Mutual Information-Based Evolution of Hypernetworks for Brain Data Analysis

Eun-Sol Kim and Jung-Woo Ha
Biointelligence Laboratory
School of Computer Sci. & Eng.
Seoul National University
Seoul 151-742, Korea
{eskim, jwha}@bi.snu.ac.kr

Wi Hoon Jung , Joon Hwan Jang and Jun Soo Kwon
College of Medicine
Seoul National University
Seoul 151-742, Korea
kwonjs@snu.ac.kr

Byoung-Tak Zhang
Biointelligence Laboratory
School of Computer Sci. & Eng.
Seoul National University
Seoul 151-742, Korea
btzhang@bi.snu.ac.kr

Abstract— **Cortical analysis becomes increasingly important for brain research and clinical diagnosis. This problem involves a combinatorial search to find the essential modules among a large number of brain regions. Despite several statistical approaches, cortical analysis remains a formidable challenge due to high-dimensionality and sparsity of data. Here we describe an evolutionary method for finding significant modules from cortical data. The method uses a hypernetwork which is encoded as a population of hyperedges, where hyperedges represent building blocks or potential modules. We develop an efficient method for evolving the hypernetwork using mutual information to generate essential hyperedges. We evaluate the method on predicting intelligence quotient (IQ) levels and finding potential significant modules on IQ from brain MRI data consisting of 62 healthy adults with over 80,000 measured points (variables). The experimental results show that our information-theoretic evolutionary hypernetworks improve the classification accuracy by 5~15%. Moreover, it extracts significant cortical modules that distinguish high IQ from low IQ groups.**

Keywords: cortical thickness, hypernetworks, mutual information, classifier, human intelligence

I. INTRODUCTION

It is well known that cortex of brain is related to human cognitive ability [1]. Several studies have identified cortical regions and investigated their functions. For example, it was shown that cortical thickness was related to genders [2], psychological diseases [3, 4], and human intelligence levels [5]. Since several cortical regions influence on determining the cognitive characteristics concurrently, it is essential to use the method based on searching combinatorial space for cortical analysis. Studying cortical thickness is typically based on statistical approaches. However, these approaches have limits of time costs due to high dimensionality of the data and higher-order relationships among several regions. For solving this limitation, evolutionary algorithms can be an effective alternative.

Here, we use hypernetworks for finding potential cortical modules influencing on the human cognitive characteristics [6]. Our method consists of two parts: (1) making a population-based classifier by using hypernetworks which are a higher-order probabilistic graphical model and (2) finding potentially significant subpatterns on determining cognitive characteristics by analyzing the evolved population. A hypernetwork classifier is a population consisting of large number of individuals which are arbitrary combinations of more than two attributes (data variables or features) and play a role of building blocks or sub-modules. By the definition of the individuals, a hypernetwork model can model higher-order relationships among several attributes. Also, each individual has its own fitness value i.e. weight, and the fitness value reflects the discriminative capability. Unlike genetic algorithms (GA) [7] and genetic programming (GP) [8], the population plays a role of solution and the individuals participate in determining to predict the data patterns, as in learning classifier system (LCS) [9]. In contrast to an LCS, however, the individuals in a hypernetwork are generated by not genetic operators such as crossover or mutation but combining randomly selected attribute values from given data instances. Hypernetworks are evolved by changing the composition of individuals through eliminating individuals with low fitness from the population and generating new ones every generation. This strategy keeps the diversity of the population.

In conventional hypernetworks, however, all attributes have the same probability to be selected for generating individuals without any prior knowledge. However, it may be inefficient in case of modeling high-dimensional data since the search space increases exponentially in the size of attributes. In this paper, we propose a novel method for evolving hypernetworks using mutual information (MI) between the data attributes and the class label. Because MI is a measure which reflects conditional independency between two random variables, attributes with higher MI value can be considered to be significantly related with the data patterns. Because MI assumes the conditional independency among attributes, however, we cannot consider higher-order relationships among attributes. Therefore, we use MI values as the probability for selecting attributes in

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (No.2010-0017734, 2010-0018950), in part by IT R&D Program of MKE/KEIT (KI002138, MARS), in part by the Industrial Strategic Technology Development Program (10035348) funded by the Korean government (MKE), and the BK21-IT Program.

978-1-4244-7833-0/11/$26.00 ©2011 IEEE

generating individuals. From this prior process, we can enhance the efficiency of evolving hypernetworks since the larger MI attributes have the higher probability to be selected in generating individuals. Although we added process of MI computation, the time cost of evolving hypernetworks does not increase since MI values are calculated as a preprocessing only once.

For experiments, we use a cortical thickness dataset consisting of 62 healthy adults with 31 inferior and 31 superior IQ values. While "inferior" means the IQ values of below 109, "superior" denotes the IQ value of over 121. Each sample consists of 81,924 attributes which are cortical thickness values of the measure points. Because a hypernetwork is a model for data with discrete variables, we conduct a preprocessing by converting the real-valued measured points into three-level discretized values from zero to two. Then, we evolve the hypernetworks with the training data set and classify the IQ levels of the test set. After that, the evolved classifiers with high classification accuracy is used to find significant cortical points by analyzing the hyperedges based on the occurrence of attributes in the individuals with high fitness values. Experimental results show that the hypernetwork classifiers outperform, by 5~15% in classification accuracy, other machine learning algorithms including support vector machines (SVMs), decision trees (J48), and naïve Bayesian networks. In addition, the hypernetworks can identify the important cortical points to distinguish IQ groups.

The rest of the paper is organized as follows. The next section describes the brain thickness data and the problem statement. In Section 3, the hypernetwork model is explained along with the techniques we have developed to make its evolution more efficient. The results of experimental results are reported in Section 4. Finally, we concluded in Section 5.

II. Brain Data Analysis

A. Intellectual ability and cortical thickness

Lots of researches have been studied to investigate the relationships between developing of cortical thickness and intellectual ability including reading, writing and arithmetic ability as well as IQ. Choi *et al.* showed that crystallized components (*gC*) of intelligence was strongly related to cortical thickness by statistical analyses of regions of interests (ROIs) with brain structural magnetic resonance imaging data [5]. Shaw *et al.* presented relationships between intellectual ability including 3R (reading, writing, and arithmetic abilities) and cortical development in children and adolescent on changing of time [10]. Narr *et al*. showed that intellectual ability is associated with variations in prefrontal and posterior temporal cortical thickness considering sex [11]. Most of studies on cortical thickness use statistical approaches and they showed regions of brain which have the difference in statistical meanings. Dissimilar to previous studies, we propose a classifier to distinguish the level of intelligence with cortical thickness using an evolutionary higher-order probabilistic graphical model i.e., hypernetwork and we find the significant regions for determining the intelligence levels by analyzing the learned models.

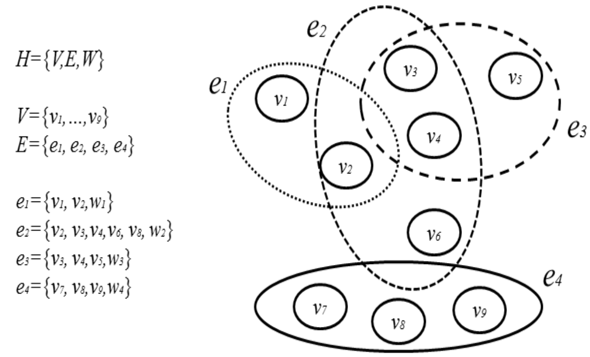

Figure 1. An example of the hypernetwork. A hypernetwork is composed of vertices set *V*, hyperedge set *E* and weight set *W*. One hyperedge can include more than two verteices and has its own weight

III. Hypernetworks

A. Conventional hypernetworks

A hypernetwork is a probabilistic graphical model for representing higher-order relationship between factors [6]. A hypernetwork *H* is formally defined as a set of tuples of (*V*, *E*, *W*) where *V, E*, and *W* are sets of vertices (in this study, a pair of cortical thickness point and its value), hyperedges, and weights respectively. A hypernetwork makes an edge with more than two vertices for representing higher-order relationship. We call these edges 'hyperedge' and the 'order' of a hyperedge denotes the number of vertices in one hyperedge. Moreover, the hyperedge represents the significance with its weight. In the viewpoint of the genetic algorithm, a hypernetwork is a population consisting of individuals which are hyperedges. Fig. 1 shows a graphical representation of a hypernetwork.

Since the hyperedge is a higher-order combination of vertices, we can consider the hyperedge as a segment of data information. For this reason, hypernetworks model is beneficial to retrieve the patterns by storing information segments. We define the amount of information to retrieve a pattern as energy function as follow:

$$\varepsilon(\mathbf{x}^{(n)}) = -\sum_{i=1}^{|E|} w_i f_i(\mathbf{x}^{(n)}) \tag{1}$$

where $\mathbf{x}^{(n)}$ i1s the *n*-th data, f_i is the feature function of the *i*-th hyperedge. The feature function is defined with an identity function which yields 1 if a hyperedge matches $\mathbf{x}^{(n)}$ and 0 in the other case. For example consider a data instance $\mathbf{x} = \{x_1=0, x_2=1, x_3=1, x_4=0, x_5=1\}$, and two hyperedges $e_1=\{x_1=0, x_2=1, x_3=1\}$ and $e_2 = \{x_2=1, x_3=1, x_5=0\}$. Then, $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ yield 1 and 0 respectively because $\mathbf{x}$ matches e_1 but does not match e_2. When a set of data $D=\{\mathbf{x}^{(n)}\}_{n=1}^{N}$ is given, the probability of retrieving the pattern in data by a hypernetwork, $P(D|h)$ is defined with (1) as follows:

$$P(D \mid h) = \prod_{n=1}^{N} P(\mathbf{x}^{(n)} \mid h) = \prod_{n=1}^{N}\left\{\frac{1}{Z(h)}\exp\left\{-\beta\varepsilon(\mathbf{x}^{(n)})\right\}\right\} = \prod_{n=1}^{N}\left\{\frac{1}{Z(h)}\exp\left\{\beta\sum_{i=1}^{|E|} w_i f_i(\mathbf{x}^{(n)})\right\}\right\} \quad (2)$$

where h is the given hypernetwork and $Z(h)$ is a partition function. The partition functions $Z(h)$ is defined as follows:

$$Z(h) = \sum_{\mathbf{x}^{(n)} \in D} \exp\left\{\beta\sum_{k=1}^{|E^*|} w_k f_k(\mathbf{x}^{(n)})\right\} \quad (3)$$

where E^* is a set of all possible hyperedges by combining vertices in a given data set.

In the case of classification, when a data set $D = \{\mathbf{x}^{(n)}, y^{(n)}\}_{n=1}^{N}$ is given with a vector of attributes $\mathbf{x}$ and a class label y, $P(D|h)$ is defined as follows:

$$P(D \mid h) = \prod_{n=1}^{N} P(\mathbf{x}^{(n)}, y^{(n)} \mid h) = \prod_{n=1}^{N} P(y^{(n)} \mid \mathbf{x}^{(n)}, h) P(\mathbf{x}^{(n)} \mid h) \quad (4)$$

Since $P(y^{(n)} \mid \mathbf{x}^{(n)}, h)$ is the probability to predict the label correctly for the given pattern, the probability is defined for all data instances as follow:

$$P(y \mid \mathbf{x}, h) = \frac{1}{N}\sum_{n=1}^{N}\left\{1-\delta(\hat{y}^{(n)}, y^{(n)})\right\} \quad (5)$$

where $\hat{y}^{(n)}$ is the predicted label for the given pattern by the model h such that $\delta(\hat{y}^{(n)}, y^{(n)}) = (\hat{y}^{(n)} - y^{(n)})^2$.

Hypernetworks have been applied to various classification problems such as bioinformatics and multimedia mining with competitive performances [6, 12-14].

B. *MI-based sampling method*

Conventional hypernetworks based on random selection are not efficient for problems which deal with high dimensional data and small number of samples. The reason is that the combinatorial space is too complex and huge due to the curse of dimension. In this study, we use mutual information between IQ level and thickness of cortical points for choosing attributes in making hyperedges to improve the efficiency of the evolution.

Mutual information which is a correlation between two variables is defined as follow:

$$MI(X,Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \quad (6)$$

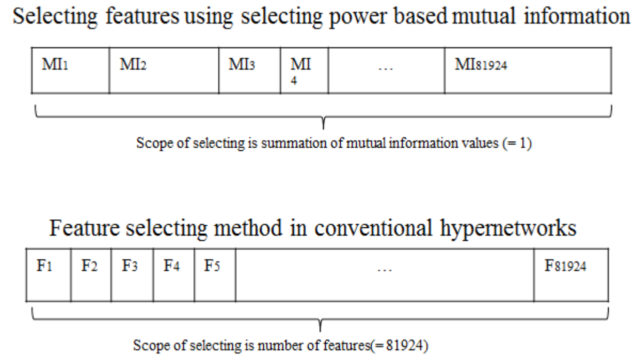

Figure 2. Difference between MI based selection and random selection in generating a hyperedge. While conventional hypernetworks choose one vertex from uniform distribution of 81,924 features (below one), hypernetworks with mutual information based sampling method choose one vertex along with mutual information value. So each vertex has different probability about choosing.

where $p(x, y)$ is the joint probability distribution function of two random variables X and Y, and $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y respectively. According to the definition, MI is 0 when two variables are conditionally independent. In this study, X is a measured point of cortical thickness, Y denotes the IQ level.

We give selecting power based on mutual information to features for improving efficiency. To do this, we define probability to whole features in proportion to mutual information for generating hyperedges. Using mutual information values of features, we make a roulette wheel. Then, we choose a feature randomly from whole roulette wheel. The difference between conventional hypernetworks and modified hypernetworks is shown in Fig. 2

C. *Region-based sampling method*

For generating hyperedges, in addition to MI values, we use a domain knowledge which specifies the cortical data. It is well known that brain regions which are in the neighborhood may be related to functional with high probability. We use this fact to generate the hyperedges. That is, when one feature is

```
GenerateHyperedge()
fi : the i-th feature
ei : i-th hyperedge
Hi : i-th hypernetwork
R : number of region
N : number of features included in a region

ei ← {};
for(j ←1 to | ei |/N)
    fj ← randomly selected feature with MI;
    ei ← ei ∪ fj ;
    for(k←1 to N)
        ei ← ei ∪ f(l +k) ;
Hi ← Hi ∪ ei ;
```

Figure 3. Algorithm for generating a hyperedge.

selected randomly, neighboring attributes are also selected. In other words, the hyperedges in hypernetworks are combinations of regions, rather than those of vertices. In Fig. 4, region-based sampling method is described. From section 3.B and 3.C, the sampling method of generating hyperedges is described as pseudo code in Figure 3.

D. Evolution of hypernetwork

Until now, we discussed the sampling method of making the hyperedges. In this section, we look at evolution process of the hypernetworks. The hyperedges were made from data samples with mutual information. For a data instance, hyperedges are generated as a specified number i.e., sampling rate (*SR*). Therefore, a hypernetwork has $N \times SR$ hyperedges where N is the size of training data set. During the evolutionary process, the hyperedges were evaluated how well the hyperedges represent the property of a data set. All hyperedges were compared with data samples for this evaluation. If indices and values of hyperedge are same with these of data sample, we examine class labels of hyperedge and data sample. If the class labels are same, a hyperedge gains a count of correctness ($\#_c$). If the class labels are different, a hyperedge gains a count of incorrectness ($\#_w$). With these values, we define a weight of hyperedge, i.e., a fitness function as

$$w = \frac{\alpha}{\#_w} + \beta \times \#_c \quad (7)$$

In (7), α is parameter related to incorrect prediction and a hyperedge with smaller $\#_w$ value gets larger weight. Also, β is a parameter concerned with correct prediction. In this study, we set α and β to 200 and 1 respectively.

The goal of evolving hypernetworks is to find the optimal composition of hyperedges. The hyperedges are individuals for recalling stored patterns from subpatterns or discriminating distinguished patterns.

The diversity of hypernetwork is guaranteed by changing individuals in every generation. To find the optimal composition of individuals in a restricted population size, low-weighted hyperedges were eliminated from the population and new hyperedges were generated in the same amount of removed hyperedges. Since the population plays a role of the solution and the composition of individuals determines the property of population in hypernetwork models, changing the composition enhance the diversity of population. We controlled the number of hyperedges of replacement according to the iteration, that is, more hyperedges are discarded in early iteration phase and are gradually decreased. Because we decrease the number of replacement hyperedges according to the iteration, replacement amount is a function about iteration number.

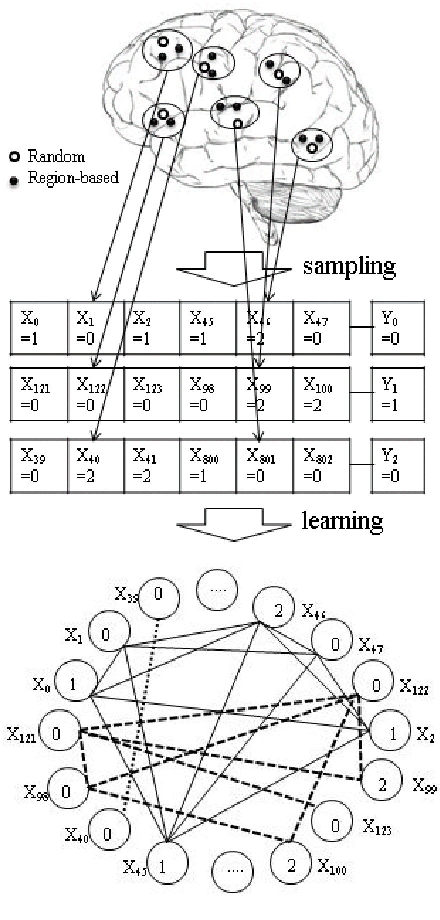

Figure 5. Process of building a hypernetwork from cortical thickness. The region-based sampling method is described above. One vertex (feature) selected randomly and neighboring vertices selected together. These vertices make the hyperedge.

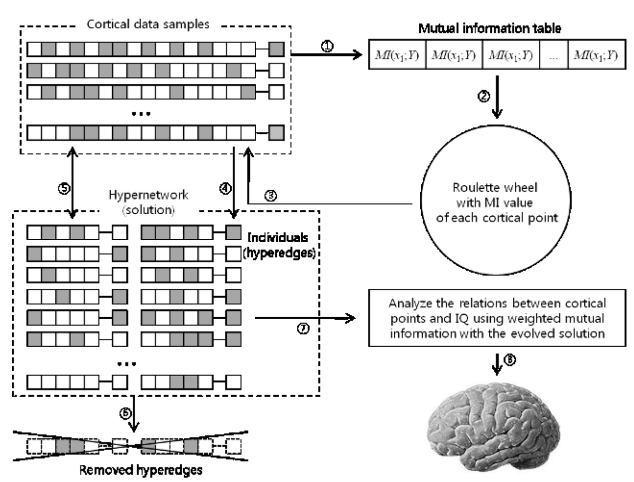

Figure 4. Flow of evolving hypernetworks and finding modules. Evolution of hypernetworks. First of all, we make the hyperedges from cortical data samples with selecting power about mutual information vales(①, ②, ③). Then we evaluate weight values of the hyperedges by comparing with data samples (④ , ⑤). After calculating weight values, hyperedges which low ranked are discarded and same number of hyperedges are made(⑥). We repeat this evolution process about 30 times and get solution (optimal population). With this hypernetworks, we find significant modules determining on IQ (⑦ , ⑧)

$$R(iter) = \frac{M - m}{\exp(\frac{iter - 1}{C})} + m \quad (8)$$

where M and m are parameters about replacement ratio and C denote convergence speed. After elimination, the same amounts of new hyperedges are generated randomly. This process of generation, evaluation and elimination is repeated for specified generations which are 30 epochs in this study.

In summary, as an evolutionary model, hypernetworks can be considered a population-based model with the hyperedges corresponding to individuals. Unlike to conventional evolutionary models such as genetic algorithm and particle swarm optimization, a hypernetwork structure of an entire population plays a role of the solution. Thus, a hypernetwork can be considered an evolutionary graphical population-based model.

Dissimilar to estimated distribution analysis (EDA) and learning classifier system (LCS), each individual is generated by selecting and combining attribute values from given data instances. Therefore, there is at least one instance matched by each hyperedge.

E. The method for finding significant modules

For finding significant modules to determining IQs, we used learned hypernetworks. After learning, hypernetworks contain hyperedges which are given high weight. Then, we evaluate each feature's weight sum as follow:

$$S(x_i) = \sum_{E_j \in P} \sum_{x_i \in E_j} w_j \quad (9)$$

where $S(x_i)$ denote feature x_i's weight sum value, P is subset of 100 high-ranked hyperedges of evolved hypernetwork, E_j is a hyperedge in P and w_j is the weight of E_j. After all, we arrange features along with weight sum values then choose 100 features of high rank as a significant module about IQ.

IV. EXPERIMENTAL RESULTS

A. Data Explanation and Experimental Condition

Data are consisted of 62 samples, including 31 samples of low IQ and 31 samples of high IQ. Each sample has 81,924 features which are cortical thickness values. While thickness values are real values, hypernetworks is a model for discrete values. So we converted these values to integer values, i.e. 0, 1 and 2. We normalized 81,924 features, then values in (-∞, -1) range are changed to 0, values about [-1, 1] are changed to 1, remaining values are changed to 2.

From implemental point of view, some parameters are needed such as the order of hyperedges, sampling rate, the number of repetition of learning and weighting parameter [8]. These conditions of this study are in Table 1.

We performed the same experiment with other machine learning methods, such as SVM with second polynomial kernel, decision tree with J48 and naïve Bayes. In all experiments, classification problems are performed with 5-fold crossover validation and repeated 10 times. As a method for comparisons, we used Waikato Environmental for Knowledge Analysis (Weka) [15].

TABLE I. PARAMETERS OF HYPERNETWORKS.

Conditions	values
Hyperedge order	6
Sampling rate	100
Repeating number	30
Weighting parameter	variable
Max replacement ratio(M)	0.6
Minimum replacement ration(m)	0.1
Convergence speed	10
Maximum number of generations	30

B. Experimental Results

Proposed method was applied to a problem classifying 62 healthy adults consisting of 31 inferior and 31 superior cases according to IQ based on 81,924 discretized values of cortical thickness.

Table 2 presents an averaged accuracy and a standard deviation of classification for each algorithm. According to the results, it is confirmed that most classifying methods cannot show good classification accuracy. We can guess that the sparsity of data is the cause that made the classifiers yield low accuracy. Nevertheless, hypernetworks showed higher performance than the other machine learning methods by 5~15%. Furthermore, modified hypernetworks shows best performance, which means that the region-based sampling method is proper to data of brain structure. From a viewpoint of a hypernetwork, the classification accuracy difference was 13% between the random sampling method and the mutual information based sampling method.

Calculating mutual information has no overhead because we can calculate before executing the algorithm. One calculating process is sufficient. With a low computation cost, the mutual information plays a great role in our problem giving selecting pressure to hypernetworks for finding an optimal solution.

In Fig.6, the learning curves about random sampling method and of mutual information based sampling method are presented. In this graph, we can make sure that as the epochs evolve, the accuracy converges to high values.

We can guess that MI-based HNs show higher

TABLE II. COMPARISON OF IQ CLASSIFICATION PERFORMANCE.

classifier	**mHN**	SVM	DT	NB	HN
Accuracy(%)	**74.36**	69.19	58.06	66.13	61.13
std	**1.93**	3.84	4.59	3.24	2.89

Accuracy values are average values from 10 repetitions. HN denotes an original hypernetworks and mHN is modified hypernetworks. DT means decision tree, and in this study, J48 is used for decision tree. In addition, second polynomial kernel is used to SVM.

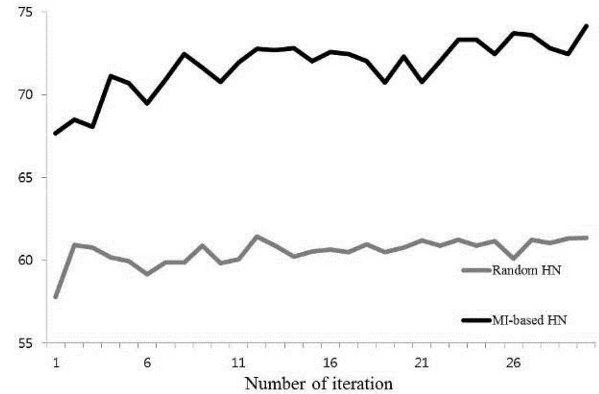

Figure 6. Learning curve on classification on test data set as increasing of generation.

classification accuracy than random HNs by analyzing the composition of hyperedges. To analyze the composition, we define a score of each attributes as follows:

$$w_{n_i c} = \sum_{j=1}^{|E|} \sum_{n_i \in E_j, y_j = c} w_j \tag{10}$$

$$s(x_i) = \frac{\left| \sqrt{w_{00} \cdot w_{21}} - \sqrt{w_{20} \cdot w_{01}} \right|}{w_{00} + w_{01} + w_{20} + w_{21}} \tag{11}$$

In (10), $c \in \{0, 1\}$, $n_j \in \{0, 1, 2\}$ where n_j is a value of the j-th feature x, and y_i is a class label in E_i. According to (10), $s(x)$ reflects the discriminating ability of each feature. Fig. 7 shows the difference of accumulated value of $s(x)$ in MI-based HNs from random HNs. According to Fig. 7, the score is concentrated on small number of significant features in MI-based HNs compared with random HNs. With the proposed method, the significant features appear more frequently in hyperedges through biased feature selection for generating hyperedges based on MI. We can guess that this property enhances the efficiency of evolution and improves classification accuracy.

For the next one, we suggest new method of feature

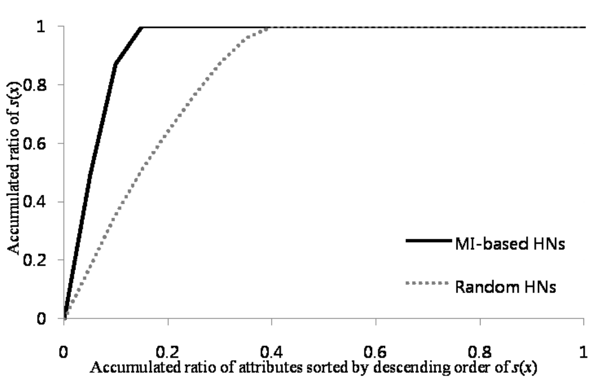

Figure 7. Comparison of accumulated attribute score of MI-based HNs with random HNs. X-axis is the ratio of attributes with high s(x) value to all attributes and Y-axis denotes ratio of accumalated value of high s(x)s to all s(x) .

TABLE III. SELECTED FEATURES FROM MUTUAL INFORMATION BASED HYPERNETWORKS.

512(34020)	35644(22036)	62923(75674)	48027(38075)
513(7211)	35645(3173)	62924(57349)	48028(25793)
1036(64938)	33606(4363)	66598(1324)	63120(10440)
1037(57359)	33607(12070)	66599(11357)	63121(10032)
8373(68510)	8962(15536)	6000(35071)	33650(6765)
8374(49613)	8963(73242)	6001(35072)	33651(16032)
9390(3186)	59225(9165)	33158(32738)	60139(8085)
9391(4280)	59226(5098)	33159(28010)	60140(75024)
12347(35184)	45443(15764)	63189(503)	25554(4527)
12348(42965)	45444(57910)	63190(115)	25555(4528)
12466(18588)	68989(4123)	73603(28521)	54431(2738)
12467(24469)	68990(4124)	73604(24330)	54432(2739)
14197(74590)	11248(32192)	71712(3999)	73055(14897)
14198(56543)	11249(32130)	71713(4042)	73056(26914)
16163(1642)	39748(52639)	11711(75693)	69279(2486)
16164(106)	39749(81259)	11712(65862)	69280(1380)
20040(6601)	28698(71417)	76039(12424)	78478(46233)
20041(22295)	28699(58960)	76040(7625)	78479(43215)
23006(553)	34184(16914)	28102(2262)	3003(33773)
23007(1636)	34185(39571)	28103(2263)	3004(61773)
23823(14396)	48391(65315)	78417(27091)	64166(70227)
23824(18367)	48392(65316)	78418(35975)	64167(70228)
24995(20653)	57591(22808)	30057(2439)	24707(3953)
24996(8321)	57592(22809)	30058(10195)	24708(1834)
25542(5318)	22977(30455)	12925(31849)	35894(23197)

Selected features are arranged. Number in the parenthesis is rank of mutual information value. Although we use mutual information value for making the hyperedges, hypernetworks explore combinatorial problem space so there are differences between selected features and high valued mutual information features.

selection. In Table 3, 100 features are selected from learned hypernetworks are arranged. Then, to validate the effect of feature selection, we remade samples with 100 features and classified the samples. Additionally, we compared with other feature selection method. Results of this classification are presented in Table 4. First, classification accuracy with the selected features from hypernetworks increased in all methods. It explained that hypernetworks has a function as feature selection. Furthermore, from Table 4, hypernetworks as a feature selection model is competitive even though other method like SVM has better performance. However, we can find meaning which is hypernetworks find optimal combinations of features during learning phase very naturally

TABLE IV. ACCURACY OF CLASSIFICATION WITH 100 SELECTED FEATURES IN TABLE 3

	mHN	SVM	DT	NB
Accuracy (%)	**83.55**	75.16	65.48	85.48

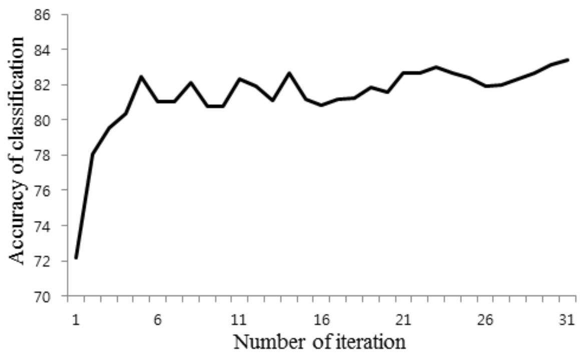

Figure 8. Learning curve of hypernetworks with selected data on classification on test data set as increasing of generation.

without extra algorithm. Also, we can guess that naïve Bayes classifier shows the best accuracy in Table 4 because most of selected attributes by hypernetworks have high MI values and both mutual information and naïve Bayes classifier assume the conditional independency among attributes.

Finally, a graph about learning process with selected features is shown in Fig.8. There are fluctuating wave in the early learning phase but as iteration unfolds, the accuracy is converged to higher value.

V. CONCLUSION

In this study, we propose the sampling method for evolving hypernetworks using MI. The proposed method is applied to analyze brain cortical data efficiently and find significant cortical module related on human IQ. Experimental results present our method not only improves the classification accuracy but also outperforms other standard classifiers. As a future works, we will extend this study to figure out significant module affected on psychological diseases such as schizophrenia, OCD (obsessive compulsive disorder) and depression. Through consecutive studies, we expect that these found modules contribute to diagnosis of brain diseases. Although hypernetworks show good performance, it has relatively heavy time complexity because it searches much larger solution space. Thus it remains to study that adding efficiency to hypernetworks.

ACKNOWLEDGMENT

E.-S. Kim thanks Soo-Jin Kim and Joon-Shik Kim for discussion.

REFERENCES

[1] G Bush, P Luu, MI Posner, "Cognitive and emotional influences in anterior cingulate cortex", *Trends in cognitive sciences*, vol. 8, pp.539-546, 2004.

[2] I. Kim *et al*, "Gender difference analysis of cortical thickness in healthy young adults with surface-based methods", *Neuroimage,* 31(2006), pp. 31-38,.

[3] S.Haller *et al.*, "Can cortical thickness asymmetry analysis contribute to detection of at-risk mental state and first-episode psychosis?: A pilot study1", *Radiology*, vol.250, pp.212-221, 2009

[4] Y. Yang, A. Raine, P. Colletti, A. W. Toga, and K. L. Narr, "Abnormal temporal and prefrontal cortical gray matter thinning in psychopaths," *Molecular Psychiatry*, vol. 14, pp. 561-562, 2009.

[5] Y. Y. Choi *et al.,* "Multiple bases of human intelligence revealed by cortical thickness and neural activation,", *The journal of neuroscience*, vol. 28(41), pp. 10323-10329, 2008.

[6] B.-T. Zhang, "Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory", *IEEE Computational Intelligence Magazine*, 3(3):49-63, 2008.

[7] J. H. Holland, *Adaptation in natural and artificial systems,* MIT Press, 1992.

[8] John R. Koza *et al., Genetic Programming IV: Routine Human-Competitive Machine Intelligence*, Kluwer Academic Publishiers, 2003

[9] J. H. Holland *et al.*,"What is a learning classifier system?", *Lecture Notes in Computer Science*, vol.1813, pp. 3-32, 2000

[10] P. Shaw *et al.* "Intellectual ability and cortical development in children and adolescents," *Nature letters*, vol. 440(30), pp.676-679, 2006.

[11] K. L. Narr *et al.*"Relationships between IQ and regional cortical gray matter thickness in healthy adults," *Phil. Cerebral Cortex*, vol. 17(9), pp. 2163–2171, 2007.

[12] J.-W. Ha, J.-H. Eom, S.-C. Kim, and B.-T. Zhang, "Evolutionary hypernetwork models for aptamer-based cardiovascular disease diagnosis", *The Genetic and Evolutionary Computation Conference (GECCO 2007)*, pp.2709-2716, 2007.

[13] J.-W. Ha *et al.*, "Gender classification with cortical thickness measurement", *IEEE International Conference on Fuzzy Systems (Fuzz IEEE 2009)*, pp. 41-46, 2009.

[14] S.-J. Kim, J.-W. Ha, B. Lee, and B.-T. Zhang, "Evolutionary layered hypernetworks for identifying microRNA-mRNA regulatory modules", *IEEE World Congress Computational Intelligence (WCCI-CEC 2010)*, pp. 2299-2306, 2010.

[15] University of Waikato New Zealand, "Waikato Environment for Knowledge Analysis (Weka) Version 3.4.11", Available: http://www.cs.waikato.ac.nz/ml/weka/index.htm