

ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ С ПОМОЩЬЮ НАБОРА НЕЧЕТКИХ ПРАВИЛ, НАСТРАИВАЕМЫХ С ПОМОЩЬЮ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ

Введение. Большинство предприятий функционируют в условиях неопределенности и подвергаются действию изменяющихся факторов. Принятие эффективных решений и прогнозирование для этого будущего состояния предприятия есть достаточно актуальной задачей. Прогнозы необходимы практически каждому предприятию, наблюдается возрастающий рост к прогнозированию важных экономических показателей для отдельных предприятий, компаний, государства. Большинство экономических показателей предприятий представляют собой временные ряды.

В случае, если прогноз необходимо представить в понятном, интерпретируемом виде, одним из наиболее понятных и популярных способов представления является представление прогноза с помощью набора правил (продукций). Нечеткие методы [1, 2] успешно используют продукционный подход, интерпретируемость результатов, полученных с помощью нечетких правил, очень высока. Однако точность автоматически получаемых правил низка, необходимо экспертное вмешательство. Исправить этот недостаток возможно применением эволюционных методов [3, 4] для построения и настройки набора нечетких правил. Вопросами построения набора нечетких правил занимается направление генетические нечеткие системы (genetic fuzzy system) [5, 6, 7]. Такая система была успешно построена и хорошо себя показала на тестовых задачах классификации в предыдущих работах авторов [8, 9]. Но применение данной системы на задачах прогнозирования временных рядов еще не производилось. Исследования в данном направлении являются актуальной научной задачей и могут дать новые важные результаты как в теоретическом так и в практическом плане.

Целью данной статьи является дальнейшее развитие системы для создания базы знаний на основе нечетких продукций путем применения данной системы к задачам прогнозирования временных рядов.

В статье рассмотрены проблемы нечетких продукций и их настройки с помощью генетических алгоритмов.

Постановка задачи

Частный оператор электрической связи (далее - Оператор) оказывает услуги по предоставлению телефонной связи населению. Существует ряд отделов Оператора в различных районах города. В некоторых отделах (узлах) оператора возможно подключение новых абонентов, в некоторых – существующие ёмкости исчерпаны. Возможно расширения ёмкостей Оператора, но проведение коммуникаций и установка нового оборудования дороги. Возможно создание новых АТС оператора (выносов) в других районах. Ведется база данных, регистрируются подключения абонентов к АТС Оператора во всех районах, существуют сведения по поводу введения в эксплуатацию новых отделов (выносов) АТС. Но количество вводов в эксплуатацию новых отделов АТС невелико. В базе данных также есть сведения о длительности всех звонков, на основании которых возможно посчитать среднюю длительность звонков, сделанных новыми абонентами, и их структуру. По структуре разговоров можно определить ориентировочный доход Оператора от подключения новых абонентов, в зависимости от стоимости времени разговоров.

Необходимо на основании имеющейся информации о ходе подключений при расширениях АТС создать прогнозную модель хода (количества) подключений при дальнейшем расширении АТС предприятия или вводе новых АТС в зависимости от различных факторов (наличия и количества конкурентов в районе, благосостояния жителей, ассортимента услуг предприятия). Также представляет интерес дальнейший ход подключений в уже существующих АТС, где количество номеров еще не достигло максимума. Прогноз важен для расчета срока окупаемости аппаратуры и прочих затрат на введение в эксплуатацию новых выносов и расширения мощностей на существующих отделах и выносах Оператора.

Для решения данной задачи необходимо получить прогноз количества подключений в различных условиях (в том числе и в новых условиях, для которых еще нет данных). Поскольку количество подключений является временным рядом, данная задача сводится к задаче прогнозирования временных рядов. При этом при прогнозировании хода подключений на уже существующих узлах Оператора можно предположить, что качественно временной ряд меняться не будет. В случае же ввода новых узлов возможны абсолютно новые условия, и прогнозны модели могут обладать значительной погрешностью, поскольку заранее хотя бы грубо определить структуру временного ряда может быть достаточно затруднительно. Так как введение в эксплуатацию новых узлов дорого, созданная прогнозная модель должна проверяться, и неоднократно, что выдвигает требование к интерпретируемости прогнозны модели. При этом необходимо построение модели, учитывающей все вышеупомянутые факторы в условиях, когда некоторые из них изменяются.

Поскольку прогнозная модель будет использоваться людьми, скорее всего не обладающими большими знаниями в области математики (директором, ориентировочно имеющим экономическое образование), выдвигается требование к простоте и интерпретируемости прогнозны модели. Также модель должна быть точной. Следует отметить, что количество существующих вариаций влияющих факторов невелико (по состоянию на январь 2008г. - 12 узлов и выносов, условия работы многих похожи). Важной особенностью данной задачи является быстрое убывание автокорреляционной функции (значащая глубина-3 лага) и большой вес влияющих факторов на прогноз.

Описание системы для получения базы знаний. В данной работе используется генетическая нечеткая система для создания базы нечетких правил для прогнозирования временных рядов. База состоит из набора правил треугольного вида на входные и выходные переменные. Хромосома состоит из двух частей. Первая часть включает в себя номера термов лингвистических переменных, на основе которых строится набор правил. Пример такого набора термов (с именами: “очень маленький” (ММ), “маленький” (М), “средний” (С), “большой” (Б), “очень большой” (ББ)) для лингвистической переменной «Размер» приведен на рис. 1. Вторая часть хромосомы уточняет вид правила и имеет вид $C=C_1C_2$, где $C_1=C_{11}, C_{21}, \dots, C_{n1}$ – выбор конкретной лингвистической переменной для каждой входной, $C_2=C_{a12} C_{b12} C_{c12} C_{a22} C_{b22} C_{c22} \dots C_{an2} C_{bn2} C_{cn2} C_{ay2} C_{by2} C_{cy2}$ – точная настройка нечетких функций правила (координаты абсцисс левого края треугольника, его центра и его правого края). С помощью части хромосомы C_1 выполняется грубая настройка правила, с помощью части C_2 – точная. Если задействована только часть хромосомы C_1 , получают лингвистические правила, которые можно объяснить словесно, например, вида «ЕСЛИ Спрос Низкий ИЛИ Зарплата Маленькая ТО Продажи Низкие». Однако точность, полученная только лингвистическими средствами, зачастую оказывается недостаточной. В случае, если необходимо задействовать и часть C_2 , которая более точно регулирует форму функций принадлежности под условий, точность результата увеличивается. Однако это происходит за счет снижения интерпретируемости результата. В таком случае создаются правила свободной семантики. Пример таких правил показан на рис. 2.

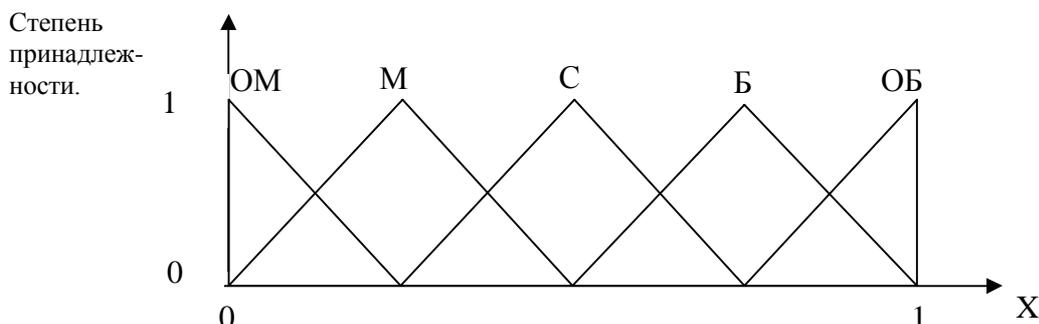


Рис.1. Построение функций принадлежности, характеризующих термы лингвистической переменной «размер».



Рис.2. Пример правил свободной семантики.

Для получения прогноза из нечетких правил используется процедура нечеткого вывода по Мамдани [2].

Система состоит из трех последовательных этапов, которые постепенно улучшают результат прогнозирования:

Первичная генерация правил (генерация).

На этом этапе создается набор нечетких правил, на основании которого и получается прогноз. Правила состоят из двух частей: части для лингвистического (дескриптивного) описания (в правиле представлен индекс элемента из терм-множеств лингвистических переменных), и части для аппроксимативного описания (содержащей в себе формы функций принадлежности подусловий правила), которая уточняет базовую форму переменных из первой части правила. Используются лингвистические переменные треугольного вида, кодируемые тремя числами - координатами левого края, центра и правого края. Создание и подстройка правил осуществляется генетическим алгоритмом, который использует как целочисленные генетические операторы для грубой настройки правил, используя первую часть правил, так и вещественные операторы для более точной их настройки с использованием второй части правил [6, 7].

Постпроцессинг включает в себя несколько процедур, которые работают с уже созданным набором правил, последовательно улучшая его, настраивая для улучшения точности. Включает в себя три процесса:

Процесс упрощения базы знаний (мультипликации) производится путем удаления ненужных или малоинформативных правил из их множества, созданного на первом этапе. Это делается с помощью второго генетического алгоритма, который работает с двоичными хромосомами, каждый бит которых включает/не включает соответствующее по порядку правило, полученное в результате работы первого этапа – генерации, в итоговое множество.

Процесс взвешивания улучшает точность прогноза путем присваивания каждому правилу своего «веса», с которым оно делает вклад в общий прогноз набора правил. Присвоение веса производится генетическим алгоритмом с вещественным кодированием хромосомы. Гены хромосомы кодируют «веса» соответствующих правил (от 0 до 2)

В процессе окончательной настройки правил базы знаний (тюнинга) форма правила корректируется для получения оптимального по точности результата. Это производится ещё одним генетическим алгоритмом, который работает с мультихромосомами, первая из которых является «склеенными» вместе правилами, полученными в результате работы второй части, остальные созданы по подобию первой, но с небольшими отклонениями в пределах

разрешенных интервалов. Для работы здесь используются те же операторы, что и при работе с второй частью хромосом (вещественной, C_2) на этапе генерации популяции.

Первые два процесса сохраняют лингвистическую понятность результата, третий – переводит правила к правилам свободной семантики.

Результатом работы системы является набор правил, по которым можно получить прогноз следующего элемента временного ряда на основании подаваемых на вход влияющих факторов. Более подробно эта система описана в [8], аналогичные системы описаны также в [6, 7]. Общий вид системы представлен на рис. 3.

Предварительная подготовка данных, отбор значащих факторов.

Из базы данных были получены временные ряды с временным шагом данных (размером лага) в 1 неделю. Поэтому логично осуществление двух видов прогнозирования: краткосрочного и долгосрочного (на максимально возможный горизонт прогноза). В результате анализа автокорреляционной функции рядов «количество подключений» была определена значащая глубина автокорреляционной функции – 3 лага (предыдущих элементов ряда). Поэтому максимально возможной глубиной прогноза является 4 недели (1 месяц). Таким образом, производилось создание двух баз правил для нахождения двух видов прогноза: краткосрочного (на 1 неделю) и долгосрочного (на 1 месяц).

Отбор значащих факторов производился с помощью генетического алгоритма (компактного генетического алгоритма). Особыми таких алгоритмов является двоичный вектор, который кодирует набор факторов (каждый аллель отвечает за вхождение/невхождение определенного атрибута в этот набор), фитнес - функцией является оценка качества такого набора. Оценкой качества набора являлась погрешность прогнозирования. Эту погрешность при отборе значащих факторов находили с помощью нейронных сетей, использовался пакет matlab [10]. Шаг изменения вероятностного вектора компактного генетического алгоритма был модифицирован в [11] для ускорения нахождения результата.

Первоначально экспертами было отобрано 20 факторов, важных для решения данной задачи. Для долгосрочного прогнозирования компактными генетическими алгоритмами были отобраны факторы Ф3 («спрос»), Ф14 («работы сумм 3 лагов»), и Ф13 («работы по расширению»). Это вполне логично, поскольку именно эти факторы из всего набора могут что-то сказать о том, каким будет состояние по подключениям за следующий месяц. Для краткосрочного прогнозирования были признаны важными факторы Ф13 («работы по расширению»), Ф8 («подключений за месяц»), Ф9 («отключений за месяц»), Ф1 («подключений за неделю»). Это вполне объяснимо: на краткосрочное изменение числа подключений особенности рынка влияние имеют слабое, но динамика подключений сильно коррелирует с предыдущим ходом подключений. На долгосрочный же прогноз влияние имеют больше особенности рынка, чем ход подключений ранее. Отсутствие фактора Ф3 («спрос») для краткосрочного прогнозирования можно объяснить тем, что для недельного прогноза величина спроса уже не играет значительной роли, т.к. не меняется быстро и не может объяснить краткосрочные колебания количества подключений.

При опытах по предварительному преобразованию данных было выяснено, что особенностями данной задачи является наличие большого количества одинаковых данных (отсутствие подключений), логарифмическое распределение величины («количество подключений») и несоответствие тестовой выборки обучающей. Несоответствие тестовой выборки обучающей является следствием изменения рынка, но, поскольку интересует прогноз в современных условиях, в качестве ТВ брались последние данные. Логарифмическое распределение данных ряда очень ухудшает качество ряда, приведенного к равномерному виду, поскольку все рабочие значения находятся в одной трети пространства поиска. В результате опытов отобранные значащие факторы были приведены к виду, представленному в таблице 1.

Таблица 1. Нормировка факторов выборки данных

| Наименование фактора | Вид нормировки | Диапазон |
|----------------------------------|--|----------|
| $\Phi 1$ – Подключений за неделю | $\Phi 1' = \ln(\Phi 1)$ | [0,5] |
| $\Phi 3$ – Спрос | Стандартная | [0,3] |
| $\Phi 8$ – Подключений за месяц | $\Phi 8' = \ln(\Phi 8)$ | [0,5] |
| $\Phi 9$ – Отключений за месяц | $\Phi 9' = \ln(\Phi 9)$ | [0,5] |
| $\Phi 13$ - Работы по расширению | Не произв | [0,1] |
| $\Phi 14$ - Работы сумм 3 лагов | Стандартная | [0,3] |
| $\Phi 19$ - Прогноз 1 нед | $\Phi 19' = \text{знак}(\Phi 19) * \ln(\text{abs}(\Phi 19))$ | [-4,5] |
| $\Phi 20$ - Прогноз 4 нед | $\Phi 20' = \text{знак}(\Phi 20) * \ln(\text{abs}(\Phi 20))$ | [-4,5] |

Фактор «Спрос» определялся экспертно, 0 – нет спроса, 3 – спрос огромный. Фактор «Подключений за неделю» брался из базы данных, «подключений за месяц», «отключений за месяц» получались суммированием 4-х лагов числа подключений и отключений за неделю. Фактор «Работы по расширению» содержит информацию о ведущихся работах по расширению оборудования в данном районе («Да» – «Нет»). Фактор «Работы сумм 3 лагов» содержит сумму 3 предыдущих значения фактора «Работы по расширению», содержа значения от 0 до 3.

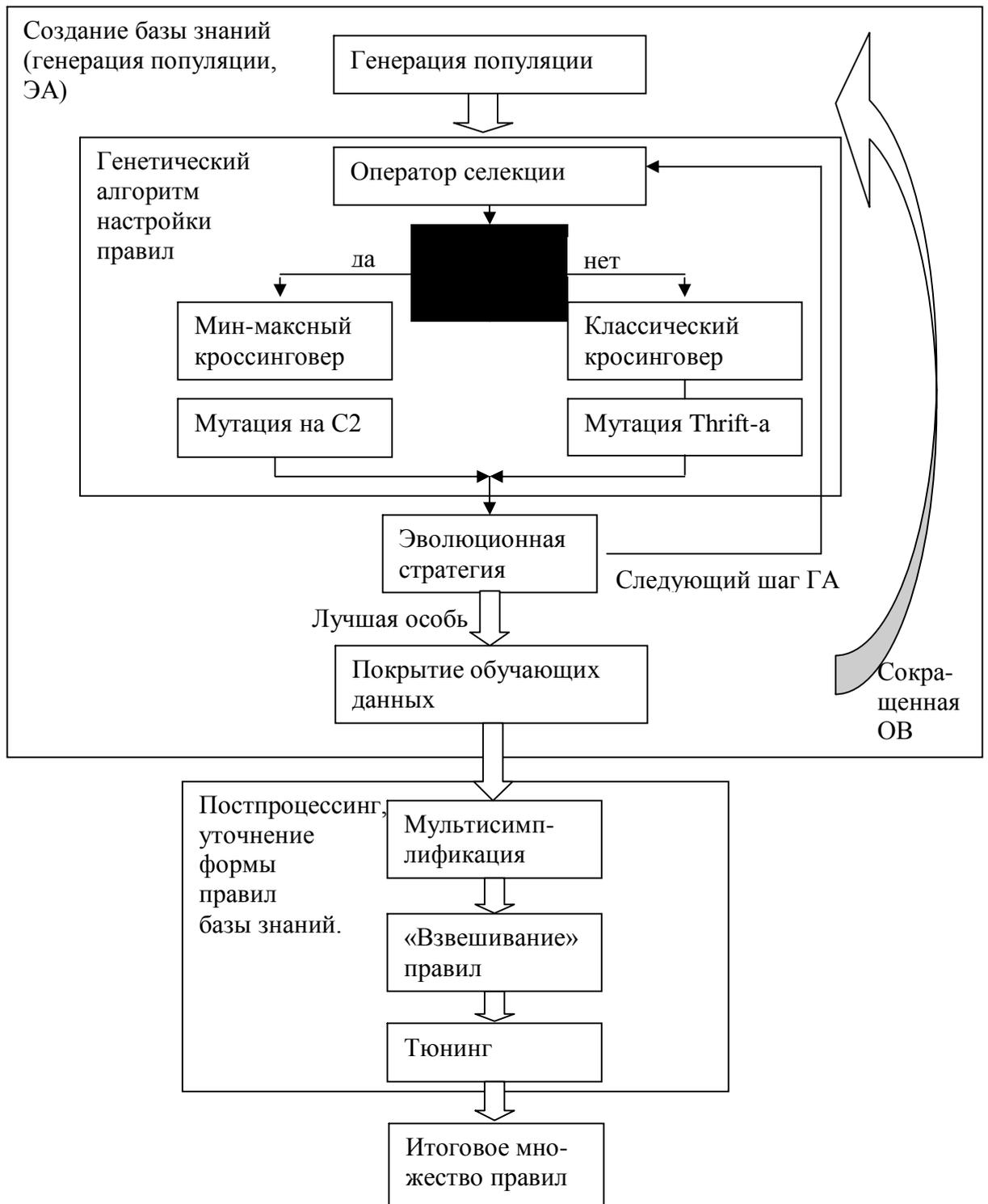


Рис. 3. Схема эволюционного алгоритма для прогнозирования временных рядов на основе продукционного подхода

Построение базы знаний для прогнозирования числа подключений абонентов.

Для каждого из видов прогнозирования с использованием описанной системы была получена своя база правил. Первоначально база правил строилась с сохранением лингвистически понятного результата (для генерации правил использовалась только часть С1 правила, для постпроцессинга – только процедуры мультипликации и взвешивания). Затем при необходимости для улучшения точности результата использовались и остальные средства системы.

Результаты опытов представлены в таблице 2. В таблице показаны среднеквадратичная (MSE) и средняя арифметическая (MAE) ошибки для ОБ и ТВ, СКО ошибок (в скобочках), а также число правил (Правил). И для долгосрочного и для краткосрочного прогноза производилось сравнение с прогнозом, полученным нейронными сетями (многослойный персептрон, 1-й слой – распределительный, скрытый слой имеет активационную функцию «гиперболический тангенс», число нейронов – 20, выходной слой имеет 1 нейрон, линейную активационную функцию). Для нейронной сети находилось среднее по 60 опытам, для генетической нечеткой системы – не менее чем по 5.

Таблица 2. Результаты опытов по прогнозированию числа подключений

| | | | ОБ | | ТВ | | Правил |
|-----------------------|---------------------------|----------------------------|----------------------|----------------------|----------------------|----------------------|---------------|
| | | | MSE | MAE | MSE | MAE | |
| Долгосрочный прогноз | НС | Несокращенные факторы | 0.771 (0.106) | 6.970 (0.737) | 3.125 (0.574) | 10.479 (1.222) | |
| | | Сокращённые факторы | 1.359 (0.029) | 7.848 (0.238) | 2.250 (0.041) | 7.878 (0.287) | |
| | Лингвистическая семантика | Генерация | 1.226 (0.549) | 9.181 (0.723) | 1.792 (0.631) | 8.632 (0.275) | 103.75 |
| | | Мультисимплификация | 1.353 (0.000) | 8.001 (0.010) | 2.204 (0.009) | 7.709 (0.051) | 50.000 |
| | | М-симпл лучшее | 1.360 | 8.022 | 2.194 | 7.676 | 47.000 |
| | Свободная семантика | Генерация | 2.027 | 11.557 | 2.665 | 11.499 | 39.000 |
| | | Мультисимплификация | 2.468 (0.052) | 11.572 (0.251) | 2.941 (0.048) | 11.659 (0.229) | 10.333 |
| Тюнинг | | 1.360 | 8.030 | 2.194 | 7.681 | 10.000 | |
| Краткосрочный прогноз | НС | Несокращенные факторы | 0.544 (0.184) | 2.232 (0.305) | 1.879 (0.458) | 4.489 (0.839) | |
| | | Сокращённые факторы | 0.602 (0.032) | 2.157 (0.102) | 1.617 (0.241) | 3.811 (0.370) | |
| | Лингвистическая семантика | Генерация | 1.438 (0.121) | 3.776 (0.120) | 1.791 (0.204) | 4.200 (0.193) | 33.000 |
| | | Мультисимплификация | 1.521 (0.010) | 3.580 (0.008) | 2.418 (0.081) | 4.508 (0.242) | 15.333 |
| | | М-симпл. лучшее | 1.519 | 3.575 | 2.378 | 4.386 | 15.000 |
| | Свободная семантика | Тюнинг | 0.493 | 1.968 | 1.496 | 3.501 | 15.000 |

Долгосрочное прогнозирование. . Результаты опытов по долгосрочному прогнозированию показаны в группе «Долгосрочный прогноз». Последовательно запускались процедуры генерации базы правил с использованием лингвистических средств ($MSE_{ОБ}=1.226$, строка «Генерация») и мультисимплификации ($MSE_{ОБ}=1.353$, 50 правил, строка «Мультисимплификация»). Полученные лингвистические правила и их веса для лучшего результата долгосрочного прогноза ($MSE_{ОБ}=1.360$, 47 правил, строка «Мультисимплификация лучшее») представлены в табл. 2. Для 1, 2 и 4-го столбцов таблицы («Спрос», «Работы сум 3 лагов», «Прогноз 1 мес.») использовались имена термов: очень низкий (ОН), низкий (Н), ниже среднего (НС), средний (С), выше среднего (ВС), высокий (В), очень высокий (ОВ). Для столбца «Работы по расшир» имена термов: «Да» и «Нет».

Правила описывают прогноз в случае, если подключений нет вне зависимости от свободной ёмкости (только в зависимости от спроса, правила 12 и 13). Часто значительное увеличение подключений наступает после незначительного – полный пуск станции наступает не сразу (правило 18). Значительный рост подключений часто продолжается несколько недель (правила 20, 22, 24, 25 таблицы 5.6). Дальше рост останавливается (правило 14), начинается слабое (правило 10) или сильное (правила 1, 2, 6, 7) падение. Значительные свободные ёмкости не стимулируют количество подключений, их отсутствие – не допускает их значительного роста.

Для одних и тех же комбинации подусловий есть правила с различными подзаклечениями. Их можно заменить одним с подзаклечением произвольной формы или описать лингвистически. Например, для 43-44 правил:

ЕСЛИ Спрос Очень Высокий И Работы по расширению ведутся 2 недели, ТО Подключения через месяц Высокие (вес 1.08) ИЛИ Очень Высокие (вес 0.98).

Но, хотя точность полученной базы знаний и сравнима с точностью нейронной сети ($MSE_{ОБ}=1.359$, строка «Сокращенный набор факторов»), её размер (47 правил) оказался достаточно большим. Поэтому была получена еще одна версия базы знаний, с использованием свободной семантики, которая получалась с целью минимизации количества правил базы правил (группа «Свободная семантика»). Полученная база правил ($MSE_{ОБ}=1.360$) содержит всего 10 правил (рис. 4). Но их интерпретируемость уменьшилась.

Для долгосрочного прогнозирования основной критерий точности правил на основе лингвистической семантики и правил на основе свободной семантики имеют близкие значения как для ОБ, так и для ТВ ($MSE_{ТВ}=2.204$ и 2.194). Число правил базы знаний лингвистической семантики, равно 47, свободной семантики – 10. О преимуществе одной базы знаний над другой, и о целесообразности уменьшения интерпретируемости за счет уменьшения размера базы правил предлагается решать экспертно.

Таблица 2. Лингвистическая база знаний для прогнозирования числа подключений

| № | Спрос | Работы сум 3 лагов | Работы по расшир | Прогноз 1 мес. | Вес | Част. |
|----|-------|--------------------|------------------|----------------|------|-------|
| 1 | ОН | ОН | Нет | НС | 0.97 | 6.92 |
| 2 | ОН | ОН | Нет | В | 0.99 | 7.01 |
| 3 | ОН | ОН | Нет | ВС | 0.94 | 3.37 |
| 4 | ОН | ОН | Нет | Н | 1.15 | 3.01 |
| 5 | ОН | ОН | Да | ВС | 1.40 | 1.72 |
| 6 | ОН | Н | Да | ВС | 1.16 | 4.18 |
| 7 | ОН | НС | Нет | ВС | 1.34 | 2.22 |
| 8 | ОН | НС | Нет | НС | 0.67 | 2.67 |
| 9 | ОН | ВС | Нет | ВС | 1.83 | 2.15 |
| 10 | Н | НС | Нет | НС | 0.79 | 1.78 |
| 11 | Н | ВС | Нет | НС | 0.34 | 1.22 |
| 12 | Н | ОВ | Нет | В | 1.23 | 1.00 |
| 13 | НС | ОН | Нет | ВС | 1.59 | 1.51 |
| 14 | НС | ОН | Нет | НС | 1.00 | 0.94 |
| 15 | НС | ОН | Нет | В | 1.35 | 3.28 |
| 16 | НС | ОН | Нет | Н | 0.61 | 1.94 |
| 17 | НС | ОН | Да | В | 0.80 | 1.40 |
| 18 | НС | ОН | Да | ВС | 0.96 | 4.23 |
| 19 | НС | Н | Нет | ВС | 0.62 | 1.00 |
| 20 | НС | Н | Да | В | 1.79 | 7.05 |
| 21 | НС | НС | Нет | В | 0.92 | 1.00 |
| 22 | НС | ВС | Нет | ВС | 1.68 | 1.16 |
| 23 | НС | В | Да | В | 1.52 | 12.84 |

| № | Спрос | Работы сум 3 лагов | Работы по расшир | Прогноз 1 мес. | Вес | Част. |
|----|-------|--------------------|------------------|----------------|------|-------|
| 24 | ВС | ОН | Нет | ОВ | 1.10 | 1.38 |
| 25 | ВС | ОН | Нет | В | 0.15 | 0.33 |
| 26 | ВС | ОН | Нет | НС | 1.07 | 0.82 |
| 27 | ВС | ОН | Да | ВС | 0.86 | 13.84 |
| 28 | ВС | НС | Да | В | 1.08 | 20.15 |
| 29 | ВС | ВС | Да | В | 0.82 | 0.67 |
| 30 | ВС | ОВ | Да | В | 1.28 | 4.00 |
| 31 | В | НС | Нет | В | 1.75 | 0.67 |
| 32 | В | В | Нет | В | 1.54 | 0.89 |
| 33 | В | ОВ | Нет | В | 0.90 | 0.60 |
| 34 | В | ОВ | Да | ОВ | 1.25 | 0.78 |
| 35 | ОВ | ОН | Нет | ВС | 0.89 | 0.56 |
| 36 | ОВ | ОН | Нет | НС | 1.49 | 37.58 |
| 37 | ОВ | ОН | Нет | В | 0.37 | 0.67 |
| 38 | ОВ | ОН | Да | В | 1.12 | 1.09 |
| 39 | ОВ | Н | Нет | В | 0.72 | 0.60 |
| 40 | ОВ | НС | Да | В | 1.53 | 4.89 |
| 41 | ОВ | НС | Да | ОВ | 1.14 | 0.83 |
| 42 | ОВ | ВС | Нет | ВС | 0.81 | 0.83 |
| 43 | ОВ | ВС | Да | В | 1.08 | 0.67 |
| 44 | ОВ | ВС | Да | ОВ | 0.78 | 0.33 |
| 45 | ОВ | В | Нет | В | 1.68 | 1.29 |
| 46 | ОВ | ОВ | Нет | В | 1.06 | 1.87 |
| 47 | ОВ | ОВ | Да | ОВ | 1.44 | 7.08 |

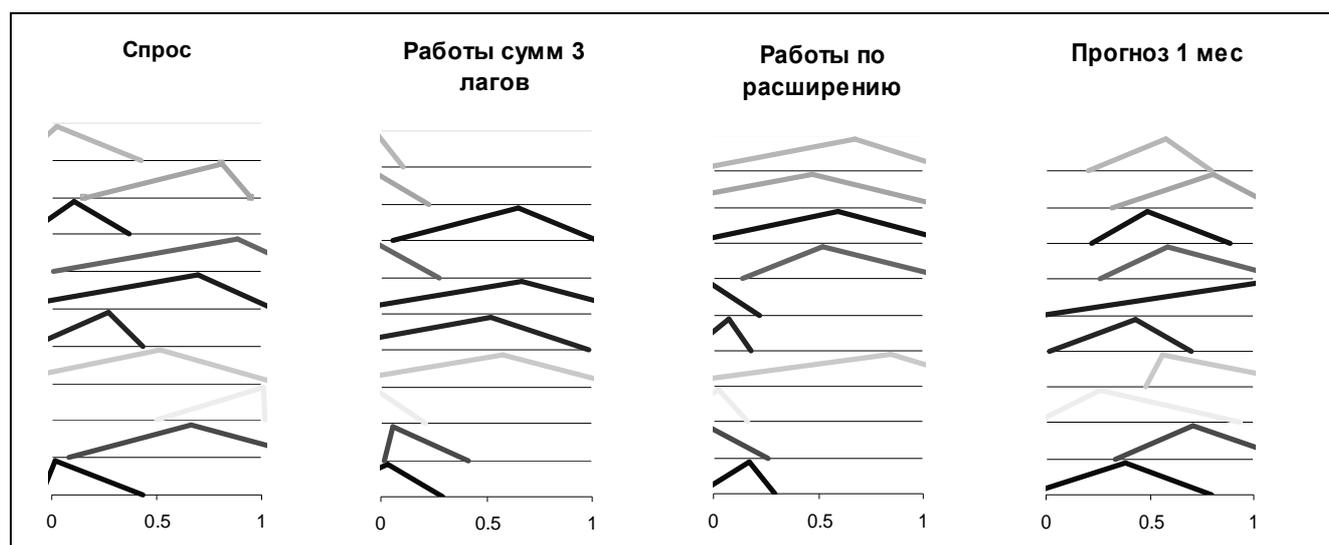


Рис. 4. База правил свободной семантики для долгосрочного прогнозирования количества подключений.

На рис. 4. и 5. по оси абсцисс отображается нормированная (диапазон [0,1]) единица измерения физической величины, по оси ординат – величина уверенности принадлежности к данной функции принадлежности.

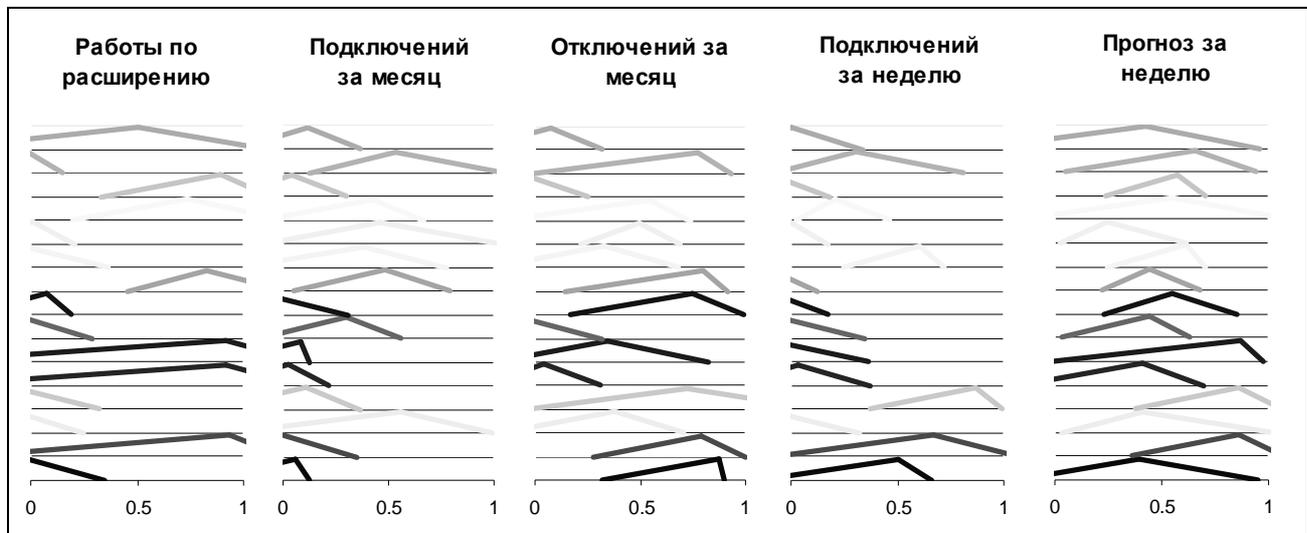


Рис. 5. База правил для краткосрочного прогнозирования количества подключений.

Краткосрочное прогнозирование. Результаты опытов по краткосрочному прогнозированию показаны в группе «Краткосрочный прогноз». На основании произведенных исследований была получена лингвистическая база знаний, содержащая 15 правил («Мультисимплификация лучшее», $MSE_{OB}=1.519$). Однако точность данной базы знаний значительно ниже точности, полученной НС (строка «Сокращённый набор факторов», $MSE_{OB}=0.602$). Поэтому была запущена процедура тьюнинга с превращением лингвистических правил в правила свободной семантики (строка «Тьюнинг» $MSE_{OB}=0.493$). Результаты оказались даже лучше, чем у НС. Полученные правила показаны на рис. 5. Для краткосрочного прогнозирования точность базы правил на основе лингвистической семантики (1.519) и правил на основе свободной семантики (0.493) существенно различаются. Число же правил для данного набора одинаково (15). Целесообразность улучшения точности за счет потери интерпретируемости предлагается также определить экспертно.

Анализируя результаты, можно сказать, что создана генетическая нечеткая система для получения набора правил, примененная для решения задачи прогнозирования временных рядов. С помощью данной системы можно получить лингвистически интерпретируемый прогноз, который может быть понятен даже неподготовленным лицам. В случае, если качество такого прогноза (количество правил или точность прогнозирования) недостаточно, существуют средства для улучшения качества за счет ухудшения интерпретируемости. Можно также сказать, что полученная точность и качество базы знаний уже достаточна для решения практических задач и получения информации по окупаемости оборудования, интересующей руководство Оператора электросвязи.

Выводы

В результате исследований можно сделать следующие выводы:

1. Произведен анализ задачи прогнозирования числа подключений к АТС для определения окупаемости установки телекоммуникационного оборудования. Выделены две подзадачи: нахождение краткосрочного (1 неделя) и долгосрочного прогноза. В результате анализа автокорреляционной функции была определена глубина значащей выборки, влияющая на прогноз, равная трем лагам. В результате максимальный горизонт долгосрочного прогноза был определен в 1 месяц (4 недели).
2. На основе экспертного заключения отобраны 20 факторов, признанных значащими для данной задачи. В результате отбора значащих факторов для задачи долгосрочного прогноза были отобраны 3 фактора из 20, для задачи краткосрочного прогнозирования – 4 фактора.
3. Созданная ранее система для автоматического получения набора нечетких правил, настраиваемых генетическими алгоритмами, была впервые применена для задачи прогнозирования временных рядов (задачи прогнозирования числа подключений к АТС).
4. Для долгосрочного прогнозирования была получена лингвистическая база правил, содержащая 49 правил. Точность базы правил сравнима с точностью, полученной нейронной сетью.
5. С целью уменьшения числа правил была создана база правил свободной семантики, которая содержала всего 10 правил и по точности также была сравнима как с нейронной сетью, так и с лингвистической базой правил.
6. Для краткосрочного прогнозирования была получена лингвистическая база правил, содержащая 15 правил. Однако точность созданной базы существенно уступала точности, достигнутой нейронными сетями.

7. С целью улучшения точности лингвистические правила были подвергнуты тунингу. Полученные правила свободной семантики (15 правил) по точности даже превосходили точность нейронной сети. Однако интерпретируемость правил уменьшилась.

8. Целесообразность уменьшения числа правил или улучшения точности за счет потери интерпретируемости предлагается решать экспертно отдельно для каждой задачи.

Литература

1. Заде Л.А. Понятие лингвистической переменной и его применение к принятию приближенных решений / Заде Л.А. – М.: Мир, 1976. – 167с.
2. Леоненков А. Нечеткое моделирование в среде Matlab и fuzzyTECH / Леоненков А. - СПб.: БХВ-Петербург, 2005. – 736с.: ил.
3. Michalewicz Z. Genetic Algorithms + Data Structures= Evolution Programs / Michalewicz Z. .- [3rd edition] .- Berlin: Springer-Verlag, 1996. – 387p.
4. Скобцов Ю.О. Основы еволюційних обчислень / Скобцов Ю.О. - Навчальний посібник.-Донецьк: ДонНТУ, 2008.- 326с.
5. Cordon O. GENETIC FUZZY SYSTEMS. Evolutionary Tuning and Learning of Fuzzy Knowledge Bases / [Cordon O. Herrera F., Hoffmann F., Magdalena.L.]. - World Scientific. - Advances in Fuzzy Systems - Applications and Theory. - Vol. 19. - 463pp.
6. Cordon O. A Three-Stage Evolutionary Process for Learning Descriptive and Approximative Fuzzy Logic Controller Knowledge Bases from Examples / [Cordon O., Herrera F.] // International Journal of Approximate Reasoning.-1997.-V.17-4.-pp.369-407.
7. Cordon O. Hybridizing Genetic Algorithms with Sharing Scheme and Evolution Strategies for Designing Approximate Fuzzy Rule-Based Systems. / Cordon O., Herrera F. // Fuzzy Sets and Systems.- 2001.- № 118:2 .- pp.235-255.
8. Хмелевой С.В. Создание и применение базы знаний на основе аппроксимативных нечетких логических контроллеров для прогнозирования internet трафика / Хмелевой С.В. // Наукові праці ДонНТУ Серія: “Обчислювальна техніка та автоматизація” .- Донецьк: ДонНТУ. – 2007.- Випуск 13 (121). - 226с.- с.132-139.
9. Хмелевой С.В. Инструментальные средства для создания базы знаний на основе нечетких продукций, настраиваемых с помощью генетических алгоритмов. / [Хмелевой С.В.,Скобцов Ю.А. Фонов А.М.] // Сучасні тенденції розвитку інформаційних технологій в науці, освіті та економіці: Матеріали II Всеукраїнської науково-практичної конференції. 8-10 квітня 2008р. М.Луганськ. – Луганськ: Альма-матер.- 2008. – 226с. – с. 103-105.
10. Дьяконов В. Математические пакеты MATLAB. Специальный справочник / Дьяконов В., Круглов В. – СПб.: Питер, 2001. – 480с.: ил.
11. Хмелевой С.В. Некоторые аспекты предварительной обработки данных в задачах нейросетевого прогнозирования и классификации/ [Скобцов Ю.А., Хмелевой С.В., Панченко З.В.] // Вестник Донецкого Национального Университета. – Серия А: естественные науки.- Донецк:ДонНУ.- 2005г.- №2,Часть 2. – 436с. – с.381-386.

УДК 681.3 /С.В. Хмелевой, А.М. Фонов, А.А. Теличко/ Прогнозирование временных рядов с помощью набора нечетких правил, настраиваемых с помощью генетических алгоритмов. В статье рассмотрен вопрос расширения сферы применения системы для создания нечетких правил, настраиваемых с помощью генетических алгоритмов. Ранее такая система применялась для задач классификации, произведено применение для реальной задачи прогнозирования временных рядов: прогнозирование определения окупаемости оборудования. Получен набор интерпретируемых нечетких правил, с помощью которых осуществляется прогнозирование.

УДК 681.3 /С.В. Хмільовий, А.М. Фонов, Г.О. Теличко/ Прогнозування часових рядів за допомогою набору нечітких правил, що настроюються за допомогою генетичних алгоритмів. В статті розглянуто питання розширення сфери застосування системи для створення нечітких правил, що настроюються за допомогою генетичних алгоритмів. Раніше така система застосовувалася для задач класифікації, проведено застосування для реальної задачі прогнозування часових рядів: прогнозування визначення окупності обладнання. Отримано набір інтерпретуємих нечітких правил, за допомогою яких здійснюється прогнозування.

/S.V. Khmilovyy, A.M. Fonotov, A.O. Telichko/ Time Series Forecasting with the fuzzy rules set, adjusted by genetic algorithms. In the article the question of expansion of application sphere of system for the fuzzy rules creation tuned by genetic algorithms is considered. Earlier this system was applied to classification problems, now application is made for a real problem of time series forecasting: forecasting of recoupment decision of the equipment. The interpretable fuzzy set for forecasting is received.

СВЕДЕНИЯ ОБ АВТОРАХ

1. Хмелевой Сергей Владимирович, ассистент каф. АСУ Донецкого национального технического университета.

Научные интересы: нечеткие методы обработки данных, эволюционные алгоритмы, прогнозирование, временные ряды, data mining, нейронные сети

Тел. (0622) 304-90-20 (сл.), (050) 614-29-40 (моб.)

E-mail: hmelevoy@kita.dgtu.donetsk.ua

2. Фонов Анастас Михайлович, доцент каф. АСУ Донецкого национального технического университета.

Научные интересы: нечеткие методы обработки данных, эволюционные алгоритмы, теория вероятностей, системный анализ

Тел. (0622) 304-90-20 (сл.), (050) 567-96-68 (моб.)

E-mail: fonotov@kita.dgtu.donetsk.ua

3. Теличко Анна Александровна, к.т.н., доцент кафедры Прикладной математики и информатики Донецкого национального технического университета.

Научные интересы: генетические алгоритмы, управление персоналом предприятия, методы оптимизации, моделирование экономических процессов, системный анализ.

Тел. 8(062)301-07-29, 301-08-56

E-mail: anna@pmi.dgtu.donetsk.ua