

Исследование выполнено на базе Томского политехнического университета при финансовой поддержке Министерства образования и науки Российской Федерации в рамках выполнения научно-исследовательских работ по направлению «Оценка и улучшение социального, экономического и эмоционального благополучия пожилых людей», договор No 14.Z50.31.0029.

ЛИТЕРАТУРА

1. Stanley Colcombe and Arthur F. Kramer. Fitness effects on the cognitive function of older adults: a meta-analytic study. *Psychological Science* March 2003 vol. 14 no. 2 125-130
2. C.B. Chan, E. Spangler, J. Valcour, C. Tudor-Locke, "Cross-sectional Relationship of Pedometer-Determined Ambulatory Activity to Indicators of Health," *Obesity Research*, Vol.11, No. 12, (Dec 2003).
3. F.A. Treiber, T. Baranowski, D.S. Braden, W.B. Strong, M. Levy, W. Knox, "Social Support for Exercise: Relationship to Physical Activity in Young Adults," *Preventative Medicine*, Vol. 20, (1991), pp.737-50

ПРИМЕНЕНИЕ НЕЧЕТКИХ ДЕРЕВЬЕВ РЕШЕНИЙ ДЛЯ ОЦЕНКИ ПОТЕНЦИАЛА СТУДЕНТОВ

Т.А. Пискунова, О.Г. Берестнева
(г. Томск, Томский политехнический университет)
e-mail: tana.alex.a42@yandex.ru

THE IMPLEMENTATION OF FUZZY DECISION TREES FOR STUDENTS' POTENTIAL ESTIMATION

T.A. Piskunova, O.G. Berestneva
(Tomsk, Tomsk Polytechnic University)

The following article presents the application of fuzzy decision tree to students' potential estimation for creating DSS engine. The linguistic terms and membership functions are defined and the resulting rule base is given.

Keywords: fuzzy decision trees, classification, DSS, data mining, fuzzy logic

Введение. В связи с увеличением объема информации, поступающей к руководителям и усложнением решаемых задач необходимо создание информационных систем, предназначенных для решения слабоструктурированных и неструктурированных. В них используются научные методы, ищущие оптимальное решение путем обработки больших массивов информации с использованием современных технических средств.

Одной из задач, которые должна выполнять ИСППР является задача классификации студентов по учебно-научному потенциалу, т.е. ставится цель определения предпочтительной дальнейшей траектории развития студента или же прекращения обучения.

Задача классификации. Основная цель классификации состоит в построении правила, в соответствии с которым устанавливается, к какому из классов объектов может быть отнесен классифицируемый объект. При этом под классом понимается некоторая совокупность объектов, обладающих близкими свойствами.

Задача классификации решается в два этапа: построение классификационной модели и ее использование. В данной работе для этого применяются нечеткие деревья решений. На этапе построения модели строится дерево классификации или создается набор неких правил. На этапе использования модели построенное дерево, являющееся набором правил для конкретного объекта, применяется к новым данным для дальнейшей классификации.

В отличие от простых деревьев решений, где каждый пример 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. Процесс обучения происходит быстро, а результат прост для интерпретации. Конечный результат представляется в форме нечетких чисел.

Определение участвующих в решении переменных. СППР предназначены для решения слабоструктурированных и неструктурированных задач, которые содержат как количественные, так и качественные переменные, причем качественные аспекты проблемы имеют тенденцию доминировать. Тем не менее, такие неструктурированные данные все же возможно описать при помощи разных типов шкал. Выделим из таких данных необходимые переменные.

Успешность в образовании, прежде всего, складывается из успеваемости и участия в олимпиадах и конкурсах, позволяющих выявить учебные достижения студента. Но чтобы провести всесторонний анализ студента, необходимо учитывать не только его достижения в учебе и науке, а также знать его личностные качества. На успешность влияют и психологические параметры личности, такие как Логический интеллект, вербальный интеллект, коммуникабельность, консерватизм, тревожность, мотивация и др.

Также интересно посмотреть, существует ли зависимость успехов и достижений студента от типа населенного пункта, в котором он обучался ранее. В оценке стоит учесть и то, является ли студент какой-либо группы ее старостой или профоргом.

Теперь сведем все необходимые элементы с их допустимыми значениями (табл. 1)

Таблица 1. Характеристики для классификации

Достижение	Тип шкалы	Значения	Весовой коэффициент	Обозначение
Тип нас пункта	Номинальная	1,2,3,4	0,25	locality
Институт			0,25	Inst
Уровень обучения			0,25	Dgree
Курс			0,25	year
Научный потенциал	интервальная	[0,10000]	4	publications
Участие в грантах, научные стипендии, патенты	интервальная	[0,10000]	6	sciach
Средний балл	отношений	[0,100]		avgScore
оценка	ранговая	2,3,4,5		mark
цикл ГСЭ	отношений	[0,5]	1	HSE
цикл ЕН	отношений	[0,5]	2	SCI
цикл ОПД	отношений	[0,5]	2	GP
цикл СД	отношений	[0,5]	3	SP
ФТД	отношений	[0,5]	1	Fac
Успевал/не успевал вовремя закрыть сессию	интервальная	[0,1]	0,5	intime
Пропуски	интервальная	[0,1]	0,5	miss
Олимпиады, конкурсы	интервальная	[0,10000]	5	olympiad
Спорт	ранговая	0, 0.5, 1	1	sport
Творческая деятельность	Ранговая	0, 0.5, 1	1	art
Общественная деятельность	Ранговая	0, 0.5, 1	1	socact
Участие в организации группы	Номинальная	0,1,2	1	lead
Креативность	Ранговая	[0,12]	1	cr

Логический интеллект	Ранговая	[0,12]	1	log
Вербальный интеллект	Ранговая	[0,12]	1	spint
Коммуникабельность	Ранговая	[0,12]	1	com
Консерватизм	Ранговая	[0,12]	1	cons
Тревожность	Ранговая	[30,60]	1	wor
Мотивация	Ранговая	[-25,55]	1	mot

При описании объектов и явлений с помощью нечетких множеств используются лингвистические переменные (β , T, X, G, M), где β – наименование лингвистической переменной; T – множество ее значений (термов); G – синтаксическая процедура для оперирования элементами терм-множества T,

определяется научный потенциал студента с помощью понятий «Низкий», «Средний» и «Высокий»

science = (β , T, X, G, M),

где β = научный потенциал студента;

T = {«Низкий», «Средний», «Высокий»};

X = [0,1000];

G – процедура образования новых термов с помощью связок «и» «или» и модификаторов типа «очень», «слегка», «не», «почти», «не совсем», и т.п.

M – процедура задания на X = [0, 1000] нечетких подмножеств: A1 = «Низкий», A2 = «Средний» и A3 = «Высокий», а также нечетких множеств для термов из G(T) в соответствии с правилами трансляции нечетких связок и модификаторов.

Для каждого терма из T создается функция принадлежности

Функции принадлежности, определяющие A1 = «Низкий», A2 = «Средний» и A3 = «Высокий» (рис. 1), представлены на рисунке ниже:

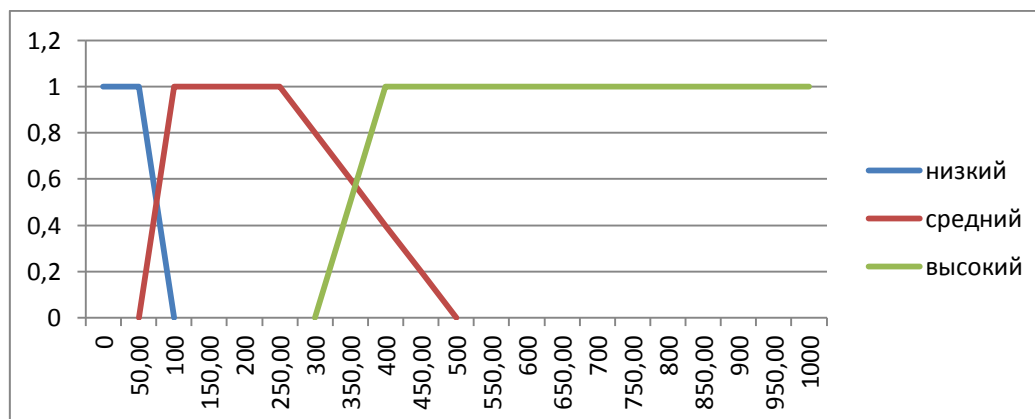


Рис. 1. Функции принадлежности для термов ЛП «научный потенциал студента»

Аналогично определяются лингвистические переменные для всех остальных показателей.

ЛП для классификации является один из классов: «Идеальный студент», «Хорошист», «Средний», «Проблемный»

Построение дерева решений. Формирование базы правил основано на пошаговом режиме обучения, в результате которого получают нечеткое дерево решений. Обучение происходит на тестовых данных. 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. При достижении целевого значения считается, что построение нечеткого дерева решений завершено и создана база нечетких правил.

Построим дерево по набору минимальных характеристик: успеваемости в разных предметных циклах и достижений в науке и учебе. Пример использованных данных для построения и тестирования дерева представлены в таблице 2.

Таблица 2 – Данные для обучения и тестирования дерева решений

Научный потенциал	ГСЭ	ЕН	ОПД	СД	ФТД	Класс
-1	0,6	0,8	0,86	0,94	-1	В
-1	0,8	0,8	0,88	0,95	-1	В
-1	0,7	1	0,95	0,92	-1	В
0,018	1	1	1	0,96	1	А
0,023	1	1	1	0,95	-1	А
-1	0,8	0,9	0,92	0,94	-1	В
-1	1	1	0,86	0,8	0,8	В
0,196	1	1	1	1	-1	А
-1	0,85	0,74	0,8	0,74	-1	С
-1	-1	0,74	0,96	0,9	-1	В
-1	-1	0,9	0,92	0,98	0,975	В
-1	-1	1	1	1	0,975	А
-1	-1	0,74	0,74	0,76	0,8	С
-1	-1	0,8	0,88	0,92	0,925	В
-1	-1	0,8	0,88	0,96	1	В

Алгоритм построения дерева решений допускает существование пропущенных значений. Поэтому в приведенной таблице ячейки, значения которых неизвестны, отмечены числом «-1».

В итоге получаем следующую структуру (рис. 2).

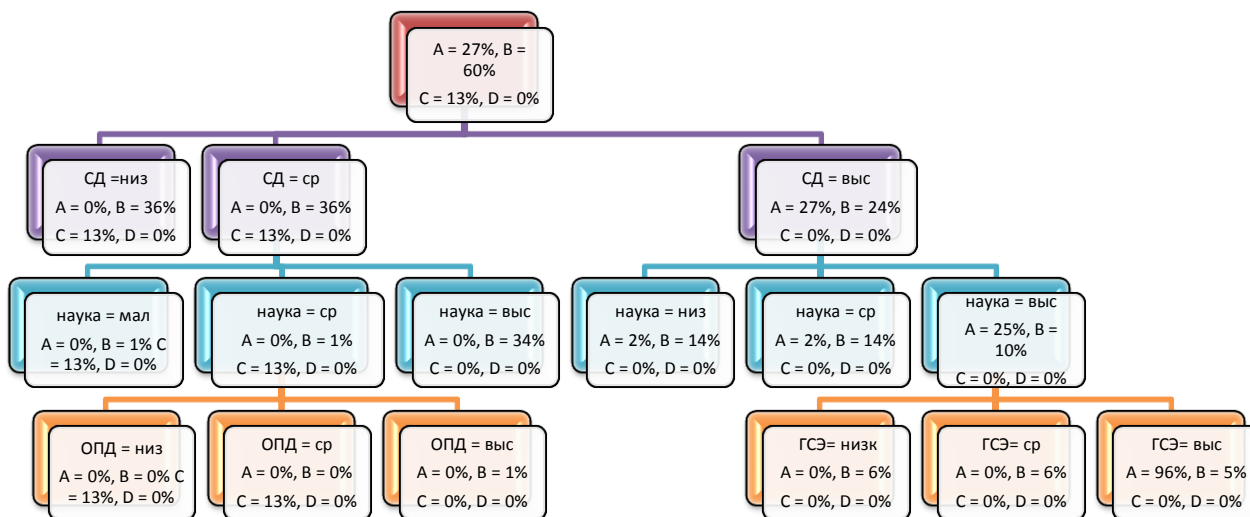


Рис. 2. Дерево решений

Такая структура является базой правил в виде нечеткого дерева решений. Здесь при формировании правил происходит расщепление по разным показателям на разных уровнях. Результат расщепления в листах принадлежит не одному конкретному классу, а отображает их степень принадлежности, что дает более полную и качественную оценку результатов классификации.

Заключение. Построена база правил в виде нечетких деревьев решений. Проведена классификация объектов из базы данных. Полученные результаты продемонстрировали адекватность и работоспособность выбранных методов для решения задач классификации и формирования рейтинга анализируемых объектов. Построенные деревья решений отображают структуру связей и закономерностей между свойствами объектов и целевыми значениями. В дальнейшем они используются в качестве базы правил для классификации и построения рейтинга для новых объектов (студентов).

Исследование выполнено при финансовой поддержке РФФИ в рамках проекта № 14-06-00026

ЛИТЕРАТУРА

1. Hand D., Mannila H., Smyth P. - Principles of data mining, MIT, 2001
2. В. Г. Чернов. Нечеткие деревья решений (нечеткие позиционные игры), Информационно-управляющие системы №5, 2010
3. Пискунова, Татьяна Александровна. Применение интеллектуального анализа данных для создания системы решающих правил / Т. А. Пискунова, О.Г. Берестнева // XIII Всероссийская научно-практическая конференция студентов, аспирантов и молодых ученых «Технологии Microsoft в теории и практике программирования», г. Томск, 2016 / Национальный исследовательский Томский политехнический университет (ТПУ) . — Томск: Изд-во ТПУ, 2016.
4. Janikow C.Z. Fuzzy Decision Trees: Issues and Methods / C.Z. Janikow – IEEE Trans Syst Man Cybern, 2008. – P. 1 – 14.
5. Жиров В.Г. Графическое представление и анализ нечеткой модели логического вывода в базе знаний информационной системы. – Самара, 2010.

ИСПОЛЬЗОВАНИЕ ПАКЕТА WIZWHY ДЛЯ ФОРМИРОВАНИЯ БАЗЫ ЗНАНИЙ ЭКСПЕРТНЫХ СИСТЕМ

*A.S. Seidova, O.V. Marukhina
(г. Томск, Томский политехнический университет)
e-mail: aysel4421@mail.ru*

USE OF A WIZWHY PACKAGE FOR FORMATION OF THE KNOWLEDGE BASE EXPERT SYSTEMS

*A.S. Seidova, O. V. Marukhina
(Tomsk, Tomsk Polytechnic University)*

In the early eighties the last century in researches on artificial intelligence the independent direction which has received the name "expert systems" was created. Modern expert systems represent computer programs which can analyze on the basis of certain basic data and capable to replace specialists of a narrow profile in problem situations. Use of modern information technologies becomes a critical factor of development of the majority of industries of knowledge and areas of practical activities therefore development and deployment of information systems is one of the most actual tasks.

Key words: medicine, expert system, diseases, artificial intelligence, information technologies.

В начале восьмидесятых годов прошлого века в исследованиях по искусственному интеллекту сформировалось самостоятельное направление, получившее название "экспертные системы" (ЭС). Первыми предпосылками развития этой научной отрасли стали так