

УДК 681.3

С.В. Хмелевой, Т.А. Васяева  
Донецкий национальный технический университет  
hmelevoy@cs.dgtu.donetsk.ua, vasyaeva@gmail.com

## Информационная технология извлечения знаний для прогнозирования временных рядов

*В статье описывается информационная технология извлечения набора нечетких правил для прогнозирования временных рядов, а именно временного ряда числа подключений абонентов к строящимся АТС. Приведены описание математического, программного, организационного и других обеспечений технологии. Приведен пример использования технологии для задачи прогнозирования числа подключения абонентов к АТС.*

**Нечеткие правила, прогнозирование временных рядов, подключение абонентов к строящимся АТС, UML**

### 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]. Можно сказать, что мы имеем дело с процессом извлечения знаний, где знания представлены в форме набора нечетких правил. Процесс создания и применения базы знаний был описан в [8]. Инструментальные средства для извлечения знаний были описаны и протестированы в [9]. Однако последовательность использования данных инструментальных средств для извлечения знаний, представляющий собой

информационную технологию, описана не была. 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 описание информационной технологии для извлечения знаний в виде набора нечетких продукций применимых к решению задач прогнозирования временных рядов.

### 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 ёмкости 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 прогнозные 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 прогнозной 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 прогнозной 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.

Необходимо описать процесс получения результата с помощью созданных инструментальных средств. Данное описание также должно быть понятно людьми, не обладающими большими знаниями в области математики и программирования (операторами). Выполняться данное описание должно в рамках создаваемой информационной технологии.

Успешное решение задачи прогнозирования подразумевает построение в том или ином виде модели  $g$  поведения временного ряда, используемой для прогнозирования. Как уже говорилось ранее, при решении задач, требующих построения прогнозной модели, требующейся для обоснования принятия решений лицам, не имеющим специальных математических знаний, выдвигается требование интерпретируемости полученной модели  $g$ . А для успешного решения задачи сравнения моделей между собой необходимо решить задачу извлечения знаний из прогнозной модели временного ряда. Как уже указывалось, построение модели производится на основе системы нечеткого вывода (Fuzzy Rule-Based System) *FRBS*, с помощью которой осуществляется *нечеткий вывод* (Fuzzy Reasoning, *FRSN*). Нечеткий вывод *FRSN* можно определить как результат принятия решения на основе базы нечетких правил и значения входных переменных. Нахождение прогноза, с учетом дифференциации входных переменных, можно представить как

$$Y_{t+1} = FRSN(R, \bar{Y}, \bar{V}), \quad (1)$$

где  $R$  - база нечетких правил,

$\bar{Y}$  - лаги временного ряда,

$\bar{V}$  - другие влияющие факторы.

При этом интересует как улучшение (уменьшение) критерия точности ( $AC(FRSN)$ ), так и критерия интерпретируемости ( $IC(FRSN)$ ) системы нечеткого вывода. Необходимо, чтобы выполнялась минимизация одного из этих двух критериев (выбор критерия производится экспертно):

$$\begin{aligned} AC(FRSN(R, \bar{Y}, \bar{V}, \bar{C})) &\rightarrow \min \text{ или} \\ IC(FRSN(R, \bar{Y}, \bar{V}, \bar{C})) &\rightarrow \min \end{aligned} \quad (2)$$

При этом должно выполняться: 1) условие конечности времени поиска решения; 2) ограничение роста критерия интерпретируемости (понятность результата должна сохраняться).

В результате анализа предметной области и применимости различных методов было принято решение об уместности использования для получения интерпретируемого прогноза аппарат нечетких множеств. В таком случае прогноз будет определяться на основании полученного набора нечетких правил. При подаче на вход конкретных данных для получения прогнозного значения используется классический аппарат нечеткого вывода. Неподготовленное лицо может визуально проверить, какие правила срабатывают для

конкретных данных, поданных на вход. Каждое правило может быть объяснено лингвистически, и озвучено.

В таком случае основным критерием интерпретируемости создаваемой БЗ при использовании лингвистической семантики является число подусловий правила и число правил. Можно свести его к общему числу подусловий всех правил базы знаний:

$$IC = \sum_{i=1}^m \sum_{j=1}^n isExist(\mu_{A_{ij}}(x_j)), \quad (3)$$

где функцию  $isExist$  можно определить как

$$isExist = \begin{cases} 1, & \text{если } \max_{x \in X}(\mu_A(x)) > 0 \\ 0, & \text{иначе} \end{cases} \quad (4).$$

В случае использования свободной семантики при создании базы знаний о существовании лингвистических переменных говорить можно весьма условно, о выполнимости условий 2, 3, 4, 8 нельзя говорить однозначно. Степень выполнимости этих условий предлагается выразить одним коэффициентом -  $K_{free}$ . Тогда выражение (3) принимает вид:

$$IC_{free} = K_{free} * \sum_{i=1}^m \sum_{j=1}^n isExist(\mu_{A_{ij}}(x_j)) \quad (5).$$

В качестве критерия точности используется средняя квадратичная ошибка:

$$AC = MSE = \frac{1}{N} \sum_{l=1}^N err_l^2 = \frac{1}{N} \sum_{l=1}^N (y^l - FRSN(x_0^l))^2, \quad (6)$$

где  $err_l$  – погрешность нечеткого вывода на  $l$ -м примере,  $err_l = y^l - FRSN(x_0^l)$ ;

$y_l$  – реальное прогнозное значение (эталонный выход);

$FC(x_0^l)$  - прогноз системы нечеткого вывода для этого примера (нечеткий вывод системы).

По имеющейся информации на данный момент информационных технологий, осуществляющих прогнозирование временных рядов с помощью набора нечетких правил, еще не существует.

Поскольку входной информацией являются временные ряды, то среди всех входных факторов присутствуют и различные

отсчеты временных рядов (лаги). После анализа предметной области в качестве входной информации для получения долгосрочного прогноза из 12 первоначальных были выделены следующие факторы: Ф3 («спрос»), Ф14 («работы сумма последних 3-х лагов»), и Ф13 («работы по расширению»). Фактор Ф3 («спрос») – определяет, есть ли в конкретном регионе спрос на данные услуги. Фактор Ф14 («работы сумма последних 3-х лагов») содержит время, которое ведутся работы по установке нового оборудования. И фактор Ф13 («работы по расширению») – наличие работ по установке нового оборудования в настоящее время.

Данная информационная технология при построении основывается на том, что существует автокорреляция между будущим значением ряда, и его прошлыми значениями. Поэтому можно сказать, что данной информационной технологией используется авторегрессионная прогнозная модель.

**Выходной информацией** является значение прогноза подключений. Горизонт прогноза логично устанавливать максимальным. Анализ автокорреляционной функции ряда подключений показал, что максимальная значимая глубина автокорреляции составляет не более одного месяца. Поэтому горизонт прогноза выбран равным одному месяцу. Пример лингвистического правила, получаемого с помощью данной информационной технологии:

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).

### Информационное и математическое обеспечения технологии

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  $S=C_1C_2$ , 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  $C_1$ , получаются 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.

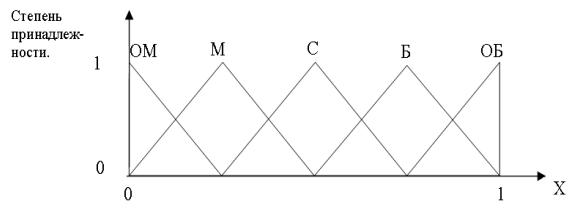


Рисунок 1. Построение функций принадлежности, характеризующих термы лингвистической переменной «размер»

ЕСЛИ	$X \in$		ТО $Y \in$	
ЕСЛИ	$X \in$		ТО $Y \in$	

Рисунок 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].

**Программное обеспечение  
информационной технологии**

Программное обеспечение целесообразно реализовать, используя модульный принцип с применением объектно-ориентированного программирования. Основные объекты системы и их взаимодействие показано диаграммой компонентов (в терминах UML [14]), представленной рис. 3. Использование модульного построения позволяет получить гибкость настройки, высокую универсальность, возможность расширения функциональности системы.

В структуре инструментальных средств можно выделить модуль ввода и предварительной обработки данных, модули получения базы знаний и постпроцессинга. Кроме этого, необходимо создание модуля управления и визуализации (в рамках создания интерфейса взаимодействия с пользователем).

Для реализации *модуля создания базы знаний* необходимо создание иерархии объектов. Разработанная диаграмма классов в терминах UML [14] представлена на рис. 5.

Для создания базы правил предусмотрены следующие типы и классы:

- TExample: одна точка из ОВ или ТВ;
- TMemberFunc: одно условие правила (или терм);
- TRule: одно правило;
- TPrimaryFuzzy: объект для терм-множества лингвистических переменных;
- TDataSet: класс для множества обучающих и тестовых данных;
- TPopulation: класс для нижнего уровня генетического алгоритма (содержащий популяцию правил);
- TGenerGa: верхний уровень генетического алгоритма создания базы знаний.
- TMultiSimpGa, TTuningGa – объекты для реализации мультисимплификации и тюнинга.

Возможно применение различных видов функций принадлежности: треугольных, гауссовых, трапецевидных и т.д.. Объекты TRule, TPrimaryFuzzy, TGenerGa рассчитаны на треугольную функцию принадлежности. Для возможности её модификации предусмотрено создание абстрактных классов TBaseRule, TBasePrimaryFuzzy, TBaseGenerGa, на которых могут основываться и классы-потомки с другими функциями принадлежности. Более подробно программное обеспечение информационной технологии описано в [9].

Также для убыстрения работы генетического алгоритма выполнена его

параллельная реализация (объекты TCoverServer, TCoverClient). Более подробно параллельная реализация показана в [15]

**Лингвистическое обеспечение  
информационной технологии**

Лингвистическое обеспечение информационной технологии выполнено в рамках модуля управления и визуализации. На рис. 4 показана экранная форма визуализации процесса получения прогноза.

Организационное обеспечение технологии может быть представлено диаграммой вариантов использования (в терминах UML [14]). Вводятся следующие роли:

- директор. Дает команду на производство прогноза для определенного района. Принимает решение о расширении мощностей Оператора на основании произведенного прогноза.
- администратор базы данных. Выполняет запуск запросов на считывание данных из БД Оператора.
- оператор. Управляет программой. Получает данные от администратора БД. Собирает сведения о районах, для которых будет производиться прогноз. Получает непосредственный прогноз.

Для получения прогноза может выбирать различные функции принадлежности (треугольная, гауссова и т.д.), в зависимости от задачи. Для ускорения работы генетического алгоритма может запускать параллельный генетический алгоритм.

Для улучшения полученной базы знаний используют процедуры постпроцессинга. Они включают в себя мультисимплификацию (для сокращения числа правил), взвешивание (для улучшения точности), а также тюнинг (улучшает точность за счет ухудшения интерпретируемости). Тюнинг применяется при неприемлимой точности даже после применения всех остальных методов улучшения точности.

– программист. Модифицирует программное обеспечение информационной технологии. Это происходит в случае, если полученный результат все равно не устраивает руководство, несмотря на все примененные процедуры по улучшению.

Диаграмма вариантов использования показана на рис. 6. С созданием диаграммы вариантов использования и созданием организационного обеспечения задачу описания информационной технологии извлечения знаний для прогнозирования временных рядов можно считать выполненной.

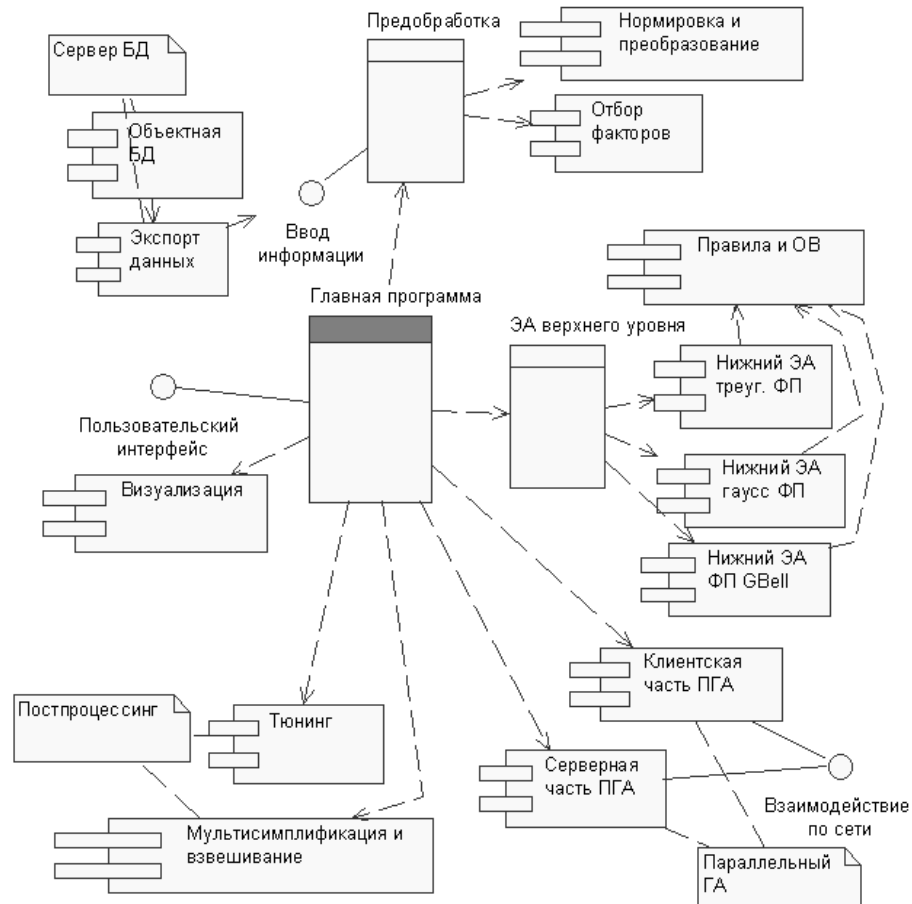


Рисунок 3. Диаграмма компонентов создаваемой системы

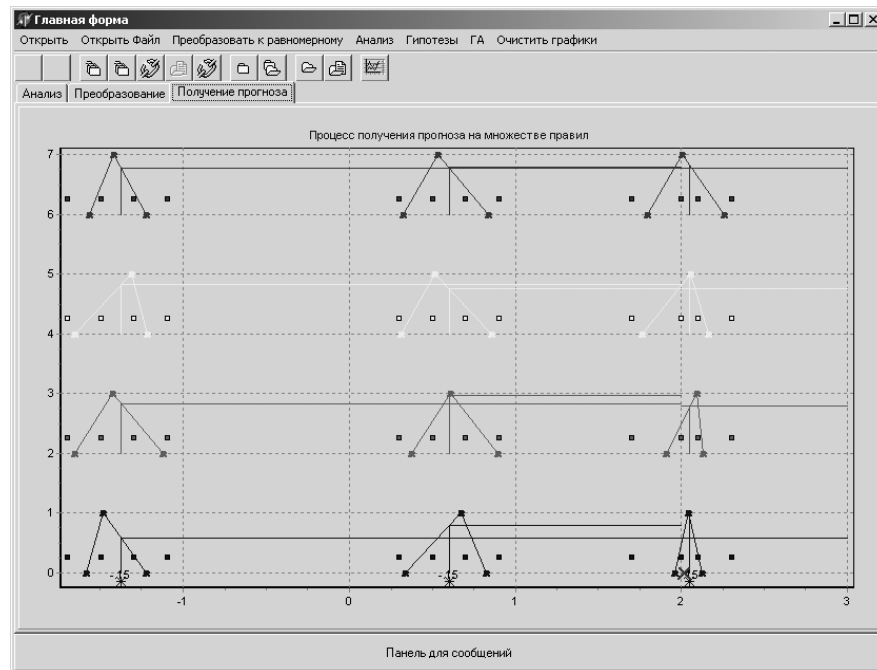


Рисунок 4. Визуализация процесса получения прогноза для 2 входных факторов

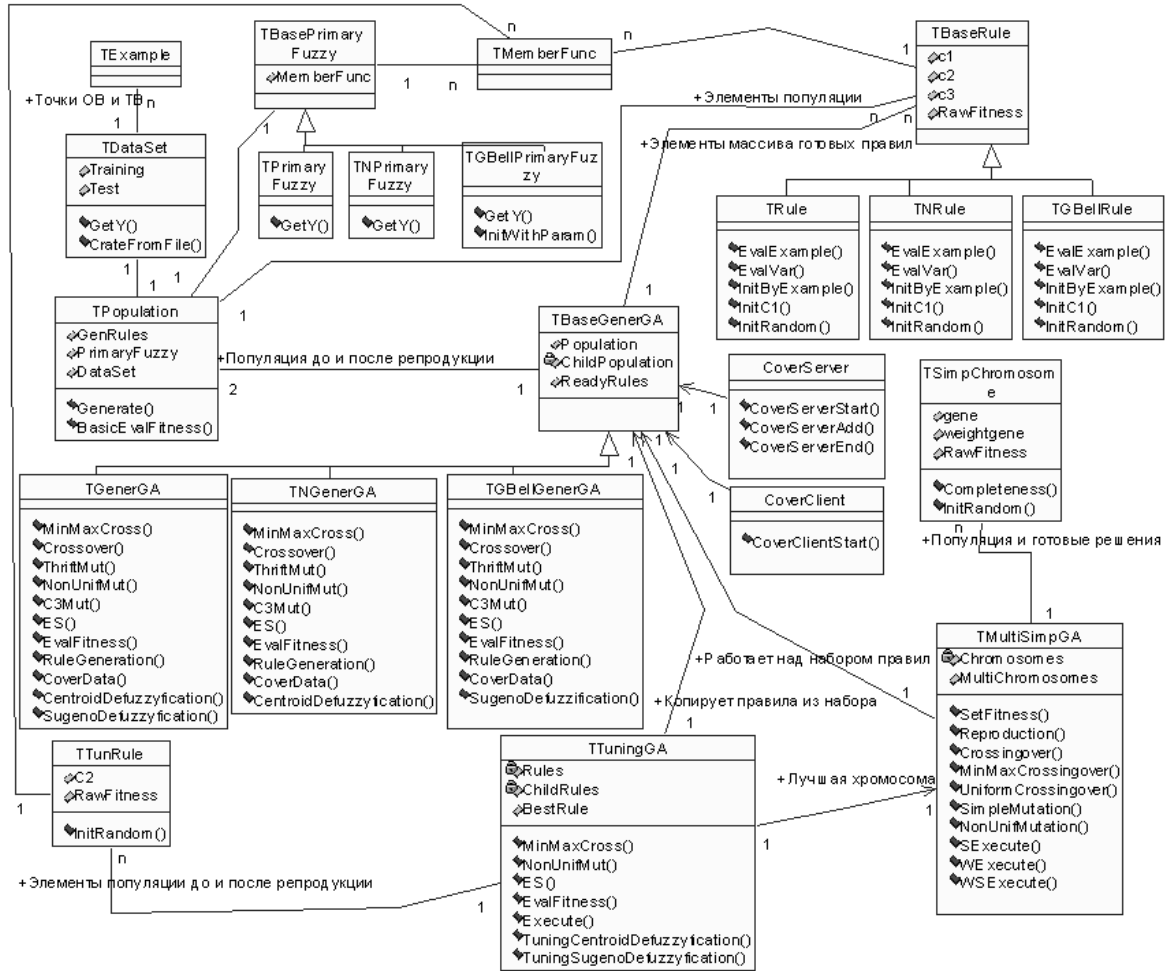


Рисунок 5. Разработанная диаграмма классов объектов

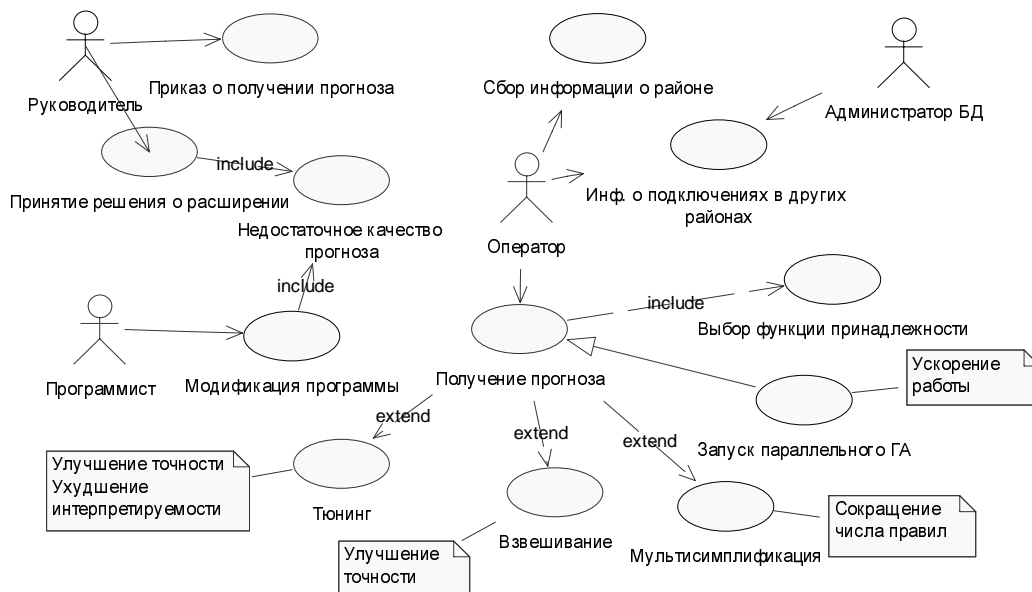


Рисунок 6. Диаграмма вариантов использования программы по получению прогноза

### Анализ полученных правил

Полученные лингвистические правила и их веса для долгосрочного прогноза представлены в табл. 1. Для 1, 2 и 4-го столбцов таблицы («Спрос», «Работ 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) (см. рис. 1). 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 подзаключениями. Их можно заменить одним с подзаключением произвольной формы или описать лингвистически. Например, для 43-44 правил:

Таблица 1. Лингвистические правила БЗ для прогнозирования числа подключений

№	Спрос	Работ 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
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

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).

Правила базы знаний свободной семантики для прогноза показаны на рис.7.

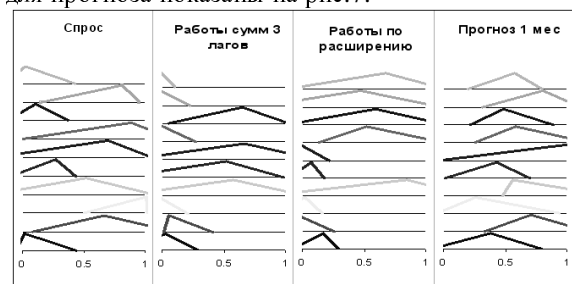


Рис. 7. База знаний для прогнозирования количества подключений.

### Анализ результатов

Отобрано оптимально краткое множество, на основании которого можно решить задачу прогнозирования числа подключений, не ухудшая критерий точности АС. Для долгосрочного прогнозирования число факторов=3, к множеству предыдущих значений ряда  $\bar{Y}$  не принадлежит ни



один из них, к множеству характеристик, все факторы принадлежат к множеству параметров  $\bar{V}$ . Таким образом, основное соотношение (1), определяющее прогноз, для долгосрочного прогнозирования приобретает следующий вид:

$$Y_{t+l} = g(\Phi 3, \Phi 13, \Phi 14), \quad (7)$$

где  $\bar{V} = (\Phi 3, \Phi 13, \Phi 14)$ ;  $\bar{Y} = ()$ .

Для долгосрочного прогнозирования основной критерий точности АС правил лингвистической семантики (2.204) и правил свободной семантики (2.194) имеют близкие значения. Число подусловий базы знаний лингвистической семантики, являющимся основным составляющим критерия интерпретируемости  $IC$ , равно 141, свободной семантики  $IC_{free} = 30 * K_{free}$ . Целесообразность уменьшения числа правил за счет ухудшения интерпретируемости предлагается решать экспертно.

### Выводы

В результате исследований можно сделать следующие выводы:

- Показано преобразование входной информации к выходной, производимое с помощью информационной технологии. Входная информация – временные ряды, а также их лаги. Выходная – временной ряд прогноза числа подключений через месяц.

- Получение прогноза производится на основании набора нечетких правил, получаемых информационной технологией. Интерпретация таких правил очень проста и не требует специальных знаний.
- В рамках математического и информационного обеспечений приведены в словесной форме алгоритмы и этапы получения интерпретируемого прогноза.
- В рамках программного обеспечения технологии приведены диаграмма классов и диаграмма компонентов системы для получения базы знаний, включающей в себя набор интерпретируемых правил.
- В рамках организационного обеспечения показана последовательность использования различных процедур программного обеспечения. Приведена диаграмма вариантов использования системы.
- Показано применение описанной информационной технологии для примера прогнозирования числа подключений к АТС.
- С помощью описанной информационной технологии можно создавать как набор правил, обладающих большей понятностью, за счет увеличения числа правил, так и набор правил, имеющих меньшее число правил, за счет ухудшения их понятности. Целесообразность ухудшения интерпретируемости для уменьшения числа правил в рамках данной технологии предлагается решать экспертно.

### Список литературы

1. Заде Л.А. Понятие лингвистической переменной и его применение к принятию приближенных решений / Л.А. Заде. – М.: Мир, 1976. – 167с.
2. Леоненков А. Нечеткое моделирование в среде Matlab и fuzzyTECH / А. Леоненков. – С-Пб.: БХВ-Петербург, 2005. – 736с.: ил.
3. Michalewicz Z. Genetic Algorithms + Data Structures= Evolution Programs. [3rd edition] Berlin: Springer. – Verlag, 1996. – 387p.
4. Скобцов Ю.О. Основы эволюційних обчислень: навч. посiб. / Ю.О. Скобцов. – Донецьк: ДонНТУ, 2008.- 326с.
5. 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. Three-Stage Evolutionary Process for Learning Descriptive and Approximate Fuzzy Logic Controller Knowledge Bases from Examples / O. Cordon, F. Herrera // 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 / O. Cordon, F. Herrera // Fuzzy Sets and Systems. – 2001. – № 118:2. – 235-255 pp.
8. Хмелевой С.В. Создание и применение базы знаний на основе аппроксимативных нечетких логических контроллеров для прогнозирования интернет траффика / С.В. Хмелевой // Науковi працi ДонНТУ. – 2007. – Вип. 13 (121). – С. 226, 132-139.
9. Хмелевой С.В. Инструментальные средства для создания базы знаний на основе нечетких продукций, настраиваемых с помощью генетических алгоритмов / С.В. Хмелевой, Ю.А. Скобцов, А.М. Фонотов // Сучаснi тенденцiї розвитку iформацiйних технологiй в науцi, освiтi та екoномiцi: Матерiали II Всеукраїнської науково-практичної конференції. М.Луганськ. – Луганськ: Альма-матер. – 2008. – С. 226, 103-105.

10. Информация. Информационные технологии. Информационные системы. [Электронный ресурс] / ITstan.ru. – Интернет-сайт.- [М]. – 2007. – Режим доступа: <http://www.itstan.ru/> .- Загл. с экрана.
11. Уткин В.Б. Информационные системы в экономике: учебник для студ. высш. учеб. заведений / В.Б. Уткин, К.В. Балдин. – М.: Издательский центр «Академия», 2004. — 288 с.
12. Иванова Г.С. Технология программирования: учебник для вузов / Г.С. Иванова. – М.: Изд-во МГТУ им. Н.Э.Баумана, 2002. – 320 с.
13. Титоренко Г.А. Информационные технологии в маркетинге: учебник для вузов / Г.А. Титоренко. – М.: ЮНИТИ-ДАНА, 2000. – 335с.
14. Фаулер М. UML Основы / М. Фаулер, К. Скотт ; пер. с англ. – СПб: Символ-Плюс, 2002. – 192 с., ил.
15. Хмелевой С.В. Параллельная реализация эволюционного алгоритма для создания базы знаний на основе нечетких логических контроллеров / С.В. Хмелевой // Информационно-управляющие системы на железнодорожном транспорте. – 2007. – №4 (66). – С. 128, 120-123.

Надійшла до редколегії 10.05.2011

**С.В. ХМІЛЬОВИЙ, Т.О. ВАСЯЄВА**

Донецький національний технічний університет

**Інформаційна технологія видобування знань  
для прогнозування часових рядів**

В статті описується інформаційна технологія видобування набору нечітких правил для прогнозування часових рядів, а саме часового ряду числа підключень абонентів до АТС, що будуються. Приведені описи математичного, програмного, організаційного та інших забезпечень технології. Опис всіх видів забезпечення проводився з використанням термінів UML

*Нечіткі правила, прогнозування часових рядів,  
підключення абонентів до АТС, UML*

**S.V. KHMILOVYY, T.A. VASJAEVA**

Donetsk national technical university

**Information technology of knowledge discovery for  
time series forecasting**

There is described the information technology of fuzzy rules set generation for time series forecasting, namely a time series of subscribers connections number before automatic switching center construction. Mathematical, program, organizational and other support descriptions of technology are present. Descriptions of all kinds of supports was made with use of UML

*Fuzzy rules, time series forecasting, subscribers  
connections number before automatic switching  
center construction, UML*