



Моделирование сети магистральных авиалиний



Виктор БАЛАШОВ
Victor V. BALASHOV

Андрей СМЕРНОВ
Andrey V. SMIRNOV



Татьяна ЦЕЙТЛИНА
Tatiana O. ZEITLINA

Балашов Виктор Васильевич – кандидат технических наук, доцент, ведущий научный сотрудник Центрального аэрогидродинамического института имени Н.Е. Жуковского (ЦАГИ).
Смирнов Андрей Валентинович – кандидат технических наук, доцент, начальник отдела ЦАГИ.
Цейтлина Татьяна Олеговна – младший научный сотрудник ЦАГИ (г. Жуковский).

Решение задачи формирования модели «условий существования» авиалиний (УСА) – правил, в соответствии с которыми возникают, действуют и исчезают авиалинии (прямая авиасвязь между двумя городами). При построении нечёткой нейросетевой модели УСА используются методы генетического алгоритма и самоорганизующихся карт Кохонена, а также программные средства пакета Fuzzy Logic Toolbox и нейросетевого пакета среды программирования MatLab. Модель адекватно отражает структуру коммуникационного ядра сети авиалиний и дает возможность прогнозировать ее развитие.

Ключевые слова: сеть авиалиний, условия существования авиалиний, нечёткое моделирование, нейросетевая модель.

При создании новой авиационной технологии необходима оценка её вклада в улучшение эксплуатационных характеристик перспективных самолётов. Причем важно не только оценить, насколько та или иная технология повлияет на летные характеристики, но и проанализировать, как её использование в конструкции судна отразится на состоянии авиационной транспортной системы страны в целом. Нужны такие модели, которые позволят сформировать сценарии развития пассажирских авиаперевозок: прогнозируемую интенсивность движения самолётов различных классов на сети авиалиний, степень удовлетворения распределённого спроса населения на услуги гражданской авиации.

Исходя именно из этого, нами была определена задача сформировать модель «условий существования» авиалиний (УСА) – правил, в соответствии с которыми возникает, действует и исчезает прямая авиасвязь между двумя пунктами сообщения.

ПОНЯТИЕ «АВИАЛИНИЯ СУЩЕСТВУЕТ»

Выбор технологии нечёткого моделирования при разработке модели УСА обо-

снован присутствием в описании развития авиатранспортной сети ряда неопределенностей. Они связаны с неточностью описания границ сети магистральных авиалиний или отдельных её состояний, неполнотой модельных представлений об авиатранспортной сети, неопределённостью появления и исчезновения тех или иных авиалиний, а также неоднозначностью семантики (смысла понятий) некоторых терминов, используемых при формулировке задачи. Выходной параметр нейронной сети, характеризующий УСА, представляет собой лингвистическую переменную, определяемую словесно, а не в цифровом виде. Переход к цифровому виду осуществляется путём введения в рассмотрение функции принадлежности выходного параметра нейронной сети.

Для описания УСА наилучшим образом подходит нечёткая продукционная модель. Нечёткие модели данного класса позволяют отражать плохо формализуемые сложные системы и процессы. Формально они могут быть реализованы в виде адаптивных нечётких нейронных продукционных сетей [7], по структуре идентичных многослойным нейронным сетям с прямым распространением сигнала. На элементах каждого слоя (или совокупности слоев) сети осуществляется отдельный этап нечёткого вывода нечёткой продукционной модели:

- первый слой нейронов выполняет функцию введения нечёткости (fuzzification);
- скрытые слои отображают совокупность нечётких правил и реализуют алгоритм нечёткого вывода;
- последний слой выполняет функцию приведения к чёткости (defuzzification).

Основное преимущество нейросетевого подхода — возможность выявления закономерностей в данных (их обобщение), а основной недостаток — трудность формирования представительной выборки для обучения сети. В то же время обобщающие свойства нейросетевой модели в значительной степени определяются качеством обучающей выборки — составом ее элементов (каждый представляет собой совокупность входных и выходного параметров модели), адекватностью значений выходного параметра значениям входных параметров.

ЗАВИСИМЫЕ ЗНАЧЕНИЯ ВЫХОДНОГО ПАРАМЕТРА

Одним из возможных способов задания значений выходного параметра является их определение в зависимости от стадии развития авиалинии.

Для характеристики состояния авиалиний в течение года используется информация о регулярности и интенсивности полётов, содержащаяся в расписании движения. В общем случае на одной и той же линии регулярность и интенсивность могут меняться. На основе данных о навигационном периоде, суммарном количестве дней в году, когда осуществляются полёты, и суммарном количестве рейсов за год определяются осреднённые показатели регулярности и интенсивности полётов воздушных судов (ВС). На основе показателей состояния авиалинии предлагается классификация авиалиний по стадиям развития. Было выделено пять групп авиалиний. Каждую группу объединяет своя стадия развития.

В зависимости от стадии развития авиалинии возможно задание различных значений выходного параметра. Каждой, относящейся к той или иной группе, резонно поставить в соответствие определённое значение выходного параметра (например, 0,6; 0,75; 0,9; 1,0). Для несуществующих авиалиний ставится в соответствие единственное значение выходного параметра (0).

Однако осреднение показателей интенсивности и регулярности полётов ВС не позволяет достаточно обоснованно судить о стабильности состояния авиалинии или направлении и темпах её развития. Иначе говоря, отсутствие информации о динамике развития не дает в полной мере представления об объекте и, следовательно, мешает адекватно раскрыть и содержание понятия «авиалиния существует».

ВЫХОДНОЙ ПАРАМЕТР И ДИНАМИКА РАЗВИТИЯ

При определении выходного параметра, характеризующего наличие или отсутствие прямого авиасообщения между двумя городами России в «базовом» 2006 году, были использованы данные расписаний полётов «соседних» 2005 и 2007 годов. Получены необходимые сведения по каждой авиалинии, существовавшей хотя бы в одном



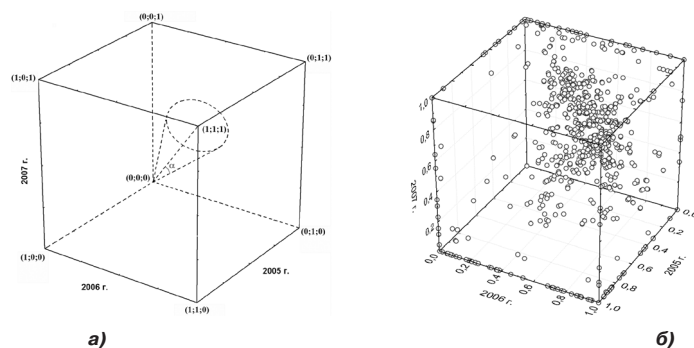


Рис. 1. Нормированное количество вылетов на авиалинии в 2005, 2006 и 2007 гг. (а). Элементы трёхмерного куба нормированных значений числа полётов (б).

из трёх лет, а также суммарное количество вылетов. На основании этих данных делается соответствующее заключение. В частности, если суммарное количество вылетов за год менее 12, то авиалиния рассматривается как эпизодическая и в дальнейшем считается, что она отсутствует. Если же вылетов за год более или равно 12, то фиксируется факт ее наличия.

Если данные о количестве полётов ВС подвергнуть нормировке (отнести к максимальному числу в рассматриваемые годы), то их можно распределить по поверхности трёхмерного куба, на рёбрах которого отложены нормированные значения числа полётов на каждой авиалинии в 2005, 2006, 2007 годах (рис. 1а). Такое представление позволяет визуализировать изменения интенсивности полётов.

На рис. 1б показаны элементы куба. Близость точки, соответствующей авиалинии, к вершине с координатами (0;0;0) означает практическое отсутствие авиалинии в течение трёх лет, а близость точки к вершине с координатами (1;1;1) свидетельствует о том, что число полётов на маршруте в тот же период было достаточно стабильным. Те авиалинии, чьи точки расположены на рёбрах куба, выходящих из его вершины с координатами (0;0;0), присутствуют только в одном из трёх лет. Те авиалинии, чьи точки на гранях куба, проходящих через его вершину с координатами (0;0;0), присутствуют в двух годах из трёх.

Близость точки, соответствующей авиалинии, к пространственной диагонали куба, соединяющей вершины с координатами (0;0;0) и (1;1;1), свидетельствует о стабильности состояния объекта. Степень

этой близости можно оценить углом α между диагональю куба $\vec{a}$ (1;1;1) и вектором, соединяющим начало координат (0;0;0) с точкой $\vec{b}$ (x_1 ; x_2 ; x_3). Здесь x_1 , x_2 , x_3 — нормированные значения числа полётов на авиалинии в 2005, 2006 и 2007 годах. Угол α определяется соотношением $\cos \alpha = \cos(\vec{a} \wedge \vec{b}) = (\vec{a} \times \vec{b}) / (|\vec{a}| \times |\vec{b}|) =$

$$= (1 \cdot x_1 + 1 \cdot x_2 + 1 \cdot x_3) / (\sqrt{1^2 + 1^2 + 1^2} \times \sqrt{x_1^2 + x_2^2 + x_3^2}).$$

По результатам качественного анализа расположения точек авиалиний на поверхности куба можно выделить ряд групп с различной степенью стабильности состояния объектов. Основанием для этого является степень близости точки, соответствующей авиалинии, к пространственной диагонали куба, характеризуемой углом α . Этот угол — показатель стабильности авиалинии на протяжении трёх лет.

В итоге получаем основание построить функцию принадлежности для значения выходного параметра нейронной сети, являющуюся лингвистической переменной, характеризующей нечёткое понятие «авиалиния существует». В рассматриваемой модели функция принадлежности имеет вид кусочно-постоянной функции $y(\alpha) = (90^\circ - \alpha_n(\alpha)) / 90^\circ$, где $\alpha_n(\alpha)$ — начальный угол диапазонов изменения угла α , определённых для сформированных групп авиалиний. Кусочно-постоянная функция принадлежности $y(\alpha)$ приведена на рис. 2.

Поскольку авиалинии в базах расписания рассматриваются как однонаправленные (в одном направлении), производится

симметризация значений выходного параметра нейронной сети для авиалинии в двух направлениях (туда и обратно) путём выбора для обоих наибольшего из двух значений выходного параметра.

Таким образом, в модели каждой авиалинии ставится в соответствие (в зависимости от степени стабильности их состояния в течение трёх лет и темпа изменения интенсивности полётов ВС) определенное значение выходного параметра (0,4; 0,6; 0,8; 0,9; 1,0). Для отсутствующих линий на выход сети в процессе обучения подаётся единственное значение (0). Переход от анализа осредненных показателей состояния авиалинии в данном году к анализу степени стабильности её состояния в течение трёх лет позволяет более обоснованно судить и о самом объекте, и о содержании понятия «авиалиния существует».

ПАРАМЕТРЫ, ХАРАКТЕРИЗУЮЩИЕ АВИАЛИНИЮ

Изначально рассматривается 18 входных параметров, которые, как предполагается, могут оказывать влияние на существование прямого авиасообщения между двумя городами. Затем исключается ряд параметров, образованных путем суммирования или перемножения однотипных параметров обоих городов авиалинии. В таблице 1 представлены в качестве примера для ряда авиалиний значения оставшихся 12 входных параметров (F1-F12), а также значения выходной значимой переменной (выходного параметра) (E1).

В качестве входных параметров фигурируют:

- F1: класс авиалинии; определяется наибольшим значением класса аэродрома вылета/прилёта. Последние определяются в зависимости от типов принимаемых ВС (таблица 2);
- F2: длина авиалинии – расстояние между городами вылета и прилёта (длина ортодромии, соединяющей аэропорты городов);
- F3: наличие (1) или отсутствие (0) железнодорожного сообщения между городами авиалинии;
- F4: валовый региональный продукт (ВРП') субъекта федерации, пересчитанный в соответствии с минимальным прожиточным минимумом для города вылета;

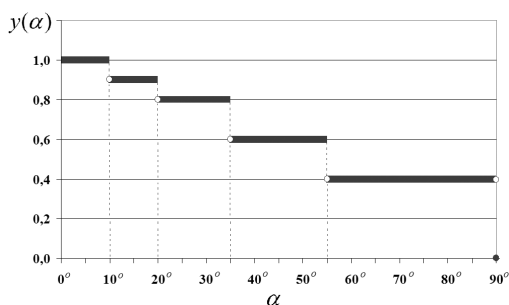


Рис. 2. Кусочно-постоянная функция принадлежности $y(\alpha)$.

– F5: валовый региональный продукт, пересчитанный в соответствии с минимальным прожиточным минимумом для города прилёта;

– F6: численность населения, отнесенного к городу вылета;

– F7: численность населения, отнесенного к городу прилёта;

– F8: число мест размещения, отнесенное к городу вылета;

– F9: число мест размещения, отнесенное к городу прилёта;

– F10: статус города вылета – экспертный параметр, используются четыре категории и соответствующие им значения переменной:

- 1 – столица страны – МОС (г. Москва).
- 2 – столицы ФО либо их заменители, либо федеральные курорты: АНА (г. Анапа), КЛЦ (г. Екатеринбург), КПА (г. Краснодар), НСБ (г. Новосибирск), ПЛК (г. Санкт-Петербург), СКЧ (г. Самара), СОЧ (г. Сочи).

3 – столицы субъектов федерации (области, края и т.д.), либо промышленные и курортные центры, но не принадлежащие п. 2.

4 – остальные города.

– F11: статус города прилёта (см. выше);

– F12: направление авиалинии (определяется по статусам городов вылета и прилёта, федеральным округам и субъектам федерации).

Согласно целевой концепции формирования обучающей выборки для создания нейросетевой модели УСА проводится «десимметризация» **OD**-матрицы. Множество элементов исходной выборки, на базе которой создаются обучающая, тестовая и проверочная выборки, формируется по результатам сравнения целевых возможностей двух городов авиалинии.





Таблица 2

Класс аэропорта по типу принимаемых ВС

Класс аэропорта*	Принимаемые ВС	
	Масса, т	Типы ВС
1	>300	A-380, A-340, B-747, B-777
2	150÷300	A-330, A-310, B-767, Ил-62М, Ил-86, Ил-96
3	75÷150	A-320, A-321, B-737, B-757, Ту-154, Ту-204
4	30÷75	A-319, Ан-12, Ту-134, Як-42, АTR-72, Е-190, Е-170
5	10÷30	Ан-24, Ан-26, Ан-72, Ан-74, Ан-140, Як-40, АTR-42
6	< 10	Ан-2, Ан-3, Ан-28, Ан-38, L-410, Е-120

* Не совпадает с классификацией ICAO.

Таблица 1

Параметры, характеризующие авиалинию (индекс 1 соответствует городу вылета, индекс 2 — городу прилёта)

№ п/п	Код авиалинии	1 Код города	2 Код города	F1 - Класс авиалинии	F2 - Длина авиалинии, км	F3 - Наличие ж/д сообщения	F4 - 1_ВРП субъекта РФ, млрд. руб.	F5 - 2_ВРП субъекта РФ, млрд. руб.	F6 - 1_Числ. насел., тыс. чел.	F7 - 2_Числ. насел., тыс. чел.	F8 - 1_Число мест, тыс. коек	F9 - 2_Число мест, тыс. коек	F10 - 1_Статус города	F11 - 2_Статус города	F12 - Направление авиалинии	E1 - Авиасвязь в 2006г.
1	1	АБН	АЛД	5	2097	1	53,5	6,4	536,6	46,7	3,5	0,2	3	4	16	0,0
2	2	АБН	АМД	5	2228	0	53,5	0,5	536,6	0,6	3,5	0,0	3	4	16	0,0
3	3	АБН	АНА	4	3914	1	53,5	58,1	536,6	664,0	3,5	109,6	3	2	7	0,0
4	4	АБН	АНЫ	3	4564	0	53,5	2,2	536,6	18,4	3,5	0,2	3	3	10	0,0
5	5	АБН	АПХ	4	3254	0	53,5	22,5	536,6	189,4	3,5	1,6	3	3	10	0,0
6	6	АБН	АРХ	3	2951	1	53,5	100,8	536,6	996,1	3,5	6,1	3	3	10	0,0
7	7	АБН	АСР	3	3151	1	53,5	101,0	536,6	994,1	3,5	5,0	3	3	10	0,0
8	8	АБН	БАН	3	533	1	53,5	199,2	536,6	2728,7	3,5	18,2	3	3	9	0,4
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
13738	15002	ЯКТ	ЧКД	5	1238	0	68,0	0,4	499,6	3,0	2,3	0,0	3	4	14	1,0
13739	15003	ЯКТ	ЧЛБ	3	3824	0	68,0	389,6	499,6	2645,4	2,3	27,4	3	3	10	0,0
13740	15004	ЯКТ	ЧРС	5	1612	0	68,0	1,8	499,6	13,0	2,3	0,1	3	4	14	0,8
13741	15005	ЯКТ	ЭЛИ	4	5370	0	68,0	15,9	499,6	287,2	2,3	0,3	3	3	10	0,0
13742	15006	ЯКТ	ЮЖХ	2	1877	0	68,0	93,6	499,6	521,2	2,3	3,4	3	3	9	0,0

В качестве главного города авиалинии выбирается тот, что имеет более высокие значения целевых параметров. Это осуществляется последовательным сравнением значений числа мест размещения двух городов (с прилегающими к ним территориями), а при одинаковом количестве мест размещения — сравнением значений статусов городов. Целевыми параметрами элементов исходной выборки принимаются целевые параметры главного города, а генерационных параметров — генерационные параметры другого города (с более низкими целевыми ресурсами). В случае невозможности определения главного города в число элементов исходной выборки включаются оба элемента. Сформированная таким образом исходная выборка содержит 6963 элемента (из 13742 до десимметризации). После десимметризации

каждый элемент исходной выборки содержит 8 входных параметров из 12 возможных (F1-F4, F6, F9, F11, F12).

ФОРМИРОВАНИЕ МНОЖЕСТВ ДАННЫХ. САМООРГАНИЗУЮЩИЕСЯ КАРТЫ КОХОНЕНА

Важнейшим этапом разработки нейронной сети является этап ее обучения. Качество процесса и, следовательно, качество нейросетевой модели определяются информативностью обучающей выборки. Она формируется в результате целенаправленного отбора данных из общей исходной их совокупности. Для этих целей была разработана методика формирования обучающей выборки с использованием самоорганизующихся карт Кохонена [8] (далее — сетей Кохонена).

Для корректного обучения нейросетевой модели и проверки её на данных, от-

личных от использованных при обучении, исходная выборка разделяется на три части: обучающую (training), тестовую (testing) и проверочную (checking). Обучающая используется при построении нейронной сети. Тестовая — для тестирования построенной сети с целью проверки качества её функционирования. Проверочная выборка контролирует переобучение.

Разделению исходной выборки на обучающую, тестовую и проверочную предназначен алгоритм обучения сетей Кохонена, реализованный в пакете STATISTICA. Формирование выборок для существующих и отсутствующих авиалиний выполняется раздельно, затем одноимённые выборки объединяются. Такой подход позволяет создать ограниченную по объёму обучающую выборку.

Сети Кохонена рассчитаны на неуправляемое обучение, при котором обучающие данные содержат только значения входных переменных: сеть Кохонена «учится понимать» структуру данных. Она может распознавать в них кластеры, устанавливая их близость. Фигурируют два слоя: входной и выходной, составленный из радиальных элементов (его также называют слоем топологической карты). Элементы топологической карты располагаются в некотором пространстве, как правило, двумерном.

Обучается сеть Кохонена методом последовательных приближений. Начиная со случайным образом выбранного исходного расположения центров, алгоритм постепенно улучшает его, чтобы обеспечить рациональную кластеризацию обучающих данных. В результате итеративной процедуры сеть организуется таким образом, что элементы, соответствующие центрам, расположенным близко друг к другу в пространстве входов, будут располагаться столь же близко и на топологической карте. Топологический слой можно представить себе как двумерную решетку, которую нужно так отобразить в N -мерном пространстве входов, чтобы по возможности сохранить исходную структуру данных. Хотя при любой попытке представить N -мерное пространство на плоскости теряются многие детали, такой приём позволяет визуализировать данные, которые никак иным способом «понять» невозможно.

При формировании элементов исходной выборки рассматривались векторы, включа-

ющие 8 входных переменных (параметры F1-F4, F6, F9, F11, F12), а также выходную переменную (параметр E1), которая может принимать значения 0; 0,4; 0,6; 0,8; 0,9 и 1,0. Поскольку примеров, имеющих отличное от нуля значение параметра E1 (условное обозначение группы «>0»), во много раз меньше примеров с нулевым значением параметра (условное обозначение группы «=0»), при формировании обучающей и проверочной выборок эти две группы примеров отделены друг от друга. Для группы «>0» берутся все входные параметры, а также параметр E1, а для группы «=0» — только входные.

Вначале определяется наилучшее процентное распределение для группы «>0», поскольку она малочисленная (501 пример) по сравнению с группой «=0» (6462 примера). Для построения адекватной модели «условий существования» авиалиний необходимо, чтобы в обучающей выборке количество «существующих» авиалиний было примерно равно количеству «отсутствующих». Исходя из этого, после определения количества примеров для группы «>0» определяется количество примеров для группы «=0». При этом учитывается, что первая группа уже содержит некоторое количество авиалиний, интерпретируемых как «отсутствующие» (значение параметра E1 равно 0,4). Поиск оптимальных репрезентативных обучающих выборок варьирует параметр «seed for sampling», отвечающий за инициализацию элементов выборок. Размерность топологической карты подбирается так, чтобы, с одной стороны, карта содержала как можно больше элементов, а с другой — чтобы количество пустых элементов было невелико. В итоге размерность топологической карты выбирается равной 15×15 .

Таким образом, «отсутствующих» авиалиний в обучающей выборке $214 + 68 = 282$, «существующих» — 282. Сравнение топологических карт попарно показывает, что обучающие выборки в обеих группах репрезентативны. В результате создаются репрезентативные выборки — обучающая, тестовая и проверочная. Обучающая имеет существенно меньший размер по сравнению с исходной выборкой (всего 564 элемента из 6963), при этом обучающая сохраняет уровень информативности исходной. Множества обучающей, тестовой и проверочной выборок не пересекаются.





РАЦИОНАЛЬНЫЙ СОСТАВ ПАРАМЕТРОВ

Определён рациональный (близкий к оптимальному) состав входных переменных модели УСА. Для понижения размерности задачи был использован генетический алгоритм отбора входных переменных (Genetic Algorithm Input Selection), реализованный в пакете STATISTICA Neural Networks [3]. Его можно считать «интеллектуальной» формой метода проб и ошибок. Этот алгоритм анализирует большое число комбинаций входных переменных с помощью вероятностных и обобщенно-регрессионных нейронных сетей.

Для обучения нейронной сети доступен большой объём исходной информации, однако неизвестно, какие из входных переменных наиболее «полезны» или, более того, обязательны. Задача осложняется наличием взаимозависимостей или корреляций между некоторыми входными переменными. Как правило, увеличение числа входов нейронных сетей требует существенного увеличения числа обучающих примеров. Качество сети может быть улучшено сокращением числа входов, иногда даже ценой некоторой потери исходной информации.

Имеются две основные возможности разрешить эту проблему. Первая (выделение признаков) состоит в том, чтобы, сохранив все исходные переменные, уменьшить их количество так, чтобы оставить максимум информации. Методом, обычным для этого, принят анализ главных компонент. Вторая возможность (понижение размерности) связана с выявлением переменных, которые значимо не влияют на результат.

Генетический алгоритм является оптимизационным, эффективным при анализе двоичных строк. В данном случае двоичные строки представляют собой маски, которые определяют, какие входные переменные могут быть пригодны при создании нейронной сети: 0 означает, что переменная не должна быть использована, 1 — что переменную стоит включить в рассмотрение. Например, для шести входных переменных маска 001101 означает, что первая, вторая и пятая переменные должны быть отброшены, а третья, четвертая и шестая — оставлены.

Случайным образом генетический алгоритм создает совокупность таких строк

и затем использует процесс, аналогичный процессу естественного отбора, чтобы выбрать лучшие строки. Затраты времени велики, но практически не зависят от количества переменных, в то время как последовательные алгоритмы затрачивают время, пропорциональное квадрату числа переменных. То есть генетический алгоритм целесообразно использовать в случаях, когда количество переменных велико (например, более пятидесяти). Он также пригоден при принятии окончательного решения о включении параметра в список при меньшем числе переменных.

В процессе использования генетического алгоритма было взято $256 \left(\sum_{k=0}^8 C_8^k = 256 \right)$ вариантов состава вектора входных переменных (произведен исчерпывающий перебор). Из них к рассмотрению приняты варианты, не содержащие одновременно параметры F11 и F12, поскольку F12 определяется через F11.

В результате определен вектор значимых переменных, содержащий 6 входных параметров — F1, F3, F4, F6, F9 и F11, для которого ошибка обучения является минимальной. Для других вариантов состава вектора входных переменных ошибка не более чем на 7% превышает минимальное значение. Однако ошибки носят оценочный характер и служат прежде всего выбору оптимального состава вектора.

Множество исходных данных содержит 8 возможных входных переменных, в обучающей выборке $282+282=564$ элемента.

В [9] указывается, что для хорошего обобщения достаточно, чтобы размер обучающей выборки N удовлетворял следующему соотношению, полученному эмпирически:

$$N = O(W/\epsilon),$$

где W — общее количество свободных параметров (синаптических весов) сети; ϵ — допустимая точность ошибки аппроксимации; $O(\cdot)$ — порядок заключенной в скобки величины. В частности, для ошибки в 10% количество примеров обучения должно в 10 раз превосходить количество свободных параметров.

В [1] подчеркивается, что качество обучения гибридной нейронной сети зависит от объёма обучающей выборки и что возможны проблемы с работоспособностью



Рис. 3. Общая схема адаптивной системы нейро-нечёткого вывода (ANFIS).

системы MATLAB, если количество входных переменных превышает 5–6. Следовательно, полученное с помощью генетического алгоритма оптимальное количество переменных вектора входных переменных, равное 6, удовлетворяет необходимым требованиям как с точки зрения количества входных переменных, так и соотношения количества входных переменных и количества примеров обучения.

НЕЧЁТКАЯ НЕЙРОСЕТЕВАЯ МОДЕЛЬ «УСЛОВИЙ СУЩЕСТВОВАНИЯ»

Формирование модели УСА осуществляется с использованием программных средств пакета Fuzzy Logic Toolbox и нейросетевого пакета Neural Network Toolbox среды программирования MATLAB [1, 2]. При этом используется адаптивная система нейро-нечёткого вывода (ANFIS) (рис. 3), типа Сугено. Ее сопровождает процедура субтрактивной кластеризации (Sub. clustering).

Данная процедура позволяет создать достаточно компактную структуру нейронной сети, требующую приемлемых вычислительных ресурсов. Параметры алгоритма кластеризации ($[C, S] = \text{subclust}(X, \text{radii}, x\text{Bounds}, \text{options})$) подбираются с учетом получения оптимальной структуры нейронной сети. Для рассматриваемой модели в результате субтрактивной кластеризации было определено, что оптимальная структура включает 6 нейронов в скрытом слое, что соответствует 6 нечётким кластерам. Структура нейронной сети показана на рис. 4.

На основе элементов трёх выборок, полученных с помощью сети Кохонена, сформировано три множества: обучающее (training), тестовое (testing), проверочное (checking). Для обучения нейронной сети использован стандартный метод обратного распространения ошибки. Обученная сеть является основным элементом адаптивной системы нейро-нечеткого вывода (ANFIS).

Получено, что значения выходного параметра, моделируемые нейронной сетью, находятся в диапазоне значений выходного параметра обучающего множества.

АНАЛИЗ АДЕКВАТНОСТИ МОДЕЛИ

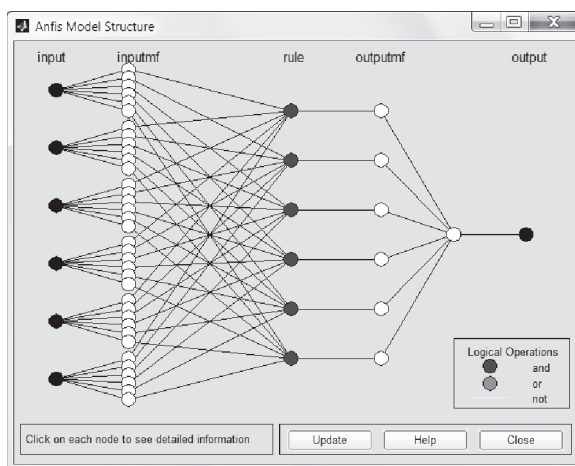
С использованием сформированной УСА проведено моделирование сети внутрироссийских магистральных авиалиний для 2006 года. Если авиалиния моделируется как «существующая», то ей присваивается значение «авиалиния существует», в противном случае — «авиалиния отсутствует». Результаты моделирования представляются в виде модельной **OD**-матрицы. Она поэлементно сравнивается с исходной **OD**-матрицей, содержащей авиалинии, реально существовавшие в том году.

При поэлементном сравнении модельной матрицы с исходной выделяются четыре класса авиалиний. Для обозначения классов в рассмотрение вводится двузначный индекс, в котором первая цифра относится к реальной авиалинии, а вторая — к моделируемой: 11 — авиалиния существует и моделируется; 00 — авиалиния не существует и не моделируется; 01 — авиалиния не существует, но моделируется; 10 — авиалиния существует, но не моделируется. Первые два класса демонстрируют совпадение результатов моделирования с реальной сетью, третий и четвертый — отличие моделируемой сети от реальной.

Сформированная модель адекватно моделирует 73% (291 объект) существовавших в 2006 году авиалиний, при этом годовой пассажиропоток, приходящийся на них, составляет 94%. Модель правильно моделирует 91% (5950 объектов) отсутствовавших в 2006 году авиалиний. Модельная **OD**-матрица имеет четко выраженную линейчатую структуру столбцов для авиалиний классов 11, 10 и 01. Однако достаточно большое их количество класса 01 (617) не позволяет использовать данную модель при прогнозировании развития сети внутрироссий-



Рис. 4. Структура нейронной сети для модели «условий существования» авиалиний.



ских магистральных авиалиний. Поэтому по ходу интерпретации результатов моделирования была введена так называемая «зона нечувствительности» модели.

На рис. 5 для обучающей выборки представлено распределение количества «существовавших» и «отсутствовавших» в 2006 году авиалиний по моделируемым значениям выходного параметра. Из рассмотрения этих данных можно сделать вывод, что некоторая часть авиалиний моделируется неоднозначно: для них моделируемые значения выходного параметра близки к значению 0,5, что не даёт достаточных оснований отнести авиалинии к «существующим» или «отсутствующим».

На основании анализа результатов моделирования обучающей выборки выделена «зона нечувствительности», соответствующая моделируемым значениям выходного параметра в интервале (0,3; 0,7]. На рис. 5 эта зона выделена серым цветом. Сравнение результатов моделирования без учёта и с учётом «зоны нечувствительности» показывает, что число неоднозначно моделируемых авиалиний (находящихся в пределах «зоны нечувствительности») составило от их общего числа 32% (2239 объектов, из них 131 существующие и 2108 отсутствующие). Среди остальных авиалиний (которые условно отнесены к «однозначно моделируемым») правильно моделируются 83% существующих (221 из 265) и 96% отсутствовавших (4291 из 4459). На «правильно моделируемые» авиалинии приходится 90% годового пассажиропотока. Количество их класса 01 с учётом «зоны нечувствительности»

(168 объектов) позволяет использовать модель при прогнозировании развития сети внутрироссийских магистральных сообщений. Разработанная схема адекватно отражает структуру сложившегося в стране сетевого коммуникационного ядра.

Созданная нами модельная **OD**-матрица получена с учётом наличия «зоны нечувствительности» при интерпретации результатов моделирования. Среди неправильно моделируемых авиалиний можно выделить две группы. Первая, относящаяся к классу 10, в основном включает линии, расположенные внутри субъектов федерации, то есть их можно считать местными. Группа вторая, относящаяся к классу 01, включает «перспективные» авиалинии.

ВЫВОДЫ

1. Переход от анализа осреднённых показателей состояния авиалинии к анализу степени стабильности её состояния в течение трёх лет позволяет сформировать более адекватную функцию принадлежности для нечёткого понятия «авиалиния существует». Каждой авиалинии (в зависимости от стабильности её состояния в течение трёх лет и темпа изменения интенсивности полётов ВС) поставлено в соответствие определённое значение выходного параметра нейронной сети, характеризующего существование авиалинии.

2. Качество моделирования сети авиалиний в значительной степени зависит от выбора параметров, определяющих «условия существования» авиалиний. Рассмотрена совокупность из 12 параметров авиалиний, которые можно было бы ис-

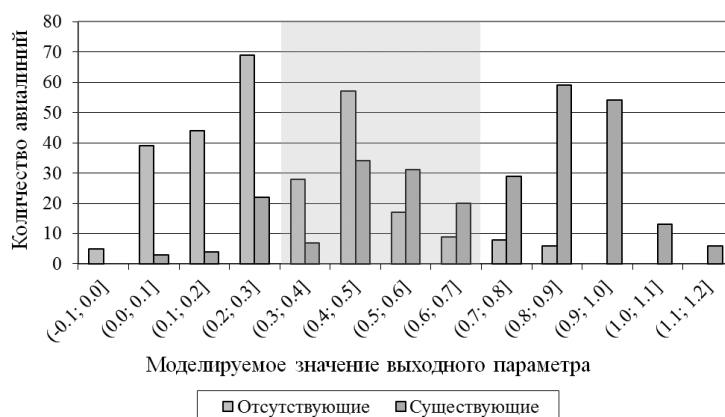


Рис. 5. Распределение моделируемых авиалиний для обучающей выборки.

пользовать в качестве входных переменных модели. Определен рациональный состав входных переменных из 6 параметров.

3. Обобщающие свойства нейросетевой модели во многом зависят от состава обучающей выборки. Применение метода самоорганизующихся карт Кохонена позволило сократить объём обучающей выборки более чем на порядок, сохранив информативность исходных данных на уровне, близком к уровню информативности всей совокупности имеющихся в распоряжении статистических данных.

4. Рассматриваемая модель «условий существования» авиалиний обладает хорошими топологическими свойствами (формирует ярко выраженную структуру столбцов *OD*-матрицы), что подтверждает возможность качественного моделирования. Модель адекватно отражает структуру коммуникационного ядра внутрироссий-

ских магистральных авиалиний, что позволяет использовать ее для прогнозирования развития сети.

ЛИТЕРАТУРА

1. Леоненков А. В. Нечёткое моделирование в среде MATLAB и fuzzy TECH. — СПб., 2005.
2. Штовба С. Д. Проектирование нечётких систем средствами MATLAB. — М., 2007.
3. Нейронные сети. STATISTICA Neural Networks: Методология и технологии современного анализа данных. — М.: Горячая линия—Телеком, 2008.
4. Социальное положение и уровень жизни населения России/Федеральная служба государственной статистики (Росстат). — М., 2007.
5. Туризм и туристские ресурсы в России: Стат. сб. Росстата. — М., 2004.
6. Численность населения Российской Федерации по городам, поселкам городского типа и районам на 1 января 2008 года/Росстат. — М., 2008.
7. Борисов В. В., Круглов В. В., Федулов А. С. Нечёткие модели и сети. — М., 2007.
8. Кохонен Г. Самоорганизующиеся карты. — М., 2008.
9. Хайкин С. Нейронные сети. Полный курс. — М. — СПб. — Киев, 2006.

DESIGNING OF THE NETWORK OF LONG-HAUL AIRLINES

Balashov, Victor V. — Ph.D. (Tech), associate professor, leading researcher of Central Aero-hydrodynamics Institute named after N. Zhukovsky (TSAGI).

Smirnov, Andrey V. — Ph.D. (Tech), associate professor, head of the section of TSAGI.

Zeitlina, Tatiana O. — junior researcher at TSAGI.

The authors describe the solution of a problem of engineering of a model of "conditions of existence of airlines" (CEA). The model means a rule that determines creation, operation and disappearance of airlines (direct air links between two cities). They create fuzzy neuronet model of CEA using genetic algorithm, Kohonen self-organizing map, Fuzzy Logic Toolbox software and neuronal network package for Matlab programming environment. The model adequately reflects the structure of communication core of the airlines' network and allows to forecast its future developments.

Key words: airlines network, conditions of existence of airlines, fuzzy simulation, neuronal model.

Координаты авторов (contact information): Балашов В. В. — balashov@tsagi.ru, Смирнов А. В. — smirnov@tsagi.ru, Цейтлина Т. О. — ttseytlna@gmail.com.

