

УДК 004.832.3

ПРИМЕНЕНИЕ НЕЙРО-НЕЧЕТКОЙ СИСТЕМЫ ВЫВОДА ДЛЯ УСТАНОВЛЕНИЯ ДИСЦИПЛИН, СООТВЕТСТВУЮЩИХ КОМПЕТЕНЦИИ ОБРАЗОВАТЕЛЬНОГО СТАНДАРТА

Найханова Л.В., Дышенов Б.А.

*ФГБОУ ВПО «Восточно-Сибирский государственный университет технологий и управления»,
Улан-Удэ, e-mail: obeka_nlv@mail.ru, dyshenov@gmail.com*

В данной статье рассматривается одна из задач формирования учебного плана – задача поиска дисциплин, соответствующих заданной компетенции. Решение осуществляется на основе применения латентно-семантического анализа и нечетких нейронных сетей. Для проведения анализа формируются две коллекции текстов с содержанием компетенции и с содержанием дисциплин. На основе латентно-семантического анализа определяется расстояние от дисциплины до заданной компетенции. Для повышения качества полученных результатов в работе создается гибридная модель в виде нейро-нечеткой сети. В статье показано, что применение адаптивной нейро-нечеткой сети позволяет определить степень принадлежности дисциплины заданной компетенции. Эксперименты проведены для общепрофессиональных компетенций федерального государственного образовательного стандарта (ФГОС 3+) направления подготовки 02.03.03 «Математическое обеспечение и администрирование информационных систем».

Ключевые слова: гибридная модель, нейронные сети, системы нечеткого логического вывода, обучение, компетентностная модель образовательного стандарта, дисциплины учебного плана

APPLICATION OF NEURO-FUZZY INFERENCE SYSTEM FOR THE DEFINITION OF DISCIPLINES RELEVANT A COMPETENCE OF EDUCATIONAL STANDARD

Naykhanova L.V., Dyshenov B.A.

*Federal State Educational Institution of Higher Professional Education East-Siberian State University
of Technology and Management, Ulan-Ude, e-mail: obeka_nlv@mail.ru, dyshenov@gmail.com*

This article discusses the task of finding the relevant disciplines competence. The solution is based on the use of latent semantic analysis and fuzzy neural networks. For analysis formed two collections of texts with the content of the competence and the contents of disciplines. The distance from the subjects to the desired competence determine by latent semantic analysis. Neuro-fuzzy network allows us to refine the relevant disciplines competence. Consequently is calculated degree of membership of discipline given competence. Experiments were carried out for the preparation of competencies direction 02.03.03 «Software and administration of information systems».

Keywords: hybrid model, neural networks, fuzzy logic inference systems, training, competence model of the educational standard, discipline curriculum

Более десяти последних лет отечественная система высшего профессионального образования находится на этапе внедрения компетентностного подхода. По всей вероятности, большинству участников образовательного процесса пришло понимание сути этого подхода. Однако окончательное внедрение компетентностного подхода произойдет еще не скоро. Естественно, что автоматизация хотя бы некоторых процессов создания компетентностной модели и выбора дисциплин учебного плана в какой-то мере будет способствовать решению этой проблемы.

В данной статье рассматривается задача поиска дисциплин, соответствующих заданной компетенции. Решение осуществляется на основе применения латентно-семантического анализа и нечетких нейронных сетей.

Методика решения задачи

Для того чтобы система имела понятие о сути компетенции, построим для каждой

компетенции ее иерархическую структуру. На первом уровне компетенция разбивается на составляющие, которые на следующем уровне снова декомпозируются. Декомпозиция заканчивается на уровне «простых» терминов. Для примера рассмотрим общепрофессиональную компетенцию ОПК-4 «Способность применять в профессиональной деятельности основные методы и средства автоматизации проектирования, производства, испытаний и оценки качества программного обеспечения» по направлению подготовки 02.03.03 «Математическое обеспечение и администрирование информационных систем». В ней можно выделить три составляющие:

1) способность применять в профессиональной деятельности основные методы и средства автоматизации проектирования программного обеспечения;

2) способность применять в профессиональной деятельности основные методы и средства производства программного обеспечения;

3) способность применять в профессиональной деятельности основные методы и средства испытаний и оценки качества программного обеспечения.

Каждая из приведенных составляющих разбивается на компоненты. Например, третья составляющая имеет следующие компоненты:

а) способность применять в профессиональной деятельности основные методы и средства испытаний программного обеспечения:

– способность применять в профессиональной деятельности основные методы испытаний программного обеспечения;

– способность применять в профессиональной деятельности основные средства испытаний программного обеспечения;

б) способность применять в профессиональной деятельности основные методы и средства оценки качества программного обеспечения:

– способность применять в профессиональной деятельности основные методы оценки качества программного обеспечения;

– способность применять в профессиональной деятельности основные средства оценки качества программного обеспечения.

Минимум четырехуровневая иерархия компетенции ОПК-4 имеет форму дерева. Листочки дерева содержат термины (например, «основные методы испытаний программного обеспечения»), по которым можно осуществлять запросы в Wikipedia. Трассировка статей Wikipedia позволяет выделить нужную информацию об объекте поиска (термине, содержащемся в листе дерева). Будем считать, что коллекция статей с найденной информацией по всем терминам листочков составляет содержание компетенции.

С другой стороны, по многим направлениям подготовки в интернете имеются рабочие программы по различным дисциплинам. При формировании коллекции рабочих программ по соответствующему направлению подготовки можно рассмотреть включение дисциплин и по смежным направлениям под-

готовки. Коллекция текстов с содержанием дисциплин по направлению подготовки создается посредством трассировки выбранных рабочих программ дисциплин.

Сформированные коллекции позволяют провести над ними латентно-семантический анализ (ЛСА), который определяет взаимосвязь между коллекциями текстов по дисциплинам и текстами с содержанием компетенции, а также между встречающимися в них терминами на основе сопоставления некоторых факторов (тематики) по всем текстам и терминам. В основе метода латентно-семантического анализа лежат принципы факторного анализа, в частности выявление латентных связей изучаемых явлений или объектов [6]. Латентно-семантический анализ позволил получить расстояния дисциплин до центров компетенций [2]. При этом достоверность результатов была недостаточно высокой.

Для уточнения этих результатов в работе предлагается использовать гибридную модель нейронных сетей и нечеткой логики – ANFIS-сеть. Акроним ANFIS (Adaptive Neuro-Fuzzy Inference System) получил свое название от адаптивной нейро-нечеткой системы вывода типа Сугено. Системы, разработанные по нейро-нечеткой технологии ANFIS, обладают хорошей сходимостью, а субтрактивная кластеризация позволяет синтезировать компактное множество нечетких продукций на основе данных обучающей выборки [1, 5].

Посредством нейро-нечеткой технологии ANFIS в работе смоделирована нечеткая система вывода (FIS) с тремя входами, заданными множеством $X = \{x_1, x_2, x_3\}$ и одним выходом – y . FIS состоит из четырех компонентов (рис. 1).

Фазификатор, применяя функцию принадлежности Гаусса, преобразует значение входного сигнала $x_j \in X$ в нечеткое значение \tilde{x}_j . Нечеткий логический вывод отображает нечеткие множества входного пространства X на нечеткое множество выходного пространства \tilde{Y} .

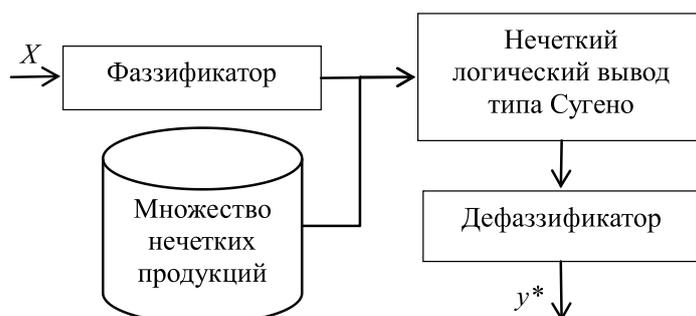


Рис. 1. Схема системы нечеткого вывода

Дефаззификатор представляет собой отображение нечеткого множества \tilde{Y} в точку y^* [3]. Таким образом, каждое из m правил определяется выходным нечетким множеством $y \subset Y$, полученным с помощью вычисления правила вида:

$$\text{если } (x_1 = \tilde{x}_i) \wedge (x_2 = \tilde{x}_i) \wedge (x_3 = \tilde{x}_i), \quad (*)$$

$$\text{то } (y^* = \tilde{y}_i),$$

где \tilde{x}_i – центр нечеткого множества, являющийся центром i -го кластера, найденного при субтрактивной кластеризации.

Множество входов X составляют:

а) x_1 – расстояние дисциплины от центра компетенции;

б) x_2 – относительная частота встречаемости терминов определения компетенции в содержании дисциплин;

в) x_3 – относительная частота встречаемости терминов текста самой компетенции в содержании дисциплин.

Выход y^* – степень принадлежности дисциплины компетенции.

Экспериментальная часть

Для проведения экспериментов выбраны семь общепрофессиональных компетенций ФГОС 3+ по направлению подготовки

02.03.03 «Математическое обеспечение и администрирование информационных систем». Для формирования коллекции текстов по дисциплинам использован перечень дисциплин, представленный в табл. 1.

Для обучения сети по каждой компетенции были сформированы наборы данных. Пример набора для компетенции ОПК-5 «Владение информацией о направлениях развития компьютеров с традиционной (нетрадиционной) архитектурой; о тенденциях развития функций и архитектур проблемно-ориентированных программных систем и комплексов» показан в табл. 1.

Значения x_1 получены в результате выполнения латентно-семантического анализа; x_2 и x_3 – спектрального оценивания лексических единиц текстов; значения y сформированы экспертом. Обучающая выборка составила 88 образцов (от ОПК-5 до ОПК-8), проверочная и тестовая выборки – матрицы по ОПК-9 и ОПК-10 соответственно.

При проведении субтрактивной кластеризации заданы следующие значения параметров: Range of Influence = 0.4; QuashFactor = 0,95; для AcceptRatio и RejectRatio использованы значения по умолчанию. В результате синтезированы шесть правил вида (*), представленные на рис. 2. Также определена структура нейро-нечеткой сети (рис. 3).

Таблица 1

Набор данных по компетенции ОПК-5

№ п/п	Дисциплина	x_1	x_2	x_3	y_3
1	Архитектура вычислительных и компьютерных систем	0,58	0,111	0,152	1,0
2	Базы данных и СУБД	0,8	0,016	0,061	0,0
3	Дискретная математика	1,04	0,016	0,000	0,0
4	Информатика	0,78	0,095	0,030	0,1
5	Компьютерное моделирование	0,75	0,095	0,030	0,0
6	Методы прикладного системного анализа	0,72	0,016	0,030	0,2
7	Методы и средства проектирования ПО	0,82	0,032	0,030	0,0
8	Метрология, стандартизация и сертификация	0,93	0,111	0,000	0,0
9	Объектно-ориентированное программирование	0,65	0,000	0,030	0,1
10	Операционные системы	0,83	0,159	0,061	0,3
11	Организация ЭВМ и систем	0,67	0,079	0,182	1,0
12	Проектирование и архитектура программных систем	0,83	0,032	0,212	1,0
13	Рекурсивно-логическое программирование	0,65	0,016	0,000	0,0
14	Структуры и алгоритмы обработки данных	0,76	0,095	0,000	0,0
15	Системы искусственного интеллекта	0,75	0,016	0,061	0,2
16	Система реального времени	0,73	0,016	0,030	0,2
17	Теория вероятностей и математическая статистика	1,04	0,016	0,000	0,1
18	Теория систем	0,78	0,016	0,030	0,0
19	Теория формальных грамматик и автоматов	0,69	0,016	0,000	0,1
20	Технология разработки программного обеспечения	0,77	0,000	0,061	0,3
21	Функциональное программирование	0,66	0,016	0,000	0,0
22	Экономико-правовые основы рынка ПО	1,06	0,032	0,000	0,0

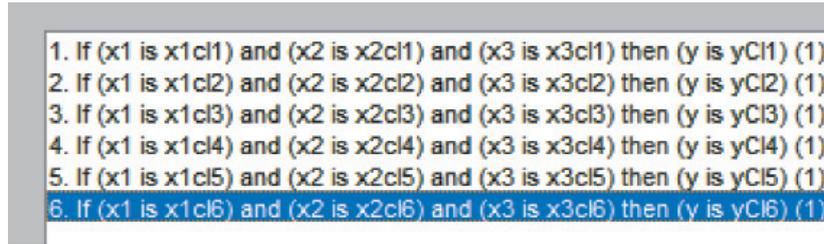


Рис. 2. Множество нечетких продукций

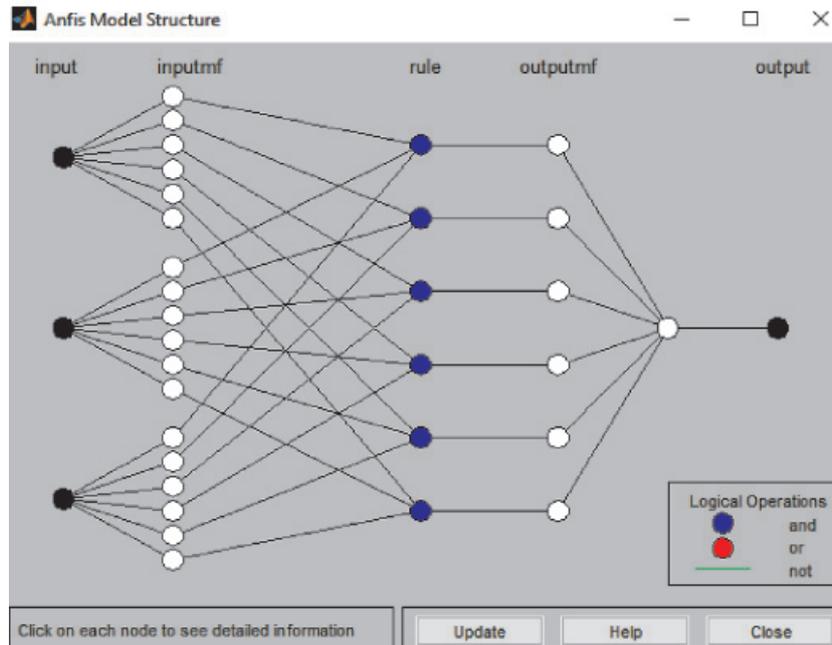


Рис. 3. Нейро-нечеткая сеть

Затем выполнено обучение нейронной сети, ошибки обучения и проверки на пятом шаге составили 0,002 и 0,009 соответственно и больше не изменялись. ANFIS создает систему нечеткого вывода типа Сугено в виде четырехслойной нейронной сети прямого распространения сигнала. На рис. 3 изображена ANFIS-сеть с тремя входами $X = \{x_1, x_2, x_3\}$ и шестью нечеткими правилами. Согласно рисунку входные сигналы X (*input*) соединены с узлами первого слоя (*inputmf*), в котором расположены термы типа « x_i около кластера j ».

Каждый входной сигнал x_i со своими термами составляют пучок (один столбец шести правил, рис. 2). Соединение x_i с термом j есть одна посылка j -го правила. Поэтому выходом этого слоя является степень принадлежности d_{ij} значения входа x_i j -му терму, вычисленная с помощью функции принадлежности Гаусса. Во втором слое расположены нечеткие продукции, их количество равно количеству кластеров (m), обнаруженных в процессе субтрактив-

ной кластеризации. Посылки, вычисленные в первом слое, поступают в соответствующую нечеткую продукцию (*rule*, r). В этом слое вычисляется степень выполнения правила d_r ($r = \overline{1, m}$), которая рассчитывается как t -норма d_{ij} и выполняется нормализация степеней выполнения правил r или вычисление относительной степени выполнения правила r :

$$d_r^* = \frac{d_r}{\sum_{j=1}^m d_j}$$

Третий слой – заключения правил, в узлах которого рассчитывается вклад нечеткого правила в выход сети:

$$y_r = d_r^* \cdot (b_{0,r} + b_{1,r}x_1 + b_{2,r}x_2 + b_{3,r}x_3),$$

где $b_{i,r}$ – коэффициенты линейной зависимости выходного параметра.

В четвертом слое осуществляется агрегирование результатов всех правил. Этот

слой имеет один узел, в котором вычисляется результирующее значение y :

$$y = \sum_{r=1}^m y_r.$$

Для обучения ANFIS-сети применен гибридный метод, состоящий из градиентного спуска в виде алгоритма обратного распространения ошибки и метода наименьших квадратов. Алгоритм обратного распространения ошибки настраивает параметры функций принадлежности antecedентов правил. Метод наименьших квадратов применяется для настройки коэффициентов линейной зависимости в заключениях правил.

Каждая итерация процедуры настройки выполняется в два этапа. На первом этапе

на входы подается обучающая выборка, и по невязке между желаемым и действительным поведением сети итерационным методом наименьших квадратов находятся оптимальные параметры узлов третьего слоя. На втором этапе остаточная невязка передается с выхода сети на входы, и методом обратного распространения ошибки модифицируются параметры узлов первого слоя. При этом найденные на первом этапе коэффициенты заключений правил не изменяются. Итерационная процедура настройки продолжается, пока невязка превышает заранее установленное значение [5]. Таким образом, обучение завершается настройкой векторов коэффициентов линейной зависимости и параметров функции принадлежности (табл. 2).

Таблица 2

Результирующие данные обучения

Правило Кластер	Вектор параметров функции Гаусса			Вектор коэффициентов линейной зависимости
	x_1	x_2	x_3	y
1/c11	[0.1358 0.3795]	[0.0378 0.0206]	[0.03544 -0.00193]	[0.6461 6.024 13.35 -0.1114]
2/c12	[0.1347 0.6593]	[0.0314 0.01178]	[0.03474 0.03078]	[4.953 -7.413 -2.563 -2.785]
3/c13	[0.1345 0.8502]	[0.03073 0.001926]	[0.03896 0.000531]	[1.668 -10.04 -10.23 -1.608]
4/c14	[0.1336 0.8307]	[0.0331 0.0624]	[0.0558 0.0695]	[0.8125 -0.8705 2.218 0.2758]
5/c15	[0.1316 0.42]	[0.0311 0.06122]	[0.03658 0.03686]	[3.804 2.856 1.039 -2.108]
6/c16	[0.1336 0.8307]	[0.02797 0.06264]	[0.04077 0.02637]	[0.824 0.3527 4.688 -0.7646]

Таблица 3

Степени принадлежности дисциплины компетенции ОПК-4

№ п/п	Наименование дисциплины	y_{FIS}
1	Методы и средства проектирования ПО	0,936
2	Проектирование и архитектура программных систем	0,758
3	Технология разработки программного обеспечения	0,736
4	Метрология, стандартизация и сертификация	0,569
5	Методы прикладного системного анализа	0,337
6	Объектно-ориентированное программирование	0,192
7	Компьютерное моделирование	0,167
8	Теория систем	0,158
9	Экономико-правовые основы рынка ПО	0,130
10	Функциональное программирование	0,107
11	Теория вероятностей и математическая статистика	0,105
12	Дискретная математика	0,103
13	Информатика	0,098
14	Архитектура вычислительных и компьютерных систем	0,096
15	Базы данных и СУБД	0,089
16	Системы искусственного интеллекта	0,076
17	Система реального времени	0,073
18	Операционные системы	0,054
19	Организация ЭВМ и систем	0,033
20	Структуры и алгоритмы обработки данных	0,032
21	Рекурсивно-логическое программирование	0,029
22	Теория формальных грамматик и автоматов	0,010

Как видно из табл. 2, четвертый и шестой кластеры можно объединить, тогда количество правил будет равно пяти, что незначительно влияет на погрешность вычислений.

Результаты

Для проверки работоспособности были обработаны данные по компетенции ОПК-4 «Способность применять в профессиональной деятельности основные методы и средства автоматизации проектирования, производства, испытаний и оценки качества программного обеспечения», не участвовавшие в создании ANFIS-сети. В табл. 3 данные отсортированы по убыванию вычисленного значения y_{FIS} . Результаты показывают, что первые четыре дисциплины, перечисленные в табл. 3, являются основными для компетенции ОПК-4, т.е. это межпредметная компетенция, и результаты обучения должны быть распределены между этими четырьмя дисциплинами.

Таким образом, данные проведенных экспериментов показали, что предложенный подход можно использовать для уточнения результатов латентно-семантического анализа.

Заключение

В работе исследована часть компетенций, по которым достаточно легко построить их иерархические структуры. Анализ компетенций по направлению подготовки 02.03.03 – «Математическое обеспечение и администрирование информационных систем» показал, что все компетенции условно можно разбить на три группы по отношению к сложности построения иерархической структуры в автоматическом режиме. Первая группа компетенций рассмотрена в данной статье. Построение иерархий для второй и третьей групп пока возможно только в автоматизированном

режиме. В дальнейшем требуется провести эксперименты для компетенций всех трех групп компетенций.

Список литературы

1. Гушин А.В. Теория и алгоритмы: нечеткие арифметика, кластеризация, синтез знаний и принятие решений в условиях лингвистической неопределенности. – Самара: СамГУПС, 2012. – 96 с.
2. Найханова Л.В. Интеллектуальные средства автоматизированного построения основной образовательной программы / Л.В. Найханова, Б.А. Дышенов Б.А., Н.В. Найханов // Теоретические и прикладные вопросы современных информационных технологий: Материалы XII Всерос. научно-техн. конф. (Улан-Удэ, 11–18 авг. 2015 г.). – Улан-Удэ, 2015. – С. 60–65.
3. Штовба С.Д. Введение в теорию нечетких множеств и нечеткую логику: монография. – Винница: Континент-Прим, 2003. – 198 с.
4. Yager R. Essentials of Fuzzy Modeling and Control / R. Yager, D. Filev – USA: John Wiley & Sons, 1984. – 387p.
5. Jang J.-S. R. ANFIS: Adaptive-Network-based Fuzzy Inference Systems // IEEE Transactions on Systems, Man, and Cybernetics. – 1993. – Vol. 23, № 3. – P. 665–685.
6. Deerwester S. Indexing by Latent Semantic Analysis // Journal of the American Society for Information Science. – 1990. – № 41 (6). – P. 391–407.

References

1. Gushchin A.V. Teoriya i algoritmy: nechetkie arifmetika, klasterizaciya, sintez znanij i prinyatie reshenij v usloviyah lingvističeskoj neopredelennosti. Samara: SamGUPS, 2012. 96 p.
2. Najhanova L.V. Intellectualnye sredstva avtomatizirovannogo postroeniya osnovnoj obrazovatelnoj programy / L.V. Najhanova, B.A. Dyshenov B.A., N.V. Najhanov // Teoreticheskie i prikladnye voprosy sovremennyh informacionnyh tekhnologij: Materialy XII Vseros. nauchno-tekhn. konf. (Ulan-Udeh, 11-18 avg. 2015g.). Ulan-Udeh, 2015. pp. 60–65.
3. SHtovba S.D. Vvedenie v teoriyu nechetkih mnozhestv i nechetkuyu logiku: Monografiya / S.D. SHtova. Vinnica: Kontinent-Prim, 2003. 198 p.
4. Yager R. Essentials of Fuzzy Modeling and Control / R. Yager, D. Filev USA: John Wiley & Sons. 1984. 387 p.
5. Jang J.-S.R. ANFIS: Adaptive-Network-based Fuzzy Inference Systems // IEEE Transactions on Systems, Man, and Cybernetics. 1993. Vol. 23, no. 3. pp. 665–685.
6. Deerwester S. Indexing by Latent Semantic Analysis // Journal of the American Society for Information Science. 1990. 41 (6): 391–407.