NEW GENETIC PROGRAMMING HYPOTHESIS SEARCH STRATEGIES FOR IMPROVING THE INTERPRETABILITY IN MEDICAL DATA MINING APPLICATIONS

Michael Affenzeller, Christian Fischer, Gabriel Kronberger, Stephan M. Winkler, Stefan Wagner

Upper Austria University of Applied Sciences
School for Informatics, Communications, and Media
Heuristic and Evolutionary Algorithms Laboratory
Softwarepark 11, 4232 Hagenberg, Austria

michael.affenzeller@fh-hagenberg.at, gabriel.kronberger@heuristiclab.com, stephan.winkler@fh-hagenberg.at, christian.fischer@students.fh-hagenberg.at, stefan.wagner@heuristiclab.com

ABSTRACT
In this paper we describe a new variant of offspring selection applied to medical diagnosis modeling which is designed to guide the hypothesis search of genetic programming towards more compact and more easy to interpret prediction models. This new modeling approach aims to combat the bloat phenomenon of genetic programming and is evaluated on the basis of medical benchmark datasets. The classification accuracies of the achieved results are compared to those of published results known from the literature. Regarding compactness the models are compared to genetic programming prediction models achieved without the new offspring selection variant.

Keywords: Medical data mining, Genetic programming, Offspring selection.

1. INTRODUCTION

Genetic Programming (GP) plays an outstanding role among the various data-mining techniques from the field of machine learning and computational intelligence. Due to its model representation, GP is able to produce human interpretable models without taking any assumptions about the nature of the relationship. Also GP-based data analysis has quite good generalization properties. Furthermore, GP is able to simultaneously evolve the structure and the parameters of a model with implicit feature selection. The combination of these aspects makes GP a very powerful and also robust method for various data analysis tasks.

Nevertheless, there are still some aspects in the practical application of GP-based data analysis which leave room for improvement:

GP-based data analysis suffers from the fact that – even if the models are interpretable – the results are often quite complex and far from being unique. Often the models are still quite complex because of the tendency of GP to bloat and also because of introns which is counterproductive in terms of interpretability as well.

One of the reasons for genetic bloat is identified in the tendency of GP to favor more complex hypothesis structures for explaining equivalent correlations (Luke and Panait, 2006). The new proposed offspring selection variant aims to counteract this phenomenon by including additional offspring selection criteria: Instead of only considering the error measure, the enhanced offspring selection (OS) criteria also consider the complexity as well as the number of variables of the candidate hypothesis in order to decide, whether or not a new candidate hypothesis is accepted for the next generation. By this means the hypothesis search should be lead not only to models with more predictive power, but also to more compact and more unique models which are easier to interpret. Especially the latter aspects are considered as important in the field of medical data mining where the domain expert should be able to analyze not only the statistical properties of the prediction models but also their usefulness in the medical context.

The effects of the new introduced extended offspring selection formulation for data based modeling are discussed for medical benchmark datasets from the UCI machine learning repository[1].

The rest of the paper is organized as follows: Section 2 describes standard offspring selection with its parameters, its main characteristics, and how it can be integrated into genetic programming. Section 3 discusses specific extensions of offspring selections designed for data based modeling which aim to guide hypothesis search to simpler and easier to interpret models. In section 4 the characteristics of the extended offspring selection variant are discussed exemplarily for medical benchmark data sets. Finally, section 5 summarizes the achieved results and points out future perspectives for future research.

[1] http://archive.ics.uci.edu/ml/

2. OFFSPRING SELECTION

The basic principles of offspring selection have been described in (Affenzeller and Wagner 2005). In the meanwhile, offspring selection has been discussed for several benchmark problems from the field of combinatorial optimization, function optimization and data based modeling. The following description of standard offspring selection is taken from (Wagner et al 2010) where the aspect of mutation in offspring selection has been discussed in further detail.

In general, offspring selection consists of the following steps:

At first parents are selected for reproduction either randomly or in any other well-known way of genetic algorithms (e.g., fitness proportional selection, linear rank selection, tournament selection). After crossover and optionally mutation have been applied to create a new child solution, another selection step is introduced which considers the success of the applied reproduction procedure. The goal of this second selection step is to continue the search process with offspring which surpass their parents' quality. Therefore, a new parameter called success ratio (*SuccRatio*) is introduced. The success ratio defines the relative amount of members in the next population that have to be generated by successful mating (crossover, mutation).

Additionally, it has to be defined when a solution is considered to be successful: Is a child solution better than its parents, if it surpasses the fitness of the weaker, the better, or some kind of mean value of both? For this purpose a parameter called comparison factor (*cf*) is used to define the success criterion for each created solution as a weighted average of the quality of the worse and the better parent (i.e., if the comparison factor is 0, successful solutions at least have to be better than the worse parent, and if it is 1 they have to outperform the better parent).

For steering the comparison factor, the authors decided to introduce a cooling strategy which is similar to simulated annealing. Following the basic principle of simulated annealing, an offspring only has to surpass the fitness value of the worse parent in order to be successful at the beginning of the search process (*cf* is initialized with 0 or a rather small value). While evolution proceeds solutions have to be better than a fitness value continuously increasing between the fitness of the weaker and the better parent (*cf* is increased in each generation until it reaches 1 or a rather high value). As in the case of simulated annealing, this strategy leads to a broader search at the beginning, whereas at the end the search process becomes more and more directed.

After the amount of successful solutions in the next generation has reached the success ratio, the remaining solutions for the next generation (i.e., (1-*SuccRatio*)·|*POP*|) are taken from the pool of solutions which were also created by crossover and mutation but did not necessarily reach the success criterion. The actual selection pressure *ActSelPress* at the end of a single generation is defined by the quotient of individuals that had to be created until the success ratio was reached and the number of individuals in the population:

$$ActSelPress = \frac{|POP| \cdot SuccRatio + |POOL|}{|POP|} \quad (1)$$

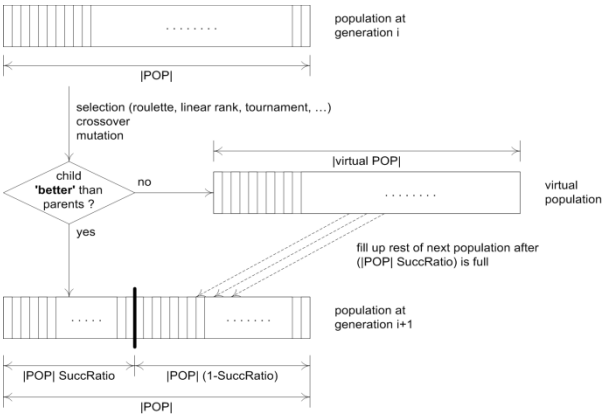

Figure 1: Flowchart of Offspring Selection

Figure 1 shows these basic steps of offspring selection and how they are embedded into a classical genetic algorithm.

Furthermore, an upper limit for the selection pressure (*MaxSelPress*) can be defined as another parameter which states the maximum number of children (as a multiple of the population size) that might be created in order to fulfill the success ratio. With this additional parameter offspring selection also provides a precise detector for premature convergence: If the algorithm cannot create a sufficient number of successful solutions (*SuccRatio*·|*POP*|) even after *MaxSelPress*·|*POP*| solutions have been created, convergence has occurred and the algorithm can be stopped.

If OS is applied with the parameters *cf* = 1 and *SuccRatio* = 1, it is commonly referred to as strict OS. Strict OS has the property that children with worse quality compared to its better parent are automatically discarded and therefore the overall quality of the population steadily increases.

3. NEW OFFSPRING SELECTION FOR DATA ANALYSIS

The standard variant of offspring selection as discussed in Section 2 implements the offspring selection criterion purely on the basis of solution quality. For data based modeling the offspring selection criterion is usually based on the mean squared error (MSE) for classification problems and on the coefficient of correlation R^2 or MSE for regression problems. This means that an offspring solution candidate is considered successful if the MSE or R^2 fitness measure of the candidate offspring is better than the respective fitness measure of the parent solutions. This means that only the quality of the models is considered and not the

simplicity of interpretability of the involved solution candidates.

The main idea of the here discussed offspring selection extension is that not only the quality of the candidate models should be considered but also its compactness in order to combat the bloat. From theoretical bloat analyses (Luke and Panait, 2006) it is known that genetic programming based hypothesis search tends to find rather more complex models in order to achieve the same model quality. Therefore, it seems reasonable to include also model complexity measures into the offspring selection criterion. In that sense, an offspring candidate model is considered successful if it surpasses not only the model quality of its own parents but is also not more complex than its parent models. As model complexity measures we have introduced the number of nodes as well as the number of used input features (variables) of the involved structure trees. In that sense an offspring solution is considered successful not only if it surpasses the model quality of its own parents; additionally, the offspring model must not be more complex than its parent model. Similar to the standard case of offspring selection we have to decide if the criterion compares the resulting offspring to the better, the weaker, or to some intermediate value. In order to handle this aspect we introduce new model complexity comparison factors: Let q_b and q_w be the model qualities of the better and the weaker model, l_b and l_w the length of the shorter (better) and the more complex (worse) model. As a model complexity measure we here use the number of nodes of the two parent structure trees. For the number of variables or the two parent models let v_b be the model using less variables (better) and v_w the model using more variables (worse). Similar to the standard case of offspring selection comparison factors cf_q , cf_l , and $cf_v \in [0, 1]$ define the certain thresholds which distinguish a successful offspring from an unsuccessful offspring based on the characteristic features of the parents. But in contrast to original offspring selection a candidate offspring has to fulfill three criteria instead of one in order to be accepted; it does not only have to be better but also less complex and use less variables. Similar to the standard case a comparison factor of 0 means that it is sufficient to surpass the certain characteristics of the worse parent whereas a comparison factor of 1 means that the candidate offspring has to be better than the better of the two parents.

Obviously it becomes harder to evolve successful offspring solution candidates which results in higher selection pressures on the one hand; on the other hand due to the preference to simpler and more compact models genetic diversity can hardly emerge. Therefore, the additional offspring selection criteria concerning the model complexities and the number of variables should better not be activated from the start. First studies have shown that the new OS variant works a lot better if the additional criteria are activated not until genetic diversity can emerge which usually happens after about one or two dozen of iterations. Algorithmically we have considered this aspect by introducing further parameters which specify two time windows tw_l and tw_v which specify when the additional length and number of variables criterion should be active.

Summarizing the above mentioned aspects, the here discussed first version of a new offspring selection criterion dedicated to the reduction of bloat can be stated as follows (in the minimization variant for MSE as quality):

$$\begin{gathered}\boldsymbol{isSuccessful}(\boldsymbol{co}, \boldsymbol{p1}, \boldsymbol{p2}, \boldsymbol{gen}) \Leftrightarrow \\ \left[\boldsymbol{q}(\boldsymbol{co}) < \boldsymbol{q_w} + \boldsymbol{cf_q}(\boldsymbol{q_w} - \boldsymbol{q_b})\right] \boldsymbol{and} \\ \left[\left(\boldsymbol{l}(\boldsymbol{co}) < \boldsymbol{l_w} + \boldsymbol{cf_l}(\boldsymbol{l_w} - \boldsymbol{l_b})\right) \boldsymbol{or} (\boldsymbol{gen} \notin \boldsymbol{tw_l})\right] \boldsymbol{and} \\ \left[\left(\boldsymbol{v}(\boldsymbol{co}) < \boldsymbol{v_w} + \boldsymbol{cf_v}(\boldsymbol{v_w} - \boldsymbol{v_b})\right) \boldsymbol{or} (\boldsymbol{gen} \notin \boldsymbol{tw_v})\right]\end{gathered}$$

This means that in order to be considered as successful, a candidate offspring (*co*) has to be better than some intermediate fitness value of its own parents (defined by cf_q) in any case. Additionally, in some predefined time window tw_l an offspring does not only have to be better but also at least as compact than some intermediate compactness value of its own parents and in the same sense there is a time window tw_l where the candidate offspring have to use not more variables than some intermediate value of variables used by its parent models. Therefore, also the actual generation *gen* has to be considered in order to decide if one of the two time windows is active at the moment.

The empirical discussion of the next section compares the achieved results on the basis of standard classification benchmark datasets for generating prediction models for breast cancer, thyroid, and melanoma.

4. RESULTS

The configurations used for the test runs in table 2, 4 and 6 with Melanoma, Thyroid and Wisconsin datasets are shown in table 1. If not otherwise stated the time windows include all generations. The maximum solution length is 100, maximum solution height is 12. Up to 1000 Generations are created with a maximum permitted selection pressure of 555.

#	Configuration
1	cfq=0,cfl=0
2	cfq=1,cfl=0
3	cfq=0..1,twq=1..100,cfl=0
4	cfq=0,cfl=-100..1,twl=20..100
5	cfq=1,cfl=-100..1,twl=20..100
6	cfq=0..1,cfl=-100..1,twl=20..100
7	cfq=0,cfl=1,twl=20..100
8	cfq=1,cfl=1,twl=20..100
9	cfq=0..1,cfl=1,twl=20..100
10	cfq=0,cfh=0,twh=20..100
11	cfq=1,cfh=0,twh=20..100
12	cfq=0..1,twq=1..100,cfh=0,twh=20..100
13	cfq=0,cfh=-100..1,twh=20..100
14	cfq=1,cfh=-100..1,twh=20..100
15	cfq=0..1,cfh=-100..1,twh=20..100

16	cfq=0,cfh=1,twh=20..100
17	cfq=1,cfh=1,twh=20..100
18	cfq=0..1,cfh=1,twh=20..100
19	cfq=0,cfv=0,twv=20..100
20	cfq=1,cfv=0,twv20..100
21	cfq=0..1,twq=1..100,cfv=0,twv=20..100
22	cfq=0,cfv=-100..1,twv=20..100
23	cfq=1,cfv=-100..1,twv=20..100
24	cfq=0..1,cfv=-100..1,twv=20..100
25	cfq=0,cfv=1,twv=20..100
26	cfq=1,cfv=1,twv=20..100
27	cfq=0..1,cfv=1,twv=20..100
28	cfq=0,cfl=cfh=cfv=-100..1,twl=twh=twv=20..100
29	cfq=1,cfl=cfh=cfv=-100..1,twl=twh=twv=20..100
30	cfq=0..1,cfl=cfh=cfv=-100..1,twl=twh=twv=20..100

Table 1: Configurations for all datasets

For comparison purposes the same datasets were used in regular Offspring Selection Genetic Algorithms (OSGA) as seen in table 3, 5 and 7. A strict configuration with a comparison factor of 1 is used. Maximum allowed length is 50, 100 and 200; maximum allowed height is 7, 12 and 17 for configurations a, b, and c respectively.

4.1. Results Melanoma Dataset

The results of the performed test runs with the Melanoma dataset are shown in table 2. The regular OSGA results for Melanoma are shown in table 3.

#	Acc.(Tr.)	Acc.(Te.)	Height	Length	Nr.OfVar	MSE(Tr.)	MSE(Te.)	Exec.Time
1	0.924	0.925	9.8	31.6	5.6	0.091	0.074	03:04
2	0.934	0.934	8.2	37.8	6.6	0.069	0.166	03:24
3	0.923	0.907	9.4	36.6	5.8	0.070	0.075	02:30
4	0.935	0.929	9.8	49.8	7.2	0.062	0.068	39:03
5	0.916	0.913	9.0	32.2	5.4	0.078	0.083	02:48
6	0.931	0.926	11.8	68.0	8.8	0.068	0.077	08:37
7	0.927	0.924	8.6	30.4	5.8	0.080	0.075	01:22
8	0.929	0.929	11.0	40.8	7.2	0.063	0.069	03:53
9	0.927	0.914	9.6	32.8	6.6	0.075	0.096	01:12
10	0.924	0.920	7.8	18.2	4.0	0.073	0.077	01:04
11	0.920	0.921	10.6	56.8	8.6	0.072	0.249	02:46
12	0.917	0.921	9.4	32.0	5.4	0.081	0.081	01:55
13	0.937	0.927	12.0	66.4	8.0	0.060	0.066	09:25
14	0.931	0.911	11.4	42.0	7.2	0.065	0.071	09:15
15	0.925	0.914	10.4	50.8	6.4	0.065	0.072	05:33
16	0.921	0.914	11.2	46.2	7.4	0.075	0.077	01:41
17	0.919	0.921	11.2	36.8	5.8	0.078	0.153	11:14
18	0.918	0.915	8.4	34.6	6.2	0.123	0.087	01:10
19	0.927	0.921	11.8	49.8	6.8	0.071	0.073	02:19
20	0.926	0.907	10.6	41.0	6.6	0.071	0.078	02:47
21	0.930	0.918	11.4	45.4	7.2	0.075	0.075	03:12
22	0.944	0.927	12.8	85.0	10.8	0.054	0.074	15:39
23	0.921	0.915	9.4	29.6	4.8	0.072	0.071	03:00
24	0.925	0.918	11.8	67.8	10.0	0.091	0.070	12:02
25	0.920	0.917	11.0	45.8	8.8	0.077	0.074	01:28
26	0.924	0.902	8.6	33.6	6.0	0.111	0.095	03:32
27	0.911	0.911	9.8	41.2	6.8	0.086	0.082	02:12
28	0.930	0.917	10.8	43.2	5.8	0.064	0.070	10:41
29	0.923	0.915	10.2	31.2	6.0	0.104	0.125	03:07
30	0.932	0.922	11.6	65.8	8.8	0.077	0.064	07:07

Table 2: Results with Melanoma dataset

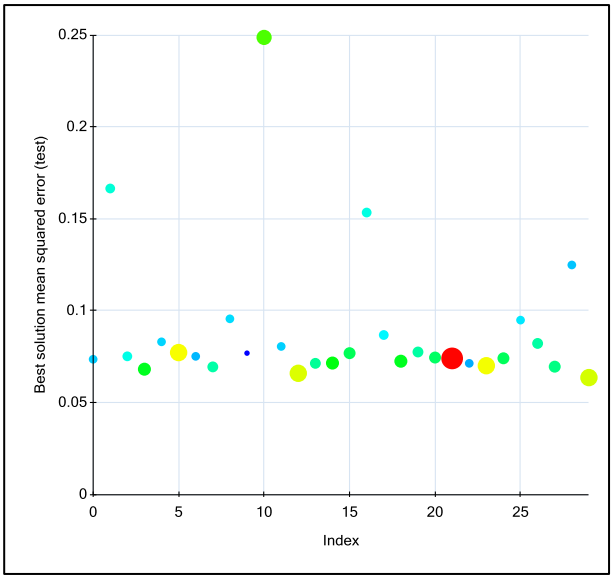

Figure 2: Melanoma Results: Configuration vs. Quality; little–high complexity (blue–red and bubble size)

#	Acc.(Tr.)	Acc.(Te.)	Height	Length	Nr.OfVar	MSE(Tr.)	MSE(Te.)	Exec.Time
a	0.919	0.917	6.8	15.6	3.2	0.086	0.100	02:49
b	0.931	0.919	9.0	35.6	7.0	0.075	0.072	03:46
c	0.927	0.915	13.2	71.0	11.6	0.071	0.078	04:30

Table 3: Regular OSGA results Melanoma

4.2. Results Thyroid Dataset

The results of the performed test runs with the Thyroid dataset are shown in table 4. The regular OSGA results for Thyroid are shown in table 5.

#	Acc.(Tr.)	Acc.(Te.)	Height	Length	Nr.OfVar	MSE(Tr.)	MSE(Te.)	Exec.Time
1	0.970	0.962	6.6	14.2	2.2	90.87	220.17	04:08
2	0.983	0.976	10.6	49.8	6.8	57.37	78.24	08:55
3	0.989	0.987	7.8	22.8	3.2	61.17	74.60	06:48
4	0.991	0.988	12.6	85.0	6.6	32.81	39.22	58:05
5	0.983	0.980	12.4	67.2	7.6	69.52	87.90	13:18
6	0.990	0.986	12.6	76.4	6.0	36.74	52.99	24:02
7	0.941	0.935	9.0	30.2	3.4	220.92	207.82	02:53
8	0.984	0.983	11.4	61.8	6.2	63.17	63.53	05:00
9	0.950	0.945	8.8	23.2	3.0	183.75	194.81	02:17
10	0.943	0.943	7.4	29.6	3.8	197.61	192.17	02:21
11	0.983	0.975	10.4	61.2	6.8	65.27	94.61	05:46
12	0.952	0.947	9.0	32.8	4.8	150.90	170.13	03:19
13	0.992	0.990	12.2	88.0	6.2	36.00	59.93	17:11
14	0.988	0.987	11.4	82.2	7.4	46.20	60.26	33:37
15	0.993	0.991	11.8	76.4	7.4	34.88	45.13	22:04
16	0.938	0.938	10.2	37.4	4.0	199.67	186.36	01:50
17	0.982	0.980	11.4	60.0	6.4	63.38	76.04	03:14
18	0.935	0.936	10.6	48.4	6.2	216.02	236.95	01:34
19	0.942	0.939	9.2	34.2	1.4	161.51	160.31	01:06
20	0.987	0.984	11.8	67.6	6.6	53.26	65.30	13:02
21	0.953	0.952	10.4	44.8	2.2	136.96	171.29	02:10
22	0.993	0.987	12.8	91.6	6.2	32.45	50.64	18:20
23	0.989	0.987	12.8	73.0	7.4	48.39	54.27	45:34
24	0.993	0.989	12.6	92.0	7.0	41.28	46.95	21:06
25	0.943	0.940	8.6	24.0	3.0	230.51	235.51	01:10
26	0.979	0.979	11.2	60.2	7.0	73.97	96.89	03:13

27	0.938	0.935	10.0	40.4	2.0	189.46	189.00	01:43
28	0.993	0.991	11.8	86.0	5.8	39.09	223.87	37:29
29	0.986	0.982	12.8	88.0	7.6	51.26	66.97	09:51
30	0.985	0.983	12.2	75.6	7.6	56.67	90.64	10:13

Table 4: Results with Thyroid dataset

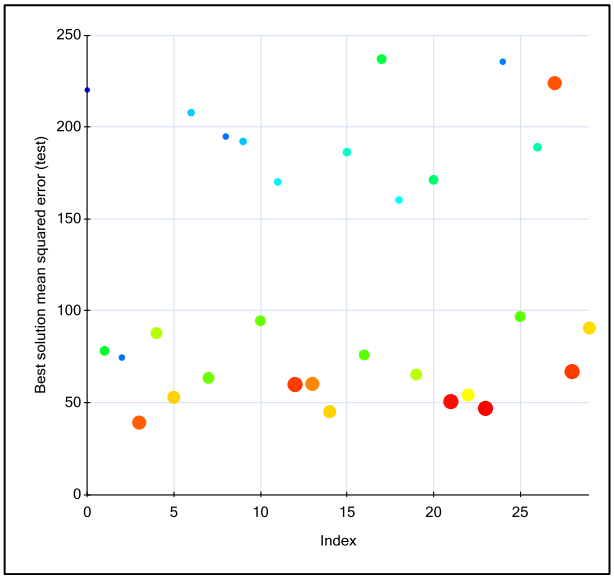

Figure 3: Thyroid Results: Configuration vs. Quality; little–high complexity (blue–red and bubble size)

#	Acc.(Tr.)	Acc.(Te.)	Height	Length	Nr.OfVar	MSE(Tr.)	MSE(Te.)	Exec.Time
a	0.982	0.977	7.8	36.2	4.8	50.67	70.50	08:33
b	0.977	0.972	12.0	68.0	7.6	311.36	98.74	09:36
c	0.990	0.986	17.4	115.4	9.0	53.14	67.22	09:45

Table 5: Regular OSGA results Thyroid

4.3. Results Wisconsin Dataset

The results of the performed test runs with the Wisconsin dataset are shown in table 6. The regular OSGA results for Wisconsin are shown in table 7.

#	Acc.(Tr.)	Acc.(Te.)	Height	Length	Nr.OfVar	MSE(Tr.)	MSE(Te.)	Exec.Time
1	0.956	0.955	9.2	26.4	4.2	0.188	0.204	02:25
2	0.959	0.947	12.4	59.0	6.0	0.168	0.236	03:03
3	0.960	0.955	11.4	39.8	4.4	0.190	0.174	03:28
4	0.960	0.943	12.6	71.4	4.4	0.149	0.211	14:09
5	0.967	0.959	12.0	63.0	6.8	0.151	0.185	13:39
6	0.961	0.934	12.2	67.6	5.6	0.162	0.197	11:21
7	0.944	0.925	8.4	23.6	3.4	0.225	0.238	01:10
8	0.968	0.959	11.2	50.0	6.4	0.159	0.166	04:17
9	0.950	0.952	9.4	23.2	3.2	0.239	0.209	02:20
10	0.947	0.946	7.6	20.4	3.2	0.205	0.232	01:33
11	0.967	0.955	11.8	62.0	5.8	0.153	0.180	03:17
12	0.942	0.934	9.0	31.4	3.8	0.208	0.241	02:12
13	0.963	0.953	12.0	71.6	6.6	0.155	0.189	14:08
14	0.966	0.949	12.0	60.0	6.2	0.146	0.172	04:41
15	0.960	0.937	11.8	58.4	5.6	0.158	0.214	08:53
16	0.947	0.937	8.8	34.8	4.4	0.220	0.239	01:26
17	0.959	0.946	12.4	59.8	6.6	0.183	0.198	02:49
18	0.946	0.925	7.4	15.6	2.8	0.256	0.285	00:59
19	0.944	0.937	10.8	36.4	3.2	0.206	0.202	01:20
20	0.964	0.959	12.0	55.8	6.8	0.161	0.191	02:41
21	0.948	0.938	10.4	30.4	3.2	0.241	0.239	01:28
22	0.965	0.950	12.4	55.4	5.6	0.158	0.217	09:11
23	0.966	0.947	12.2	69.0	6.4	0.144	0.200	03:18
24	0.962	0.950	12.6	79.0	6.0	0.149	0.155	11:06
25	0.940	0.930	9.2	30.8	2.8	0.229	0.255	00:57
26	0.963	0.958	12.2	65.0	6.4	0.175	0.193	03:27
27	0.935	0.921	9.6	30.2	2.8	0.266	0.240	00:53
28	0.966	0.955	11.4	52.8	5.4	0.141	0.213	16:04
29	0.967	0.960	11.8	51.0	6.6	0.148	0.170	02:59
30	0.965	0.960	10.4	53.4	5.0	0.155	0.160	11:01

Table 6: Results with Wisconsin dataset

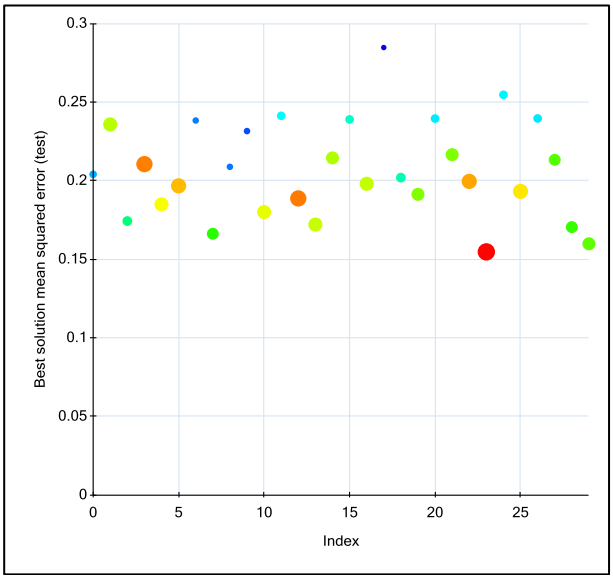

Figure 4: Wisconsin Results: Configuration vs. Quality; little–high complexity (blue–red and bubble size)

#	Acc.(Tr.)	Acc.(Te.)	Height	Length	Nr.OfVar	MSE(Tr.)	MSE(Te.)	Exec.Time
a	0.976	0.966	8.0	36.6	5.6	0.109	0.157	53:56
b	0.979	0.966	13.0	92.8	6.6	0.087	0.139	138:34
c	0.981	0.960	17.8	181.8	8.4	0.085	0.151	64:17

Table 7: Regular OSGA results Wisconsin

5. CONCLUSION

In this paper we have considered the aspects of model interpretability and uniqueness in genetic programming based medical data mining. Due to introns and the bloat phenomenon GP models tend to produce more complex than necessary (Luke and Panait, 2006). In contrast to the so called bloat free GP (Silva, 2011) which allows only those models which do not exceed a certain model complexity we have adapted the concept of offspring selection in a way that the hypothesis search process should he guided towards simple and good prediction models. For this purpose the offspring selection criterion has been extended in a way that it considers not only the model quality in order to decide whether or not a candidate hypothesis should be accepted; in addition also the complexity in terms of number of nodes and the interpretability in terms of number of used variables are considered for the offspring selection criterion. The effects of this approach have been analyzed for some well-known

benchmark problems from the field of medical data mining. The results show that the new offspring selection criterion is quite sensitive in terms of causing premature convergence due to the loss of genetic diversity caused by the complexity limiting aspects in the OS-criterion. Therefore, it remains as a topic for future research to further develop this new way of hypothesis search in a way that a sufficient amount of genetic diversity is maintained in the GP population. One possible way of achieving such kind of behavior might an automated switch on/off of the additional criteria depending on the average model complexity or the diversity in the actual population.

ACKNOWLEDGMENTS

The work described in this paper was done within the Josef Ressel Centre for Heuristic Optimization *Heureka!* (http://heureka.heuristiclab.com/) sponsored by the Austrian Research Promotion Agency (FFG).

REFERENCES

Affenzeller, M. and Wagner, S., 2005. Offspring selection: A new self-adaptive selection scheme for genetic algorithms. *Adaptive and Natural Computing Algorithms*, Springer Computer Science, pp. 218-221.

Affenzeller, M., Winkler, S., Wagner, S., A. Beham, 2009. *Genetic Algorithms and Genetic Programming - Modern Concepts and Practical Applications*. Chapman & Hall/CRC. ISBN 978-1584886297. 2009.

Luke S. and Panait L., 2006. *A Comparison of Bloat Control Methods*. Journal of Evolutionary Computation, Vol. 14, No.3 pp. 48-48.

Silva S., 2011 *Reassembling Operator Equalisation - A Secret Revealed*. Proceedings of GECCO 2011, pp. 1395-1403.

Wagner, S., Affenzeller M., Beham A., Kronberger G., and Winkler S.M., 2010. Mutation Effects in Genetic Algorithms with Offspring Selection Applied to Combinatorial Optimization Problems. *Proceedings of EMSS 2010*, pp. 48-48.

AUTHORS BIOGRAPHIES

MICHAEL AFFENZELLER has published several papers, journal articles and books dealing with theoretical and practical aspects of evolutionary computation, genetic algorithms, and meta-heuristics in general. In 2001 he received his PhD in engineering sciences and in 2004 he received his habilitation in applied systems engineering, both from the Johannes Kepler University of Linz, Austria. Michael Affenzeller is professor at the Upper Austria University of Applied Sciences, Campus Hagenberg, and head of the Josef Ressel Center *Heureka!* at Hagenberg.

CHRISTIAN FISCHER received his BSc in software engineering in 2009 from the Upper Austria University of Applied Sciences, Campus Hagenberg. He is currently pursuing studies for his master's degree. In the course of his studies he is involved in the project team for the prediction of blood demands in a hospital in cooperation with the Josef Ressel Centre Heureka! and the General Hospital Linz.

GABRIEL KRONBERGER authored and co-authored numerous papers in the area of evolutionary algorithms, genetic programming, machine learning and data mining. Currently he is a research associate at the Research Center Hagenberg of the Upper Austria University of Applied Sciences working on data-based modeling algorithms for complex systems within the Josef-Ressel Centre for Heuristic Optimization Heureka!.

STEPHAN M. WINKLER received his MSc in computer science in 2004 and his PhD in engineering sciences in 2008, both from Johannes Kepler University (JKU) Linz, Austria. His research interests include genetic programming, nonlinear model identification and machine learning. Since 2009, Dr. Winkler is professor at the Department for Medical and Bioinformatics at the Upper Austria University of Applied Sciences, Campus Hagenberg.

STEFAN WAGNER received his MSc in computer science in 2004 and his PhD in engineering sciences in 2009, both from Johannes Kepler University (JKU) Linz, Austria; he is professor at the Upper Austrian University of Applied Sciences (Campus Hagenberg). Dr. Wagner's research interests include evolutionary computation and heuristic optimization, theory and application of genetic algorithms, machine learning and software development.