

## БАЗИ ДАНИХ, БАЗИ ЗНАНЬ ТА ИНЖЕНЕРИЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕНИЯ

УДК 004.8

**Н.А. Новоселова, И.Э. Том**

**Объединенный институт проблем  
информатики**

**Национальной академии наук Беларуси**

# ГЕНЕТИЧЕСКОЕ ОБУЧЕНИЕ НЕЧЕТКИХ СИСТЕМ, ОСНОВАННЫХ НА ПРАВИЛАХ

*Рассматривается общая схема эволюционного обучения нечетких систем, основанных на правилах, где генетические процедуры применяются для обучения и настройки различных компонентов системы. Предлагаются схемы генетической настройки базы данных, генетического обучения базы правил и два варианта обучения базы знаний НС. Рассматривается вопрос интерпретируемости нечеткой системы*

*Розглядається загальна схема еволюційного навчання нечітких систем, заснованих на правилах, де генетичні процедури застосовуються для навчання та налаштування різних компонентів системи. Пропонуються схеми генетичної налаштування бази даних, генетичного навчання бази правил і два варіанти навчання бази знань НС. Розглядається питання інтерпретованості нечіткої системи.*

*The paper presents the general scheme of the evolutionary learning of the rule-based fuzzy systems, where the genetic procedures are used for the learning and tuning the different system components. Several schemes for the genetic tuning of the data base, genetic learning of the rule base and two variants for the fuzzy system' knowledge-base learning are proposed. The problem of the interpretability of the fuzzy system is considered.*

**Ключевые слова:** нечеткая система, база знаний, генетическое обучение

### Введение

Методы мягких вычислений или вычислительного интеллекта (computational intelligence), такие как искусственные нейронные сети, нечеткая логика и генетические алгоритмы (ГА) являются в настоящее время одними из наиболее востребованных, потому что позволяют найти решение сложных практических задач, которые не могут быть решены стандартными методами [1].

В последнее время активно разрабатываются новые подходы к построению гибридных моделей решения задач в различных прикладных областях.

Наиболее распространенным подходом является гибридизация нечеткой логики и генетических алгоритмов, результатом чего являются генетические нечеткие системы (ГНС), которые представляют собой нечеткие системы, обучаемые на данных с использованием генетических процедур [2].

Нечеткая система [3] – математическая модель, включающая нечеткие множества, правила вывода типа "если - то", систему нечеткого

вывода и, при необходимости, преобразователи (фаззификатор и дефаззификатор) входных и выходных переменных.

Нечеткие системы (НС), являясь наиболее важным прикладным инструментом теории нечетких множеств, успешно используются для решения задач классификации [3], моделирования [4] и управления [5] в различных областях знаний. Разработка алгоритмов генетического обучения нечетких систем расширила возможности их применения путем настройки на исходные данные, возможности динамической модификации структуры нечеткой системы в процессе ее использования [6-9].

### Общая схема обучения нечетких систем

Наиболее распространенным и широко исследуемым типом ГНС является генетическая нечеткая система, основанная на правилах (ГНСП) [10], где генетические процедуры применяются для обучения и настройки различных компонентов нечеткой системы, от параметров функций принадлежности отдельных лингвистических термов до набора решающих правил.

Основной причиной использования генетического обучения нечеткой системы является возможность автоматизации процесса определения и настройки базы знаний (БЗ) нечеткой системы, что может рассматриваться как оптимизационная задача.

Параметры базы знаний составляют оптимизационное пространство, которое трансформируется в соответствующее генетическое представление, где и осуществляется процесс поиска параметров. Важно различать задачи настройки (адаптации) и обучения НС. Настройка представляет собой оптимизацию уже существующей НС, в то время как обучение является процессом автоматического построения наборов нечетких правил с нуля. Процессы настройки предполагают наличие предопределенной базы правил (БП) и имеют целью найти множество оптимальных параметров для функций принадлежности и/или масштабирующих функций, т.е. параметров базы данных (БД). Обучающие же процессы осуществляют поиск в пространстве всех возможных БП и не зависят от предопределенного набора правил.

В зависимости от ряда критериев, включая имеющиеся априорные знания о системе, количества параметров, доступности и полноты входных/выходных данных, процесс обучения может применяться на различных этапах построения нечеткой системы:

- генетическое обучение на уровне БП;
- генетическое обучение на уровне БЗ;
- генетическое обучение на уровне БД.

#### **Генетическое обучение базы правил нечеткой системы**

Основные подходы к генетическому обучению БП нечеткой системы с точки зрения кодирования правил в популяции особей ГА можно разделить на две основные группы [10]:

– подход “Хромосома = набор правил”, также называемый подходом Питсбург, в котором каждая особь ГА представляет собой набор правил [11];

– подход “Хромосома = правило”, где каждая особь ГА кодирует одно правило НС, а весь набор правил представляет собой комбинацию нескольких особей популяции ГА.

В свою очередь подход “Хромосома = правило” подразделяется на три группы:

Подход Мичиган, который обычно характерен для обучаемой системы классификации, использующей ГА для обучения правил, управляющих ее эффективностью [12]. ГА в этом случае позволяет находить эффективные нечеткие правила в большом поисковом про-

странстве, однако его возможности поиска всего набора правил невысоки. Это происходит потому, что в качестве оценочной или оптимизируемой функции используется точность классификации отдельных нечетких правил, а не всего набора правил.

1. Итерационный подход к обучению нечетких правил, при котором хромосома представляет правило, а решением является наилучшая особь, полученная в процессе эволюции. Глобальное решение формируется из наилучших особей при многократном запуске ГА. Например, обучаемые нечеткие системы MOGUL [13] и SLAVE [14] относятся именно к этой категории.

2. Подход «кооперация-соревнование», при котором вся популяция или подмножество особей ГА кодирует весь набор правил НС. В частности, обучаемая нечеткая система LOGENPRO [15] относится к этой категории.

#### **Генетическое обучение базы знаний**

С точки зрения авторов для повышения эффективности генетического обучения НС необходима разработка таких подходов и алгоритмов, в которых:

– предоставляется возможность одновременного обучения, как базы данных, так и базы правил НС;

– при обучении учитываются дополнительные ограничения, которые гарантируют выполнение критериев интерпретируемости НС, о которых будет сказано ниже.

Для построения базы знаний нечеткой системы необходимо решить две задачи:

– задачу обучения БД, включающую задание области определения и количества лингвистических термов каждой из лингвистических переменных, а также настройку параметров функции принадлежности, соответствующей каждому терму;

– задачу обучения БП, включающую определение количества правил и состав каждого из них (т.е. определение предпосылок и следствий правила).

Рассмотрим четыре типа подходов, используемых для обучения базы знаний ГНСП.

#### **Генетическое обучение БП.**

Этот подход, подробнее рассмотренный выше, уделяет внимание обучению БП, используя предопределенную БД (рис. 1). Наиболее распространенным способом определения БД является выбор количества лингвистических термов для каждой лингвистической переменной (например, нечетное число между 3 и 9,

которое обычно одинаково для каждой переменной) и задание значений параметров функций принадлежности лингвистических термов, распределяемых равномерно в границах обла-

сти определения значений лингвистической переменной.

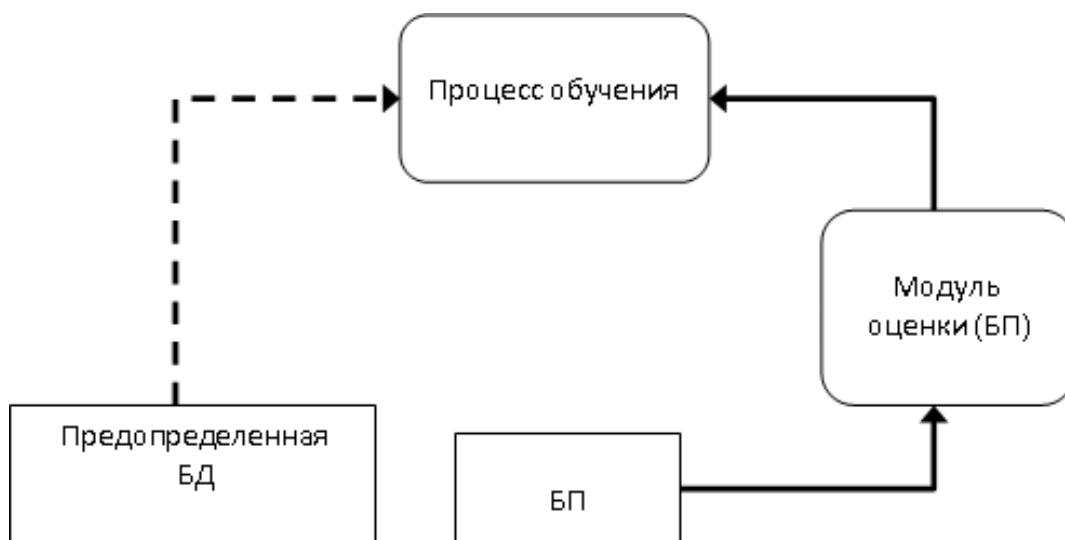


Рис. 1 – Генетическое обучение БП

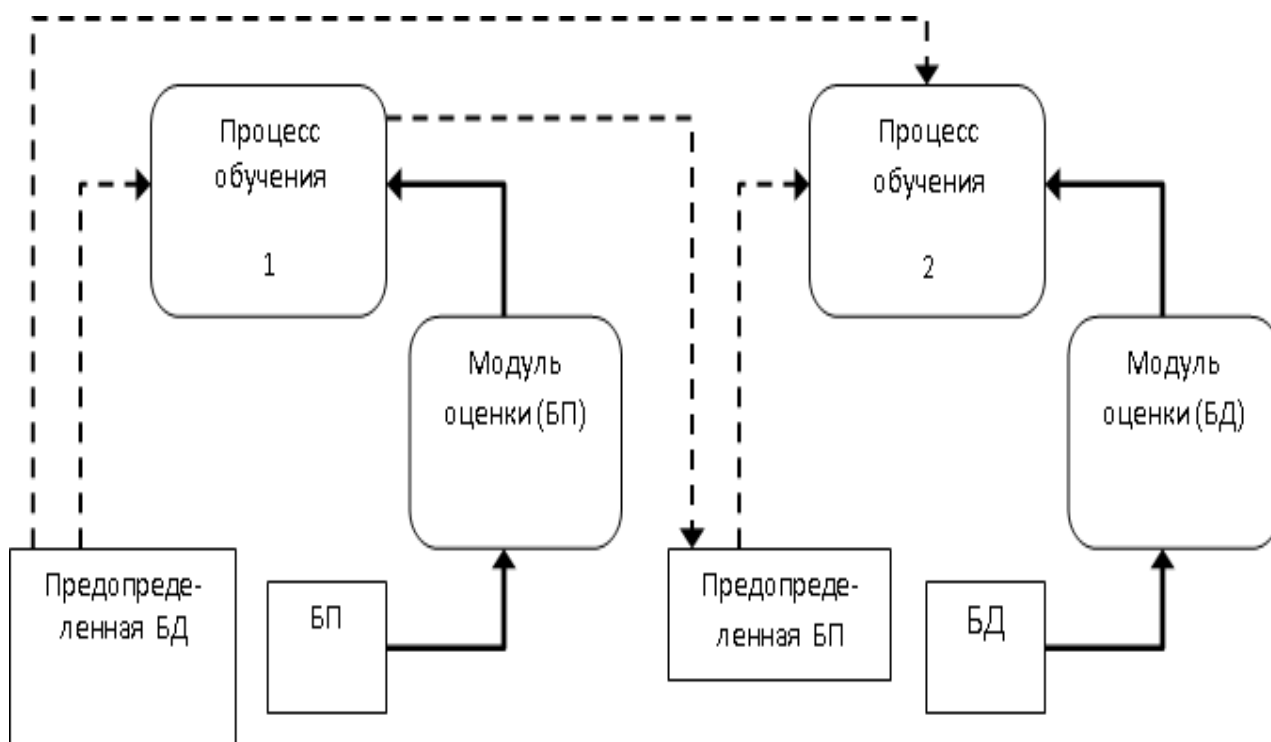


Рис. 2 – Генетическая настройка БД predetermined БП

#### Генетическая настройка БД

Для повышения эффективности ГНСП этот подход позволяет адаптировать (настроить) к имеющимся исходным данным базу данных нечеткой системы, для которой уже определена БП. В этом случае настраиваются параметры функций принадлежности.

Однако при этом подходе адаптируется только форма (вид) функций принадлежности, количество же лингвистических термов для каждой переменной остается неизменным (рис. 2).

**Генетическое обучение базы знаний.**

Этот подход предполагает одновременное обучение двух компонентов БЗ (рис. 3). В этом случае возможно получение наиболее эффективных решений, однако процесс обучения является наиболее сложным и медленным.

Для того чтобы сузить поисковое пространство при обучении и настройке всей базы зна-

ний генетическое обучение выполняется либо в условиях ограничений на изменение параметров НС [16], либо начальная популяция ГА, соответствующая базе знаний НС, инициализируется с использованием результатов нечеткой кластеризации входных объектов данных [17]. Однако в большинстве случаев это приводит к получению только субоптимальных решений.

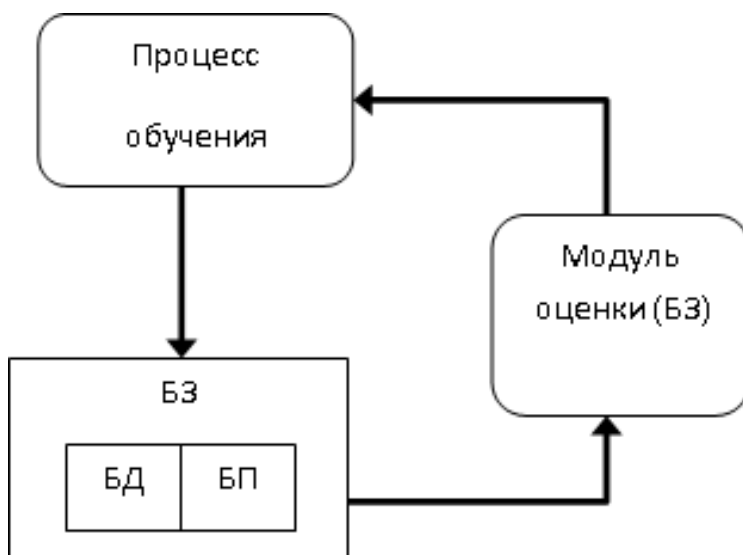


Рис. 3. Генетическое обучение всей БЗ

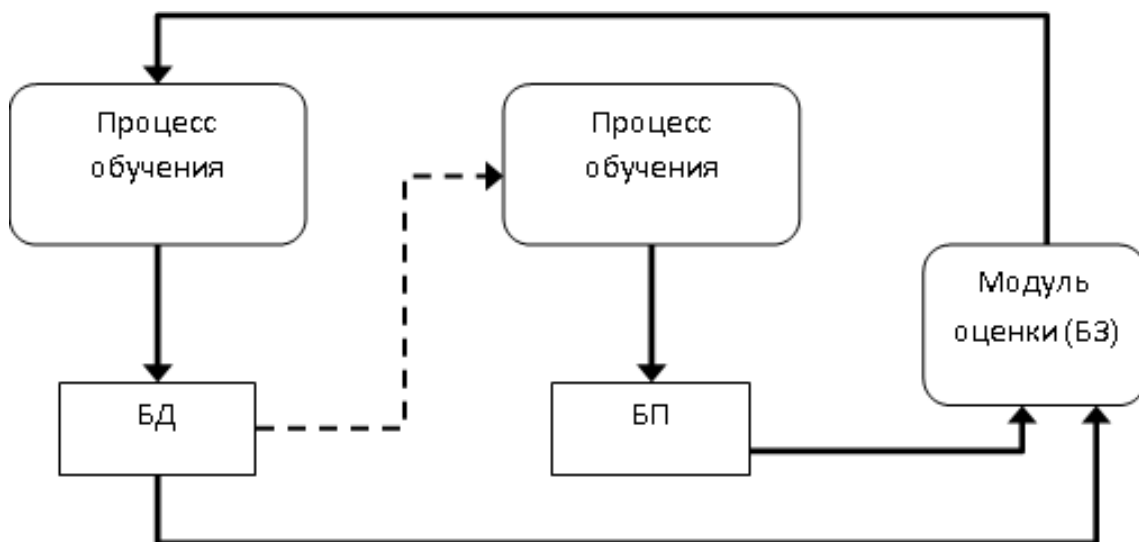


Рис.4. Генетическое обучение всей БЗ

**Предварительное генетическое обучение БД.**

Этот подход предполагает использование двух отдельных процессов для обучения двух компонентов базы знаний.

Процесс генерации БД охватывает процесс обучения БП следующим образом: каждый раз при определении БД, запускается процесс генерации БП для определения правил, при этом

рассчитывается некоторая заданная мера ошибки для оценки всей полученной базы знаний.

Таким образом, в отличие от подхода, представленного на рис. 3, где осуществляется поиск решения в глобальном поисковом пространстве (БД+БП), настоящий подход предполагает поиск в двух более простых пространствах поиска (отдельно БД и БП) для получения

общего решения оптимизационной задачи (рис.4).

Целью всех рассмотренных выше подходов является получение оптимальных параметров НС, где в качестве критерия оптимизации выступает в большинстве случаев точность НС. Однако вторым важным фактором, о котором мы говорили выше и который необходимо учитывать при построении и обучении НС, является ее интерпретируемость.

Использование автоматических подходов к построению и обучению ГНСП на данных не гарантирует сохранение свойств ее интерпретируемости, что часто приводит к тому, что ГНСП являются менее полезными при практическом использовании, чем НС, база знаний которых построена экспертом вручную.

#### **Алгоритмы обучения нечеткой системы, учитывающие ее интерпретируемость**

Свойство интерпретируемости НС является одним из важнейших для ее успешного практического использования. В работе [18] для удовлетворения ряда семантических и синтаксических критериев интерпретируемости НС совместно с генетическим обучением параметров БД используется предварительная нечеткая кластеризация данных и упрощение БП с использованием метода, основанного на расчете меры сходства [19] функций принадлежности для каждого из входных признаков НС.

Таким образом, процесс итерационного построения НС совмещается с упрощением структуры БП и генетическим обучением БД. С использованием ГА осуществляется многокритериальный поиск параметров НС с акцентом на поиск наиболее точной модели. После этого, полученный результат с избыточным количеством схожих функций принадлежности используется для упрощения и сокращения БП нечеткой системы.

Результатом выполнения нескольких итераций генетического обучения и упрощения БП является компактная НС, обладающая достаточно высокой классификационной точностью.

На заключительном шаге выполняется генетическое обучение параметров НС по двум критериям с акцентом на интерпретируемость НС.

Ограничением подхода, предложенного в [18], является его итерационный характер при построении НС, что затрудняет получение оптимального решения, а также объединение нескольких критериев оптимизации в составе единой оптимизируемой функции, где каждому из критериев соответствуют весовые коэффициенты.

Для преодоления этого ограничения авторами был разработан алгоритм многокритериальной оптимизации структуры нечеткого классификатора, представленного в виде НС, который осуществляет генетическое обучение БП, учитывая синтаксические критерии интерпретируемости НС, и позволяет получить НС с простой и компактной структурой правил. Алгоритм использует генетические процедуры, позволяющие параллельно оптимизировать несколько критериев и находить ряд компромиссных решений между точностью классификации и интерпретируемостью НК.

Важным достоинством алгоритма является возможность получения нескольких недоминируемых НС за один запуск ГА. Благодаря многокритериальной оптимизации результатом работы алгоритма являются наборы нечетких систем с различными соотношениями точности классификации и простоты системы решающих правил.

В качестве критерия интерпретируемости НС используется агрегированная мера сходства нечетких множеств, а точность оценивается на основе среднеквадратичной ошибки. В таблицах 1-3 представлены результаты сравнительного анализа разработанного алгоритма многокритериальной оптимизации структуры нечеткого классификатора, представленного в виде НС, который осуществляет генетическое обучение БП, с наиболее известными в литературе подходами на примере двух наборов данных из архива данных по машинному обучению (<http://www.ics.uci.edu/~mllearn/>).

Таблица 1. Характеристика наборов данных

Название	Количество признаков	Количество объектов данных	Количество классов
WBC	9	683	2
Pima diabetes	8	768	2

Таблица 2. Характеристика результата классификации набора данных WBC с использованием различных подходов

Название	C4.5 [20]	Нечеткий генетический подход [22]	Метод декомпозиции [23]	Генетический отбор [24]	НС, обученная предлагаемым алгоритмом
Точность классификации на тестовом множестве, %	94,9	96,88	95,3	95,59	96,79
Количество правил / средняя длина правила	более 4	4/3,44	более 4	4/2,06	4/2,28

Таблица 3. Характеристика результата классификации набора данных Pima Diabetes с использованием различных широко распространенных методов

Название	C4.5 [20]	Нечеткий классификатор [21]	Генетический отбор [24]	НС, обученная предлагаемым алгоритмом
Ошибка классификации на тестовом множестве, %	25	26,95	26,76	24,6
Количество правил	более 3	более 3	3	3

Сравнительный анализ НС, обученной рассмотренным алгоритмом, с тремя наиболее близкими зарубежными методами-аналогами (таблицы 2 и 3) на примерах наборов данных Pima Diabetes и WBC, показал его более высокую эффективность как в части компактности и интерпретируемости, так и в отношении меньшей ошибки классификации.

#### Заключение

В настоящей работе рассмотрена общая схема эволюционного обучения нечетких систем, основанных на правилах, где генетические процедуры применяются для обучения и настройки различных компонентов системы.

Предлагаются схемы генетической настройки базы данных, генетического обучения базы правил и два варианта обучения базы знаний нечеткой системы.

Рассматривается вопрос интерпретируемости нечеткой системы. На основании рассмотренных схем обучения и понятия интерпретируемости предлагается алгоритм обучения нечеткого классификатора с использованием генетических процедур, который позволяет не только обучать БП, но и БД нечеткого классификатора путем настройки параметров функций принадлежности.

#### Список использованных источников

1. Konar, A. Computational Intelligence: Principles, Techniques and Applications / A. Konar. – Springer-Verlag, 2005.
2. Cordon, O. Genetic Fuzzy Systems. Evolutionary Tuning and Learning of Fuzzy Knowledge Bases / O. Cordon, F. Herrera, F. Hoffmann, L. Magdalena. - World

Scientific, 2001.

3. Mitra, S. Data Mining. Multimedia, Soft Computing and Bioinformatics / S. Mitra, T. Acharya. – John Wiley&Sons: – New Jersey, Canada, 2003/ – 401 p.

4. Pedrycz, W. Fuzzy Modelling: Paradigms and Practice / W. Pedrycz (Ed.) - Kluwer Academic Press, Dordrecht, 1996.

5. Driankov, D. An Introduction to Fuzzy Control / D. Driankov, H. Hellendoorn, M. Reinfrank. - Springer, Berlin, 1993.

6. Том И.Э. Методы интеллектуального анализа многомерных данных для решения задач классификации/ Том И.Э. , Н.А. Новоселова, О.В. Красько. – Минск: ОИПИ НАН Беларуси, 2011. – 236 с.

7. Новоселова, Н.А. Эволюционный подход к выделению информативных признаков в задачах анализа медицинских данных / Н.А. Новоселова, И.Э. Том, А.С. Мاستыкин // Искусственный интеллект – 2008.- №3. –С. 105-112.

8. Новоселова, Н.А. Эволюционный метод нечеткой кластеризации /Н.А. Новоселова, И.Э. Том // Информатика. - 2009. № 3 (23) - С. 55-67.

9. Новоселова, Н.А. Подход к построению ансамбля классификаторов с использованием генетического алгоритма / Н.А. Новоселова, И.Э. Том // Искусственный интеллект - 2009. - № 3. - С. 81-88.

10. Cordon, O. Genetic Fuzzy Systems - Evolutionary Tuning and Learning of Fuzzy Knowledge Bases / O. Cordon, F. Herrera, F. Hoffmann, L. Magdalena. - World Scientific, Singapore, 2001.

11. De Jong, K.A. Using genetic algorithms

for concept learning / K.A. De Jong, W.M. Spears, D.F. Gordon // *Machine Learning*. – 1993. – Vol. 13. – P. 161–188.

12. Kovacs, T. *Strength or Accuracy: Credit Assignment in Learning Classifier Systems* / T. Kovacs. – Springer-Verlag, 2004.

13. Cordon, O. *MOGUL: A Methodology to Obtain Genetic fuzzy rule-based systems under the iterative rule Learning approach* / O. Cordon, M.J. del Jesus, F. Herrera, M. Lozano // *International Journal of Intelligent Systems*. – 1999. – Vol. 14. – P. 1123–1153.

14. Gonzalez, A. *SLAVE: A genetic learning system based on an iterative approach* / A. Gonzalez, R. Perez // *IEEE Transactions on Fuzzy Systems*. – 1999. – Vol. 27. – P. 176–191.

15. Wong, M.L. *Data Mining using Grammar Based Genetic Programming and Applications* / M.L. Wong, K.S. Leung. – Kluwer Academic Publishers, 2000.

16. Setnes, M. *GA-fuzzy modeling and classification: Complexity and performance* / M. Setnes, J.A. Roubos // *IEEE Transactions on Fuzzy Systems*. – 2000. – Vol. 8, №5. – P. 509–522.

17. Setnes, M. *Fuzzy relational classifier trained by fuzzy clustering* / M. Setnes, R. Babuska // *IEEE Transactions on Systems, Man, and Cybernetics–Part B: Cybernetics*. – 1999. – Vol. 29. – P. 619–625.

18. Roubos, J. A. *Compact and transparent fuzzy models and classifiers through iterative complexity reduction* / J. A. Roubos, M. Setnes // *IEEE Trans. on Fuzzy Systems*. – 2001. – Vol. 9,

№4. – P. 516–524.

19. Setnes, M. *Similarity measures in fuzzy rule base simplification* / M. Setnes, R. Babuska, U. Kaymak, and H. R. van Nauta Lemke // *IEEE Trans. Systems, Man and Cybernetics - Part B*. – 1998. – Vol. 28. – P. 376–386.

20. Elomaa, T. *General and Efficient Multisplitting of Numerical Attributes* / T. Elomaa, J. Rousu // *Machine Learning*. – 1999. – Vol. 36, № 3. – P. 201–244.

21. Abonyi, J. *Data-driven Generation of Compact, Accurate, and Linguistically-sound Fuzzy Classifiers Based on a Decision-tree Initialization* / J. Abonyi, J. Roubos, F. Szeifert // *Intern. J. of Approximate Reasoning*. – 2003. – Vol. 32, № 1. – P. 1–21.

22. Pena-Reyes C.A. *A Fuzzy-genetic Approach to Breast Cancer Diagnosis* / C.A. Pena-Reyes, M. Sipper // *Artificial Intelligence in Medicine*. – 1999. – Vol. 17, № 2. – P. 131–135.

23. Guan, S.-U. *Class Decomposition for GA-based Classifier Agents – A Pitt Approach* / S.-U. Guan, F. Zhu // *IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics*. – 2004. – Vol. 34,

№ 1. – P. 381–392.

24. Ishibuchi, H. *Multi-objective Evolutionary Design of Fuzzy Rule-based Systems* / H. 25 Ishibuchi, T. Yamamoto // *Proc. of Intern. Conf. on Systems, Man and Cybernetics (IEEE SMC 2004)*. The Hague, The Netherlands, Oct. 10–13, 2004. – Madison, Wisconsin : Omni Press, 2004. – Vol. 3. – P. 2362–2367.

## Сведения об авторах



**Новоселова Наталья Анатольевна** - к.т.н., старший научный сотрудник лаборатории биоинформатики Объединенного института проблем информатики Национальной академии наук Беларуси, кандидат технических наук. Научные интересы: методы интеллектуального анализа медико-биологических данных.

E-mail: [novosel@newman.bas-net.by](mailto:novosel@newman.bas-net.by)



**Том Игорь Эдуардович** - к.т.н., заведующий лабораторией биоинформатики Объединенного института проблем информатики Национальной академии наук Беларуси, кандидат технических наук. Научные интересы: методы обнаружения знаний в данных, информационные технологии, системы поддержки принятия решений в медицине.

E-mail: [tom@newman.bas-net.by](mailto:tom@newman.bas-net.by)

Статья поступила в редакцию 01.03.2012 г.