Use of Evolutionary Hypernetworks for Mining Prostate Cancer Data

Chan-Hoon Park*,[a] Soo-Jin Kim*,[b] Sun Kim[a] Dong-Yeon Cho[a] Byoung-Tak Zhang[a,b]
[a]School of Computer Science and Engineering, [b]Graduate Program in Bioinformatics
Seoul National University
Seoul 151-742, Korea
{chpark, sjkim, skim, dycho, btzhang}@bi.snu.ac.kr

Abstract— **Hypernetwork models have been proposed as a random graph model of learning and memory inspired by biomolecular networks in the cell. The hypergraph structure of the hypernetworks turned out to be very useful for discovering the building blocks of higher-order interaction of multiple variables. Here we use the hypernetwork model for the analysis of microarray data for cancer diagnosis. Tested on a prostate cancer problem consisting of 102 training examples and 34 test examples of 225 features, the hypernetwork models outperformed the accuracy of multilayer perceptrons and decision trees. We also analyze the modules discovered by the hypernetwork models as potential macro-biomarkers for molecular diagnosis.**

I. Introduction

A cancer is a research area of constant interest. As high-throughput data such as microarrays increase, the cancer researches tend to be more systematic based on computational methods for managing large and complex data [1, 2].

The microarray provides large and complex parallel interactions among genes. These interactions are valuable to discover the complex mechanisms of cancer development. Thus, computational approaches are required to analyze the data and collect cancer-related genes, while removing massive redundant information [3-11]. The correlations between genes and samples are frequently used for such purpose. They provide some valuable genes in cancer classification [3-5], whereas the effects of gene combinations are ignored.

Machine learning techniques are alternative approaches to analyze the microarrays [6-8]. They have shown good performances in the cancer classification tasks. However, previous studies do not consider the gene-gene interactions. Even though some approaches utilize the gene combinations by mapping into the high-dimensional space, they cannot obtain human-interpretable solutions.

Here, we propose a simple, but effective method to discover significant higher order interactions of gene pairs for cancer classification. The proposed method is based on the hypernetworks model [9, 10]. In the hypernetwork models, gene combinations are selected to build the hypernetwork. An evolutionary learning fits the hypernetworks to joint probability distribution of given examples. Since the gene combinations are explicitly considered in the hypernetworks as a unit of learning, one can get a list of meaningful decision rules which represent gene-gene interactions.

The proposed method is applied to the prostate cancer classification. The results show that our method outperforms conventional machine learning algorithms such as multilayer perceptrons and decision trees in accuracy. In addition, from the modules discovered by the hypernetworks, we can find some potential macro-biomarkers for molecular diagnosis.

The paper is organized as follows. In Section 2, the hypernetwork models are explained. Section 3 describes the evolutionary learning method to find the optimal hypernetworks. In Section 4, the experimental results are explained. Section 5 draws the conclusion and further research.

II. Hypernetwork Models

The hypernetwork models are probabilistic frameworks which are motivated from molecular computing [9]. For hypernetworks, a training set D of K label is represented as follows:

$$D = \{(\mathrm{x}_i, y_i)\}_{i=1}^{K} \quad (1)$$

$$\mathrm{x}_i = (x_{i1}, x_{i2}, ..., x_{in}) \in \{0,1\}^n \quad (2)$$

$$y_i \in \{0,1\}, \quad (3)$$

where x_i is a vector of features from a sample i and y_i is its class.

A hypergraph of the hypernetworks is a decision rule or an individual which is a conjunction of binary variables x_i and a class label y_i. The number of variables in an individual is defined as the *order* of the individual. For example, the individual $\mathrm{z} = (\mathrm{x}, y) = (x_1 = 1, x_3 = 1, x_5 = 0, x_6 = 1, y = 1)$, is a conjunction of four variables with class 1. Thus, z is the individual with the *order* of 4. The library of hypernetwork model includes multiple copies of each individual and the number of copies means the importance of the individual for the class.

Substantially, the library represents the joint probability $P(X,Y)$ of the input pattern X and the output class Y. Given an example, the class is determined by matching it against each individuals and taking the majority class. This ensemble ap-

*: Both authors have equally contributed to this work.

proach naturally makes use of huge number of individuals to make decision robust.

The classification of the hypernetworks can be explained to the conditional probability of each class on the input. Given input x, the class y^* is decided by computing the conditional probability of each class, and then selecting the class which has the highest conditional probability, as follows:

$$y^* = \arg\max_{Y\in\{0,1\}} P(Y \mid \mathrm{x}) \quad (4)$$

$$= \arg\max_{Y\in\{0,1\}} \frac{P(Y,\mathrm{x})}{P(\mathrm{x})}. \quad (5)$$

Here, $P(Y,\mathrm{x})=P(Y \mid \mathrm{x})P(\mathrm{x})$, and Y represents the candidate classes.

The empirical probability distribution $P(X,Y)$ can be represented by a set of point estimators that constitute the library L of individuals:

$$P(X,Y) \approx \frac{1}{|L|}\sum_{i=1}^{|L|} f_i^{(n)}(X_1, X_2, ..., X_n, Y), \quad (6)$$

where $f_i^{(n)}(X_1, X_2, ..., X_n, Y)$ is the ith individual of *order n* and $|L|$ is the size of library. By increasing the $|L|$, the approximation can be more arbitrarily accurate. There is a more theoretical background in [9].

The procedure of decision making is summarized in Figure 1. Given an input x, all individuals matching with x is extracted from the library L. The extraction is implemented by matching x with each and every library element of the library. The class decision of x is made by selecting the class which has the higher number of elements.

In step 3.1, the count $c(\mathrm{x})$ of x in M approximates the evidence which is the probability of observing the matching individuals:

$$c(\mathrm{x})/|L| = |M|/|L| \approx P(\mathrm{x}). \quad (7)$$

Step 3.2 computes the frequencies $c(Y|\mathrm{x})$ of each class. These are an approximation of a posteriori probabilities which is the conditional probabilities given the example:

$$c(Y \mid \mathrm{x})/|M| = |M^Y|/|M| \approx P(Y \mid \mathrm{x}). \quad (8)$$

Thus, the procedure computes the maximum a posteriori (MAP) criterion:

$$y^* = \arg\max_{Y\in\{0,1\}} c(Y \mid \mathrm{x})/|M|$$

$$\approx \arg\max_{Y\in\{0,1\}} P(Y \mid \mathrm{x}). \quad (9)$$

which validates the class decision in Eq. (4).

1. Let the library L represent the current empirical distribution $P(X, Y)$.
2. Given an input x,
3. Classify x using L as follows:
   3.1 Extract all individuals matching with x into M.
   3.2. Separate the individuals from M according to their classes:
   - Extract the individuals with label Y=0 into M^0.
   - Extract the individuals with label Y=1 into M^1.
   3.3. Compute $y^* = \arg\max_{Y\in\{0,1\}} c(Y \mid \mathrm{x})/|M|$

Figure 1. The procedure of decision making in hypernetworks.

III. Evolutionary Hypernetworks

The hypernetworks can be learned to fit the training example more accurately through evolutionary learning procedure. It is summarized as a procedure of finding a proper distribution of the library which best fits in the training examples using a gradient descent [10]. The procedure adjusts the optimal number of copies of each individual to reduce the magnitude of the classification error.

An evolutionary procedure is applied after initializing the hypernetworks. Given a training input x_i and class y_i, the total quantity c for a class Y is computed as follows:

$$c(Y \mid \mathrm{x}_i) = \sum_{j=1}^{|L|} c_j I_{z_j=(\mathrm{x},Y)}$$

$$\text{where } I_{z_j=(\mathrm{x},Y)} \begin{cases} 1 & \text{if } z_j = (\mathrm{x},Y) \\ 0 & \text{otherwise} \end{cases}, \quad (10)$$

Here, c_j is the number of the individual z_j. With a weight vector w, the conditional probability given the input x_i is represented as follows:

$$P(Y \mid \mathrm{x}_i) \approx \frac{c(Y \mid \mathrm{x}_i)}{|M|} = \sum_{j=1}^{|L|} w_j I_{z_j=(\mathrm{x},Y)}$$

$$\text{where } w_j = c_j/|M|. \quad (11)$$

An error e_i of a training example x_i and class Y is given by

$$e_i = P^*(Y \mid \mathrm{x}_i) - P(Y \mid \mathrm{x}_i), \quad (12)$$

where $P^*(Y \mid \mathrm{x}_i) = I_{y_i=Y} \in \{0,1\}$ is the target probability for the training example x_i. The error function of a training dataset D is defined as follows:

1. Let the library L represent the current empirical distribution $P(X, Y)$.

2. Given a training example (x, y),

3. Classify x using L (Figure 1).
   Let the result of classification y^*.

4. Update L if $y^* \neq y$.
   - $L_n \leftarrow L_{n-1} + \Delta c(\mathrm{x}, y)$

5. Normalize L.

6. Goto step 2 unless the termination condition is met.

Figure 2. The procedure of evolutionary learning of the hypernetworks.

$$E(\mathrm{w}) = \frac{1}{2}\sum_{i \in D} e_i^2. \tag{13}$$

The gradient descent is minimize the error function $E(\mathrm{w})$ by

$$w_j \leftarrow w_j + \Delta w_j,$$
$$\text{where} \quad \Delta w_j = -\eta \frac{\partial E}{\partial w_j}, \tag{14}$$

and η is a learning rate which controls the amount of update. The update rule for gradient search is defined as follows:

$$\Delta w_j = \eta \sum_{i \in D} (P^*(Y \mid \mathrm{x}_i) - P(Y \mid \mathrm{x}_i)) I_{z_{ji=(\mathrm{x},Y)}}. \tag{15}$$

The Eq. (15) can be modified to update weight for each training example using a stochastic gradient descent. It approximates the gradient descent by updating w incrementally, following the calculation of the error for each example. The modified equation is given by

$$\Delta w_j = \eta (P^*(Y \mid \mathrm{x}) - P(Y \mid \mathrm{x})) I_{z_{j=(\mathrm{x},Y)}} \tag{16}$$
$$\approx \Delta c_j. \tag{17}$$

$P^*(Y \mid \mathrm{x})$ and $P(Y \mid \mathrm{x})$ are the target and the predicted value. The update is implemented by controlling the number of copies c_j with a certain amount of value, Δc_j.

We update the library by increasing the count of correctly matched individuals, $c(\mathrm{x}, y)$, in case of misclassifying the input x. The update rule is defined as follows:

$$L \leftarrow L + \Delta L,$$
$$\text{where} \quad \Delta L = \Delta c(\mathrm{x}, y). \tag{18}$$

Figure 2 describes the evolutionary procedure to adjust the elements of the hypernetworks. As an example (x, y) is given, the individuals matching with x is extracted from the library. If class y* of x which is predicted by the hypernetworks is correct, no action is performed. If y* is incorrect, the library is updated to reduce the error. The number of copies of each element is normalized to keep the initial library size.

TABLE I

PERFORMANCE COMPARISON

Algorithms	*Accuracy (%)*
Hypernetworks	88.24
Neural Networks	79.41
Decision Trees	79.41
Naïve Bayes	73.53
Bayesian Networks	73.53

IV. EXPERIMENTAL RESULTS

Hypernetworks are applied to the prostate cancer microarray data obtained from [11, 12]. The training data set contains 52 tumor samples and 50 normal samples with 12,600 genes expression profiles [12] which are measured by the Affymetrix HG-U95Av2 chip. Some cancer-related genes are extracted according to the Cancer Gene Census [13] for the purpose of reducing massive unimportant genes. As a result, we remain 225 genes. To utilize hypernetworks, we also binarize each expression pattern by setting all elements to '1' if the values are greater than the average expression level of each sample, '0' otherwise. The test data (25 tumor and 9 normal samples) from [11] is also preprocessed in the same way.

The second order uniform hypernetworks are used to classify the data. Thus, all individuals are composed of two genes and a class of the sample. They are initialized by randomly selected genes from training data with the probability of 0.5. The number of individuals is 50,000. The default copies of each individual are set to 1,000 initially.

The learning rate η controls the adaptability and stability of the update. The large η causes rapid update. In our experiments, η is set to 0.005.

A. *Classification performance*

Table I is the classification accuracy of the hypernetworks and the other machine learning methods. In experiments, the hypernetworks show 88.24% classification accuracy. It is the best accuracy among tested methods. In addition, the hypernetworks outperform the similar rule generating or structure inferring methods such as decision trees and Bayesian networks. Hence hypernetwork models can be used as an effective method both in classification and discovery of valuable genes by using small parts of whole data.

B. *Mining potential macro-biomarkers*

Table II enumerates the highly weighted 22 gene combinations gathered from the 10 repeated experiments. The activation status of cell signaling pathways controls cell fate and deregulation of these pathways demonstrates carcinogenesis. The PTEN/Akt pathway and Wnt signaling pathway are reported to relate to prostate cancer. Thus, prostate cancer is affected by genes on PTEN/Akt pathway and Wnt signaling pathway. The Wnt signaling pathway dysfunction is an important component of prostatic tumorigenesis. By changing the activity of Wnt signaling pathway, prostate cancer cells upset the normal balance between formation and destruction of tumor

TABLE II

HIGH-RANKED GENE COMBINATIONS RELATED TO CANCER. THE GENES ON THE PTEN DEPENDENT PATHWAY AND WNT-SIGNALING PATHWAY ARE MARKED BY SYMBOL * AND †, RESPECTIVELY.

Gene Combinations					
	1	***2***		***1***	***2***
1	CTNNB1*	IGH@	12	BCL3	EIF4A2
2	NPM1	ITGB1†	13	MYH11	SHC1
3	IGH@	MYH11	14	CTNNB1*	EIF4A2
4	BCL3	LASP1	15	CTNNB1*	ILK†
5	EIF4A2	MAPK3†	16	NACA	ILK†
6	CEBPA	COX6C	17	CCND1*	FOXO3A†
7	EIF4A2	IGH@	18	LASP1	ITGB1
8	MYH9	ILK†	19	BCL3	CTNNB1*
9	MAF	MYH9	20	FOXO3A†	MAF
10	BCL3	PDPK1†	21	ILK†	PDPK1†
11	HDAC1*	ILK†	22	CTNNB1*	MAF

cells [14]. The PTEN/Akt signaling cascades also play critical roles in the transmission of signals from growth factor receptors to regulate gene expression and prevent apoptosis [15]. Table II shows the high-ranked gene combination found in experiments. There are many genes located in the Wnt or PTEN/Akt pathway.

To examine the meaning of selected gene combinations, we compare the functional correlations between genes in Table II with those extracted from the Gene Ontology (GO) term [16]. If the genes are closely related, they might reflect their functional relevance in a specific biological context. We examine significant terms with p-value < 0.01. The results are shown in Table III. Among our selected genes, 16 genes are annotated in a significant level. These genes belong to specific functional categories which are related to cell and development maturation, hemopoiesis, cell differentiation and regulation of transcription.

We find a gene combination ILK and PDPK1, located in the PTEN/Akt pathway. It reveals that this gene combination is significantly related to the prostate cancer. This pair can be used as a potential macro-biomarker in complex cancer pathway. Further, we can find such a highly related pair of genes in a highly weighted hypernetwork of hypernetwork model.

V. CONCLUSION

We propose an effective method to discover significant gene pairs for cancer classification. The proposed method is based on the hypernetwork models which are motivated from molecular computing. Here, gene combinations are selected to build hypernetworks, and an evolutionary learning is performed to obtain the optimal joint probability distribution of given samples. Since the combinatorial effects among genes are explicitly considered in the hypernetwork models, a meaningful decision rules can be found after the learning process.

TABLE III

BIOLOGICAL PROCESS ENRICHED IN HIGHLY WEIGHTED GENES. OVERREPRESENTED TERMS WERE CHOSEN BY HYPERGEOMETRIC TESTING AND MULTIPLE TESTING ADJUSTMENTS USING THE FALSE DISCOVERY RATE (FDR). *ADJUSTED *P*-VALUE BY FDR.

GO ID	***Biological Process***	****p-value***	***Genes***
GO:0048469	Cell maturation	1.57E-3	MAF, PDPK1, EIF4A2, CEBPA, HDAC1, ITGB1, CCND1, MAPK3, CTNNB1, COX6A, FOXO3A, ILK MYH11, LASP1, MYH9, BCL3
GO:0021700	Developmental maturation	1.6E-3	
GO:0030099	Myeloid cell differentiation	1.6E-3	
GO:0030097	Hemopoiesis	1.96E-3	
GO:0048534	Hemopoietic or lymphoid organ development	2.18E-3	
GO:0002520	Immune system Development	6.14E-3	
GO:0045944	Positive regulation of transcription from RNA polymerase II promoter	6.15E-3	

The proposed method is applied to the prostate cancer classification. The result shows that our method outperforms conventional machine learning algorithms in accuracy. By examining the hypernetworks, we can find the cancer-related genes with several candidates of macro-biomarker.

Our future research is to confirm the effect of higher-orders. These complex hypernetworks consider the combinations of more than three genes. We expect that they can contribute to the discovery and understanding of the complex relationship among the genes.

ACKNOWLEDGEMENTS

This work was supported by KOSEF through the National Research Laboratory Program (No. M10400000349-06J0000-34910) and MOCIE through the Molecular Evolutionary Computing (MEC) Project. C.-H. Park and B.-T. Zhang were also supported by the Ministry of Education and Human Resources Development under the BK21 Program. The ICT at Seoul National University provided research facilities for this study.

REFERENCES

[1] G. Russo, C. Zegar, and A. Giordano, "Advantages and limitations of microarray technology in human cancer", *Oncogene*, 2003, 22(42), pp. 6497-6507.

[2] M. Schena, D. Shalon, R.W. Davis, and P.O. Brown, "Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray", *Science*, 1995, 270 (5235), pp. 467-470.

[3] A. Cromer, A. Carles, R. Millon, G. Ganguli, F. Chalmel, F. Lemaire, J. Young, D. Dembélé, C. Thibault, D. Muller, O. Poch, J. Abecassis and Bohdan Wasylyk, "Identification of genes associated with tumorigenesis and metastatic potential of hypopharyngeal cancer by microarray analysis", *Oncogene*, 2004, 23(14), pp. 2484-2498.

[4] P. Warnat, R. Eils, and B. Brors, "Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes", *BMC Bioinformatics*, 2005, 6(265).

[5] P.T.C. Wan, M.J. Garnett, S.M. Roe, S. Lee, D. Niculescu-Duvaz, V.M. Good, C.G. Project, C.M. Jones, C.J. Marshall, C.J. Springer, D. Barford and R. Marais, "Mechanism of Activation of the RAF-ERK Signaling Pathway by Oncogenic Mutations of B-RAF", *Cell*, 2004, 116(6), pp. 855-867.

[6] I. Guyon, J. Weston, S. Barnhill and V. Vapnik, "Gene Selection for Cancer Classification using Support Vector Machines", *Machine Learning*, 2002, 46(1), pp. 389-422.

[7] A. Statnikov, C.F. Aliferis, I. Tsamardinos, D. Hardin and S. Levy, "A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis", *Bioinformatics*, 2005, 21(5), pp. 631-643.

[8] L. Xu, A.C. Tan, D.Q. Naiman, D. Geman and R.L. Winslow, "Robust prostate cancer marker genes emerge from direct integration of inter-study microarray data", *Bioinformatics*, 2005, 21(20), pp.3905-3911.

[9] B.T. Zhang, H.Y. Jang, "Molecular programming: evolving genetic programs in a test tube", *The Genetic and Evolutionary Computation Conference (GECCO 2005)*, 2005, 2, pp.1761-1768.

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

[11] J.B. Welsh, L.M. Sapinoso, A.I. Su, S.G. Kern, J. Wang-Rodriguez, C.A. Moskaluk, H.F. Frierson, Jr. and G.M. Hampton, "Analysis of Gene Expression Identifies Candidate Markers and Pharmacological Targets in Prostate Cancer", *Cancer Research*, 2001, 61, pp. 5974-5978.

[12] D. Singh, P.G. Febbo, K. Ross, D.G. Jackson, J. Manola, C. Ladd, P. Tamayo, A.A. Renshaw, A.V. D'Amico, J.P. Richie, E.S. Lander, M. Loda, P.W. Kantoff, T.R. Golub, and W.R. Sellers, "Gene expression correlates of clinical prostate cancer behavior", *Cancer Cell*, 2002, 1, pp. 203–209.

[13] P. Futreal, L. Coin, M. Marshall, T. Down, T. Hubbard, R. Wooster, N. Rahman and M. Stratton, "A CENSUS OF HUMAN CANCER GENES", *Nature Reviews Cancer*, 2004, 4, pp. 177-183.

[14] G. Yardy and S. Brewster, "Wnt signalling and prostate cancer", *Prostate Cancer and Prostatic Diseases*, 2005, 8, pp. 119–126.

[15] J. Paez and W. Sellers, "PI3K/PTEN/AKT pathway. A critical mediator of oncogenic signaling", *Cancer Treat Res*., 2003, 115, pp. 145 167.

[16] F. Al-Shahrour, R. Díaz-Uriarte and J. Dopazo, "FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes", *Bioinformatics*, 2004, 20(4), pp. 578-580.