

УДК 004

## ПРИМЕНЕНИЕ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ДЛЯ СОЗДАНИЯ СИСТЕМЫ РЕШАЮЩИХ ПРАВИЛ

Пискунова Т.А.

Научный руководитель: Берестнева О.Г., профессор каф. ПМ ИК ТПУ

*Национальный Исследовательский Томский политехнический университет,*

*634050, Россия, г. Томск, пр. Ленина, 30*

*E-mail: tana.alex.a42@yandex.ru*

*The following article presents the implementation of the data mining method –classification tree for generating decision rule base as a part of DSS and consider the main steps for its constructing. The necessity and advantages of the given fuzzy modification are shown.*

**Key words:** *DSS, fuzzy logic, classification trees, data mining*

**Ключевые слова:** *СППР, нечеткая логика, деревья классификации, data mining*

В связи с увеличением объема информации, поступающей к руководителям и усложнением решаемых задач с большим числом взаимосвязанных факторов, появился новый класс информационных систем, предназначенных для решения слабоструктурированных и неструктурированных задач – системы поддержки принятия решений (СППР). В них используются научные методы, ищущие оптимальное решение путем обработки больших массивов информации с использованием современных технических средств. Большую область таких методов представляют методы Data Mining. Их основу составляют всевозможные методы классификации, моделирования и прогнозирования. В данной работе рассматривается первая часть задачи классификации – построение классификационной модели.

### Деревья решений с нечеткими оценками

В отличие от простых деревьев решений, где каждый примера принадлежит конкретному узлу, нечеткое дерево решений использует степень принадлежности. Очевидным преимуществом данного подхода является повышение точности классификации за счет сочетания достоинств нечеткой логики и деревьев решений. В связи с наличием слабоструктурированных данных, использование нечетких деревьев решения для построения СППР оправдано. Процесс обучения происходит быстро, а результат прост для интерпретации. Также такой подход не требует принципиальных изменений в процессе прохождения дерева решений. Конечный результат представляется в форме нечетких чисел, также результатом является и распределение его истинности. Это станет дополнительной информацией для принятия решения.

Нечеткое множество определяется через некоторую базовую шкалу  $V$  и функцию принадлежности  $\mu(x)$ , принимающую значения на интервале  $[0...1]$ . При описании объектов и явлений с помощью нечетких множеств используются лингвистические переменные  $(\beta, T, X, G, M)$ , где  $\beta$  – наименование лингвистической переменной;  $T$  – множество ее значений (термов);  $G$  – синтаксическая процедура для оперирования элементами терм-множества  $T$ , в частности, генерировать новые термы (значения);  $M$  – семантическая процедура, позволяющая сформировать нечеткое множество.

В качестве примера представим одну из ЛП текущей задачи. Необходимо определить успеваемость студента с помощью понятий «низкая успеваемость», «средняя успеваемость», «высокая успеваемость». Диапазон балловых значений определяется интервалом  $[0,100]$  согласно системе 60 в семестре плюс 40 за экзамен. Формализация такого описания может быть проведена с помощью следующей ЛП.

- $\beta$  – успеваемость студента;
- $T$  – {«низкая успеваемость», «средняя успеваемость», «высокая успеваемость»};
- $X$  –  $[0,100]$ ;
- $G$  – процедура образования новых термов с помощью связей «и» «или» и модификаторов типа «очень», «слегка», «не» и т. п.
- $M$  – задает на  $X$  нечеткие подмножества из  $T$  в соответствии с правилами трансляции из  $G$ .

В силу специфики рассматриваемой задачи при построении функций принадлежности для нечетких систем принятия решений, будут применяться групповые прямые методы с привлечением экспертных оценок. Примеры таких функций для терм-множества представленной выше лингвистической переменной отображены на рис. 1.

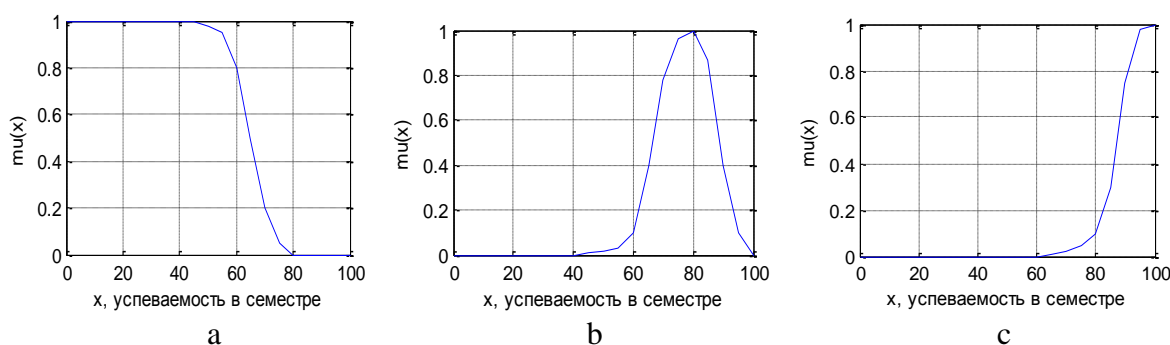


Рис. 1. Функции принадлежности для значений «низкая успеваемость» (а), «средняя успеваемость» (b) и «высокая успеваемость» (с)

Аналогично при помощи нечетких множеств формируются описания остальных характеристик объекта для составления базы правил. Ее формирование основано на пошаговом режиме обучения, в результате которого получают нечеткое дерево решений. В каждом узле дерева при обучении корректируются значения и интерполируют функции принадлежности с целью последующей фаззификации входных переменных. На этапах агрегирования получаем оптимизацию количества нечетких логических правил. Аккумуляирование заключений и дефаззификация происходят пошагово в каждом узле дерева решений. Тестирования построенного нечеткого дерева решений происходит на примерах реальных данных. Далее проверяется критерий качества обучения системы, путем его сравнивая с целевым значением. При достижении целевого значения считается, что построение нечеткого дерева решений завершено и создана база нечетких правил.

На данном этапе работы в качестве модели базы знаний выбрана структура нечетких деревьев решений, так как они способны решать поставленные задачи классификации и регрессии, имеют высокую скорость обработки данных и представляют заключение в виде, доступном для понимания любому пользователю. Определены переменные, в том числе лингвистические, их диапазоны, сформированы функции принадлежности, используемые при построении базы решающих правил.

### Список литературы

1. Бразовская Н.В., Бразовская О.В. Математические методы принятия управленческих решений: учебное пособие. – Барнаул: изд. АлтГТУ, 2004. – 153 с.
2. Hand D., Mannila H., Smyth P. – Principles of data mining, MIT, 2001.
3. Чернов В.Г. Нечеткие деревья решений (нечеткие позиционные игры) // Информационно-управляющие системы. – 2010. – № 5.
4. Нечеткие множества в моделях управления и искусственного интеллекта / под. ред. Д.А. Поспелова. – М.: Наука, 1986. – 312 с.

УДК 004

## ФОРМИРОВАНИЕ БАЗЫ ЗНАНИЙ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ МЕДИЦИНСКИХ ИССЛЕДОВАНИЙ

Шагарова М.Д.

Научный руководитель: Мокина Е.Е., старший преподаватель, каф. ОСУ ИК ТПУ

*Национальный Исследовательский Томский политехнический университет,  
634050, Россия, г. Томск, пр. Ленина, 30  
E-mail: mds1@tpu.ru*

*The paper gives the example of the description process of knowledge base formation for decision support system. This paper describes the need of data mining tools implementation to construct the rule-based for intelligent component of DSS medical studies.*

**Key words:** DSS, data mining, knowledge data, rules logic inference, rule-based, medical studies

**Ключевые слова:** СППР, интеллектуальный анализ данных, база знаний, правила логического вывода, продукционные модели, медицинские исследования

### Введение

Целью данного исследования является поиск средств для формирования базы знаний системы поддержки принятия решения в области медицинских исследований. Актуальность данного исследования определяется: с одной стороны, необходимостью разработки программного продукта, поддерживающего процесс постановки диагноза и анализа показателей здоровья в динамике, а с другой – тем, что построение СППР представляет собой наукоемкий процесс, предполагающий использование различных методов анализа данных.

### Средства интеллектуального анализа данных для СППР

Важным компонентом системы поддержки принятия решения являются знания, содержащиеся в нем, которые могут быть представлены в виде базы знаний, наборов правил. Неотъемлемой частью такой системы является интеллектуальный компонент, с помощью которого осуществляется поиск функциональных и логических закономерностей в накопленных данных, построение моделей и правил, объясняющих найденные закономерности и/или с определенной вероятностью прогнозирующих развитие некоторых процессов. Этот класс задач решается построением систем интеллектуального анализа, реализующего методы и алгоритмы Data Mining.