ldots, X_n\}, E = \{E_1, E_2 \ldots, E_m\}, W = \{w_1, w_2 \ldots, w_m\}$. An k-hypernetwork consists of a set of X of vertices, a subset of E of $X[k]$, and a set W of hyperedge weights, where $E = X[k]$ is a set of subsets of X whose elements have precisely k members. A hypernetwork H is said to be k-uniform if every edge E_i in E has cardinality k.

From the aspect of biological network, the hyperedges in a hypernetwork can be viewed as building blocks, such as modules, motifs, and circuits [21], [22]. Particularly, it is a significant discovery when the hyperedges are of large weights in the biological problem. In this sense, the hypernetwork structure can be used to identify massively-interacting biological modules.

A learning task can be regarded as storing a data set D at a model, so that the stored data can be retrieved later by an example. Formally, a hypernetwork can be used as a probabilistic memory. Let $\mathcal{E}(\mathbf{x}^{(n)}; W)$ be the energy of the

hypernetwork, where $\mathbf{x}^{(n)} \in D$ denotes the n-th data to store and W represents the parameters (hyperedge weights) for the hypernetwork model. Then, the probability of the data being generated from the hypernetwork is given as Gibbs distribution

$$P(\mathbf{x}^{(n)}|w) = \frac{1}{Z(W)} \exp\{-\mathcal{E}(\mathbf{x}^{(n)}; W)\}, \qquad (1)$$

where $\exp\{-\mathcal{E}(\mathbf{x}^{(n)}; W)\}$ is the Boltzmann factor and $Z(W)$ is the normalizing term.

In classification tasks, a data consists of a set of features x_i and a label y, i.e. $(\mathbf{x}, y) \in D$. Here, the hypernetwork classifiers can be represented by adding a vertex y to the set of vertices X. At this point, we can formulate the joint probability $P(\mathbf{x}, y)$ as

$$P(\mathbf{x}, y) = \frac{1}{Z(W)} \exp\{-\mathcal{E}(\mathbf{x}, y; W)\}. \qquad (2)$$

Given input $\mathbf{x}$, the classifier returns its class by computing the probability of each class conditional, and then determining the class whose conditional probability is the highest, i.e.

$$y^* = \arg\max_y P(y|\mathbf{x}) \qquad (3)$$

$$= \arg\max_y \frac{P(\mathbf{x}, y)}{P(\mathbf{x})}, \qquad (4)$$

where $P(\mathbf{x}, y) = P(y|\mathbf{x})P(\mathbf{x})$ and y represents the candidate classes. Since $P(\mathbf{x})$ can be omitted in the discriminative model, Equation (4) is rewritten as follows:

$$y^* = \arg\max_y \frac{P(\mathbf{x}, y)}{P(\mathbf{x})} = \arg\max_y P(\mathbf{x}, y) \qquad (5)$$

$$= \arg\max_y \frac{1}{Z(W)} \exp\{-\mathcal{E}(\mathbf{x}; W)\} \qquad (6)$$

$$= \arg\max_y \exp\{-\mathcal{E}(\mathbf{x}; W)\} \qquad (7)$$

$$= \arg\max_y -\mathcal{E}(\mathbf{x}; W) \qquad (8)$$

$$= \arg\min_y \mathcal{E}(\mathbf{x}; W). \qquad (9)$$

The energy function $\mathcal{E}(\mathbf{x}; W)$ can be expressed in many ways such as linear functions, sigmoid functions, and gaussian functions. In effect, a hypernetwork represents a probabilistic model of a data set using a population of hyperedges and their weights.

III. EVOLUTIONARY LEARNING FOR HYPERNETWORK CLASSIFIERS

The hypernetwork classifiers is to choose a label y to minimize the energy function $\mathcal{E}$, and the learning task is to adjust the weights of hyperedges to fit in with training data. We now introduce an evolutionary learning method to find the optimal hypernetworks that maximize classification accuracy.
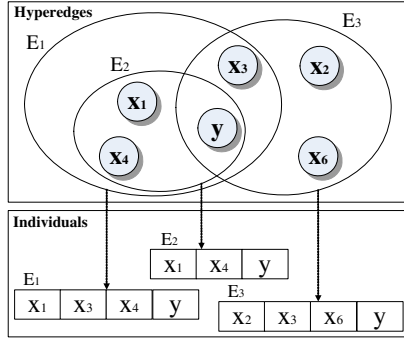
Fig. 2. An example of transforming hyperedges to individuals to be evolved by an evolutionary learning algorithm. An individual consists of a set of vertices and a label, which indicates a hyperedge.

For evolving the hypernetworks, we assume that a population represents a hypernetwork classifier, and its individuals represent hyperedges. We express the weight of a hyperedge by allowing duplicates of an individual in the population. The learning task of a hypernetwork is now changed to adjust the number of individuals towards minimizing the classification errors. Figure 2 shows an example of the individuals. Note that a hyperedge consists of a set of vertices and a label in supervised learning problems.

To make an initial population, i.e. a hypernetwork classifier, we use a random graph model, which is a graph constructed by a random procedure [15], [16]. For a k-hypergraph, the number of possible hyperedges are

$$|E| = C(n, k) = \frac{n!}{k!(n-k)!}, \tag{10}$$

where $n = |X|$. If we denote the set of all graphs as Ω, its size is

$$|\Omega| = 2^{C(n,k)}. \tag{11}$$

However, $|\Omega|$ rapidly increases when k and n becomes large, which is common in real-world problems. Hence, we use a stochastic approach based on the random graphs to solve this combinatorial explosion. A hypernetwork generated from the random graph process is called a random hypernetwork. A random graph model chooses a graph at random, with equal probabilities, from the set of all possible graphs. We consider a probability space

$$(\Omega, \mathcal{F}, \mathcal{P}), \tag{12}$$

where Ω is the set of all graphs, $\mathcal{F}$ is the family of all subsets of Ω, and to every $\omega \in \Omega$ we assign its probability as

$$\mathcal{P}(\omega) = 2^{-C(n,k)}. \tag{13}$$

The probability space can be viewed as the product of $C(n, k)$ binary spaces. It is a result of $C(n, k)$ independent tosses of a fair coin, i.e. Bernoulli experiments.

- 1. Start with the empty hypernetwork $H = (X, E, W) = (\emptyset, \emptyset, \emptyset)$.
- 2. Get a training sample $\mathbf{x}$ with the probability p. Generate a hypernetwork $H' = (X', E', W')$ as follows:
  - Generate hyperedges (individuals), E_i, of cardinality k from $\mathbf{x}$ by a random hypergraph process.
  - $E' \leftarrow E' \cup \{E_i\}$.
  - $W' \leftarrow W' \cup \{w_i | w_i = w_{init}\}$.
  - $X' \leftarrow X' \cup \{x_j | x_j \in E_i\}$.
- 3. $H \leftarrow H \cup H'$.
- 4. Go to step 2 unless the termination condition is met.

Fig. 3. The procedure for building an initial population.

- 1. Generate a population by the random hypernetwork process (Fig. 3).
- 2. Get a training example $(\mathbf{x}, y)$.
- 3. Evaluate the fitness by classifying $\mathbf{x}$ as described in the previous section. Let this class be y^*.
- 4. Update the population if $y^* \neq y$,
  - $c_{E_i} \leftarrow c_{E_i} + \Delta c_{E_i}$, where c_{E_i} is the number of individuals corresponding the hyperedge $E_i \in E(\mathbf{x}, y)$
  - Normalize the duplicates of all individuals for the current population.
- 5. Goto step 2 unless the termination condition is met.

Fig. 4. The evolutionary algorithm to adjust the weights of hyperedges in hypernetwork classifiers.

The random hypernetworks can be generated by a binomial random graph process. Given a real number $p, 0 \leq p \leq 1$, the binomial random graph $\mathcal{G}(n, p)$ is defined by taking as Ω and setting

$$P(\mathcal{G}) = p^{|E(\mathcal{G})|}(1 - p)^{C(n,k) - |E(\mathcal{G})|}, \tag{14}$$

where $|E(\mathcal{G})|$ stands for the number of edges of $\mathcal{G}$. The random hypernetworks are generated by repeating the random hypergraph process.

Figure 3 denotes the procedure for building an initial population based on the random hypergraph process. Starting with the empty hypernetwork, new hypernetwork H' is repeatedly generated from a training sample $\mathbf{x}$ with the probability p. Alternatively, a random H' (not from $\mathbf{x}$) can be generated with the probability $(1-p)$. This alternate case helps to give a diversity in the population. For every H', the duplicates of the hyperedge E_i are added to the initial population, where the number of duplicates is w_{init}. The procedure is terminated if the population reaches a predefined size m. The random hypernetwork results in reducing the population size, while maintaining its classification performance.

Figure 4 presents the evolutionary algorithm to adjust the number of individuals of the population. We start with a random hypernetwork. As a new training example $(\mathbf{x}, y)$ is observed, the population is evaluated by classifying $\mathbf{x}$. The class y^* of $\mathbf{x}$ is determined by the classification procedure

described in the previous section. If the label y^* is correct, no action is performed because the current population correctly classifies the example. If the label y^* is incorrect, the population is modified by adding a number of hyperedges, Δc_{E_i}, where $E_i \in E(\mathbf{x}, y)$.

It is interesting that the evolutionary learning performs gradient search to find an optimal hypernetwork for the training examples. Given $\mathbf{x}$ and y, where $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \{0, 1\}^n$ and $y \in \{0, 1\}$, let us assume that the energy function $\mathcal{E}(\mathbf{x}^{(n)}; W)$ is a sigmoid function

$$\mathcal{E}(\mathbf{x}; W) = \frac{1}{1 + \exp(-f(\mathbf{x}, W))}, \quad (15)$$

where

$$f(\mathbf{x}, W) = \sum_{i=1}^{|E|} w_{i_1 i_2 \ldots i_{|E_i|}} x_{i_1} x_{i_2} \ldots x_{i_{|E_i|}}. \quad (16)$$

Note that $x_{i_1} x_{i_2} \ldots x_{i_{|E_i|}}$ is a combination of k elements of the data $\mathbf{x}$ which is represented as a k-hyperedge in the network. We can then write down the error function

$$\begin{aligned} G(W) &= -\sum_{n=1}^{N} (y^{(n)} \ln \mathcal{E}(\mathbf{x}^{(n)}; W) + \\ &\quad (1 - y^{(n)}) \ln(1 - \mathcal{E}(\mathbf{x}^{(n)}; W))). \end{aligned} \quad (17)$$

Here, the derivative $\mathbf{g} = \partial G / \partial W$ is given by

$$g_i = \frac{\partial G}{\partial w_i} = \sum_{n=1}^{N} -(y^{(n)} - y^{*(n)}) \mathbf{x}^{(n)}. \quad (18)$$

Since the derivative $\partial G / \partial W$ is a sum of $\mathbf{g}^{(n)}$, we can obtain an online algorithm by putting each input at a time, and adjusting W in a direction opposite to $\mathbf{g}^{(n)}$. $(y^{(n)} - y^{*(n)})$ is the error on an example, and W is changed only if the classifier is incorrect. According to Equation (18), we show that the evolutionary algorithm in Figure 4 is a simplified version of the on-line gradient search. More details related to the derivation can be found in [23].

IV. EXPERIMENTAL RESULTS

For experiments, we performed the miRNA expression classification using the microarray dataset in [19]. It includes the expression profiles of 151 miRNAs on 89 samples, which consists of 68 multiple human cancer tissues and 21 normal tissues. We use a set of data $(\mathbf{x}, y)$, where $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \{0, 1\}^n$ and $y \in \{0, 1\}$. i.e. a binary dataset. Although the hypernetwork classifiers can accept any attribute such as integers or real numbers, the discretized expression data provides flexibility for extending to molecular computation [14]. Moreover, the hypernetwork classifiers are easily implemented *in silico* with binary numbers. Hence, we divide the expression levels of the miRNA data into binary numbers based on medians on each sample.
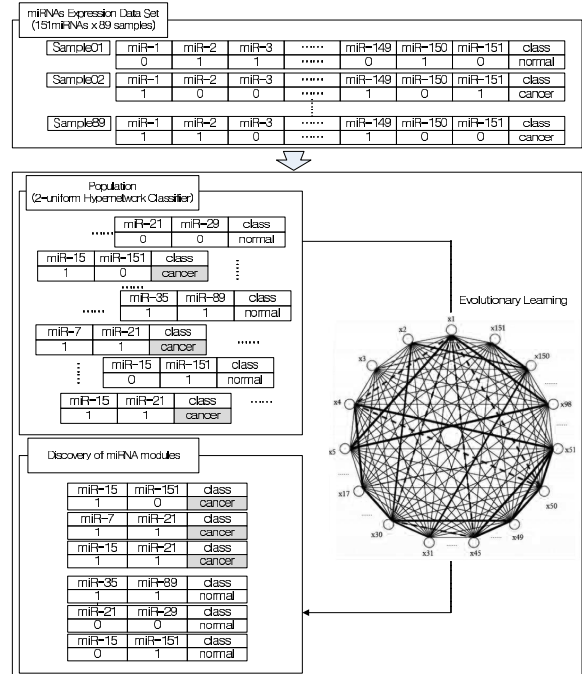

Fig. 5. The procedure for building a hypernetwork classifier from miRNA expression data set. The hypernetwork is represented as a collection of hyperedges which are then encoded as a population for evolutionary learning. A population represents the hypernetwork, where the weights of hyperedges are encoded as the number of duplicates of the individuals.

Figure 5 presents the whole procedure for building a hypernetwork classifier, i.e. a population. We use a 2-uniform hypernetwork to classify the miRNA expression profiles. The initial population is generated using the random hypernetwork process. The individuals are selected from the training examples with the probability $p = 0.5$. Unless the training examples are selected, the individuals are sampled from random examples. The individuals are set to $50,000$, and the number of duplicates is initialized to $1,000$. We use a sigmoid function as the energy function $\mathcal{E}(\mathbf{x}; W)$ of the hypernetwork classifier.

Setting the learning parameter $\eta = \Delta c_{E_i} / c_{E_i}$ in Figure 4 is important to balance the adaptability and stability of the population. The larger η is, the larger gets the distribution changes of the population. In the experiments, learning parameter η is started from 0.01, and decreased to $\eta = 0.75 \times \eta$ when the whole accuracy of current epoch drops compared than that of previous epoch. The learning procedure is stopped after 40 epochs.

Figure 6 depicts the performance evolution of the population as generation goes on. Since the evolution progresses in on-line manner, we present the performance evolution by taking the classification accuracy at each epoch. Note that the actual fitness is measured every time a training example is observed. The performance curves are increased gradually, and stabilized after 20th epoch. The early generations are
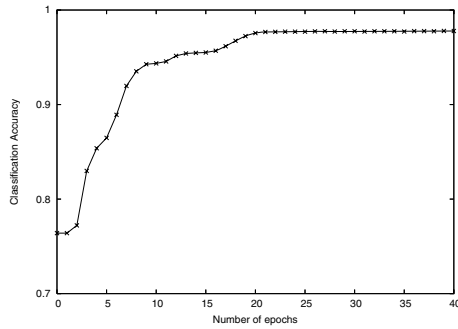
Fig. 6. Performance evolution of the population representing the hypernetwork for the miRNA expression dataset. Shown are the average classification rates of leave-one-out cross validation.

Algorithms	Accuracy (%)
Backpropagation Neural Networks	92.13
Hypernetworks	91.46
Support Vector Machines	91.01
Decision Trees	88.76
Naive Bayes	83.14

the process to explore candidate hypernetworks for better miRNA classification. As the generation progresses further, the increment of the performance falls down because the population is converged to the optimal hypernetwork.

A. miRNA Expression Classification

Table I presents the performance comparison of the hypernetworks and other machine learning methods, backpropagation neural networks (BPNNs), support vector machines (SVMs), decision trees, and naive Bayes. Using leave-one-out cross validation, the hypernetwork classifier shows 91.46% of accuracy. It is better than decision trees and naive Bayes, while providing competitive performance to the SVM or BPNNs. Compared to the SVM or BPNNs, the hypernetwork classifiers feature the ability of analyzing significant gene modules.

As mentioned before, the hypernetwork classifiers can be used in molecular computation, which allows huge population size. Therefore, higher-order hypernetwork classifiers can be implemented by the molecular computing for better classification performance and analysis of more sophisticated gene interactions.

B. miRNA Module Discovery

The hypernetwork classifiers naturally can be used for microarray analysis to discover significant gene modules. Table II shows the high-ranked miRNA modules among ten experiments. *hsa-miR-215* is located in the region with DNA copy number gains in ovarian and breast cancers [24]. It is because DNA copy number alterations may be a critical factor affecting expression of miRNAs in cancers.

miRNA modules	
a	b
hsa-miR-147	*hsa-miR-296*
hsa-miR-215	*hsa-miR-7*
hsa-miR-130b	*hsa-miR-23b*
hsa-miR-105	*hsa-miR-133a*
hsa-miR-147	*hsa-miR-206*

GO ID	Term	Ontology	*p-value	Genes
GO:0050794	Regulation of cellular physiological process	BP	2.63E-18	BCL3, BCL6, CCND1, CCND2, CDH1, DDX6, ETV6, FGFR1, MYCL1, IRF4,NF2, NRAS, PDGFB
GO:0050789	Regulation of physiological process	BP	6.43E-18	
GO:0005634	Nucleus	CC	1.52E-17	
GO:0065007	Biological regulation	BP	1.60E-16	
GO:0031323	Regulation of cellular metabolic process	BP	3.73E-16	
GO:0045449	Regulation of transcription	BP	3.91E-16	
GO:0005515	Protein binding	MF	4.36E-16	
GO:0019219	Nucleobase, nucleotide and nucleic acid metabolism	BP	7.22E-16	

hsa-miR-23b is located in one of two regions on 9q, where genomic deletion is found [25]. It is known that there is the genomic alteration in a human cancer. Furthermore, *hsa-miR-147* is located near (<2Mb) to the markers with the highest rate of LOH (loss of heterozygosity) [25]. The LOH is a major mechanism in the genomic alteration that transforms a normal cell into an unregulated tumor cell.

To examine the discovered miRNA modules, we found the functional correlations between mRNAs by extracting the gene ontology (GO) terms. The GO has become a standard to validate the functional coherence of genes. This project aims to develop three structured, controlled vocabularies that describe gene products in terms of their associated biological processes (BP), cellular components (CC), and molecular functions (MF) in a species-independent manner. Typically, the validation is accompanied by a statistical significance analysis. If the discovered miRNA modules are closely related, the target mRNAs corresponding the miRNAs might reflect their functional relevance. The analysis using target genes can be biologically significant because miRNAs determine the target gene functions in a specific biological context. We examined significant terms with p-value < 0.01 for the module I, *hsa-miR-147* and *hsa-miR-296*. The results are shown in Table III. Among common target genes of two miRNAs, 13 genes (*BCL3, BCL6, CCND1, CCND2, CDH1, DDX6, ETV6, FGFR1, MYCL1, IRF4, NF2, NRAS,*

miRNA	Chr.	Start - End Position	Strand
hsa-miR-147	Chr9	122047078-122047149	-
hsa-miR-296	Chr20	56826065-56826144	-

Target mRNA	Description
BCL3	B-Cell Leukemia/Lymphoma-3
BCL6	B-Cell Lymphoma-6 (zinc finger protein 51)
CCND1	Cyclin D1
CCND2	G1/S-specific cyclin D2
CDH1	cadherin 1, type 1, E-cadherin (epithelial)
DDX6	DEAD(Asp-Glu-Ala-Asp) box polypeptide 6
ETV6	ets variant gene 6 (TEL oncogene)
FGFR1	fibroblast growth factor receptor 1, fms-related tyrosine kinase 2, Pfeiffer syndrome
IRF4	interferon regulatory factor 4
MYCL1	v-myc myelocytomatosis viral oncogene homolog 1, lung carcinoma derived (avian)
NF2	neurofibromin 2 (bilateral acoustic neuroma)
NRAS	neuroblastoma RAS viral oncogene homolog
PDGFB	platelet-derived growth factor beta polypeptide, (simian sarcoma viral (v-sis) oncogene homolog)

and *PDGFB*) are annotated in a significant level. Overall, the target genes in the module I belong to characteristic functional categories, which are related to transcription, protein binding, regulation of cellular, physiological or biological process. Also, these are all related to cancer progression.

Table IV describes the miRNA module I in detail. It shows the chromosomal location information of the module I, and functional description of their shared putative target mRNAs, which are annotated GO terms with $p < 0.01$. As is stated above, *hsa-miR-147* is located at 9q.22 with high frequency of LOH, and the sequence of *hsa-miR-296* maps to human chromosome 20. All annotated target genes are actively involved in tumorigenesis. For instance, *BCL3* is inducible by DNA damage and is required for the suppression of persistent *p53* activity which regulates the cell cycle and hence functions as a tumor suppressor [26]. The human proto-oncogene *BCL6* suppresses the expression of the *p53* tumor suppressor gene and modulates DNA damage-induced apoptotic responses in germinal-centre B cells [27]. Thus, altered expressions of *BCL3* and *BCL6* lead to tumorigenic potential and it is functionally essential for cancer growth and survival. As a result, we conclude that the hypernetwork classifiers find cancer-related miRNA modules, which apparently interact with each other.

V. CONCLUSIONS

We proposed a method for detecting gene modules from microRNA data using hypernetwork classifiers. An evolutionary approach is designed to find best hypernetworks without exhaustive search in limited resources.

The proposed method was applied to the miRNA expression profiles on multiple human cancers. The experimental results show that the hypernetwork classifiers outperform decision trees and naive Bayes, while providing comparable performance to neural networks and support vector machines.

It also shows that the hypernetwork classifiers find biologically significant miRNA blocks. The hypernetwork structures are effective since it provides interpretable solutions, as well as producing good classification performance. Future study includes the analysis of the order-effect in hypernetworks and a more detailed analysis of the gene modules discovered.

REFERENCES

[1] S. Ramaswamy and T.R. Golub, DNA microarrays in clinical oncology, *Journal of Clinical Oncology*, 20, pp. 1932–1941, 2002.

[2] M. Schena, D. Shalon, R.W. Davis, and P.O. Brown, Quantitative monitoring of gene expression patterns with a complementary DNA microarray, *Science*, 270, pp. 467–470, 1995.

[3] D.J. Lockhart, H. Dong, M.C. Byrne, M.T. Follettie, M.V. Gallo, M.S. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Norton, and E.L. Brown, Expression monitoring by hybridization to high-density oligonucleotide arrays, *Nature Biotechnology*, 14, pp. 1675–1680, 1996.

[4] E. Segal, N. Friedman, N. Kaminski, A. Regev, and D. Koller, From signatures to models: understanding cancer using microarrays, *Nature Genetics*, 37, s38–s45, 2005.

[5] E. Segal, N. Friedman, D. Koller, and A. Regev, A module map showing conditional activity of expression modules in cancer, *Nature Genetics*, 36, pp. 1090–1098, 2004.

[6] M.P.S. Brown, W.N. Grundy, D. Lin, N. Cristianini, C.W. Sugnet, T.S. Furey, M. Ares, Jr., and D. Haussler, Knowledge-based analysis of microarray gene expression data by using support vector machines, *Proceedings of the National Academy of Sciences*, 97(1), pp. 262–267, 2000.

[7] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, Gene selection for cancer classification using support vector machines, *Machine Learning*, 46, pp. 389–422, 2002.

[8] M. Dettling and P. Buhlmann, Boosting for tumor classification with gene expression data, *Bioinformatics*, 19, pp. 1061–1069, 2003.

[9] P.M. Long, V.B. Vega, Boosting and microarray data, *Machine Learning*, 52, pp. 31–44, 2003.

[10] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, and J.P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles, *Proceedings of the National Academy of Sciences*, 102, pp. 15545–15550, 2005.

[11] J. Lamb, S. Ramaswamy, H.L. Ford, B.Contreras, R.V. Martinez, F.S. Kittrell, C.A. Zahnow, N. Patterson, T.R. Golub, and M.E. Ewen, A mechanism of cyclin D1 action encoded in the patterns of gene expression in human cancer, *Cell*, 114, pp. 323–334, 2003.

[12] D.R. Rhodes, J. Yu, K. Shanker, N. Deshpande, R. Varambally, D. Ghosh, T. Barrette, A. Pandey, and A.M. Chinnaiyan, Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression, *Proceedings of the National Academy of Sciences*, 101, pp. 9309–9314, 2004.

[13] E. Huang, S. Ishida, J. Pittman, H. Dressman, A. Bild, M. Kloos, M. D'Amico, R.G. Pestell, M. West, and J.R. Nevins, Gene expression phenotypic models that predict the activity of oncogenic pathways, *Nature Genetics*, 34, pp. 226–230, 2003.

[14] B.-T. Zhang and J.-K Kim, DNA hypernetworks for information storage and retrieval, *Lecture Notes in Computer Science*, DNA12, 4287, pp. 298–307, 2006.

[15] B.-T. Zhang, Random hypergraph models of learning and memory in biomolecular networks: shorter-term adaptability vs. longer-term persistency, *The First IEEE Symposium on Foundations of Computational Intelligence*, 2007.

[16] S. Janson, T. Luczak, and A. Rucinski, Random graphs, Wiley, 2000.

[17] D.B. Fogel, Evolutionary computation. IEEE Press, 1995

[18] T. Back, Evolutionary algorithms in theory and practice, Oxford University Press, 1996.

[19] J. Lu, G. Getz, E.A. Miska, E. Alvarez-Saavedra, J. Lamb, D. Peck, A. Sweet-Cordero, B.L. Ebert, R.H. Mak, A.A. Ferrando, J.R. Downing, T. Jacks, H.R. Horvitz, and T.R. Golub, MicroRNA expression profiles classify human cancers, *Nature*, 435, pp. 834–838, 2005.

[20] C. Berge, Graphs and hypergraphs, North-Holland Publishing, Amsterdam, 1973.

[21] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashitan, D. Chklovskii, and U. Alon, Network motifs: simple building blocks of comples networks, *Science*, 298, pp. 824–827, 2002.

[22] D.M. Wolf and A.P. Arkin, Motifs, modules and game in bacteria, *Current Opinion Microbiology*, 6(2), pp. 125–134, 2003.

[23] S. Kim, M.-O. Heo, and B.-T. Zhang, Text classifiers evolved on a simulated DNA computer, *IEEE Congress on Evolutionary Computation*, pp. 9196–9202, 2006.

[24] L. Zhang, J. Huang, N. Yang, J. Greshock, M.S. Megraw, A. Giannakakis, S. Liang, T.L. Naylor, A. Barchetti, M.R. Ward, G. Yao, A. Medina, A.O. Brien-Jenkins, D. Katsaros, A. Hatzigeorgiou, P.A. Gimotty, B.L. Weber, and G. Coukos, MicroRNAs exhibit high frequency genomic alterations in human cancer, *Proceedings of the National Academy of Sciences*, 103, pp. 9136–9141, 2006.

[25] G.A. Calin, C. Sevignani, C.D. Dumitru, T. Hyslop, E. Noch, S. Yendamuri, M. Shimizu, S. Rattan, F. Bullrich, M. Negrini, and C.M. Croce, Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers, *Proceedings of the National Academy of Sciences*, 101(9), pp. 2999–3004, 2006.

[26] D. Kashatus, P. Cogswell, and A.S. Baldwin, Expression of the Bcl-3 proto-oncogene suppresses p53 activation, *Genes and Development*, 20, pp. 225–235, 2006.

[27] R.T. Phan and R. Dalla-Favera, The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells, *Nature*, 432(7017), pp. 635–639, 2004.