DOI: 10.6084/m9.figshare.9783212
УДК 517.11+519.92

Минаева Юлия Ивановна
Кандидат технических наук, доцент, *orcid.org/0000-0001-9548-1959*
Киевский национальный университет строительства и архитектуры, Киев
Минаев Юрий Николаевич
Доктор технических наук, профессор
Национальный авиационный университет, Киев
Филимонова Оксана Юрьевна
Кандидат технических наук, доцент, *orcid.org/0000-0001-9548-1959*
Киевский национальный университет строительства и архитектуры, Киев

ПРЕДСТАВЛЕНИЕ НЕОПРЕДЕЛЕННОСТИ НА УРОВНЕ ТЕНЗОРНЫХ МОДЕЛЕЙ ДАННЫХ

***Аннотация**. Теория нечетких множеств, являющаяся эффективным аппаратом для решения задач в условиях неопределенности, в ряде случаев может иметь определенные ограничения, обусловленные сложностью технологического процесса (или объекта управления), высокой размерностью и суперогромным объемом исходного множества данных, усугубляемой пропуском данных, противоречивостью требований, предъявляемых к функции принадлежности и др. Предложен способ построения подмножества упорядоченных пар, который может быть использован вместо функции принадлежности (при невозможности его формирования) при решении задачи в условиях неопределенности или как возможность получения предварительного решения, позволяющего оценить рациональность экспертно сформированного нечеткого множества (функции принадлежности). В основу построения подмножества упорядоченных пар положена нечетко-множественная интерпретация сингулярной декомпозиции преобразованного в 2D-тензор исходного множества данных. Приведены примеры, иллюстрирующие эффективность предложенного метода.*

***Ключевые слова**: неопределенность; нечеткое множество; функция принадлежности; тензор; тензорная декомпозиция; Фробениусова норма; тензорная модель данных; подмножество упорядоченных пар*

Введение

Современная наука, занимающаяся проблемами управления в условиях неопределенности, столкнулась в последнее 10-летие с проблемой больших данных (BIG DATA-Big DATA), смысл которой может быть определен как «…обозначение структурированных и неструктурированных данных огромных объёмов и значительного многообразия, эффективно обрабатываемых горизонтально масштабируемыми программными инструментами, появившимися в конце 2000-х годов и альтернативных традиционным системам управления базами данных и решениям класса Business Intelligence». Основными характеристиками больших данных были признаны так называемые ”4V” – объем, скорость, разнообразие, достоверность.

К серии технологий стали относить разнообразные информационно-технологические решения, в той или иной степени обеспечивающие сходные по характеристикам возможности по обработке сверхбольших массивов данных.

Кроме названных *технических* характеристик, большие данные обладают рядом специфических качеств и свойств, которые делают обработку таких данных проблематичной. В соответствии с работой [1], укажем, что большие данные *могут быть переполнены ошибками или шумом, следовательно их анализ может стать беспредметным*.

Изложение основного материала

Нечеткие множества и проблема больших данных

Способность нечетких множеств (НМ) представлять и количественно оценивать массивы данных неопределенности может быть использована для обработки BIG DATA. Текущее состояние исследований и перспективы дальнейших разработок представлены в работе [2], в частности отмечено, что более сложные дополнения НМ и их интеграция с другими инструментами могли бы предложить новую перспективную среду обработки.

В работе [2] показано, что неопределенности не только существуют в самих данных, но происходят на каждой фазе обработки BIG DATA. Методы НМ принципиально могут быть одними из наиболее эффективных инструментов для обработки различных типов неопределенностей. Обработка неопределенностей связана с различными потребностями, задачами и финансовыми возможностями. Рационально использовать методы НМ и другие методы гранулярного компьютинга (ГрК) [3], чтобы реконструировать проблему на определенном гранулярном уровне

Анализ тенденции роли нечетких множеств в комплексе проблем BIG DATA, выполненный в [2], показал, что наиболее часто применяются интеллектуальные алгоритмы, основанные на НМ. Однако алгоритмы на основе НМ в силу их простоты не являются достаточно удобными для представления неопределенностей в комплексных задачах обработки и анализа BIG DATA. Согласно [2] решающей тенденцией в анализе и обработке BIG DATA будет использование сложных моделей, таких как нечеткие множества типа 2 и нечеткие грубые множества, т.к., например, НМ 2 типа могут моделировать сложные случаи неопределенностей, т.е. моделировать многомерную неопределенность.

В работе [4] утверждается, что ТНМ в общей проблеме BIG DATA будет эффективной только тогда, когда она может касаться задач извлечения знаний. В основном это связано с преимуществами, вытекающими из извлечения знаний из большого объема информации. Учитывая, что проблема BIG DATA решается комплексно с использованием программно-аппаратных средств, среди различных подходов, используемых в Data Mining, модели, основанные на нечетких системах, в среде MapReduce/Hadoop-моделей и их различных расширений, могут быть достаточно эффективными. Среди преимуществ нечетких моделей следует подчеркнуть использование представления, близкого к естественному языку. Кроме того, они используют модель вывода, которая позволяет адаптироваться к различным ситуациям. Несмотря на успех такого рода систем, миграция нечетких моделей в среду BIG DATA в разных областях находится на предварительном этапе.

В работе [5] показано, что теория НМ и более конкретно интуиционистские НМ обеспечивают эффективное решение для моделирования с учетом неточностей в реляционных базах данных. В работах, опубликованных за последние 30 лет, использовали ТНМ, с целью расширения модели реляционных данных для обеспечения представления и поиска неточных данных. Однако такие подходы не были рассчитаны на масштабирование до уровня супербольших наборов данных. В работе [6] представлен комплекс задач анализа и обработки BIG DATA, которые могут быть успешно решены при использовании методов и алгоритмов ТНМ, среди которых выделены следующие:

– классификация BIG DATA с нечеткими моделями;

– анализ надежности методов ТНМ при отсутствии или деформации данных.

Тензорные модели в обработке и анализе BIG DATA

Тензоры – это многомерные массивы, являющиеся естественным и компактным представлением огромных массивов многомерных данных с высокой мерностью. В работе [7] показана эффективность и рациональность представления таких массивов многомерных данных с помощью подходящих приближений низкого ранга. В работе [8] показано, что тензоры, как естественное расширение матриц и их разложений, предоставляют важные инструменты *современных подходов к крупномасштабным тензорным данным, что является важной частью BIG DATA*. Для наборов данных, собранных в многомерной форме, они могут представляться многопоточными массивами, которые называются ранг-1 тензорами. Тензорные разложения позволяют видеть различные структуры, лежащие в основе BIG DATA, и тем самым обеспечивают эффективные инструменты для анализа, сжатия и понимания данных.

Отдельно следует остановиться на модели жизненного цикла (ЖЦ) BIG DATA, т.к. на каждой фазе этого цикла способы анализа и обработки BIG DATA различны. Модель ЖЦ BIG DATA, принятая на основании работы [1], представляет собой структуру для формулировки задач и мероприятий, связанных с управлением данными в рамках проекта или организации. Лабораторная модель ЖЦ данных (DLCL) оперирует идеями проекта, данными, полученными измерениями, наблюдениями, экспериментами и моделированием и используется для генерации гипотез. При этом необработанные данные должны быть сохранены. В данной работе авторами особое внимание уделено фазе анализа данных, которая содержит такие процедуры, как хранение, компьютинг.

Известно, что матричные факторизации и их расширения на тензорные разложения стали основой для известных методов, используемых в интеллектуальном анализе данных (Brain Data Analysis) [9]. Более того, тензорные разложения имеют много других потенциальных применений, однако до настоящего времени не рассмотрены вопросы нечетко-множественной интерпретации тензорных декомпозиций, использование которой позволяет по-новому увидеть возможности ТНМ и ее использования для анализа BIG DATA.

Необходимость учета ограниченных возможностей назначения ФП

Теория нечетких множеств возникла как ответ на многочисленные запросы практики, вызванные необходимостью решения задач (прежде всего задач принятия решений (управление)) в условиях неопределенности [10]. Предложенный Л. Заде математический объект – подмножество упорядоченных пар (ПмУП) – $\left(x_1\mu^{x1};.....;x_n\mu^{xn}\right)$, 1-я компонента которого $\{x_i\} \in X \subset \nabla$ – универсальное множество (УМ), 2-ая – $\{\mu^{x_i}\} \to [0,1]$ – виртуальная значимость (членство) каждой отдельной компоненты УМ в составе, например, некоторого утверждения, оказалась (в значительной степени) универсальным.

Однако принятая конструкция, несмотря на широкую сферу применения, с самого начала вызвала ряд вопросов, ответы на которые не даны до настоящего времени. С тех пор, как Л. Заде [11] ввел понятие НМ, до настоящего времени [12; 13] одна из основных трудностей связана с функцией принадлежности (ФП), точнее, с возможными в целом ряде случаев ограничениями ее получения, что привело к тому, что за последние 50 лет сформулированы различные интерпретации ФП. В работе [13] приведена таксономия интерпретаций ФП для НМ. В зависимости от отдельной интерпретации метод извлечения ФП и теория, касающаяся приложений, являются различными.

Укажем, что под термином "интерпретация" скрывается не что иное, как изменение *точки зрения на объект*, роль которой убедительно показана в работе [14] (аномалии рационального поведения). Таким образом, ставится под сомнение любое решение, полученное на основании ТНМ *при выборе* и использовании *только одного* типа ФП, предложенной экспертом, хотя это решение может отвечать *реальности*.

Извлечение ФП на основе имеющихся данных – один из фундаментальных вопросов, связываемых с приложением ТНМ, в частности, относительно их применения в BIG DATA, т.к. нет руководящих принципов или правил, на основании которых можно обычно выбирать соответствующую генерацию техники принадлежности. Другая проблема, которая делает генерации ФП нетривиальным заданием, – недостаток согласованности в определении и интерпретации ФП, т.к. в зависимости от интерпретации можно рассматривать ряд методов генерирования величины принадлежности.

В частности, ряд источников ориентируют по отношению к определению ФП, только на отражение субъективных восприятий о неопределенных или неточных понятиях. К сожалению, эти методы не могут быть непосредственно приложены ко многим практическим проблемам, в частности анализу BIG DATA. Этот недостаток в большой степени устраняется методологией Канемана и Тверски (изменение точки зрения на объект), которая требует обязательного введения эталона, не зависящего от точки зрения. Такую возможность предоставляют массивы исходных данных, рассматриваемые как тензорные модели. При этом важно учитывать взаимодействие метода извлечения и предполагаемой интерпретации НМ. В целом, работая с неопределенностью, применяя для анализа ТНМ, даже в рамках решения одной задачи, следует учитывать наличие различных (иногда конкурирующих) интерпретаций НМ и методов их применения для практических приложений.

ТНМ, решив ряд задач, поставила новые, проблема решения которых не уступает по своей сложности исходной. Многие из нерешенных задач изложены в работе [15]. Несмотря на то, что уже в течении пятидесяти лет НМ и нечеткая логика убедительно показывают возможность обработки сложной и неопределенной информации через *призму человеческого знания и субъективность (здравый смысл)*, остро ощущается необходимость дальнейшего изучения мультимодального аспекта нечеткой логики из-за необходимости учета также и др. познавательных, эмоциональных и поведенческих сторон оценивания истинности. Отметим, что рассматриваемая в [15] методология, корреспондируется с предложенной в работе [14] теорией принятия решений, основанной на методологии аномалий рационального поведения парадигме смены точки зрения на объект. При этом новую точку зрения можно рассматривать как дополнительный источник знаний по отношению к ТНМ.

Постановка задачи

В ТНМ формирование НМ возможно на основании некоторого исходного множества данных (массива данных) – ИМД, неформальный анализ которого позволяет принять некоторую ФП и формальный анализ (как правило, статистический – min-max интервал значений) позволяет вычислить универсальное множество, ФП и УМ образуют подмножество упорядоченных пар – НМ. ПмУП является тем инструментарием, который формализует представление эксперта о данном объекте. При этом не всегда учитывается, что таких представлений может быть несколько, но математико-логический аппарат ТНМ, корректный в своей основе, позволяет работать только с одним объектом.

Если результат, по мнению эксперта, является *близким к реальности*, то решение считается удовлетворительным. Такой подход обладает некоторыми особенностями, которые могут в ряде случаев препятствовать принятию решения (или в более общем виде – обработке данных) в условиях неопределенности при помощи аппарата ТНМ. Такими особенностями являются:

– невозможность проверки результата и его переноса на новые, даже незначительно отличающиеся условия;

– в массивах данных, отличающихся высокой размерностью и суперогромным объемом данных (BIG DATA), где, кроме прочего, могут быть пропущены или деформированы данные, назначить ФП, аналогично тому, как это выполняется в одномерных задачах с обозримым объемом данных, невозможно;

– попытка применить цилиндрические расширения для одномерных ФП с целью получения трехмерных ФП, реализуема в ограниченном количестве случаев;

– появился ряд новых задач, отличающихся не только супербольшими объемами данных, но также имеющих новую форму хранения (жизненный цикл BIG DATA) – тензорные декомпозиции в виде ранг-1 тензоров, которые можно трактовать как подмножества упорядоченных последовательностей; в таких форматах формирование НМ как подмножеств упорядоченных пар наталкивается на серьезные (часто непреодолимые) трудности.

Главная цель работы – показать возможность формирования подмножества упорядоченных пар, обладающего свойствами НМ, на основе выявления скрытых знаний, присущих структурированному множеству исходных данных, исключив эвристики при назначении ФП, другими словами, фадзифицировать ИМД при помощи методов интеллектуального анализа в виде тензорного представления данных.

Основные случаи реализации процедуры фадзификации – формирования ПмУП (аналога НМ) $\tilde{a}=\{a/\mu^a\},\ a\in \mathrm{A}=\{a_c\},\ c=1,n;\ \mu^a \to [0,1]$, рассматриваемые в работе, в общем случае предполагают:

(а) исходное множество (массив) данных (ИМД) известно, выполняется фадзификация ИМД на основе правил ТНМ, назначается ФП и таким образом формируется НМ; параллельно выполняется тензоризация ИМД – формируется тензор $\mathbf{T}^a$, путем сингулярной декомпозиции формируется ПмУП, для НМ и ПмУП вычисляются *критериальные параметры* (Фробениусовы (F-) нормы и дефадзифицированные значения (ДЗ)), на основании которых выполняется сравнение указанных объектов с целью определения их эквивалентности (близость F-норм и ДЗ); при этом ИМД тензоризуется в 2D- и 3D-тензоры;

(б) известен только интервал, рассматриваемый как УМ, на котором сформировано эвристически НМ, выполняется погружение УМ (интервала, представленного в виде совокупности подинтервалов) в специальную матрицу (Теплица или Ганкеля), таким образом формируется тензор $\mathbf{T}^a$; далее реализуются процедуры п.(а).

Ограниченный объем статьи позволяет привести в данной работе только 1-й случай, хотя случаи, когда ИМД тензоризуется в 3D-тензор и фадзификация выполнена путем погружения УМ в специальную матрицу, также рассмотрены в отдельной статье.

Тензорное представление данных – многомерный (или одномерный) массив – это ИМД, которое после преобразования становится 2D- или 3D-тензорами. ТНМ предусматривает обязательный этап *фадзификации* данных – формирование НМ – {множество значений УМ / множество ФП}. Тензоризация ИМД дает возможность получить скрытые знания, позволяющие таким образом хотя бы оценить рациональность применения выбранного типа ФП.

Сформированное в процессе фадзификации новое формальное ПмУП – объект, в котором компоненты обладают свойствами:

– сформированное ПмУП близко (или совпадает) по критериальным показателям с ранее сформированным НМ;

– сформированное ПмУП позволяет применять математический и логический аппарат ТНМ без исключений;

– на основании сформированного ПмУП возможно восстановление ИМД;

– возможность использования скрытых знаний, содержащихся в ИМД.

Схема решения сформулированных задач:

– неопределенность (многомерный масив) $\to$ тензор $\to$ {совокупность ПмУП (или подмножество упорядоченных последовальностей)} $\to$ применение аппарата ТНМ.

Обратим внимание, что {совокупность ПмУП}, реализованная в виде Кронекерова произведения, должна асимптотически приближаться к исходной неопределенности (ИМД), представленной в виде тензора.

Еще раз отметим, что в общем случае, задача, сформулированная в работе, состоит в том, чтобы показать возможность применения нового подхода к решению задач в условиях неопределенности, состоящего в учете необходимости введения дополнительного узла контроля (или сравнения) решения, полученного на основании ТНМ (здравый смысл + корректная математика), с результатом формального решения, не основанного на использовании эксперта в процессе формирования НМ.

Нечетко-множественная интерпретация тензорных декомпозиций

Необходимость такой интерпретации тензорных декомпозиций обусловлена необходимостью поиска ближайших к заданному нечеткому множеству других нечетких множеств, аналогично тому, как выполнен поиск ближайшего четкого множества. Это принципиально, т.к. позволяет распространить результаты, полученные при использовании НМ с одной формой ФП на НМ с другой ФП, отличающейся от исходной (табл. 1).

В работе [16] рассмотрен вопрос определения обычного (четкого) *ближайшего* к нечеткому в смысле расположения на наименьшем расстоянии (евклидовом, в частности) от заданного нечеткого (пример из работы [16]. Другими словами, определено четкое множество с наименьшей нормой (по отношению к нечеткому), т.е.

$$\left\| \tilde{A} - \tilde{A}^{nearest} \right\|_2 \to 0,$$

где $\tilde{A}^{nearest} = \left\{ x_i \ \mu_{\tilde{A}nearest}(x_i) \right\}$.

Таким будет множество $\mu_{\tilde{A}^{nearest}}(x_i)$, обладающее свойствами (*): *например,* приводимое в работе [16] обычное (четкое) подмножество (ν^{χ}), ближайшее к нечеткому (μ^{χ}), $\tilde{x} = \{\mathrm{x}\ \mu^{x}\}$, $x_0^1 = \{\mathrm{x}\ \upsilon^{x}\}$, являются такими, что их Ф-нормы и дефадзифицированные значения практически совпадают:

$$\|\tilde{x}\|_F = 14.3871, \left\| x_0^1 \right\|_F = 14.3875;$$

$$def(\tilde{x}) = 4.6341;$$

$$def\left(x_0^1\right) = 4.6366.$$

С учетом того, что ПмУП ${}^{\pi}\chi = \left\{\chi, \xi^{\chi}\right\}$ с характеристической функцией ξ^{χ} имеет Ф-нормы:

$$\left\| {}^{\pi}x(0) \right\|_F = 14{,}2829;$$

$$\left\| {}^{\pi}x(1) \right\|_F = 14.5602;$$

$$\mathrm{def}(\mathrm{x}) = 4.500;$$

справедливо неравенство

$$\left\| {}^{\pi}x(0) \right\|_F < \|\tilde{x}\|_F < \left\| {}^{\pi}x(1) \right\|_F.$$

В связи с необходимостью вычисления НМ, ближайших к заданному нечеткому множеству, рассмотрим проблему ближайшего Кронекерова произведения как математического аппарата решения задач подобного типа. *Проблему ближайшего Кронекерова произведения* покажем так, как она представлена в работе [17]. Дано $\mathrm{A} \in \nabla^{m \times n}$ с $m = m_1 m_2$ и $n = n_1 n_2$. Найти $\mathrm{B} \in \nabla^{m1 \times n1}$ и $C \in \nabla^{m2 \times n2}$ такие, что $\phi(B, C) = \|\mathbf{A} - \mathbf{B} \otimes \mathbf{C}\|_F = min;$

Эта оптимизационная задача решается как билинейная проблема наименьших квадратов, фиксируя **B** (или **C**), проблема становится линейной по отношению к **C** (или **B**). Процедура решения: минимизировать

$$\phi(B, C) = \|\mathbf{A} - \mathbf{B} \otimes \mathbf{C}\|_F = \left\| \widehat{\mathbf{A}} - vec(\mathbf{B}) \cdot vec(\mathbf{C})^T \right\|_F,$$

где $\widehat{\mathbf{A}}$ – векторизованная матрица **A,** $\otimes$ – символ Кронекерова произведения; пара векторов

$$\left\{ vec(\mathbf{B})^T, vec(\mathbf{C})^T \right\}$$

рассматривается в дальнейшем как ПмУП. Решение задачи – процедура сингулярной декомпозиции [**U S V**] =svd(**A**) и формирование множества

$$vec\left(B^{(opt)}\right) = \left(\sigma_1\right)^{1/2} \mathbf{U}(:,1),$$

$$vec\left(C^{(opt)}\right) = \left(\sigma_1\right)^{1/2} \mathbf{V}(:,1).$$

Поиск ближайшей ранг-1 матрицы – проблема сингулярной декомпозиции. Если учесть, что НМ $\tilde{\mathbf{A}} = \{\mathrm{a}/\mu_{\tilde{A}}\}$ – это матрица размерностью $2 \times n$, или матрицы (КП-гранулы) размерностью $n \times n$ $\left(a_i \otimes \mu_{\tilde{A}}^i\right)_n^n$, то можно найти ближайшую к НМ $\tilde{\mathbf{A}}$ матрицу $\tilde{\mathbf{B}}$, которую, как показано ниже, можно представить после сингулярной декомпозиции в виде

$$\tilde{\mathbf{B}} = \mathbf{b}_1 \otimes \boldsymbol{\mu}_{\tilde{B}}^1 + \dots + \mathbf{b}_r \otimes \boldsymbol{\mu}_{\tilde{B}}^r \ \text{или}\ \tilde{\mathbf{B}} \approx \tilde{\mathbf{B}} \approx \mathbf{b}_1 \otimes \boldsymbol{\mu}_{\tilde{B}}^1,$$

где $\boldsymbol{\mu}_{\tilde{B}}^1 \to [0,1]$, $\mathbf{b}_1 \in \nabla$, т.е. НМ.

Таблица 1 – ***Результаты использовании НМ с одной формой ФП и отличающейся от исходной***

x	1	2	3	4	5	6	7	8	
μ^x	0.20	0.80	0.50	0.30	1.00	0	0.90	0.40	$\mu_{\tilde{A}^{nearest}}(x_i) = \begin{cases} 0, \text{если}\ \mu_{\tilde{A}}(x_i) < 0.5 \\ 1, \text{если}\ \mu_{\tilde{A}}(x_i) > 0.5 \\ 0\ \text{или}\ 1, \text{если}\ \mu_{\tilde{A}}(x_i) = 0.5 \end{cases},$
ν^x	0	1	0	0	1	0	1	0	
ξ^x	1	1	1	1	1	1	1	1	
	0	0	0	0	0	0	0	0	

Формульное представление ПмУП в виде $\mathbf{b}_1, \boldsymbol{\mu}^1_{\tilde{B}}$ – результат усечения исходного разложения.

Аппроксимация многомерных массивов с использованием Кронекерова произведения. В работе [18] рассмотрены вопросы аппроксимации с использованием Кронекерова произведения. Известно, что решение задачи $\min\left\|\mathbf{A}-\mathbf{B}\otimes\mathbf{C}\right\|_F^2$, – это проблема rank-1 аппроксимации матриц, она использует оператор векторизации (*vec*) – постолбцовое представление матрицы. В связи с этим матрицу **A** реструктуризуют таким образом, чтобы сумма квадратов в $\left\|\mathbf{A}-\mathbf{B}\otimes\mathbf{C}\right\|_F^2$ совпадала бы с суммой квадратов в $\left\|\breve{\mathbf{A}}-vec(\mathbf{B})\cdot vec(\mathbf{C})^{\mathrm{T}}\right\|_F^2$ или $\left\|\breve{\mathbf{A}}-vec(\mathbf{B})\otimes vec(\mathbf{C})\right\|_F^2$. Доказаны теоремы, одна из которых приводится ниже [C.F.Van Loan], в них использовано понятие матричного следа: $\mathbf{X}\in \mathbb{R}^{q\times q} \Rightarrow tr(\mathbf{X})=\sum_{i=1}^{q} x_{ii}$.

Теорема [18]. Предположим, $m=m_1m_2$, $n=n_1n_2$ и $\mathrm{A}\in\nabla^{m\times n}$. Если $C\in\nabla^{m2\times n2}$ фиксировано, тогда матрица $\mathrm{B}\in\nabla^{m1\times n1}$, определенная через $b_{ij}=\mathrm{tr}(\mathrm{A}_{ij}^T\mathrm{C})/\mathrm{tr}(\mathrm{C}^T\mathrm{C}),\ 1\le j\le n1,$ где tr – след матрицы, минимизирует $\left\|\mathbf{A}\text{-}\mathbf{B}\otimes\mathbf{C}\right\|_F$, $A_{ij}=A(i-1)\times m_2+1$, где $\mathrm{A}_{ij}=\mathrm{A}\left(i-1\right)\cdot m_2+1{:}\,i\cdot m_2.$

Аналогично, если $\mathrm{B}\in\nabla^{m1\times n1}$ фиксировано, тогда матрица $C\in\nabla^{m2\times n2}$, определенная через $c_{ij}=\mathrm{tr}(\widehat{\mathrm{A}}_{ij}^T\mathrm{B})/\mathrm{tr}(\mathrm{B}^T\mathrm{B}),\quad \mathrm{tr}(\mathrm{B}^T\mathrm{B}),\quad 1\le j\le n2,$ $1\le i\le m2,$ минимизирует $\left\|\mathbf{A}\text{-}\mathbf{B}\otimes\mathbf{C}\right\|_F$, где $\widehat{A}_{ij}=A\left(i{:}m2{:}m,\, j{:}n2{:}n\right).$

Это обстоятельство использовано для того, чтобы сформулировать задачу вычисления ПмУП путем решения задачи $\min\left\|\widehat{X}-\widehat{x}\cdot\left(\widehat{\mu}^x\right)^T\right\|_F^2$ или $\min\left\|\widehat{X}-\widehat{x}\otimes\widehat{\mu}^x\right\|_F^2$, где $\widehat{X}$ – матрица размерностью $n\times n$, представляющая собой ИМД или реструктурированное универсальное множество $\widehat{X}=reshape(\,X,n,m\,),X=\left\{x_i\right\},\qquad i=1,n;$ $\widehat{x}\in\widehat{X},\qquad \widehat{\mu}^x\rightarrow\left[0{,}1\right]$ – аналог ФП. Если $\tilde{x}$-НМ, сформированное на основании ИМД по правилам ТНМ, ${}^{\pi}x$ – ПмУП, вычисленное на основании СД ИМД, то в качестве критериев близости НМ и ПмУП приняты: $\left\|\tilde{x}\right\|_F\cong\left\|{}^{\pi}x\right\|_F$, $\mathrm{def}\left(\tilde{x}\right)\cong\mathrm{def}\left({}^{\pi}x\right)$, где *def* – дефадзификация по методу ЦТ.

Как правило, $\mu(x)$ формируется эвристически путем выбора из стандартных наборов ФП, входящих в библиотеки прикладных программ (пример, MatLab). Признание за $\mu(x)$ смысла весового параметра позволяет в ряде случаев предлагать разноообразную интерпретацию ФП вместе со способами получения ФП, как это реализовано в работах [9; 19]. Разнообразие интерпретаций ФП предполагает различные методы ее извлечения, в частности, применение тензорных декомпозиций, как это предложено в настоящей работе.

Поставим задачу найти такие векторы **u**, $\mathbf{v}^{\mathrm{T}}$, чтобы $\mathbf{u}\in\mathbf{X}$ представлял бы УМ, $\mathbf{v}^{\mathrm{T}}$ отображал бы значимость каждой компоненты **u**, сформируем оптимизационную задачу

$$\min_{\mathbf{u},\mathbf{v}}\left\|\mathbf{X}\text{-}\mathbf{u}\otimes\mathbf{v}^T\right\|_F^2,$$

где **X** – ИМД, преобразованное в квадратную матрицу. Желательно иметь в результате:

– пара { **u, v** } близка по F-норме к {**x**, $\boldsymbol{\mu}^x$}, где {**x**, $\boldsymbol{\mu}^x$}-НМ – аппроксимация **X**; – величина $\sum_{i=1}^{n}u_iv_i\,/\sum_{i=1}^{n}v_i$ (дефадзифицированное значение {**u,v**}) близка по величине к дефадзифицированному значению {**x**, $\boldsymbol{\mu}^x$}.

Известно, что если, например, отыскивают наилучшую ранг-1 матричную аппроксимацию матрицы **X,** для которой также известно, что любая $M\times N$ ранг-1 матрица может быть записана в виде $\mathbf{uv}^{\mathrm{T}}$, где **u** – норм-1 *M*-вектор и **v** – *N*-вектор, то эта проблема может быть сформулирована как $\min_{\mathbf{u},\mathbf{v}}\left\|\mathbf{X}\text{-}\mathbf{uv}^T\right\|_F^2$. Процедура сингулярной декомпозиции SVD(**X**) позволяет получить разложение $\mathrm{X}\cong\sum_{k=1}^{l}d_ku_kv_k$.

В процедуре [**U S V**]= SVD (**X**) принято: $\mathbf{U}=[\mathbf{u}_1,\ldots,\mathbf{u}_r]$, $\mathbf{V}=[\mathbf{v}_1,\ldots,\mathbf{v}_r]$, $\mathbf{D}=\mathrm{diag}\{\mathbf{d}_1,\ldots,\mathbf{d}_r\}$, $l\le r$.

Соответственно решение исходной задачи получено в виде: $\mathbf{u}^*=\mathrm{d}_1\mathbf{u}_1$, $\mathbf{v}^*=\mathbf{v}_1$.

Совокупность пар ($\mathbf{v}_k$, $\mathrm{d}_k\cdot\mathbf{u}_k$), $k>1$, обеспечивает наилучшую ранг-1 аппроксимацию соответствующих остаточных матриц. Например, $d_2\mathbf{u}_2\mathbf{v}_2^T$ – наилучшая ранг-1 аппроксимация $\mathbf{X}-d_1\mathbf{u}_1\mathbf{v}_1^T$.

Подмножество упорядоченных пар, тензорное произведение компонент которого обеспечивает наибольшую близость к исходному 2D-тензору

(ИМД), определим на основании следующего алгоритма:

1. Реализация сингулярной декомпозиции: [u s v] = svd (**A**'); где **A** – 2D-тензор размерностью n×n, представляющий ИМД; **u, v** – левый и правый сингулярные векторы; s – диагональная матрица сингулярных величин, в которой s (1,1) > s (2,2) >,…, > s(n,n).

2. Формирование ПмУП:

pmup=sort([abs(u(:,1)*s(1,1))*max(abs(v(:,1))), abs(v(:,1))/max(abs(v(:,1)))]).

3. Вычисление F-нормы ПмУП: n_pmup=norm([pmup(:,1),pmup(:,2)],'fro');

4. Вычисление дефадзифицированного значения:

def_pmup=sum(pmup(:,1).*pmup(:,2))/sum(pmup0(:,2)).

Укажем, что основные положения использования тензорных декомпозиций, в т. ч. в условиях неопределенности, рассмотрены в [20;21]. Показано, что тензорные декомпозиции (элемент тензорного анализа) позволяют обнаруживать скрытые структуры сложных данных и выполнять определенные обобщения. Сингулярная декомпозиция является уникальной благодаря ортогональности ограничений:

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^T = \sum_{r=1}^{R} \mathbf{s}_{rr} \mathbf{u}_r \mathbf{v}_r^T,$$

где **U, V –** ортогональны**; S –** диагональ, кроме того, ($\forall i$) abs (**v**(i,1)) ∈ [0,1], sum(**v**(:,1)^2)=1. Это позволяет подмножество упорядоченных пар {abs(**u**(:,1)* s(1,1)), abs(**v**(:,1))} после соответствующей нормализации рассматривать как ПмУП – аналог НМ.

Пример. Для произвольно сформированной матрицы случайных чисел размерностью 5×5, обладающих свойством ($\forall ij$)(3< x_{ij} < 10), реализована процедура сингулярной декомпозиции [**u s v**]=svd(**X**); левый и правый сингулярный векторы и сингулярные числа приведены ниже в табл. 2.

Прямым вычислением можно убедиться, что для левого и правого сингулярных векторов справедливы соотношения:

sum(u(:,1)^2)=1,

sum(v(:,1)^2)=1;

abs(u(1:5,1))∈ [0,1],

abs(v(1:5, 1)) ∈ [0,1].

Это позволяет рассматривать величины abs(u(1:5,1)) как показатели виртуальной значимости, формально аналогичные тем, которые рассматриваются в составе НМ $\tilde{x} \square \left\{ x / \mu^x \right\}$, $\mu^x \in [0,1]$. Отметим, что интерпретация SVD зависит от приложения.

Экспериментальные результаты

Основная цель: показать возможность выделения из ИМД некоторого множества значений, интерпретируемого семантически как ФП (в нотации ТНМ), в контексте выбора и обоснования новой математической модели – ПмУП**,** учитывающей взаимосвязь интерпретации ФП и способа ее извлечения (построения, назначения, учета структуры ИМД и др.).

Реализованы процедуры: (i) построение стандартного НМ: $\tilde{x} = \{x/\mu_x\}_{i=1}^n, \mu_x \in [0,1], x \in U = \{u_i\}_{i=1}^n$, в котором ФП задана экспертом, УМ $U = \{u_i\}_{i=1}^n \to [u_{min}, u_{max}]$ сформировано с учетом статистических характеристик массива ИД; (ii) формирование на основе ИМД 2D-тензоров (формальная процедура), на базе тензорных моделей вычисляется ПмУП $\hat{x} = \{x/\upsilon_x\}_1^n$, $\upsilon_x \in [0,1]$, $x \in \hat{U} = \{\hat{u}_i\}_{i=1}^n$ – объект, аналогичный НМ; (iii) если ИМД не задано, но известно УМ, реализуется процедура погружения УМ

$$\mathbf{I}^U = [\ \min(\{u_i\}_{i=1}^n),\ \max(\{u_i\}_{i=1}^n)]$$

в специальный тензор (Теплица или Ганкеля), тензорная декомпозиция которых позволяет вычислить ПмУП.

Таблица 2. – ***Процедура сингулярной декомпозиции [u s v]=svd(X)***

x =					***U(:,1) =***	***V(:,1) =***	***Diag(s)***
9.6509	8.3347	7.3080	5.8399	3.4052	-0.4496	-0.4760	34.8594
4.6180	6.1953	8.5436	9.5483	5.4701	-0.4466	-0.4114	7.2824
7.2479	3.1295	9.4527	9.4183	8.6922	-0.4868	-0.4863	4.8431
6.4019	8.7499	8.1675	5.8719	3.0690	-0.4195	-0.5142	3.1646
9.2391	6.1129	4.2339	9.2555	3.9722	-0.4305	-0.3214	0.1084

Необходимо сравнить объекты: НМ

$$\{x/\mu_x\}_{i=1}^{n}, \mu_x \to [0,1], x \in X,$$

ПмУП

$$\{x/\upsilon_x\}_{i=1}^{n},\ \upsilon_x \to [0,1],\ x \in X$$

с целью выявления их эквивалентности. Критерии: близость по F-норме $\min\|\tilde{x}-\hat{x}\|_F^2 \to 0,$ близость дефадзифицированных значений $\mathrm{def}(\tilde{x}) \approx \mathrm{def}(\hat{x})$. Ниже приведены числовые примеры, доказывающие возможность решения сформулированных задач и иллюстрирующие процесс достижения цели.

Пример 1. Влияние формы ФП на основные характеристики НМ: заданы стандартные НМ $\tilde{x} = \{x/\mu_x\}_{i=1}^{n},\ \mu_x \in [0,1],\ x \in U = \{u_i\}_{i=1}^{n},$ (trimf(), trapmf(), gaussmf()); вычислены их ДЗ (принцип ЦТ) – $def() = \sum_{i=1}^{n} x_i \cdot \mu_{x_i} / \sum_{i=1}^{n} \mu_{x_i}$, – нормы $\|\tilde{x}\|_F^2 = \mathrm{trace}\left((\tilde{\mathbf{x}}) \cdot (\tilde{\mathbf{x}})^T\right)$, где $(\tilde{\mathbf{x}}) = \begin{pmatrix} \mathbf{x} & \boldsymbol{\mu}_x \end{pmatrix}$ – матрица размерностью $2 \times n$, $\mathbf{x} = \{x_i\}$, $\boldsymbol{\mu}_x = \{\mu_{x_i}\}$, $i = 1, n$; необходимо определить близость данных НМ.

И М Д: <*приблизительно 6* > $\to$ $\mathrm{A} = 3 + rand(8) * 6.$

Ниже приведены результаты выполнения (табл. 3). Сравнивая полученные результаты, можно утверждать, что все НМ, аппроксимирующие одно и то же утверждение, сформированные на одном и том же УМ, независимо от выбранной ФП с точки зрения предложенных критериальных параметров (F-норма и ДЗ) являются не просто близкими, но практически эквивалентными. Представление НМ 2D – тензор-гранулой: принято НМ $\tilde{x} = \{x/\mu_x\}_{i=1}^{n}$, $\mu_x \in [0,1]$, $x \in U = \{u_i\}_{i=1}^{n}$; тензор-гранула $\tilde{x} \to \mathbf{T}_x = x \otimes (\mu_x)^T$, где $\otimes$ – символ Кронекерова произведения – это 2D-тензор размерностью $n \times n$.

Вычисление ПмУП для КП компонент стандартных НМ реализуется согласно схеме, приведенной ниже.

$$\tilde{x} = \{x/\mu^x\},$$

$$\mu^x \to [0,1] \Rightarrow \underbrace{x \otimes \mu^x}_{\text{КП НМ}} \to \mathrm{svd}((x \otimes \mu^x)^T) \to \underbrace{\begin{pmatrix} x & \bar{\mu}^x \end{pmatrix}}_{\text{ПмУП}},$$

$$\bar{\mu}^x \to [0,1]; \quad \mu^x \neq \bar{\mu}^x,$$

где $\bar{\mu}^x$ – 2-ая компонента ПмУП, интерпретируемая семантически как аналог ФП.

Рассмотрим более конкретно пример с треугольной ФП (рисунок):

$$\{x / \mu^x\},\ x \in X,$$

1. Для некоторого (произвольно сформированного в примере 1) массива данных **A**, имитирующего множество возможных значений измерения некоторой величины с существенно большой погрешностью (неточность), сформируем НМ с треугольной ФП, вычислив max, min, mean – значения:

```
xmin=min(min(A)); xmax=max(max(A));
xs=mean(mean(A)); x=[xmin:(-xmin+xmax)/9:xmax];
y=trimf(x,[xmin xs xmax]).
```

2. Для заданного НМ вычислим дефадзифицированное значение и F-норму.

```
xAcrisp=defuzz(x,y,'centroid'); norm([x' y'],'fro').
```

3. Сформируем для НМ Кронекерово произведение xA=kron(x, y') и выполним его сингулярную декомпозицию [u s v]=svd((xA)'), на основании rank-1 тензора сформируем объект-подмножество упорядоченных пар

```
pmup0=sort([abs(u(:,1)*s(1,1))*max(abs(v(:,1))),abs(v(:,1))/max(abs(v(:,1)))]).
```

4. Поскольку матрица сингулярных величин содержит только 1 ненулевое значение (все остальные равны 0), данный объект pmup0 является таким, что pmup0(:,1)⊗pmup0(:,2)^T практически точно аппроксимирует исходный 2D-тензор, сформированный на основании ИМД.

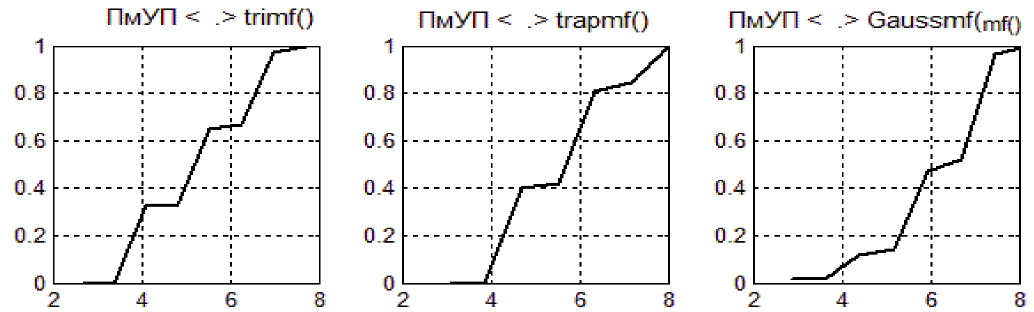
5. В ПмУП аналогично п.2 вычислим дефадзифицированное значение xAkr= sum(pmup0(:,1).*pmup0(:,2))/sum(pmup0(:,2)) и F-норму n_pmup0=norm([pmup0(:,1),pmup0(:,2)],'fro') и сравним с аналогичными значениями, полученными в п.2. НМ и ПмУП, полученные на основании SVD-процедуры КП компонент НМ, приведены в табл. 3.

*Таблиця 3 – **Влияние формы ФП на основные характеристики НМ***

A=							
7.91	8.03	7.77	8.25	4.71	5.60	7.70	5.49
6.96	6.41	8.74	7.42	5.82	4.36	7.09	4.83
5.05	5.22	6.14	3.82	3.39	6.48	5.77	8.25
4.74	7.22	8.28	3.07	8.93	7.56	6.41	3.09
5.05	6.28	4.04	8.36	6.50	6.18	7.77	7.61
6.20	5.67	8.88	4.19	5.54	6.84	3.36	8.83
7.36	7.17	4.63	4.79	6.09	4.25	6.62	8.94
4.86	6.73	4.51	6.97	5.00	5.28	3.30	7.73

Закінчення табл. 3

Стандартные НМ:	***trimf***	***trapmf***	***gaussmf***
X	μ^x	μ^x	μ^x
3.07	0	0	0.01
3.91	0.27	0.38	0.07
4.75	0.54	0.77	0.35
5.59	0.81	1.00	0.83
6.42	0.91	1.00	0.97
7.26	0.61	0.92	0.56
8.10	0.30	0.46	0.16
8.94	0	0	0.02
Характеристики стандартных НМ:	***trimf***	***trapmf***	***gaussmf***
Дефадзифицированные значения стандартных НМ:	***trimf***	***trapmf***	***gaussmf***
	5.97	***5.96***	***5.91***
Нормы стандартных НМ:	***trimf***	***trapmf***	***gaussmf***
	17.90	***17.94***	***17.89***



а *б* *в*

Рисунок – Ближайшие ПмУП, вычисленные для КП НМ с треугольной, трапециевидной и Гауссовой форм: а – НМ с треугольной ФП; б – НМ с трапециевидной ФП, в – НМ с Гауссовой ФП; НМ < приблизительно 6>

Обратим внимание: ДЗ в обоих случаях не совпадают (в пределах ± 10 %), что объясняется тем, что ПмУП получено на основе всего ИМД, а НМ (точнее КП тензор-гранула), только на его подмножестве.

Пример 2. Формирование ПмУП только на основании сингулярной декомпозиции ИМД, не прибегая к неформальному назначению ФП. Результаты сингулярной декомпозиции матрицы **А** (ИМД) приведены в табл. 4.

Таблиця 4 – ***НМ и ПмУП, полученное на основании SVD-процедуры***

Н М		***ПмУП***	
3.03	0	2.93	0
3.68	0.24	3.57	0
4.34	0.48	4.20	0.21
4.99	0.73	4.84	0.25
5.65	0.97	5.47	0.42
6.30	0.82	6.11	0.50
6.96	0.62	6.75	0.64
7.62	0.41	7.38	0.75
8.27	0.21	8.02	0.85
8.93	0	8.65	1.00
НОРМЫ:			
НМ [x y]		***ПмУП n_pmup0***	
19,89		19,29	
Дефадзифицированные значения			
НМ [x y]		***ПмУП n_pmup0***	
6,09		7,09	

Обратим внимание на матрицу сингулярных величин: только начиная с 6-го диагонального элемента значения начинают существенно уменьшаться (меньше 10 % s(1,1)). Это требует выполнения 2-х вариантов расчета:

1 – учет только 1-го rank-1 тензора, т.е.

$$A \cong s(1,1) \cdot \mathrm{kron}(u(:,1),v(:,1)^{\mathrm{T}}) \; –$$

усеченный вариант расчета;

2 – учет по крайней мере 5 rank-1 тензоров, т.е.

$$A_{pr} = u(:,1:r) * s(1:r,1:r) * (v(:,1:\mathrm{r}))^T.$$

Ниже в табл. 5, 6 приведены результаты вариантов расчета.

Резюме: НМ с треугольной ФП, сформированное на основе статистических характеристик ИМД, является практически близким (в смысле F-нормы) к ПмУП, вычисленному на основании тензорной декомпозиции ИМД: дефадзифицированные значения и F–нормы достаточно близки. Принимая во внимание, что доверие к результатам, полученным на основании ТНМ, весьма высокое, а предлагаемый авторами метод является формальным и новым, все, что связано с ПмУП, можно рассматривать как предварительный результат или базу для сравнения.

ПмУП, полученные с учетом только 1-ой компоненты сингулярной декомпозиции 2D -тензора ИМД и с учетом 5-ти компонент, дают практически полное совпадение F-норм и дефадзифицированных значений. Это подтверждает правомочность использования ПмУП в том виде, в каком оно определено в работе (с учетом только максимального значения сингулярных величин).

F–нормы стандартного НМ с треугольной ФП (кривая 1) и ПмУП (кривая *2*, рисунок) практически совпадают, аналогично близки их ДЗ, но интервалы неопределенности различны: [3; 10] для стандартного НМ и [5.75, 8.13] для ПмУП. Это является несомненным преимуществом тензорного моделирования неопределенности по сравнению с ТНМ, оно обусловлено выявлением скрытых знаний, характерных структуре ИМД.

Выводы

1. В настоящее время сложилась ситуация, когда в ряде случаев применение ТНМ для решения задач в условиях неопределенности или невозможно или сопряжено с практически непреодолимыми трудностями, обусловленными следующими причинами:

– сложность технологического процесса (или объекта управления), не позволяющая выбрать ФП;

– высокая размерность ИМД, усугубляемая пропуском данных;

– противоречивость требований, предъявляемых к возможной ФП;

– необходимость наличия альтернативного решения, свободного от предположений, принятых при назначении ФП.

*Таблица 5 – **Результаты сингулярной декомпозиции матрицы А***

abs(u(:,1))'=	[0.32	0.33	0.37	0.27	0.37	0.31	0.28	0.26	0.32	0.31]
diag(s)' =	[57.88	8.98	7.52	6.67	5.38	3.64	2.84	1.86	1.49	0.04]
abs(v(:,1))'=	[0.32	0.31	0.28	0.28	0.37	0.32	0.35	0.32	0.32	0.28]

*Таблиця 6 – **Результат расчета ПмУП***

Нечетко-множественная аппроксимация ИМД		***ПмУП на основе сингулярной декомпозиции ИМД***			
		1 rank-1 тензор		***5 rank-1 тензоров***	
3.03	0	5.50	0.76	5.88	0.71
3.68	0.24	5.76	0.77	5.97	0.74
4.34	0.48	5.99	0.77	6.01	0.77
4.99	0.73	6.50	0.86	6.68	0.84
5.65	0.97	6.57	0.87	6.75	0.85
6.30	0.82	6.80	0.87	6.77	0.88
6.96	0.62	6.84	0.88	6.81	0.88
7.62	0.41	7.08	0.89	6.90	0.91
8.27	0.21	7.74	0.95	7.39	1.00
8.93	0	7.77	1.00	7.77	1.00
F-норма	19.89	F-норма	21.42	F-норма	21.35
Дефадзифицированное значение	6.09	Дефадзифицированное значение	6.72	Дефадзифицированное значение	6.76

2. Предложен способ построения подмножества упорядоченных пар, которое может быть использовано вместо НМ (при невозможности его формирования) как возможность получения предварительного решения, позволяющего оценить рациональность экспертно сформированного НМ (ФП). Данное подмножество вычисляется на основании сингулярной декомпозиции пребразованного в 2D-тензор ИМД.

3. Выполнена экспериментальная проверка работоспособности предложенного метода, состоящая в том, что ИМД имитировалось при помощи генератора случайных величин как утверждение < *приблизительно* 6 > в виде массива значений, преобразованного в 2D тензор размерностью 8×8. Показана близость (в смысле Фробениусовой нормы и дефадзифицированного значения) подмножества упорядоченных пар, сформированного на основе ИМД при помощи формализованных процедур (сингулярной), и НМ с треугольной ФП, построенной на УМ – интервале [min, max] значений ИМД.

Список литературы

1. Pouchard, L. (2015). Revisiting the Data Lifecycle with Big Data Curation. International Jou-r-nal of Digital Curation, 10, 2, 176 – 192.

2. Wang, H., et al. (2016). An overview on the roles of fuzzy set techniques in big data processing: Trends, challenges and opportunities, Knowledge-Based Systems.

3. Yao, J., Vasilakos, A., Pedrycz, W. (2013). Granular computing: perspectives and challenges. IEEE Trans Cybern, 43(6), 1977 – 1989.

4. Fern´andez, A., et al. (2016). A View on Fuzzy Systems for Big Data: Progress and Opportunities. International Journal of Computational Intelligence Sys-tems, 9, 1, 69 – 80.

5. Chountas, P., Atanassov, V., Atanassova, E., Sotirova, S., Sotirov, K. and Roeva, O. (2018). Notes on Intuitionistic Fuzzy Sets, 24, 2, 129 – 135. DOI: 10.7546/nifs.2018.24.2.129-135 Big data, intuitionistic fuzzy sets and MapReduce operators.

6. Herrera, F., et al. (2017). Fuzzy Models for Data Science and Big Data. IEEE Intern. Conference on Fuzzy Systems. July 9-12, 2017.Naples, Italy.

7. Cichoki, A. (2017). Era of Big Data Processing: A New Approach via Tensor Networks and Tensor Decompositions.1403.2048v4 [cs.ET] 24 Aug 2014.-30.

8. Nguyen, V.D., Abed-Meraim, K. and Linh-Trung, N. (2016). Fast Tensor Decompositions for Big Data Processing 2016 International Conference on Advanced Technologies for Coommunicati-ons (ATC)

9. Cichocki, A. (2011). Tensor Decompositions: A New Concept in Brain Data Analysis? arXiv: 1305.0395v1 [cs.NA]. Control Measurement, and System Integration (SICE), special issue; Measurement of Brain Functions and Bio-Signals, 7, 507 – 517.

10. Zimmermann, H.J. (2010). Fuzzy set theory (Advanced Review). WIREs computational statistics, 2, 3, 317 – 332.

11. Zadeh, L.A. (1965). Fuzzy sets. Information and Control, 8: 3, 338 – 353.

12. Zadeh, L.A. (2005). Toward a generalized theory of uncertainty (GTU). Information sciences, 172, 1-2, 1 – 40.

13. Bilgic, T.¸ T¨rks, I.B. (1997). Elicitation of Membership Functions: How far can theory take us? IEEE International Conference on Fuzzy Systems, Proceed of the Sixth IEEE Intern. Conf., Volume: 3:1321 – 1325 vol.3 · August 1997

14. Caneman, D., Slovik, P., Twersky, A. (2005). Acceptance of solutions in non-determinind: Rules and risks. Transl. From Engl. Humanitarian cente, 632.

15. Burstein, G., et al. Kabbalah Logic and Semantic Foundations for a Postmodern Fuzzy Set and Fuzzy Logic Theory. Applied Mathematics, 5, 1375 – 1385.

16. Cofman, A. (1982). Introduction into the theory of fuzy logic.Transl. from French. M.: Raio and communication, 432.

17. Van C., Loan. (2015). The Kronecker Product SVD-Cornell Computer Science. Eletronic source: https://www. cs.cornell.edu/cv/ResearchPDF/KSVD.pdf

18. Van, Loan C.F. and Pitsianis, N. (1993). Approximation with Kronecker products. M.S.Moonen et al.(eds). Linear Algebra for large Scale and Real Time Application, 293 – 314.

19. Cichocki, A., Mandic, D., Phan, A.H., Caiafa, C., Zhao, G.,Q. & Lathauwer, L.De. (2015). Ten-sor Decompositions for Signal Processing Applications (From two-way to multiway component analysis). IEEE SIGNAL PROCESSING MAGAZINE. March 2015.-p.145 – 164.

20. Kolda, T. & Bader, B. (2009). Tensor decompositions and applications. SIAM Review, 51, 3, 455 – 500.

21. Morup, M. (2011). Applications of tensor (multiway array) factorizations and decompositi-ons in data mining. Wiley Interdisc. Rew.: Data Mining and Knowledge Discovery, 1, 1, 24 – 40.

Статья поступила в редколлегию 15.01.2019

Мінаєва Юлія Іванівна
Кандидат технічних наук, доцент, *orcid.org/0000-0001-9548-1959*
Київський національний університет будівництва і архітектури, Київ
Мінаєв Юрій Миколайович
Доктор технічних наук, професор
Національний авіаційний університет, Київ
Філімонова Оксана Юріївна
Кандидат технічних наук, доцент, *orcid.org/0000-0001-9548-1959*
Київський національний університет будівництва і архітектури, Київ

ПРЕДСТАВЛЕННЯ НЕВИЗНАЧЕНОСТІ НА РІВНІ ТЕНЗОРНИХ МОДЕЛЕЙ ДАНИХ

Анотація. *Теорія нечітких множин, що є ефективним апаратом для вирішення завдань в умовах невизначеності, в низці випадків може мати певні обмеження, зумовлені складністю технологічного процесу (або об'єкта управління), високою вимірністю та супервеликим обсягом початкової множини даних, які поглиблюються пропуском даних, суперечливістю вимог, що висуваються до функції приналежності та ін. Запропоновано спосіб побудови підмножини упорядкованих пар, яка може бути використана замість функції належності (за неможливості її формування) для розв'язання задачі в умовах невизначеності або як можливість отримання попереднього рішення, що дає змогу оцінити раціональність експертно сформованої нечіткої множини (функції належності). В основу побудови підмножини упорядкованих пар покладено нечітко-множинну інтерпретацію сингулярної декомпозиції перетвореного в 2D-тензор початкової множини даних. Наведені приклади, що ілюструють ефективність запропонованого методу.*

Ключові слова*: *невизначеність; нечітка множина; функція належності; тензор; тензорна декомпозиція; Фробеніусова норма; тензорна модель даних; підмножина впорядкованих пар

Minaeva J.
PhD, Associate professor, *orcid.org/0000-0001-9548-1959*
Kyiv National University of Construction and Architecture, Kyiv
Minaev Yu.
Doctor of Technical Sciences, Professor
National Aviation University, Kyiv
Filimonova O.
PhD, Associate professor, *orcid.org/0000-0001-9548-1959*
Kyiv National University of Construction and Architecture, Kyiv

REPRESENTATION OF UNCERTAINTY AT THE LEVEL OF TENSOR DATA MODELS

Abstract. *The theory of fuzzy sets, which is an effective tool for solving problems in conditions of uncertainty, in some cases may have certain limitations, due to the complexity of the technological process (or control object), high dimensionality and super-huge volume of the original data set, aggravated by data, contradictory requirements for the membership function, etc. A method for constructing an ordered pairs subset is proposed. This method can be used instead of using the membership function in a case when it is impossible to form it when solving a problem under conditions of uncertainty or as the possibility of obtaining a preliminary solution, which makes it possible to evaluate the rationality of an expertly formed fuzzy set (membership function). The construction of the ordered pairs subset is based on a fuzzy-set interpretation of the singular decomposition of the original data set transformed into a 2D tensor. Examples illustrating the effectiveness of the proposed method are given.*

Keywords: uncertainty; fuzzy set; membership function; tensor; tensor decomposition; Frobenius norm; tensor data model; subset of ordered pairs

---

Ссылка на публикацию

APA *Minaeva J., Minaev Yu. & Filimonova, O. (2019). Representation of uncertainty at the level of tensor data models. Management of Development of Complex Systems, 37, 93 – 104, dx.doi.org\10.6084/m9.figshare.9783212.*

ГОСТ *Минаева Ю.И. Представление неопределенности на уровне тензорных моделей данных [Текст] / Ю.И. Минаева, Ю.Н. Минаев, О.Ю. Филимонова // Управление развитием сложных систем. – 2019. – № 37. – С. 93 – 104, dx.doi.org\10.6084/m9.figshare.9783212.*