

МЕТОДЫ И ТЕХНОЛОГИИ ПРИНЯТИЯ РЕШЕНИЙ

УДК 004.891

Научная статья

DOI: 10.18287/2223-9537-2025-15-3-404-417



Методы гранулирования нечётких временных рядов для анализа данных

© 2025, Р.А. Бурнашев ✉, Я.В. Сергеев, А.Ф. Назипова

Казанский (Приволжский) федеральный университет (КФУ),

Институт вычислительной математики и информационных технологий, Казань, Россия

Аннотация

Увеличение размерности данных, вызванное множеством гетерогенных временных рядов, обуславливает разработку эффективных способов их обработки и сжатия. В данной статье рассмотрен подход к сжатию данных, представленных в виде временных рядов, основанный на гранулировании с использованием методов нечёткой логики. Проведён анализ среднесуточной температуры в Республике Татарстан в период метеонаблюдений с 1881 по 2024 годы. Гранулирование позволило значительно сжать объём данных. Для преобразования исходных числовых данных в информационные гранулы использовано нечёткое резюмирование, которое позволило осуществить автоматическую генерацию гранулярного описания поведения временного ряда. Резюмирование результатов состояния временного ряда, полученных при анализе, выполнено с помощью методов нечёткой логики, включающих набор правил, функции принадлежности для каждого временного сезона, интервальные лингвистические переменные и программный модуль логического вывода (дефазификации). Применение предложенных методов в рассмотренном примере позволило сократить объём данных с 52534 до 7504 точек (коэффициент сжатия ~85%). Разработанные методы могут быть использованы для анализа больших объёмов данных в различных прикладных областях.

Ключевые слова: анализ данных, нечёткая логика, база знаний, нечёткое резюмирование, гранулирование, временные ряды.

Цитирование: Бурнашев Р.А., Сергеев Я.В., Назипова А.Ф. Методы гранулирования нечётких временных рядов для анализа данных. *Онтология проектирования*. 2025. Т.15, №3(57). С.404-417. DOI:10.18287/2223-9537-2025-15-3-404-417.

Финансирование: работа выполнена за счёт гранта Академии наук Республики Татарстан в рамках Государственной программы Республики Татарстан «Научно-технологическое развитие Республики Татарстан» (Соглашение от 16.12.2024 №04/2024-ПД).

Вклад авторов: Бурнашев Р.А. – разработка концепции статьи и алгоритмов; Сергеев Я.В. – построение математической модели, тестирование нечётких правил; Назипова А.Ф. – обработка метео данных, гранулирование временных рядов, визуализация результатов.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Введение

Климатические изменения оказывают значительное влияние на социально-экономические системы, что обуславливает необходимость разработки новых подходов к прогнозированию и принятию решений. Рост объёмов данных, их неопределённость и сложность структуры делают традиционные методы анализа недостаточно эффективными. Актуальной задачей становится разработка методов, способных обрабатывать большие объёмы данных с учётом их неопределённости и обеспечивать достоверность данных в системах

поддержки принятия решений (СППР). Перспективным направлением в этой области является использование нечётких временных рядов, которые позволяют учитывать неопределённость и изменчивость данных, характерных для климатических и экономических процессов. В качестве предметной области рассмотрены метеоданные в одном из регионов России.

Целью данной статьи является разработка подходов к обработке больших массивов данных для СППР с использованием методов нечёткой логики и машинного обучения.

1 Нечёткая логика и нечёткие множества

Применение принципов нечёткой логики стало ключевым фактором в повышении устойчивости моделей прогнозирования.

Например, в работе [1] предложен многоступенчатый метод оперативного прогнозирования скорости ветра, который сочетает ансамблевый подход на основе модели прогнозирования погоды, нечёткую систему для извлечения данных и оптимизацию. Модель учитывает различные горизонтальные разрешения и исходные поля, что позволяет улучшить точность прогнозирования. В [2] представлен комбинированный интервал прогнозирования на основе нечёткой логики для прогнозирования ветровой энергии.

Применение методов машинного обучения и нечёткой логики позволяет повысить точность прогнозирования. Например, в [3] использована модель прогнозирования и улучшенный алгоритм кластеризации для установления функции нечёткого разделения. В [4] предложена модель прогнозирования концентраций загрязняющих веществ на основе исторических данных о погоде и загрязнении. Используются нечёткие модели для обработки неточных и неопределённых данных, что позволяет прогнозировать уровни загрязнения. В [5] предложен метод прогнозирования временных рядов на основе интуиционистских нечётких множеств, учитывающий неопределённость в данных. Метод показал высокую эффективность по сравнению с другими подходами на основе интуиционистских нечётких множеств. В [6] предложены классификация многокритериальных методов принятия решений и правила их выбора при разработке специализированного программного обеспечения для СППР. В [7] предложен подход к оценке технического состояния электрооборудования, основанный на использовании взвешенных нечётких правил. Данный подход позволяет учитывать нечёткий характер исходных данных. В [8] приведён анализ, моделирование и прогноз урожайности сельскохозяйственных культур с учётом динамики изменения природно-климатических характеристик путём применения методики нечёткой логики. Модель прогнозирования урожайности на основе нечёткой логики учитывает динамику климатических изменений. В [9] предложена методика оценки рисков снижения полевой всхожести зерновых культур, учитывающая степень кондиционности семенного материала на основе моделей нечёткой логики.

Системы прогнозирования на основе нечётких временных рядов представляют собой надёжный инструмент для работы с неопределённостью в данных временных рядов. Развитию теоретических основ и практического применения нечётких систем прогнозирования в различных областях посвящены работы [10-14].

Гранулирование временного ряда представляет собой процесс группировки данных временного ряда в более крупные временные интервалы, который позволяет: упростить структуры данных за счёт снижения их объёма; уменьшить влияния «шума» и случайных колебаний; повысить скорость обработки данных; улучшить визуализацию и интерпретацию временных рядов (см. таблицу 1).

Таблица 1 – Операции в гранулировании данных

Операция	Краткое описание
Классификация [15]	Назначение временным рядам одного из заранее определённых классов.
Кластеризация [16]	Группировка данных на основе их схожести.
Сегментация [17]	Разбиение временного ряда на значимые сегменты, расположенные последовательно.
Резюмирование [18]	Создание компактного представления данных, сохраняющего ключевые характеристики.
Ассоциативные правила [19]	Поиск правил, относящихся к временным рядам.
Оптимизация [20]	Поиск оптимальных параметров для сжатия данных.

Представленные в таблице 1 операции гранулирования данных могут включать элементы нечёткой логики. Нечёткая логика используется в следующих задачах:

- группировка данных с возможностью одному объекту принадлежать одновременно нескольким или всем кластерам, но с различной степенью принадлежности (*нечёткая кластеризация*) [21];
- разбиение временного ряда на сегменты с учётом нечётких границ между ними (*нечёткая сегментация*) [22];
- вывод на базе нечёткой онтологии, который используется для резюмирования состояния сложной технической системы (*нечёткое резюмирование*) [23] и др.

В исследовании климатических факторов, которые часто изменяются в зависимости от временных сезонов, применён метод гранулирования с использованием нечёткого резюмирования временного ряда [24, 25].

Нечёткая логика позволяет работать с неточными или неопределёнными данными, оперируя степенями принадлежности к множеству в диапазоне от 0 до 1.

Пусть X – некоторое универсальное множество. Для нечёткого множества A степень принадлежности каждого элемента $x \in X$ этому множеству является действительным числом $\mu_A(x)$, принадлежащим отрезку $[0,1]$. Функция $\mu_A(x)=0$, если $x \notin A$; $\mu_A(x)=1$, если $x \in A$; $0 < \mu_A(x) < 1$, если x принадлежит множеству A частично.

Функция, описывающая, как элементы универсального множества X относятся к нечёткому множеству A , называется *функцией принадлежности*. В настоящем исследовании использована треугольная функция принадлежности, представленная в виде:

$$\mu_A(x) = \max\left(0, \min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right)\right), \quad a \leq b \leq c.$$

Нечёткая логика позволяет оперировать лингвистическими переменными, значения которых задаются словами естественного языка. В разрабатываемой базе знаний для каждого временного сезона созданы лингвистические переменные: *холодная*, *умеренная* и *тёплая*. Логический вывод (дефазификация) выполняется с использованием метода центра тяжести:

$$y^* = \frac{\sum_i x_i \cdot \mu_A(x_i)}{\sum_i \mu_A(x_i)},$$

где y^* – лингвистическая переменная для соответствующего сезона; $\mu_A(x_i)$ – функция принадлежности нечёткого множества A ; $x_i \in X$ – нечёткие значения, принадлежащие конечному универсальному множеству.

2 Представление и обработка данных

Исходные данные для исследования включали климатические данные: средние ежедневные значения температуры воздуха и количества осадков в Республике Татарстан в период с января 1881 года по октябрь 2024 года¹. Формат записи используемых данных имеет вид:

<i>Station</i>	идентификатор метеорологической станции Всемирной метеорологической организации;
<i>Year</i>	год, в который были обработаны данные;
<i>Month</i>	месяц, в котором были собраны данные, представленный в числовом формате (01-12);
<i>Day</i>	день месяца, в который были собраны данные, в числовом формате (01-31);
<i>TMEAN</i>	среднесуточная температура воздуха, измеряемая в градусах Цельсия;
<i>R</i>	суточная сумма осадков, измеряемая в миллиметрах.

Климатические данные организованы в виде таблицы, где каждая строка соответствует определённому временному интервалу (например, день, месяц), а столбцы содержат

¹ Федеральная служба по гидрометеорологии и мониторингу окружающей среды. Всероссийский НИИ гидрометеорологической информации – Мировой центр данных. Температура и осадки: набор данных. <http://meteo.ru/data/>.

показатели, такие как температура воздуха, количество осадков и др. В ходе обработки данных выполнены: проверка на наличие пропущенных значений, выбросов и «шумов»; нормализация и очистка данных для обеспечения их качества и пригодности для дальнейшего анализа. Для использования нечёткой логики введён дополнительный категориальный признак, отражающий сезонную принадлежность каждой даты (зима, весна, лето, осень). Учёт сезонности позволяет применить метод гранулирования для сжатия набора данных, что способствует снижению его размерности.

3 Гранулирование нечётких временных рядов

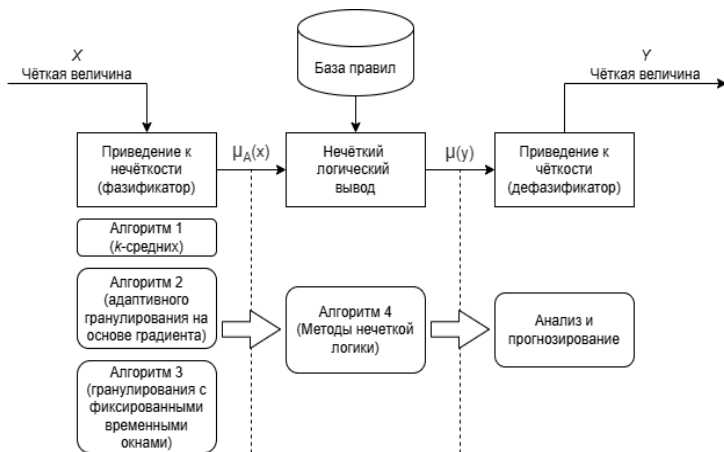
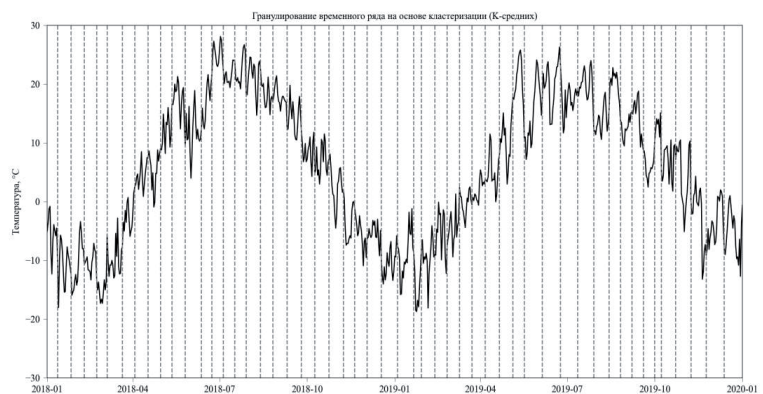


Рисунок 1 – Этапы анализа данных

схожих участков ряда. На рисунке 2 представлена схема работы алгоритма k -средних (2а) и результаты кластеризации температурного временного ряда (2б). Алгоритм зависит от значения k (количество кластеров), которое необходимо указывать для выполнения кластеризации. Средняя сложность алгоритма оценивается как $O(k*n*t)$ где k – заданное количество кластеров, n – количество выборок, t – количество итераций.



а)



б)

Рисунок 2 – Схема алгоритма (а) и результат гранулирования на основе кластеризации с использованием алгоритма k -средних (б) (временной ряд температур с 01.01.2018 по 01.01.2020, границы гранул выделены прерывистыми вертикальными линиями)

При разных значениях k изменяется скорость работы алгоритма, но может происходить и его остановка в точках локальных минимумов. В этом случае все оставшиеся значения будут отнесены к этому кластеру. Кластеры обозначаются c_i , общее количество кластеров – q , длина временного ряда – n , T_i – индекс исходного временного ряда, с которого начинается гранула i .

В качестве старта алгоритма берётся временной ряд $\{x_k\} (k=1, \dots, n)$ и определяются: границы гранул ($\forall i = 1, \dots, q, T_i = [T_i^L, T_i^R]$, где: T_i^L (индекс левой границы) – начальная точка гранулы, T_i^R (индекс правой границы) – конечная точка гранулы (условия: $T_i^L \leq T_i^R$ и $T_i^L, T_i^R \in \{1, \dots, n\}$); гранулы: $\Omega_i = \{x_k | k = T_i^L, \dots, T_i^R\}$).

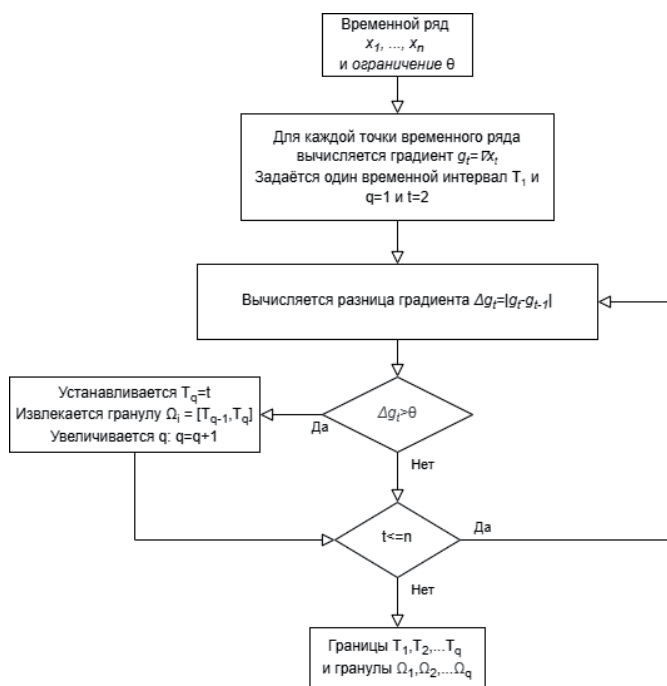


Рисунок 3 – Алгоритм адаптивного гранулирования на основе градиента

Алгоритм 2. Алгоритм адаптивного гранулирования на основе градиента (см. рисунок 3).

В ходе работы алгоритма определяются переменные, отвечающие за вычисление градиента по временным промежуткам – g_t . За основной критерий берётся разница данных градиентов Δg_t , которая оценивается относительно заранее установленного ограничения θ . Параметром, отвечающим за накопление количества временных интервалов, соответствующих количеству гранул является переменная q . Для прохода по всем временным интервалам используется переменная t , отражающая начальное количество временных интервалов. Алгоритма отличается простотой реализации и низкими вычислительными затратами на этапе инициализации. Недостатками являются самостоятельное определение порогового значения для

производной, что может повлиять на результат гранулирования, а также сложность вычислений. Результат обработки данных этим алгоритмом представлен на рисунке 4.

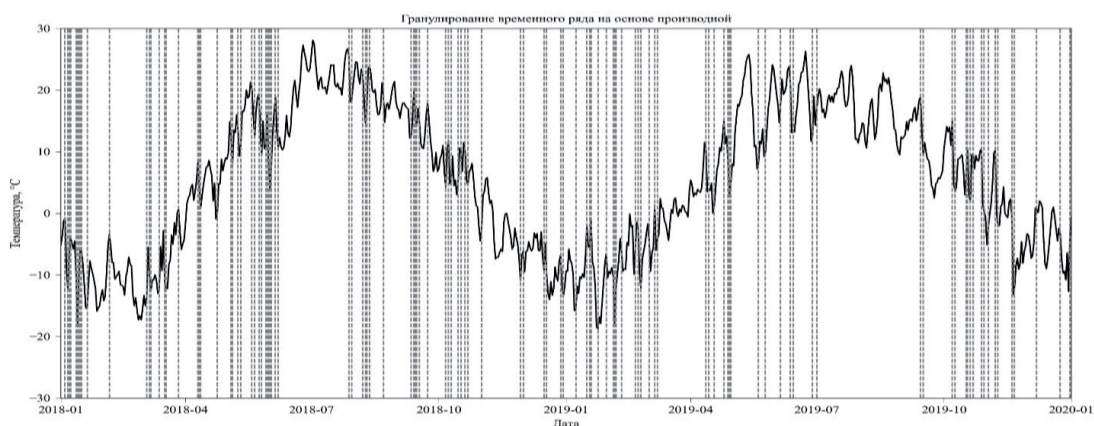


Рисунок 4 – Гранулирование временного ряда на основе градиента (границы гранул выделены прерывистыми вертикальными линиями)

Алгоритм 3. Алгоритм гранулирования с фиксированными временными окнами (рисунок 5). Он отражает выделение подпоследовательностей исходного возрастающего временного ряда в виде отдельных кластеров, содержащих основную информацию. Подобно первому алгоритму берутся границы T_i . Итерации для временного ряда происходят по $t \in N$, границы сдвигаются на $q \in N$ для сохранения порядка следования. Метод устойчив и может работать на временных рядах большого объёма. Результаты представлены на рисунке 6.

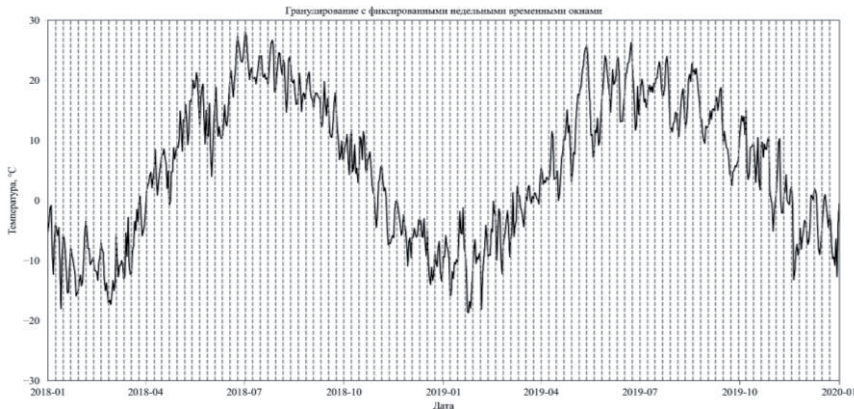


Рисунок 6 – Гранулирование с фиксированными недельными временными окнами

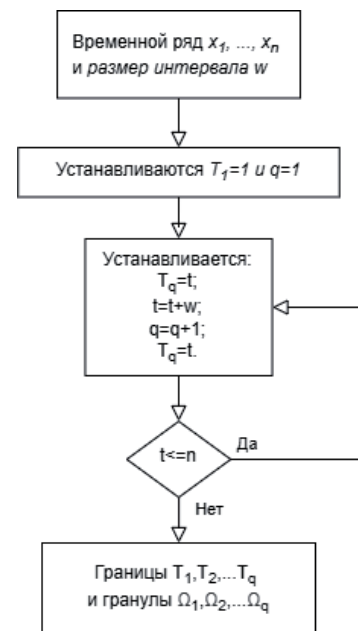


Рисунок 5 – Алгоритм гранулирования с фиксированными временными окнами

Алгоритмы 1-3 позволяют преобразовать временной ряд ежедневных температур в последовательность гранул, представляющих собой массив среднесуточных температур в течение одной недели (Ω_i). Граница T_i определяет индекс исходного временного ряда, с которого начинается гранула i . Затем каждая гранула проходит процесс фаззификации для дальнейшего её использования в формировании правил.

Алгоритм 4. Алгоритм фаззификации гранул с использованием методов нечёткой логики. Алгоритм (рисунок 7) позволяет привести чёткие численные значения к нечётким гранулам, выраженным на естественном языке [26]. Были созданы 12 лингвистических переменных, отражающих сезонные статистики S для каждого временного сезона, например, для сезона *зима* созданы лингвистические переменные: «холодная», «умеренная» и «тёплая».

В исследовании реализован гибридный алгоритм фаззификации сочетающий треугольные функции принадлежности для температурных данных (программная библиотека *scikit-fuzzy*²) и гауссовы функции для сезонных параметров. Алгоритм обработки температурных временных рядов включает следующие этапы:

- 1) ввод и предобработка данных: На вход подаётся гранула - массив из семи температурных значений (недельный временной ряд); Данные нормализуются в диапазон температур, соответствующий лингвистическим переменным.
- 2) фаззификация: для температурных значений вычисляются степени принадлежности к 12 нечётким множествам: три температурных терма («холодная», «умеренная», «тёплая») с треугольными функциями принадлежности и четыре сезонных терма «зима», «весна», «лето», «осень» (с использованием гауссовых функций принадлежности);
- 3) агрегирование: для каждого из 12 множеств вычисляется средняя степень принадлежности по всем семи точкам гранулы;
- 4) логический вывод (дефаззификация): формируется итоговый вектор F признаков гранулы. Вектор используется для активации правил из таблицы 2.

² Fuzzy logic toolkit for SciPy. DOI 10.5281/zenodo.802396. – <https://github.com/scikit-fuzzy/scikit-fuzzy?tab=readme-ov-file>.

Результатом работы *Алгоритма 4* является вектор $F = [f_1, f_2, \dots, f_{12}]$, где каждый элемент f_i представляет собой среднюю степень принадлежности гранулы к соответствующему лингвистическому терму (например, f_1 - «холодная зима», f_2 - «умеренная зима», f_3 - «тёплая зима», f_4 - «холодная весна» и т.д. по порядку сезонов и температурных термов). Этот вектор F является формализованным признаковым описанием гранулы, количественно отражающим её неопределённость и недельную изменчивость. Он непосредственно используется для активации продукционных правил из таблицы 2 в процессе нечёткого вывода.

Предложенный подход к гранулированию, основанный на методах нечёткой логики (*Алгоритм 4*), обеспечил сжатие данных – объём выборки сократился с 52 534 до 7 504 точек (коэффициент сжатия $\approx 85\%$). Это реализовано путём вычисления для каждой недельной гранулы вектора F из 12 признаков. Таким образом, вектор F формализует нечёткие характеристики исходного недельного интервала, включая неопределённость (значения признаков в $[0,1]$) и изменчивость (различное распределение признаков у разных гранул).

4 Пример моделирования биоклиматического потенциала

Для проверки работоспособности предложенных подходов разработан прототип СППР для моделирования биоклиматического потенциала (БКП) экспертами предметной области. Фрагмент анализа БКП представлен на рисунке 8. В таблице 2 представлен фрагмент структуры базы знаний, которая включает набор продукционных правил, условий логического вывода и адаптивные интервалы для функций принадлежности, которые можно изменять в зависимости от региона и климата.

Функции принадлежности для сезонных параметров заданы в виде гауссовских кривых, определяемых параметрами среднего значения (среднее) и стандартного отклонения (σ). Поскольку данные функции асимптотически стремятся к нулю, дополнительно вводятся ограниченные интервалы, в пределах которых рассматривается их ненулевая часть.

Сезоны представлены в виде числовых значений: «Зима» – 0, «Весна» – 1, «Лето» – 2, «Осень» – 3 и формализованы через полуинтервал $[0, 4)$, где: $[0,1)$ соответствует зиме, $[1,2)$ – весне, $[2,3)$ – лету, $[3,4)$ – осени. Такое представление обеспечивает математическую корректность работы непрерывных функций принадлежности при переходе между сезонами, исключая неоднозначности на границах годового цикла.

В реализованном прототипе СППР применён алгоритм нечёткого вывода на основе заданных параметров модели (сезон и температура) в соответствии с *Алгоритмом 4* (фаззификация $\rightarrow$ агрегирование $\rightarrow$ дефаззификация) и последующим логическим выводом на основе продукционных правил.

Пример правила R_4 (в таблице 2)

Условие: температура: $[5, 25]$ (гауссова функция, среднее = 10, $\sigma = 5$) – «умеренная»); сезон: $[2.5, 3.5]$ (гауссова функция среднее = 3, $\sigma = 0.5$) – «осень»).

Вывод: «БКП средний».

Логика: Умеренная температура в осенний сезон чаще даёт «средний БКП».

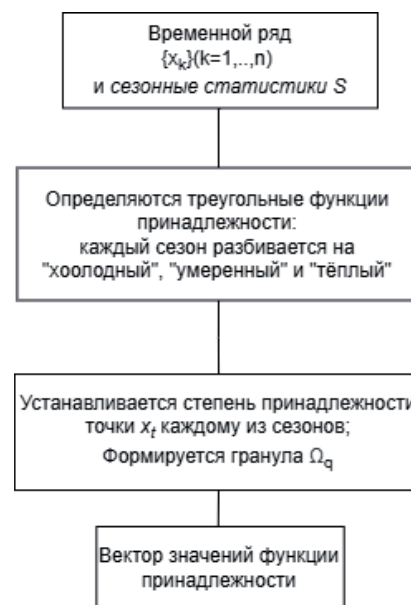


Рисунок 7 – Алгоритм гранулирования на основе нечёткого представления

Таблица 2 – Фрагмент базы правил моделирования биоклиматического потенциала (БКП)

Правило	Условие	Вывод	Интервалы и функции принадлежности
1	Если температура холодная И сезон зима	БКП низкий	Температура: $[-10, 10]$ (гауссова функция, среднее = -5 , $\sigma = 5$) Сезон: $[0, 1]$ (гауссова функция, среднее = 0 , $\sigma = 0.5$)
2	Если температура умеренная И сезон весна	БКП средний	Температура: $[5, 25]$ (гауссова функция, среднее = 10 , $\sigma = 5$) Сезон: $[0.5, 1.5]$ (гауссова функция, среднее = 1 , $\sigma = 0.5$)
3	Если температура тёплая И сезон лето	БКП высокий	Температура: $[20, 40]$ (гауссова функция, среднее = 25 , $\sigma = 5$) Сезон: $[1.5, 2.5]$ (гауссова функция, среднее = 2 , $\sigma = 0.5$)
4	Если температура умеренная И сезон осень	БКП средний	Температура: $[5, 25]$ (гауссова функция, среднее = 10 , $\sigma = 5$) Сезон: $[2.5, 3.5]$ (гауссова функция, среднее = 3 , $\sigma = 0.5$)
5	Если температура холодная И сезон осень	БКП низкий	Температура: $[-10, 10]$ (гауссова функция, среднее = -5 , $\sigma = 5$) Сезон: $[2.5, 3.5]$ (гауссова функция, среднее = 3 , $\sigma = 0.5$)

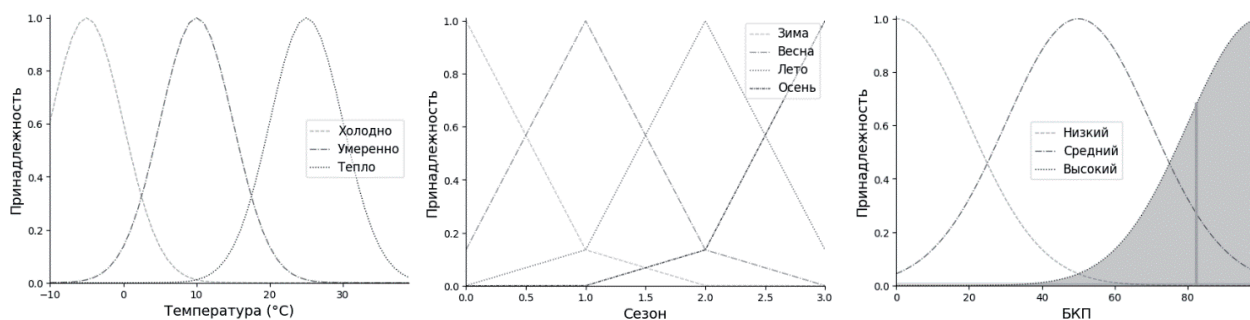


Рисунок 8 – Фрагмент анализа биоклиматического потенциала

Процесс нечёткого вывода БКП (на основе вектора признаков F) включает следующие шаги:

1) **Активация правил:** каждое продукционное правило из таблицы 2 активируется путём выбора соответствующего элемента (или комбинации элементов) из вектора F , который представляет среднюю степень принадлежности гранулы к лингвистическому терму условия правила.

- Пример активации правила R_4 (таблица 2): («Если Температура умеренная И Сезон осень → БКП средний») условие соответствует лингвистическому терму «умеренная осень». В векторе признаков F этот терм представлен отдельным элементом (f_{11} в предположенном порядке термов). Значение этого элемента (средняя степень принадлежности гранулы к терму «умеренная осень») непосредственно определяет степень истинности условия правила R_4 .
- Степень истинности антецедента (условия) правила рассчитывается как значение выбранного элемента вектора F (для правил с одним простым условием). Для правил с оператором «И» между несколькими элементами (см. реализацию ниже) степень истинности вычислялась бы как агрегация (например, $\min$ или произведение) соответствующих элементов F .

2) **Активизация заключений:** полученная для каждого правила степень истинности используется для «обрезки» (или масштабирования) функции принадлежности выходного термина БКП, указанного в заключении этого правила.

3) **Агрегирование выходов:** активизированные выходные термы БКП (результаты всех сработавших правил) объединяются в одно нечёткое множество с помощью операции максимума ($\max$).

4) **Дефаззификация:** к итоговому агрегированному нечёткому множеству выходной переменной (БКП) применяется метод центра тяжести для получения чёткого числового значения или категории биоклиматического потенциала.

Реализация правил вывода БКП с использованием вектора F (scikit-fuzzy):

В прототипе СППР с использованием библиотеки *scikit-fuzzy* продукционные правила активируются путём проверки значений конкретных элементов вектора F относительно за-

данных порогов активации θ_i . Упрощённый вид реализованных правил соответствует лингвистическим правилам таблицы 2:

- | |
|---|
| R_1 : Если $F[\text{"холодная зима"}] > \theta_1$ ТО БКП низкий; // Условие: Холодная Зима |
| R_2 : Если $F[\text{"умеренная весна"}] > \theta_2$ ТО БКП средний; // Условие: Умеренная Весна |
| R_3 : Если $F[\text{"тёплая лето"}] > \theta_3$ ТО БКП высокий; // Условие: Тёплое Лето |
| R_4 : Если $F[\text{"умеренная осень"}] > \theta_4$ ТО БКП средний; // Условие: Умеренная Осень |
| R_5 : Если $F[\text{"холодная осень"}] > \theta_5$ ТО БКП низкий; // Условие: Холодная Осень, |

где θ_i - пороговые значения, определяющие минимальную степень принадлежности для активации соответствующего правила (в коде θ_1 - θ_5).

Пример активации правил.

Пусть предполагаемый порядок термов имеет вид в первой грануле с вектором: $F = features[0] = [0.087, 0.534, 0.377, 0.510, 0.489, 0.0, 0.0, 0.0, 0.0, 0.584, 0.415, 0.0]$, где наибольшие значения имеют индексы 1, 3, 4, 9 и 10:

- $F[\text{"умеренная зима"}]$ (индекс 1) = 0.534
- $F[\text{"холодная весна"}]$ (индекс 3) = 0.510
- $F[\text{"умеренная весна"}]$ (индекс 4) = 0.489
- $F[\text{"холодная осень"}]$ (индекс 9) = 0.584
- $F[\text{"умеренная осень"}]$ (индекс 10) = 0.415

При порогах, например, $\theta_1=0.5$, $\theta_2=0.4$, $\theta_4=0.4$, $\theta_5=0.5$, активируются правила: R_1 (Умеренная Зима: $0.534 > 0.5$); R_2 (Умеренная Весна: $0.489 > 0.4$); R_4 (Умеренная Осень: $0.415 > 0.4$); R_5 (Холодная Осень: $0.584 > 0.5$).

Итоговый вывод БКП будет определяться степенями истинности активированных правил (R_1 , R_2 , R_4 , R_5) и методом дефаззификации (центра тяжести).

Предлагаемая структура продукционных правил обладает модульной архитектурой, позволяющей адаптировать параметры моделирования к различным климатическим условиям. Адаптация осуществляется путём:

- *корректировки функций принадлежности*: изменение гауссовских (сезон) и треугольных (температура) функций на этапе фаззификации (Алгоритм 4); модификация границ термов БКП;
- *динамической настройки базы правил*: добавление/изменение/удаление продукционных правил; изменение весовых коэффициентов; настройка порогов активации θ_i для элементов вектора F ;
- *регулировки диапазонов входных переменных*.

Заключение

В работе предложен подход к обработке данных метеорологических временных рядов, основанный на методах нечёткой логики. Ключевыми результатами исследования являются:

- *Алгоритм нечёткого гранулирования*. Алгоритм 4 – гибридный алгоритм, использующий треугольные функции принадлежности для температурных данных и гауссовы функции для сезонных параметров. Алгоритм преобразует исходные чёткие данные в нечёткие гранулы, представленные 12-мерными векторами признаков F . Эти векторы количественно отражают характерную неопределённость и недельную изменчивость исходных температурных колебаний.
- *Сжатие данных*. Применение алгоритма гранулирования позволило сократить объём исходных данных с 52 534 до 7 504 точек (коэффициент сжатия $\approx 85\%$) при сохранении ключевых статистических характеристик и семантического содержания данных, воплощённого в векторах F .
- *СППР для моделирования БКП*. СППР использует формализованные гранулы (векторы F) для активации набора адаптивных продукционных правил нечёткого вывода, реализующих экспертные знания о взаимосвязи температуры, сезона и БКП. Модульная

архитектура СППР может позволить настраивать функции принадлежности, пороги активации правил и саму базу правил под различные климатические условия.

Список источников

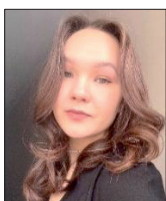
- [1] **Zhao J., Guo Z.H., Su Z.Y., Zhao Z.Y., Xiao X., Liu F.** An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed. *Applied Energy*. 2016. 162. P.808-826. DOI: 10.1016/J.APENERGY.2015.10.145.
- [2] **Kavousi-Fard A., Khosravi A., Nahavandi S.** A New Fuzzy-Based Combined Prediction Interval for Wind Power Forecasting. *IEEE Transactions on Power Systems*. 2016. V.31. №1. P.18–26. DOI: 10.1109/TPWRS.2015.2393880.
- [3] **Wen J., Yang J., Jiang B., Song H., Wang H.** Big Data Driven Marine Environment Information Forecasting: A Time Series Prediction Network. *IEEE Transactions on Fuzzy Systems*. 2021. V.29. №1. P.4-18. DOI: 10.1109/TFUZZ.2020.3012393.
- [4] **Domanska D., Wojtylak M.** Application of fuzzy time series models for forecasting pollution concentrations. *Expert Systems with Applications*. 2012. V. 39. № 9. P.7673-7679. DOI: 10.1016/j.eswa.2012.01.023.
- [5] **Kumar S., Gangwar S.** Intuitionistic Fuzzy Time Series: An Approach for Handling Nondeterminism in Time Series Forecasting. *IEEE Transactions on Fuzzy Systems*. 2016. V.24. №6. P.1270-1281. DOI: 10.1109/TFUZZ.2015.2507582.
- [6] **Чечнев В.Б.** Анализ и классификация многокритериальных методов принятия решений. *Онтология проектирования*. 2024. Т.14, №4(54). С.607-624. DOI 10.18287/2223-9537-2024-14-4-607-624.
- [7] **Колоденкова А.Е., Верещагина С.С., Фаворская Е.А., Осипова Е.А.** Подход к оценке технического состояния электрооборудования с использованием взвешенных нечётких правил. *Онтология проектирования*. 2024. Т.14, № 1(51). С.134-144. DOI 10.18287/2223-9537-2024-14-1-134-144.
- [8] **Бисочков Р.М., Суханова С.Ф.** Прогнозирование урожайности сельскохозяйственных культур методами нечёткой логики. *Вестник Башкирского государственного аграрного университета*. 2023. №1(65). С.119-125. DOI 10.31563/1684-7628-2023-65-1-119-126.
- [9] **Архипов М.В., Ермаков Р.Н.** Оценка рисков снижения полевой всхожести при посеве зерновых культур с учетом степени кондиционности семенного материала на основе моделей нечёткой логики. *Известия Санкт-Петербургского государственного аграрного университета*. 2015. № 39. С.68-70.
- [10] **Taylan O., Alamoudi R., Kabli M., AlJifri A., Ramzi F., Herrera-Viedma E.** Assessment of Energy Systems Using Extended Fuzzy AHP, Fuzzy VIKOR, and TOPSIS Approaches to Manage Non-Cooperative Opinions. *Sustainability*. 2020. V.12. №7. P.2745. DOI: 10.3390/su12072745.
- [11] **Shanchao Y., Jing L.** Time-Series Forecasting Based on High-Order Fuzzy Cognitive Maps and Wavelet Transform. *IEEE Transactions on Fuzzy Systems*. 2018. V.26. №6. P. 3391-3402. DOI: 10.1109/TFUZZ.2018.2831640.
- [12] **Khosravi A., Nahavandi S.** Load Forecasting Using Interval Type-2 Fuzzy Logic Systems: Optimal Type Reduction. *IEEE Transactions on Industrial Informatics*. 2014. V.10. №2. P.1055-1063. DOI: 10.1109/TII.2013.2285650.
- [13] **Rahman N.H.A., Muhammad H., Suhartono, Mohd Talib.** Artificial neural networks and fuzzy time series forecasting: an application to air quality. *Quality & Quantity*. 2015. V.49. №6. P.2633-2647. DOI: 10.1007/s11135-014-0132-6.
- [14] **Sulandari W., Subanar, Lee M.H., Rodrigues P.C.** Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks. *Energy*. 2020. V.190. P.116408. DOI: 10.1016/j.energy.2019.116408.
- [15] **Herbst G., Bocklisch S.F.** Online Recognition of fuzzy time series patterns. 2009 *International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference (IFSA-EUSFLAT 2009)*. 2009. P.974-979.
- [16] **Giove S.** Fuzzy logic and Clustering methods for time series analysis. 2009 *International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference (IFSA-EUSFLAT 2009)*. 2009. P.1-7.
- [17] **Graves D., Pedrycz W.** Multivariate Segmentation of Time Series with Differential Evolution. 2009 *International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference (IFSA-EUSFLAT 2009)*. 2009. P.1108-1113.
- [18] **Kacprzyk J., Wilbik A.** Using Fuzzy Linguistic summaries for the comparison of time series. 2009 *International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference (IFSA-EUSFLAT 2009)*. 2009. P.1321-1326.

- [19] **Jingjing F., Qingfei Z., Zhonglin Z.** A Method of Mining the Meta-association Rules for Dynamic Association Rule Based on the Model of AR-Markov, *2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing*, Wuhan, China, 2010. P.210-214, DOI: 10.1109/NSWCTC.2010.248.
- [20] **Wu F., Li Y., Yu F.** Fuzzy Granulation Based Forecasting of Time Series. In *Fuzzy Information and Engineering 2010. Advances in Intelligent and Soft Computing*, Springer, Berlin, Heidelberg. 2010. V.78. P.511-520. DOI: 10.1007/978-3-642-14880-4_56.
- [21] **Guo H., Wang L., Liu X., Pedrycz W.** Information Granulation-Based Fuzzy Clustering of Time Series. *IEEE Transactions on Cybernetics*. 2021. V.51. №12. P. 6253-6261. DOI: 10.1109/TCYB.2020.2970455.
- [22] **Abonyi J., Feil B., Nemeth S., Arva P.** Fuzzy clustering based segmentation of time-series. In *Advances in Intelligent Data Analysis V: 5th International Symposium on Intelligent Data Analysis, IDA 2003, Berlin, Germany, August 28-30, 2003. Proceedings 5*. P.275-285. DOI: 10.1007/978-3-540-45231-7_26.
- [23] **Hannah M.E., Geetha T.V., Mukherjee S.** Automatic Extractive Text Summarization Based on Fuzzy Logic: A Sentence Oriented Approach. In *Swarm, Evolutionary, and Memetic Computing. SEMCCO 2011. Visakhapatnam, Andhra Pradesh, India, December 19-21, 2011, Proceedings, Part I 2* P.530-538. Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-27172-4_63.
- [24] **Бурнашев Р.А., Байбикова Д.Ф., Крылов М.А., Сергеев Я.В., Ибрагимов Р.М.** Свидетельство о государственной регистрации программы для ЭВМ №2025617003. Программа экспертного анализа климатических и экономических взаимосвязей. Заявка №2025614348. Дата поступления 5 марта 2025 г. Зарегистрировано в Реестре программ для ЭВМ 21 марта 2025 г.
- [25] **Бурнашев Р.А., Назипова А.Ф.** Свидетельство о государственной регистрации программы для ЭВМ №2025618157. Программа для анализа и гранулирования нечётких временных рядов. Заявка №2025615912. Дата поступления 18 марта 2025 г. Зарегистрировано в Реестре программ для ЭВМ 02 апреля 2025 г.
- [26] **Burnashev R.A., Bagymanov R.M., Enikeeva A.I., Farahov R.R., Bolsunovskaya M.V.** Designing a Medical Fuzzy Expert System Using the JFuzzyLogic Library. *IEEE 25th International Conference of Young Professionals in Electron Devices and Materials (EDM)*, Altai, Russian Federation. 2024. P.2330-2333. DOI: 10.1109/EDM61683.2024.10615193.

Сведения об авторах



Бурнашев Рустам Арифович (1989 г. рождения). Окончил КФУ (2019). К.т.н. (2020). Доцент кафедры анализа данных и технологий программирования Института вычислительной математики информационных технологий КФУ. В списке научных трудов более 55 научных работ в области систем поддержки принятия решения, анализа данных, медицинских информационных систем. Author ID (РИНЦ): 955990; Author ID (Scopus): 57200246615; ORCID: 0000-0002-1057-0328. r.burnashev@inbox.ru ✉.



Сергеев Ярослав Владиславович (2001 г. рождения). Магистрант Института вычислительной математики и информационных технологий КФУ, направление «Прикладная математика и информатика». ORCID: 0009-0009-3370-1464. sergeevyarik7@yandex.ru.



Назипова Алиса Фанисовна (2000 г. рождения). Магистрант Института вычислительной математики и информационных технологий КФУ, направление «Прикладная математика и информатика». ORCID: 0009-0001-0560-1425. alisa.nazipova@yandex.ru.

Поступила в редакцию 17.03.2025, после рецензирования 03.07.2025. Принята к публикации 10.07.2025.



Fuzzy time series granulation methods for data analysis

© 2025, R.A. Burnashev ✉, Y.V. Sergeev, A.F. Nazipova

Kazan (Volga region) Federal University,
Institute of Computational Mathematics and Information Technology, Kazan, Russia

Abstract

The growing dimensionality of data, driven by the multitude of heterogeneous time series, requires the development of efficient methods for their processing and compression. This article presents an approach to data compression where the data are represented as time series, using granulation with fuzzy logic methods. The study analyzes average daily temperature data in the Republic of Tatarstan collected from 1881 to 2024. Data granulation enabled a significant compression of the data volume. Fuzzy summarization was applied to transform the original numerical data into information granules, facilitating the automatic generation of granular descriptions of time series behavioral patterns. The summarization of time series states was carried out using fuzzy logic methods, including a rule set, membership functions for each season, interval-based linguistic variables, and a defuzzification software module. The implementation of the proposed approach demonstrated a reduction in data volume from 52,534 to 7,504 points, achieving a compression ratio of approximately 85%. The developed methods are applicable for analyzing large datasets across various domains.

Keywords: data analysis, fuzzy logic, knowledge base, fuzzy summarization, granulation, time series.

For citation: Burnashev RA, Sergeev YV, Nazipova AF. Fuzzy time series granulation methods for data analysis [In Russian]. *Ontology of designing*. 2025; 15(3): 404-417. DOI: 10.18287/2223-9537-2025-15-3-404-417.

Financial Support: This paper is performed as part of the grant of the Tatarstan Academy of Sciences, provided to young candidates of sciences (postdoctoral fellows) for the purpose of defending their doctoral dissertation, conducting research, as well as performing their work duties in scientific and educational organizations of the Republic of Tatarstan within the framework of the State Program of the Republic of Tatarstan "Scientific and Technological Development of the Republic of Tatarstan" (Agreement No.04/2024-PD).

Authors' contributions: Burnashev R.A. – article concept and algorithms development; Sergeev Y.V. – mathematical model construction, fuzzy rules testing; Nazipova A.F. – meteorological data analysis and processing, time series granulation, results visualization.

Conflict of interest: The authors declare no conflict of interest.

List of figures and tables

Figure 1 – Stages of data analysis

Figure 2 – Algorithm flowchart (a) and the result of granulation based on clustering using the k-means algorithm (b) (temperature time series from 01.01.2018 to 01.01.2020, with granule boundaries marked by dashed vertical lines)

Figure 3 – Adaptive gradient-based granulation algorithm

Figure 4 – Gradient-based time series granulation (granule boundaries marked by dashed vertical lines)

Figure 5 – Granulation algorithm with fixed time windows

Figure 6 – Granulation with fixed weekly time windows

Figure 7 – Granulation algorithm based on fuzzy representation

Figure 8 – Fragment of bioclimatic potential analysis

Table 1 – Operations in data granulation

Table 2 – Fragment of the rule base for modeling bioclimatic potential

References

- [1] **Zhao J, Guo ZH, Su ZY, Zhao ZY, Xiao X, Liu F.** An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed. *Applied Energy*. 2016; 162: 808-826. DOI: 10.1016/J.APENERGY.2015.10.145.
- [2] **Kavousi-Fard A., Khosravi A., Nahavandi S.** A New Fuzzy-Based Combined Prediction Interval for Wind Power Forecasting. *IEEE Transactions on Power Systems*. 2016; 31(1): 18–26. DOI: 10.1109/TPWRS.2015.2393880.
- [3] **Wen J, Yang J, Jiang B, Song H, Wang H.** Big Data Driven Marine Environment Information Forecasting: A Time Series Prediction Network. *IEEE Transactions on Fuzzy Systems*. 2021; 29(1): 4-18. DOI: 10.1109/TFUZZ.2020.3012393.
- [4] **Domanska D, Wojtylak M.** Application of fuzzy time series models for forecasting pollution concentrations. *Expert Syst. Appl.* 39, 7673-7679. 2012. DOI: 10.1016/j.eswa.2012.01.023.
- [5] **Kumar S, Gangwar S.** Intuitionistic Fuzzy Time Series: An Approach for Handling Nondeterminism in Time Series Forecasting. *IEEE Transactions on Fuzzy Systems*. 2016; 24(6): 1270-1281. DOI: 10.1109/TFUZZ.2015.2507582.
- [6] **Chechnev VB.** Analysis and classification of multi-criteria decision-making methods [In Russian]. *Ontology of Designing*. 2024; 14(4): 607-624. DOI: 10.18287/2223-9537-2024-14-4-607-624.
- [7] **Kolodenkova AE, Vereshchagina SS, Favorskaya EA, Osipova EA.** An approach to assessing the technical condition of electrical equipment using weighted fuzzy rules [In Russian]. *Ontology of Designing*. 2024; 14(1): 134-144. DOI: 10.18287/2223-9537-2024-14-1-134-144.
- [8] **Bischokov RM, Sukhanova SF.** Forecasting crop yields using fuzzy logic methods [In Russian]. *Bulletin of Bashkir State Agrarian University*. 2023; 1(65): 119-125. DOI: 10.31563/1684-7628-2023-65-1-119-126.
- [9] **Arkhipov MV, Ermakov RN.** Risk assessment of reducing field germination when sowing grain crops taking into account the quality of seed material based on fuzzy logic models [In Russian]. *News of St. Petersburg State Agrarian University*. 2015; 39: 68-70.
- [10] **Taylan O, Alamoudi R, Kabli M, AlJifri A, Ramzi F, Herrera-Viedma E.** Assessment of Energy Systems Using Extended Fuzzy AHP, Fuzzy VIKOR, and TOPSIS Approaches to Manage Non-Cooperative Opinions. *Sustainability*. 2020; 12(7): 2745. DOI: 10.3390/su12072745.
- [11] **Shanchao Y, Jing L.** Time-Series Forecasting Based on High-Order Fuzzy Cognitive Maps and Wavelet Transform. *IEEE Transactions on Fuzzy Systems*. 2018; 26(6): 3391-3402. DOI: 10.1109/TFUZZ.2018.2831640.
- [12] **Khosravi A, Nahavandi S.** Load Forecasting Using Interval Type-2 Fuzzy Logic Systems: Optimal Type Reduction. *IEEE Transactions on Industrial Informatics*. 2014; 10(2): 1055-1063. DOI: 10.1109/TII.2013.2285650.
- [13] **Rahman NHA, Muhammad H, Suhartono, Mohd Talib.** Artificial neural networks and fuzzy time series forecasting: an application to air quality. *Quality & Quantity*. 2015; 49(6): 2633-2647. DOI: 10.1007/s11135-014-0132-6.
- [14] **Sulandari W, Subanar, Lee MH, Rodrigues PC.** Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks. *Energy*. 2020; 190: 116408. DOI: 10.1016/j.energy.2019.116408.
- [15] **Herbst G, Bocklish SF.** Online Recognition of fuzzy time series patterns. *2009 International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference (IFSA-EUSFLAT 2009)*. 2009. P.974-979.
- [16] **Giove S.** Fuzzy logic and Clustering methods for time series analysis. *2009 International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference (IFSA-EUSFLAT 2009)*. 2009. P.1-7.
- [17] **Graves D, Pedrycz W.** Multivariate Segmentation of Time Series with Differential Evolution. *2009 International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference (IFSA-EUSFLAT 2009)*. 2009. P.1108-1113.
- [18] **Kacprzyk J, Wilbik A.** Using Fuzzy Linguistic summaries for the comparison of time series. *2009 International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference (IFSA-EUSFLAT 2009)*. 2009. P.1321-1326.
- [19] **Jingjing F., Qingfei Z., Zhonglin Z.** A Method of Mining the Meta-association Rules for Dynamic Association Rule Based on the Model of AR-Markov, *2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing*, Wuhan, China. 2010. P.210-214. DOI: 10.1109/NSWCCTC.2010.248.
- [20] **Wu F, Li Y, Yu F.** Fuzzy Granulation Based Forecasting of Time Series. In *Fuzzy Information and Engineering 2010. Advances in Intelligent and Soft Computing*, V.78. Springer, Berlin, Heidelberg. 2010. P.511-520. DOI: 10.1007/978-3-642-14880-4_56.
- [21] **Guo H, Wang L, Liu X, Pedrycz W.** Information Granulation-Based Fuzzy Clustering of Time Series. *IEEE Transactions on Cybernetics*. 2021. V.51. №12. P. 6253-6261. DOI: 10.1109/TCYB.2020.2970455.

- [22] **Abonyi J, Feil B, Nemeth S, Arva P.** Fuzzy clustering based segmentation of time-series. In *Advances in Intelligent Data Analysis V: 5th International Symposium on Intelligent Data Analysis, IDA 2003, Berlin, Germany, August 28-30, 2003. Proceedings 5*. P.275-285. DOI: 10.1007/978-3-540-45231-7_26.
- [23] **Hannah ME, Geetha TV, Mukherjee S.** Automatic Extractive Text Summarization Based on Fuzzy Logic: A Sentence Oriented Approach. In *Swarm, Evolutionary, and Memetic Computing. SEMCCO 2011. Visakhapatnam, Andhra Pradesh, India, December 19-21, 2011, Proceedings, Part I 2* (p.530-538). Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-27172-4_63.
- [24] **Burnashev RA, Baibikova DF, Krylov MA, Sergeev YaV, Ibragimov RM.** Certificate of state registration of computer program No.2025617003. Program for expert analysis of climate and economic relationships [In Russian]. Application No.2025614348. Received March 5, 2025. Registered in the Computer Programs Register March 21, 2025.
- [25] **Burnashev RA, Nazipova AF.** Certificate of state registration of computer program No.2025618157. Program for analysis and granulation of fuzzy time series [In Russian]. Application No.2025615912. Received March 18, 2025. Registered in the Computer Programs Register April 2, 2025.
- [26] **Burnashev RA, Bagymanov RM, Enikeeva AI, Farahov RR, Bolsunovskaya MV.** Designing a Medical Fuzzy Expert System Using the JFuzzyLogic Library. *IEEE 25th International Conference of Young Professionals in Electron Devices and Materials (EDM)*, Altai, Russian Federation. 2024. P.2330-2333. DOI: 10.1109/EDM61683.2024.10615193.

About the authors

Rustam Arifovich Burnashev (b. 1989) graduated from Kazan Federal University (2019), PhD (2020). He is an associate professor of the Department of data analysis and programming technologies of Institute of computational mathematics and information technologies at Kazan Federal University. The list of scientific papers includes more than 55 scientific papers in the field of decision support systems, data analysis, and medical information systems. Author ID (РИНЦ): 955990; Author ID (Scopus): 57200246615; ORCID: 0000-0002-1057-0328. r.burnashev@inbox.ru ✉.

Yaroslav Vladislavovich Sergeev (b. 2001) is a first-year master's student at the Institute of Computational Mathematics and Information Technologies at Kazan Federal University in the field of Applied Mathematics and Informatics. ORCID: 0009-0009-3370-1464. sergeevyarik7@yandex.ru.

Alisa Fanisovna Nazipova (b. 2000) is a second-year master's student at the Institute of Computational Mathematics and Information Technologies at Kazan Federal University in the field of Applied Mathematics and Informatics. ORCID: 0009-0001-0560-1425. alisa.nazipova@yandex.ru.

Received March 17, 2025. Revised July 3, 2025. Accepted July 10, 2025.