Understanding Microarray Data through Applying Competent Program Evolution

Moshe Looks
Department of Computer Science and Engineering
Washington University
Saint Louis, MO 63130, USA
moshe@metacog.org

Ben Goertzel, Lucio de Souza Coelho,
Mauricio Mudado, Cassio Pennachin
Biomind LLC
1405 Bernerd Place
Rockville, MD
ben@goertzel.org,
{lucio, mauricio, cassio}@vettalabs.com

ABSTRACT

A number of researchers have used genetic programming (GP) to analyze gene expression microarray data; here it is asked whether the use of an alternate program evolution technique, MOSES (meta-optimizing semantic evolutionary search), can improve upon GP's results in this domain. Based on our results so far, the answer appears to be a resounding yes. We first consider standard supervised classification for two microarray datasets: one involving aging vs. young brains, and the other involving lymphoma types. On these datasets, MOSES is shown to learn classifiers with test accuracies comparable to those obtained by SVMs, and significantly superior to those of GP with a similar configuration (and even more superior to those of GP run under a more standard configuration). Furthermore, unlike SVMs, the classification rules produced by MOSES are still simple and comprehensible to humans – often strikingly simple. Finally, biological interpretation is presented regarding the "important genes" used most often in the ensemble of MOSES models learned; we show that many of the genes identified in this way reflect knowledge available in the literature, and others constitute plausible hypotheses.

Categories and Subject Descriptors

I.2.2 [**Artificial Intelligence**]: Automatic Programming – *Program synthesis*

General Terms

Algorithms, Design, Experimentation

Keywords

Empirical Study, Heuristics, Optimization, Representations

1. OVERVIEW

A variety of statistical and machine learning methodologies and algorithms have been applied to the study of gene expression microarray data. Among the more powerful methodologies applied has been supervised categorization. A wide variety of microarray analysis problems can be cast in the format of supervised categorization, but the most common sort of problem to be analyzed this way is the classification of individuals into pre-defined categories based on their gene expression profiles (e.g. the categories may represent Case vs. Control, or may represent different forms of cancer, etc.). Among other applications, supervised categorization models learned according to this methodology can be used as diagnostic rules to predict whether an individual possesses a disease or other phenotypic condition.

Within the supervised categorization approach, a number of categorization algorithms have been used, including genetic programming (GP). Each algorithm has its advocates, and different algorithms have their strengths and weaknesses. Overall, however, it is fair to say that the reported results of gene expression microarray classification using Support Vector Machines (SVMs), for example, have generally been stronger than those reported using GP. However, it has been found that GP is sometimes preferable to SVMs and other methodologies in this context because of the transparency of the models it produces. Studying the GP models themselves, can give valuable insights into exactly how the classification is being performed, which can lead to biological insights [12, 13, 14].

Along these lines, Goertzel et al. [14] introduced a novel methodology in which ensembles of GP classification models are statistically analyzed to yield biological insight. More simply, one may index the "important genes" that occur most often across the model ensemble. These genes often possess particular biological importance relative to the category under study.

This paper reports results obtained analyzing microarray data using the MOSES (meta-optimizing semantic evolutionary search) algorithm [22], an alternative to GP, still falling within the category of evolutionary algorithms but utilizing a number of ideas going beyond standard GP, including an estimation-of-distribution algorithm, a novel approach to program tree normalization, and a deme-management approach that conceptually hybridizes evolutionary learning with multistart stochastic local search. The motivation for this work was the hypothesis that MOSES might outperform GP in this context, both in classification

GECCO '07 London, UK

accuracy and model simplicity; and that MOSES might also succeed in providing "important gene" lists with substantial biological relevance.

This hypothesis was validated via experiments on a number of microarray datasets. Here we report results on two of these datasets in detail. One of the datasets poses the problem of distinguishing between types of lymphoma based on gene expression data [34, 36, 13]; the other poses the problem of classifying between young and old human brains [23, 13, 14]. In our discussion, we aim to illustrate not only the quantitative merit of the results, but also their apparent biological meaningfulness.

2. CONCERNING MICROARRAY GENE EXPRESSION DATA

Microarray profiling technology is a means of measuring expression levels across thousands of genes; a gene's expression level is a rough indicator of its activity level (i.e., translation into protein). Microarray gene expression profiling thus generates, for some particular cell type in a patient, a noisy high-dimensional numerical vector profiling activity in the particular genes chosen to be sampled. Similar technology can also be used in other variations – e.g. to detect single nucleotide polymorphisms (SNPs, the smallest possible variations in the genome), or to measure ESTs (expressed sequence tags) rather than genes, or to measure protein rather than gene expression. A general review of this area may be found in [31].

Three general machine-learning issues are highly relevant to any supervised classification approach to microarray analysis: (1) how to deal with a huge number of problem variables (potentially tens of thousands); (2) how to deal with noisy continuous data; (3) how to avoid overfitting to the data.

1. A number of approaches to handling large feature spaces have been proposed in conjunction with program evolution for learn classifiers, most commonly either preliminary runs to identify promising features [19, 1], or inexpensive preprocessing heuristics [18, 13]. We follow the later, selecting simply the most-differentiating features to use in all experiments (similar to the signal-to-noise approach found to perform well in [18]).

2. Past studies have applied GP to learn classifiers that operate directly on gene expression levels [16, 18, 1, 19, 13], which are manipulated with arithmetic and other operators. However, microarray data are very noisy [20, 39]; we conjecture that, by and large, considering the actual gene expression levels without restriction will increase the danger of overfitting and make the learning of good models more difficult (by expanding the hypothesis space is a non-productive manner). Accordingly, we consider gene expression levels as Boolean features determined by median-thresholding. This also eliminates concerns regarding data scaling and outliers.

3. Most microarray datasets contain relatively few cases (19 and 77 for the datasets studied in this paper, for instance), often comparable to or significantly less than the number of features. We thus adapt an aggressive approach to avoiding overfitting, by incorporating high parsimony pressure into the fitness function. Observe that in the limiting case where the number of features (n) is much greater than the number of training cases (m), one may obtain a "perfect" classifier g_1 OR g_2 OR ... OR g_m by ensuring that every feature g_i is highly expressed for only the ith case. Such a model (containing $2m - 1$ nodes) will of course have negligible explanatory and predictive value. Accordingly, our fitness function for classifiers is $TP + TN - s/2$, where s is the number of nodes in the classifier, TP is the number of true positives, and TN is the number of true negatives. Ties are broken by preferring shorter programs. This ensures that all complexity in classifiers is "earned"; an overfit classifier such as the one given above will be assigned negligible fitness.

3. MOSES

Meta-Optimizing Semantic Evolutionary Search (MOSES) is a recently developed program evolution system distinguished by two key mechanisms: (1) exploiting semantics (what programs actually mean) to restrict and direct search; and (2) limiting the recombination of programs to occur within bounded subspaces (constructed on the basis of program semantics). This has been shown to lead to superior performance and scalability in comparison to current purely syntactic techniques (local search, genetic programming, etc.) [22]. Furthermore, the evolved programs do not suffer from any kind of "bloating", and are generally quite comprehensible. This is of particular interest for applications such as microarray analysis, where it is useful to know not only *that* a method achieves good performance, but to understand *how.*

Recombination in MOSES occurs within parameterized program subspaces called *representations.* A specialized representation-building process is used which heuristically exploits semantics (e.g., $\forall x,\ x + 0 \rightarrow x$) to create meaningful parameters to vary, based on some initial *exemplar program.* This process has two essential steps.

In the first step, the exemplar program is reduced to a normal form to eliminate redundancy. For Boolean formulae for example, we use an extended version of Holman's elegant normal form [17, 22]. An example of the reduction process is shown in Figure 1. It is important to note that elegant normal form is *hierarchical*, consisting of alternating nested levels of conjunction and disjunction. The familiar conjunctive and disjunctive normal forms are flat; reduction to one of them destroys a formula's structure and may cause an exponential blowup in formula size.

In the second step, many possible parameters, defined based on making small variations to the exemplar program, are considered. For example, a new variable might be added to an existing conjunction, or an existing variable removed or negated. These are structured in such a way that formulae can grow or shrink unboundedly. To limit the number of parameters actually created, each possible parameter setting is experimentally set (with all other parameters kept unmodified), and the resultant program reduced to normal form. If the number of symbols in the reduced program is less than in the original program, then the parameter is eliminated as redundant. This is because any large reduction in program length can be expressed as a set of smaller length reductions. Of course, this is a heuristic; a transformation that had a redundant effect on its own might lead to a unique program if applied in conjunction with other transformations.

Parameter variation is directed by the hierarchical Bayesian optimization algorithm (hBOA) [28], an advanced estimation-of-distribution algorithm that dynamically learns problem decompositions encoded as Bayesian networks with local structure. The hBOA has been well-studied both theoretically and empirically, and found to be highly scalable and robust. Strong results have been obtained for hard problems such as Ising spin glasses and MaxSAT [27], as well as a real-world telecommunication problem, even with a "bad" encoding [32].

A population of programs associated with a common representation is called a *deme*, and a set of demes (together spanning an arbitrary area within program space in a patchwork fashion) will be referred to as a *metapopulation*. MOSES operates on a metapopulation, adaptively creating, removing, and allocating optimization effort to various demes. Deme management is the second fundamental *meta* aspect of MOSES, after (and above) representation-building; it essentially corresponds to the problem of effectively allocating computational resources to competing regions, and hence to competing programmatic organizational-representational schemes.

A basic sketch of MOSES is as follows:

1. Construct an initial representation of very small programs (i.e., with the empty program as the exemplar) and use it to generate an initial random sampling. Add this deme to the metapopulation.

2. Select a deme from the metapopulation and iteratively update its sample, as follows:

   (a) Select some promising programs from the deme's existing sample to use for probabilistic modeling by hBOA, according to the fitness function. Ties in the fitness function are broken by preferring smaller programs.

   (b) Considering the promising programs as collections of parameter settings, generate new collections of parameter settings by applying hBOA optimization.

   (c) Convert the new collections of parameter settings into their corresponding programs, evaluate their fitness scores, and integrate them into the deme's sample, replacing less promising programs.

3. For each of the new programs that meet the criteria for creating a new deme, if any:

   (a) Construct a new representation centered around the program (the deme's exemplar), and use it to generate a *new* random sampling of programs, producing a new deme.

   (b) Integrate the new deme into the metapopulation, possibly displacing less promising demes.

4. Repeat from step 2.

The details of MOSES (e.g., the deme creation criterion) are omitted for brevity, and may be found in [22]. All of the parameters for MOSES and their settings may be found in Table 1; none were varied across the problems studied herein. The extensive theory and practice of "competent optimization" [15, 26] allow reasonable parameter settings to be straightforwardly extrapolated to MOSES.

Table 1: Parameters for MOSES and their settings.

Description	Value
population size (N) is $c{\cdot}n^{1.05}$ (derived in [26]), where n is the size of the representation in bits	$c = 5$
window size (w) for restricted tournament replacement in the hBOA, which implements niching	$w = min(0.05N, n)$, based on [26]
tournament size (selection pressure) used for model-building in the hBOA	2, based on [26]
complexity penalty for hBOA model-building (to avoid overfitting)	$log_2(N)$, derived in [26]
a deme is terminated after t_1 generations of hBOA, or t_2 generations with no improvement in the best fitness	$t_1 = n$, $t_2 = \lceil\sqrt{N/w}\rceil$, based on [29]

4. CLASSIFICATION RESULTS

Experimental results are presented for SVMs, GP operating on floating-point gene-expression levels (henceforth "standard GP"), GP on Boolean features (henceforth "Boolean GP"), and MOSES on Boolean features. All results are given for test data only, and based on 10-fold cross-validation.

The SVM results for aging brains are taken from [13]. These essentially used a meta-tasking framework where algorithm parameters (e.g., the SVM kernel) were varied across a large number of runs (on portions training data) to determine the best configuration, which was then applied on all of the training data. SVM results for lymphoma are taken from [36], which used a similar approach based on applying 10-fold cross-validation *within* the training set to determine the best SVM parameters, which were then used to learn a model based on all of the training data.

The function set of "standard GP" was the traditional $\{+, -, *, /, sin, cos, log, exp\}$ (adding logical operators as well did not improve performance). The function set of "Boolean GP" and MOSES was simply the Boolean connectors *AND*, *OR*, and *NOT*. The feature preprocessing and fitness function for MOSES and Boolean GP were as described in Section 2 (taking the top 50 features based on differentiation) – runs of GP where the number of features was varied between 10 and 1000 did not produce improved performance. Standard GP used the same preprocessing, but a more standard fitness function – maximization of the number of cases correct classified, with ties broken by preferring the smaller program (i.e., lexical parsimony pressure).

The maximal number of fitness function evaluations allowed per run of MOSES and GP was $100,000$ (GP was run for 50 generations with a population size of 2000). For all algorithms and configurations ten independent runs we carried out per fold. GP used a maximum tree depth of 5, a 1% mutation rate, unitary elitism (the best program is cloned to the next generation, all other programs are generated via

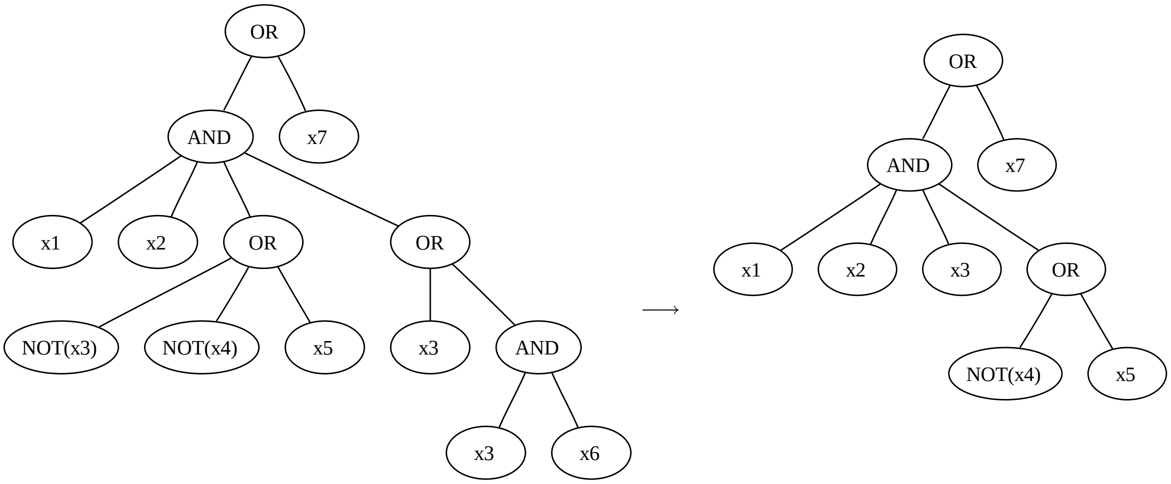

Figure 1: A redundant Boolean formula (left) and its equivalent in elegant normal form (right).

Table 2: Supervised classification average test set accuracy for max-prior (the frequency of the most common classification), support vector machines, genetic programming, MOSES, and MOSES with voting.

Technique	Lymphoma	Aging Brain
Max-prior	75.3%	52.9%
SVM	97.5%	95.0%
Standard GP	84.3%	85.8%
Boolean GP	86.9%	91.1%
MOSES	93.5%	95.3%
MOSES+Voting	98.6%	100%

crossover), and tournament selection with tournaments of size 2. Note that tournament selection was similarly used in MOSES within the hBOA algorithm for selecting instances for probabilistic modeling, also with size 2 tournaments.

The results discussed below are summarized in Table 2, which shows average test set performance over all folds and runs for the smallest best[1] programs found (i.e., averages of $10 \cdot 10 = 100$ runs based on 100,000 evaluations per run). To attempt to exploit the diversity of classification models produced both within and across different runs by MOSES, a simple averaging scheme was tried whereby the top classifiers in a fold were given a single vote each for classifying the test data, with ties broken by choosing the most frequent category in the dataset. This approach (as well as other more sophisticated ensemble methods) was applied to GP classifiers in [18]. The row in Table 2 for "MOSES+Voting" is thus an average of ten sets of runs (each set consisting of ten runs).

4.1 Diagnosing Lymphoma Type

Lymphoma is a broad term for a number of particular kinds of cancer occurring in lymphocytes (a type of white blood cell). Various subtypes require different treatments, but are difficult to distinguish clinically. Discovering good classifiers between lymphoma types based on microarray expression levels is of interest for yielding tests, as well as understanding which genes in fact have roles in lymphoma and how they may interact. We report here on work with the dataset of [34] containing 58 cases of diffuse large B-cell lymphoma, and 19 cases of follicular lymphoma (77 total). This dataset has a dimensionality of 6817. Additional results are reported on this dataset for K-nearest neighbor and two kinds of neural network [36].

Table 3: The simplest best-performing classifier models learned by MOSES for the lymphoma data.

Model	Accuracy
ALDOA OR NOT(GPR18) OR PGAM1)	98.7%
MT2A OR TubulinBeta2 OR not(GPR18)	98.7%
ENO1 OR LDHA OR NOT(GPR18)	98.7%
NOT(GPR18) OR PGAM1 OR GM2A	98.7%
TubulinBeta2 OR NOT(GPR18) OR MT2A	98.7%

Average test accuracy for MOSES was 93.5%, slightly worse that for SVMs (97.5%), and significantly better than K-nearest neighbor (87%), backpropagation neural networks (89.6%), standard GP (84.3%), and probabilistic neural networks (67.7%). Boolean GP achieves results comparable to K-nearest neighbor and backprop, but markedly better than probabilistic neural networks and somewhat better than standard GP. When voting is used, the MOSES results become slightly better than for SVMs (98.6% vs. 97.5%).

Unlike with SVMs, we can very easily look inside and across the classifiers learned by MOSES to determine which features are used in classification, how, and how often. On this dataset, the best classifiers found by MOSES, taken in aggregate, used 33 out of the 50 possible features. Many of the MOSES models found are extremely simple; Table 3 gives the five top scoring MOSES models over the entire dataset (the simplest models achieving the top score are shown).

[1] Best on the training set, of course.

Table 4: The simplest best-performing classifier models learned by MOSES for the aging brains data.

Model	Accuracy
HBB OR SPON1	100%
HBB OR AF009767	100%
not(RAB6) OR SickleCellBetaGlobin	100%
S100B OR KIAA0878	100%
NOT(HSPKCB2A) OR NOT(HSHEPLF)	100%

4.2 Classifying Aging Brains

The next dataset to be examined here concerns around 11,000 genes, and was gathered to study the effects of aging on the human brain [23]. A subset of the data consisting of 11 "old" samples (over 72 years old) and 9 "young" samples (under 42 years old) was used for supervised classification with GP and SVMs in [13]. The configuration and findings here are essentially the same as for the lymphoma experiments described above; MOSES significantly outperforms standard GP (95.3% vs. 85.8%), and achieves parity with SVMs (95.0%). Boolean GP slightly trails MOSES and SVMs (possibly not a significant gap, given the small size of the dataset), but is better than standard GP. MOSES with voting achieves 100% accuracy on all folds for this this dataset (some of the individual classifiers learned by MOSES without voting were 100% accurate as shown in Table 4, but *average* accuracy was lower).

Fully 46 out of the 50 possible features were used by MOSES in constructing high-quality classifiers of small size. This is unsurprising; it should not be very difficult to distinguish between the brains of people under 42 and those over 72! What *is* of interest here is understanding the different expression patterns active in young vs. old brains, a topic that we address in Section 5.2. Table 4 gives the top five models, which are quite simple (again here, we show the simplest of the top models).

5. IMPORTANT FEATURES ANALYSIS

As well as providing diagnostic power, classification rules may yield significant biological insight in themselves. However, generally speaking each individual classification rule only addresses one or a few of the many aspects distinguishing the categories under study. A more holistic understanding of the difference between the categories may be obtained by performing various kinds of statistical analysis on model ensembles.

Perhaps the simplest form of model ensemble analysis is simply to calculate those features that occur most frequently in the model ensemble. These are called the "important features" and are often found to have particular biological relevance. In this section we describe the most important features found by MOSES in the two datasets under study, and explore the relevance of these features to the categories under study according to the available biological literature. As compared to accuracy statistics, this is a more qualitative way of exploring the meaningfulness of the MOSES results, but it is equally important, as one of the main uses of microarray data at present is to stimulate qualitative insight on the part of research biologists. What we see in the results elaborated here is that many of the important features identified by MOSES are already known to be extremely relevant to the categories under study, whereas some others are plausibly likely to be relevant based on known biology, and hence form the basis of hypotheses potentially worthy of further biological investigation.

Table 5: Most important features for lymphoma data based on MOSES classification.

Feature	Frequency
Homo Sapiens G protein-coupled receptor 18 (GPR18)	93%
Tubulin, Beta 2	35%
Homo Sapiens aldolase A, fructose-bisphosphate (ALDOA)	30%
Homo Sapiens phosphoglycerate mutase 1, brain (PGAM1)	23%
Homo Sapiens pyruvate kinase, muscle (PKM2)	20%
Homo Sapiens lactate dehydrogenase A (LDHA)	15%
Homo Sapiens integrin, alpha 4 (antigen CD49D, alpha 4 subunit of VLA-4 receptor) (ITGA4)	13%
Homo Sapiens enolase 1, alpha (ENO1)	13%
Homo Sapiens GM2 ganglioside activator (GM2A)	10%
Homo Sapiens interferon, gamma-inducible protein 30 (IFI30)	8%

5.1 Important Features for Lymphoma Data

Table 5 lists the ten most important features for the lymphoma data; we discuss the top five features here in depth.

Feature 1: Homo Sapiens G protein-coupled receptor 18 (GPR18). A variety of G protein coupled receptors have been implicated in several types of human cancer [21].

Feature 2: Tubulin, Beta 2. Microtubules are found to have roles in apoptosis and cell division. It has been shown that a protein c-Myc, which interacts with alpha-tubulin, is mutated in Burkitt lymphoma and this somatic mutation may have roles in oncogenic activity [30, 25].

Feature 3: Homo Sapiens aldolase A, fructose-bisphosphate (ALDOA) Aldolase B is downregulated in liver hepatocarcinomas [35]. Aldolase A has been shown to be upregulated [4]. Thus it seems plausible that they may have roles in lymphomas: this is a hypothesis worthy of investigation.

Feature 4: Homo Sapiens phosphoglycerate mutase 1 (brain) (PGAM1). It has been recently shown that the oncogene P53 may regulate metabolic pathways by several novel genes, including PGAM. Also, PGAM expression levels are altered in brain, colon, breast, lung and liver cancers [6, 9, 8, 7].

Feature 5: Homo Sapiens pyruvate kinase, muscle (PKM2). Pyruvate kinase PK has been shown to have its ex-

Table 6: Most important features for aging brain data based on MOSES classification.

Feature	Frequency
Homo Sapiens hemoglobin, beta (HBB)	19%
HSU01828 Human microtubule-associated protein 2 (MAP2) mRNA	11%
HSU24152 Human p21-activated protein kinase (Pak1)	11%
Homo Sapiens ELAV (embryonic lethal, abnormal vision, Drosophila)-like 2 (Hu antigen B) (ELAVL2)	10%
Homo Sapiens Thy-1 cell surface antigen (THY1)	10%
Homo Sapiens spondin 1, extracellular matrix protein (SPON1)	9%
AF009767 Homo Sapiens cDNA	9%
Human calmodulin (CALM1) gene	9%
Homo Sapiens mRNA for KIAA0878 protein	9%
Human sickle cell beta-globin mRNA	9%

pression altered in human and canine lymphomas [42, 41].

The features utilized in the best MOSES models (33 out of the 50 total features presented to MOSES, as mentioned above) corresponded to 18 distinct genes.[2] We accessed the Gene Ontology (GO) [38] to find biological process information for this set of genes by using GO::TermFinder [3]. Examining the most frequent process label, five out of the 18 genes (27.8%) belonged to "glucose metabolic process" (LDHA, ALDOA, ENO1, PGAM1, and PKM2), with a very low p-value (2.06e-05). As discussed above, it is known that the oncogene P53 regulates several metabolic pathways. Also, as recently reviewed by Shaw [33], it is known that cancers produce changes in glucose metabolism.

5.2 Important Features for Aging Brain Data

Table 6 lists the ten most important features for the aging data. Again, will examine the top five features in depth.

Feature 1: Homo Sapiens hemoglobin, beta (HBB). Hemoglobin levels are well known to decline with age [11, 10].

Feature 2: HSU01828 Human microtubule-associated protein 2 (MAP2). Di Stefano et al. [37] note that "data suggest that in areas of the brain involved in memory acquisition and consolidation, MAP2-dependent neuroplasticity and structural integrity are significantly decreased in aging."

Feature 3: HSU24152 Human p21-activated protein kinase (Pak1). Bell and Sharpless [2] note that "inhibition of p21 seems to protect against the age-promoting effects of telomere dysfunction on stem cells, without an attendant rise in tumorigenesis."

Feature 4: Homo Sapiens ELAV (embryonic lethal, abnormal vision, Drosophila)-like 2 (Hu antigen B) (ELAVL). HuR is a protein from the ELAV family, that has been found to be related to senescence [40]. Is HuB related? This is an interesting hypothesis.

Feature 5: Homo Sapiens Thy-1 cell surface antigen (THY1). Here no direct relationship was found in the literature. However, speculative evidence for a relationship exists, forming an interesting hypothesis. It is known that anti-Thy-1 is involved in reducing neurite outgrowth [5], which is impaired in old people. One is thus to the hypothesis that THY1 could be involved in problems with neurite outgrowth in old peoples' brains.

The features utilized in the best MOSES models (46 out of the 50 total features) corresponded to 30 distinct genes. As with the important lymphoma genes, we again searched for processes involving these genes according to the Gene Ontology. The "apotosis" subset of the GO biological process contained 6 out of the 30 genes (20%) – ANXA4, S100B, PAK1, VDAC1, THY1, and PRKCZ – and scored the lowest p-value: 0.02372. Apoptosis is known to be a critical process that intermediates senescence and death and also is known to have several roles in age related diseases and neurodegenerative processes (reviewed by [24]).

6. CONCLUSION

We began the research reported here with the hypothesis that MOSES would provide higher classification accuracy and more compact models, as compared to GP, when applied to gene expression microarray data. The results described here validate this hypothesis significantly. MOSES provides classification accuracy close to that of SVMs, and also extremely compact classification models. As seen by the direct and Gene Ontology-based analyses of the most important features, the models learned by MOSES are using biologically relevant features for the respective cases (lymphoma and aging brain).

In the future, we plan to apply MOSES to additional microarray datasets. It will be interesting to see if the performance boost that simple voting provided herein caries over to other datasets as well, or whether more sophisticated ensemble methods are required. We are also planning on studying MOSES with other function sets (e.g., containing arithmetic functions), and are applying MOSES to difficult real-world problems other domains as well.

7. REFERENCES

[1] A. A. Almal, A. P. Mitra, R. H. Datar, P. F. Lenehan, D. W. Fry, R. J. Cote, and W. P. Worzel. Using genetic programming to classify node positive patients in bladder cancer. In *Genetic and Evolutionary Computation Conference*, 2006.

[2] J. F. Bell and N. E. Sharpless. Telomeres, p21 and the cancer-aging hypothesis. *Nature Genetics*, 2007.

[3] E. I. Boyle, S. Weng, J. Gollub, H. Jin, D. Botstein, J. M. Cherry, and G. Sherlock. GO::TermFinder – open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*, 2004.

[2] It is not uncommon for different features in a microarray dataset to map to the same gene, as they may represent alternative genetic sequences or different subsets of a long genetic sequence.

[4] G. Castaldo, G. Calcagno, R. Sibillo, R. Cuomo, G. Nardone, L. Castellano, C. D. V. Blanco, G. Budillon, and F. Salvatore. Quantitative analysis of aldolase a mRNA in liver discriminates between hepatocellular carcinoma and cirrhosis. *Clinical Chemisty*, 2000.

[5] C. H. Chen, Y. J. Chen, C. J. Jeng, S. H. Yang, P. Y. Tung, and S. M. Wang. Role of PKA in the anti-Thy-1 antibody-induced neurite outgrowth of dorsal root ganglionic neurons. *Journal of Cellular Biochemistry*, 2006.

[6] C. A. Corcoran, Y. Huang, and M. S. Sheikh. The regulation of energy generating metabolic pathways by p53. *Cancer biology and therapy*, 2006.

[7] N. Durany, J. Joseph, E. Campo, R. Molina, and J. Carreras. Phosphoglycerate mutase, 2,3-bisphosphoglycerate phosphatase and enolase activity and isoenzymes in lung, colon and liver carcinomas. *British Journal of Cancer*, 1997.

[8] N. Durany, J. Joseph, F. F. Cruz-Sanchez, and J. Carreras. Phosphoglycerate mutase, 2,3-bisphosphoglycerate phosphatase and creatine kinase activity and isoenzymes in human brain tumours. *British Journal of Cancer*, 1997.

[9] N. Durany, J. Joseph, O. M. Jimenez, F. Climent, P. L. Fernández, F. Rivera, and J. Carreras. Phosphoglycerate mutase, 2,3-bisphosphoglycerate phosphatase, creatine kinase and enolase activity and isoenzymes in breast carcinoma. *British Journal of Cancer*, 2000.

[10] R. Eisenstaedt, B. W. Penninx, and R. C. Woodman. Anemia in the elderly: current understanding and emerging concepts. *Blood Reviews*, 2006.

[11] W. B. Ershler, S. Sheng, J. McKelvey, A. S. Artz, N. Denduluri, and J. T. *et al.* Serum erythropoietin and aging: a longitudinal analysis. *Journal of the American Geriatric Society*, 2005.

[12] R. J. Gilbert, J. J. Rowland, and D. B. Kell. Genomic computing: explanatory modelling for functional genomics. In *Genetic and Evolutionary Computation Conference*, 2000.

[13] B. Goertzel, L. Coelho, C. Pennachin, I. Goertzel, M. Queiroz, F. Prosdocimi, and F. Lobo. Learning comprehensible classification rules from gene expression data using genetic programming and biological ontologies. In *Computational Intelligence Methods for Bioinformatics and Biostatistics*, 2006.

[14] B. Goertzel, C. Pennachin, L. Coelho, and M. Mudado. Identifying complex biological interactions based on categorical gene expression data. In *Congress on Evolutionary Computation*, 2006.

[15] D. E. Goldberg. *The Design of Innovation*. Kluwer Academic, 2002.

[16] S. Hengpraprohm and P. Chongstitvatana. Diffuse large b-cell lymphoma classification using genetic programming classifier. In *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, 2005.

[17] C. Holman. *Elements of an Expert System for Determining the Satisfiability of General Boolean Expressions*. PhD thesis, Northwestern University, 1990.

[18] J. H. Hong and S. B. Cho. The classification of cancer based on DNA microarray data that uses diverse ensemble genetic programming. *Artificial Intelligence in Medicine*, 2006.

[19] W. B. Langdon and B. F. Buxton. Genetic programming for mining DNA chip data from cancer patients. *Genetic Programming and Evolvable Machines*, 2004.

[20] M. T. Lee, F. C. Kuo, G. A. Whitmore, and J. Sklar. Importance of replication in microarray gene expression studies: Statistical methods and evidence from repetitive cdna hybridizations. *Proceedings of the National Academy of Sciences*, 2000.

[21] S. Li, S. Huang, and S. B. Peng. Overexpression of G protein-coupled receptors in cancer cells: involvement in tumor progression. *International Journal of Oncology*, 2005.

[22] M. Looks. *Competent Program Evolution*. PhD thesis, Washington University in St. Louis, 2006.

[23] T. Lu, Y. Pan, S. Y. Kao, C. Li, I. Kohane, J. Chan, and B. A. Yankner. Gene regulation and DNA damage in the aging human brain. *Nature*, 2004.

[24] M. P. Mattson. Neuronal life-and-death signaling, apoptosis, and neurodegenerative disorders. *Antioxidants and Redox Signaling*, 2006.

[25] J. Niklinski, G. Claassen, C. Meyers, M. A. Gregory, C. J. Allegra, F. J. Kaye, S. R. Hann, and M. Zajac-Kaye. Disruption of myc-tubulin interaction by hyperphosphorylation of c-Myc during mitosis or by constitutive hyperphosphorylation of mutant c-Myc in Burkitt's Lymphoma. *Molecular and Cellular Biology*, 2000.

[26] M. Pelikan. *Hierarchical Bayesian Optimization Algorithm*. Springer, 2005.

[27] M. Pelikan and D. E. Goldberg. Hierarchical BOA solves Ising spin glasses and MAXSAT. Technical report, University of Illinois at Urbana-Champaign, 2003.

[28] M. Pelikan and D. E. Goldberg. A hierarchy machine: Learning to optimize from nature and humans. *Complexity*, 2003.

[29] M. Pelikan and T. Lin. Parameter-less hierarchical BOA. In *Genetic and Evolutionary Computation Conference*, 2004.

[30] L. A. Porter and J. M. Lee. Alpha-, beta-, and gamma-tubulin polymerization in response to dna damage. *Experimental Cell Research*, 2001.

[31] J. Quackenbush. Computational analysis of microarray data. *Nature Reviews Genetics*, 2001.

[32] F. Rothlaud, D. E. Goldberg, and A. Heinzl. Bad codings and the utility of well-designed genetic algorithms. Technical report, University of Illinois at Urbana-Champaign, 2000.

[33] R. J. Shaw. Glucose metabolism and cancer. *Current Opinions in Cell Biology*, 2006.

[34] M. A. Shipp, K. N. Ross, P. Tamayo, A. P. Weng, J. L. Kutok, and R. C. T. A. *et al.* Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nature Medicine*, 2002.

[35] H. Song, S. L. Xia, C. Liao, X. L. Li, Y. F. Wang,

T. P. Li, and M. J. Zhao. Genes encoding Pir51, Beclin 1, RbAp48 and aldolase b are up or down-regulated in human primary hepatocellular carcinoma. *World journal of gastroenterology*, 2000.

[36] A. Statnikov, C. F. Aliferis, I. T. D. Hardin, and S. Levy. A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. *Bioinformatics*, 2004.

[37] G. D. Stefano, T. Casoli, P. Fattoretti, N. Gracciotti, M. Solazzi, and C. Bertoni-Freddari. Distribution of map2 in hippocampus and cerebellum of young and old rats by quantitative immunohistochemistry. *Journal of Histochemistry and Cytochemistry*, 2001.

[38] The Gene Ontology Consortium. Gene Ontology: Tools for the unification of biology. *Nature Genetics*, 2000.

[39] Y. Tu, G. Stolovitzky, and U. Klein. Quantitative noise analysis for gene expression microarray experiments. *Proceedings of the National Academy of Sciences*, 2002.

[40] W. Wang, X. Yang, V. J. Cristofalo, N. J. Holbrook, and M. Gorospe. Loss of HuR is linked to reduced expression of proliferative genes during replicative senescence. *Molecular and Cellular Biology*, 2001.

[41] T. Washizu, D. Azakami, M. Bonkobara, M. Washizu, and T. Arai. Changes in activities of enzymes related to energy metabolism in canine lymphoma cells. *Journal of veterinary medical science*, 2005.

[42] H. L. Yilmaz, A. Tanyeli, H. Ozusaglam, L. Kayrin, B. Antmen, and H. I. Sasmaz. Erythrocyte pyruvate kinase activity during chemotherapy in children with leukemia and lymphoma. *Bratislavské lekárske listy*, 2002.