

Подход к идентификации типа неопределенности исходной информации на основе теории нечеткой ЛОГИКИ

В. С. Симанков

Кубанский государственный технологический университет
vs@simankov.ru

П. Ю. Бучацкий¹, В. В. Бучацкая²,
А. В. Шопин³, С. В. Теплоухов⁴

Адыгейский государственный университет
¹butch_p99@mail.ru, ²buch_vic@mail.ru,
³ashop409@gmail.com, ⁴mentory@mail.ru

Аннотация. В статье рассматривается подход к идентификации неопределенности исходной информации на основе теории нечеткой логики. Предложена система критериев исходной информации, вычисляемая на основе входной выборки, и характеризующая меру неопределенности, присутствующей в системе. Указаны основные положения по выбору функций принадлежности системы нечеткого вывода и получена итоговая интегрированная выходная функция принадлежности, описывающая тип неопределенности исходной информации.

Ключевые слова: неопределенность информации, нечеткая логика, функция принадлежности

I. ВВЕДЕНИЕ

Анализ современных интеллектуальных информационных технологий позволяет отметить ряд особенностей предметных областей, в которых такие системы функционируют: нечеткость целей, слабая формализуемость, уникальность, нестереотипность ситуаций, неясность входных данных [1, 2]. Это приводит к необходимости учета такого важного фактора, как неопределенность исходной информации. Так, выбор и применение математических методов без учета неопределенности приводит к неадекватности и недостоверности получаемых результатов. Для решения этой проблемы необходимо осуществить процесс идентификации, то есть определить уровень и тип неопределенности, присутствующей в системе.

II. ФОРМУЛИРОВКА ПРОБЛЕМЫ

Существующие классификации типов неопределенности исходной информации являются слишком разнородными, опираются на множество отдельных параметров, которые зачастую являются трудноформализуемыми. Это приводит к необходимости разработки классификации неопределенности, которая позволит учесть, как объективные показатели, так и субъективные предпочтения лица, принимающего решения. Поэтому в работе [3] была предложена классификация, в основе которой лежит деление по степени неопределенности и учитывает наличие субъективной интерпретации, основанной на оценках

экспертов. Такое деление позволяет идентифицировать тип неопределенности при рассмотрении сложных систем различной природы.

В рамках этой классификации выделяют следующие типы неопределенности:

- детерминированная – позволяет учесть особенности системы и знаний ЛПР только при наличии однозначной причинной связи между выбором некоторой альтернативы и исходом эксперимента;
- стохастическая – зависит от наличия статистической информации и вероятностной природы параметров системы и сведений, имеющихся у ЛПР;
- нечеткая – опирается на экспертную или лингвистическую информацию, а также на интуитивные рассуждения.

Таким образом, необходимо разработать подход, позволяющий формально описать процедуру идентификации типа неопределенности на основе анализа характеристик исходной информации.

III. ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

При формальном описании сложной системы можно выделить следующие характеристики исходной информации, характеризующие ту или иную степень неопределенности системы.

Размер выборки. Описывает количество измерений в эксперименте или исследовании. При недостаточном или малом количестве измерений детерминированные и стохастические методы являются неадекватными и дают большую погрешность [4, 5].

Доля текстовых данных. Этот параметр отражает относительную долю текстовых признаков в выборке и измеряется в процентах. Также, помимо текстовых данных, определяются основные категориальные признаки: дата и время, пол, день недели и месяц. Эти признаки выбраны по причине их наибольшей распространенности [6, 7].

Доля выбросов. Выбросами считают значения, резко отличающиеся от других в исходной выборке, и которые указывают на аномалии в распределении данных или на ошибки при измерениях. Измеряется в процентах от размера исходной выборки [8].

Стационарность. Под стационарностью будем понимать инвариантность математических ожиданий относительно сдвига времени. В этом случае для определения стационарности используется тест Дики-Фуллера [9].

Для осуществления процедуры идентификации типа неопределенности исходной информации на основе выбранных характеристик необходимо воспользоваться системой нечеткого вывода. Это позволит: учитывать особенности ЛПР и экспертов, оперировать приближенными оценками, характеризующими различные параметры исходной информации; работать с лингвистической переменной; адаптировать результаты к особенностям предметной области [10, 11].

В качестве основной логической переменной в данной системе нечеткого вывода выступает «степень неопределенности», а в качестве основных термов:

«детерминированный тип», «стохастический тип» и «нечеткий тип». Результатом является вывод о том, какому терму соответствует неопределенность системы, представленной конкретной выборкой.

При этом необходимо отметить ряд особенностей:

- 1) значение функции принадлежности равно 0 должно соответствовать полной определенности системы;
- 2) значение функции принадлежности равно 1 должно соответствовать полной неопределенности системы;
- 3) функции принадлежности должны принимать значения из области вещественных чисел.

Для выбранных характеристик исходной информации необходимо определить функции принадлежности [12, 13]. Для этого часто используются сигмоидальные (S) функции принадлежности вида: $\mu(x) = \frac{1}{1+e^{-a(x-b)}}$.

На рис. 1 представлены графики функций принадлежности для указанных выше характеристик исходной информации. Так, $\mu(x_1)$ – характеризует размер выборки, $\mu(x_2)$ – долю текстовых данных, $\mu(x_3)$ – долю выбросов, $\mu(x_4)$ – стационарность.

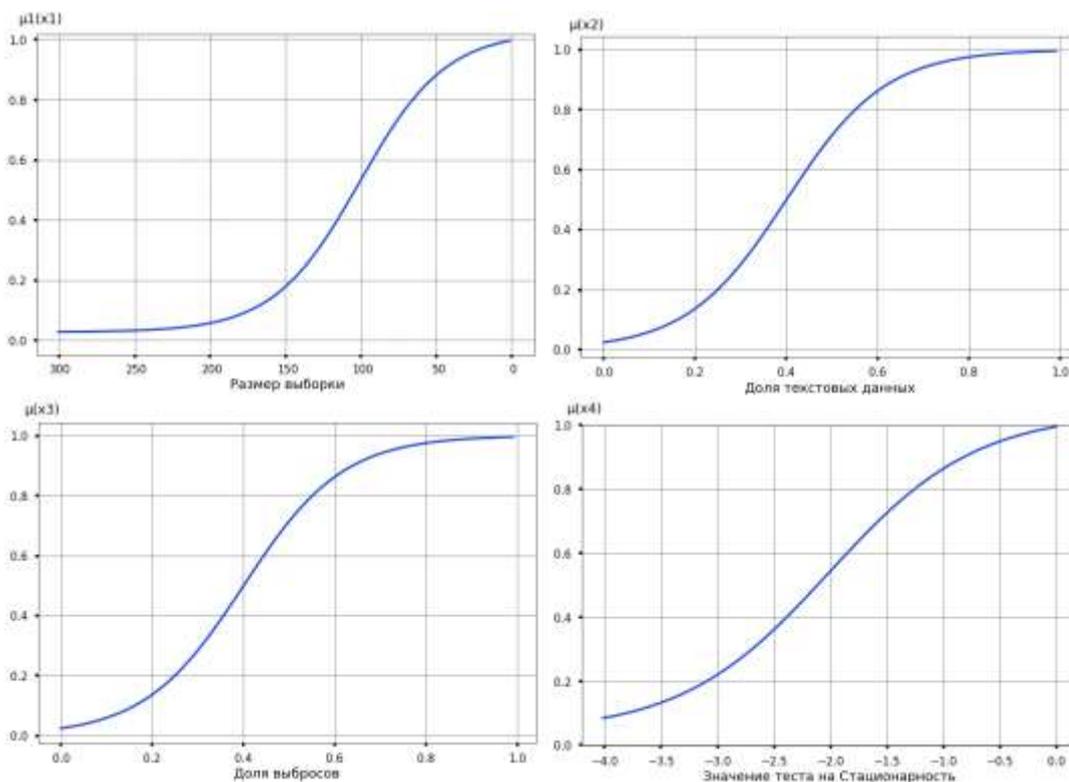


Рис. 1. Функции принадлежности для характеристик выборки

Основные особенности S-функций следующие:

1. Использование сигмоидальных функций обеспечивает получение гладких, непрерывно дифференцируемых гиперповерхностей.

2. Такие функции используются для представления нечетких множеств, которые характеризуются значениями: «большое количество», «значительная величина», «высокий уровень» и т. д.

- Такая функция имеет неограниченный носитель, что означает, что любой элемент x области определения будет принадлежать нечеткому множеству, задаваемому такой функцией.

Для описания симметричных значений непрерывного параметра x целесообразно использовать Гауссову функцию, которая описывается выражением: $\mu(x) = e^{-\left(\frac{x-b}{a}\right)^2}$. Особенности использования гауссовой функции принадлежности:

- Получается гладкая кривая для неограниченного носителя x .
- Такая функция позволяет описывать симметричные значения параметра x .
- Часто на практике необходимо ограничивать ширину этой функции.

Указанные формы функций принадлежности не являются исчерпывающими, однако применимы во многих задачах. В случае же наличия субъективных предпочтений лица, принимающего решения, особенностей задачи или выборе другого набора параметров исходной информации, возможно применить функции принадлежности другого типа.

Затем выполняется процедура агрегирования полученных значений, используя t -норму. В этом случае результирующее значение будет вычисляться по следующей формуле:
 $R = \text{MIN}(\mu(x_1); \mu(x_2); \mu(x_3); \mu(x_4))$.

Параметр R поступает на вход функций принадлежности, характеризующих термы лингвистической переменной. Для этого рассмотрим возможные виды функций принадлежности для термов:

«детерминированный тип», «стохастический» и «нечеткий».

– В условиях детерминированной определенности предполагается отсутствие любых случайных воздействий, структура системы полностью известна и может быть представлена в виде «белого ящика». В таких задачах нет необходимости доопределять проблемную ситуацию, что соответствует ситуации, когда при малых значениях переменной неопределенность системы отсутствует или мала. Таким образом, для описания этого термина можно использовать левостороннюю сигмоидальную функцию принадлежности вида: $\mu_A(R) = \frac{1}{1+e^{a*(R-b)}}$.

– В условиях стохастического типа неопределенности для каждого фактора вероятность известна или может быть оценена, т.е. неопределенность информации учитывается путем рассмотрения случайных событий и процессов. Для описания функции принадлежности в этом случае можно применять гауссову функцию: $\mu_B(R) = e^{-\left(\frac{R-b}{a}\right)^2}$.

– При нечетком типе неопределенности возможно оперировать непрерывно изменяющимися во времени значениями, неформализованными данными. В условиях такого типа система часто представляется в виде «черного ящика». Поэтому значение функции принадлежности в этом случае возрастает с ростом неопределенности системы. Это может быть формализовано правосторонней сигмоидальной функцией: $\mu_C(R) = \frac{1}{1+e^{-a*(R-b)}}$.

Указанные виды функций принадлежности были выбраны в результате анализа научных источников, как общепринятые и используемые для такого типа лингвистических переменных. На рис. 2 представлены графики этих функций принадлежности для термов лингвистической переменной.

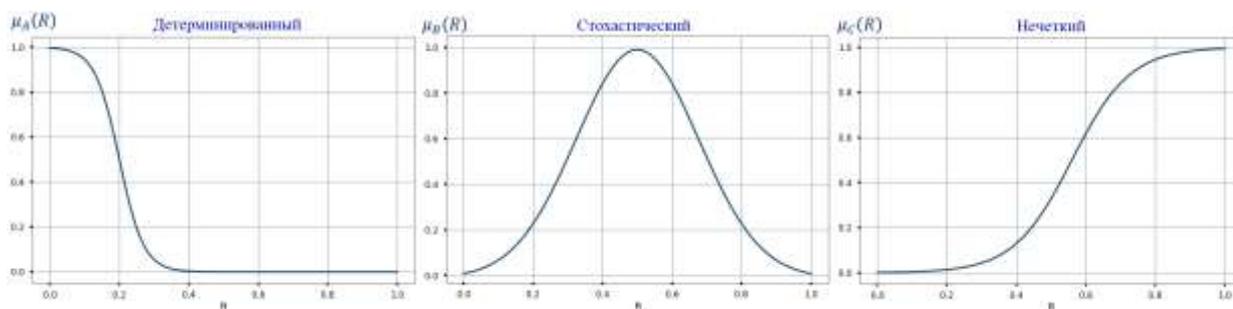


Рис. 2. Функции принадлежности для термов лингвистической переменной

Вычисленные значения $\mu_A(R)$, $\mu_B(R)$, $\mu_C(R)$ необходимы для построения интегрированной выходной функции принадлежности, характеризующей тип неопределенности исходной информации. Для этого выбран алгоритм нечеткого вывода Мамдани [14], в рамках которого необходимо осуществить аккумуляцию заключений с использованием s -меры ($aggr$) по следующей формуле:

$$aggr = \text{MAX}(\mu_A(R); \mu_B(R); \mu_C(R))$$

Осуществив процедуру построения интегрированной выходной функции принадлежности получим следующий график (рисунок 3).

Для получения четкого значения выходной переменной выполним дефазификацию с использованием метода центра тяжести [15]: $y^* = \frac{\sum \mu_{aggr}(aggr) * aggr}{\sum \mu_{aggr}(aggr)}$.

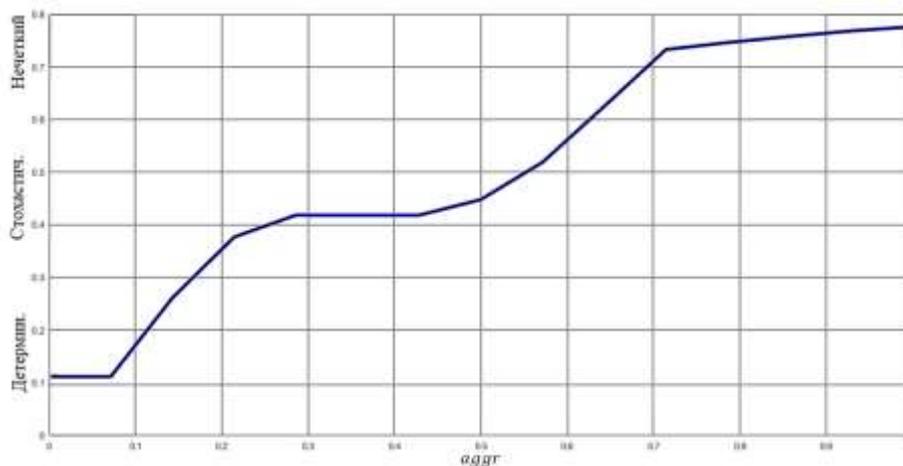


Рис. 3. График интегрированной выходной функции принадлежности, описывающей тип неопределенности исходной информации

В результате на графике можно выделить три области, соответствующие типам неопределенности: меньше 0,4 соответствует детерминированному типу; от 0,4 до 0,7 – стохастическая неопределенность; свыше 0,7 – нечеткий тип.

IV. ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ

Применение разработанного подхода позволяет численно оценить, имеющуюся в системе неопределенность, что в свою очередь позволяет осуществлять более эффективный выбор математических методов описания систем и методов искусственного интеллекта в различных прикладных задачах.

Одной из форм современных информационно-аналитических систем, которые применяют в социальных, экономических, производственных и других предметных областях и используют множество различных математических методов, в том числе и искусственного интеллекта, ситуационные центры (СЦ) [16], которые позволяют эффективно обрабатывать информацию и принимать необходимые решения. При этом выбор математического метода с учетом неопределенности исходной информации является важной задачей, решаемой во всех подсистемах ситуационного центра.

На основе разработанного подхода была спроектирована отдельная подсистема идентификации и учета неопределенности исходной информации в рамках ситуационного центра [17]. Результаты работы этой подсистемы используются в процессе функционирования других подсистем ситуационного центра, например, для решения задачи прогнозирования [18] доли вовлечения возобновляемых источников энергии (ВИЭ) в энергобаланс региона [19] и управления автономными ветроэнергетическими комплексами на их основе [20].

V. ЗАКЛЮЧЕНИЕ

Выявленная взаимосвязь между характеристиками исходной информации и типами неопределенности, имеющимися в системе, является основой для предлагаемого подхода идентификации типа

неопределенности. Таким образом, для идентификации типа неопределенности на основе исходной информации о системе необходимо рассчитать ряд количественных характеристик выборки, таких как: размер, доля текстовых данных, доля выбросов, стационарность. Затем на основе этих характеристик и предложенных функций принадлежности можно вычислить один из следующих типов: детерминированная определенность, стохастическая или нечеткая неопределенность.

Применение теории нечеткой логики для идентификации типа неопределенности исходной информации дает возможность адаптации к изменению входных данных и предметных областей, а также учету субъективных положений, имеющих у ЛПР.

СПИСОК ЛИТЕРАТУРЫ

- [1] Madera A., Risks and odds. Uncertainty, forecasting and assessment, Moscow, Krasand, 2014, 448 p.
- [2] Singh P., Tewari R. Create an Automatic Uncertainty Elimination Tools in Software Engineering, "2019 International Conference on Automation, Computational and Technology Management (ICACTM)", London, 24-26 April, 2019, pp. 237-240.
- [3] Simankov V.S., Buchatskaya V.V., Buchatskiy P.Y., Teploukhov S.V. Classification of information's uncertainty in system research // Proceedings of 2017 20th IEEE International Conference on Soft Computing and Measurements, SCM 2017.
- [4] Richard Larsen, Morris Marx. Introduction to Mathematical Statistics and An Its Applications, 6th ed., Pearson, 2017, 752 p.
- [5] Гмурман В.Е. Теория вероятностей и математическая статистика: учебник для среднего профессионального образования, 12-е изд., Москва, Юрайт, 2018, 479 с.
- [6] Дьяконов А.Г. Методы решения задач классификации с категориальными признаками, Москва: Прикладная математика и информатика, 2014, 46 с.
- [7] Mohamed Z. Kurdi. Natural Language Processing and Computational Linguistics: speech, morphology, and syntax, volume 1, Wiley, 2016, 296 p.
- [8] Manish Gupta, Jing Gao, Charu Aggarwal, Jiawei Han. Outlier Detection for Temporal Data, Morgan & Claypool, 2014, 129 p.
- [9] Enders Walter. Applied econometric time series, third ed., New York, Wiley, pp. 53–57.
- [10] Zadeh L.A. Fuzzy sets, Information and Control, №8 (3), 1965, 338–353 pp.

- [11] Fogel D.B., Liu D., Keller J.M. Basic Fuzzy Set Theory. Fundamentals of Computational Intelligence, Wiley-IEEE Press, 2016, 101–126 pp.
- [12] Пегат А. Нечеткое моделирование и управление, 4-е изд., Москва, Лаборатория знаний, 2020, 801 с.
- [13] Preuss H.P., Tresp V. Neuro-Fuzzy. Automatisierungstechnische Praxis, 1994, vol. 36, № 5, 10–24 pp.
- [14] Mamdani E.H. Applications of fuzzy algorithms for control of simple dynamic plant, Proceedings IEEE, 1974, № 121 (12), 1585–1588 pp.
- [15] Manuela Sechilariu, Fabrice Locment. Photovoltaic Source Modeling and Control. Urban DC Microgrid, Butterworth-Heinemann, 2016, 35-91 pp.
- [16] Ilyin N., Malinetsky G., Kolin K., Zatsarinny A., Raikov A., Lepskiy V., Slavin B. Distributed situational centres system of cutting edge development // International Conference Management of Large-Scale System Development (MLSD). Moscow, 2014, 3 p.
- [17] Simankov V.S., Cherkasov A.N., Buchatskiy P.Y., Teploukhov S.V. and Buchatskaya V.V. Synthesis of a Decision Support System Based on an Intelligent Situational Center // 2020 XXIII International Conference on Soft Computing and Measurements (SCM), St. Petersburg, Russia, 2020, 182-185 pp.
- [18] Buchatskaya V.V., Buchatskiy P.Y., Teploukhov S.V. Forecasting Methods Classification and its Applicability // Indian Journal of Science and Technology, 2015, № 8 (30), 275-296 pp.
- [19] Simankov V.S., Buchatskiy P.Y. Methodological Basis for Assessing Effectiveness of Involvement of Renewable Energy in Regional Energy Balance // 2019 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), Sochi, Russia, 2019, pp. 1-5
- [20] Buchatskiy P., Simankov V., Shopin A. Approach to Managing an Autonomous Energy Complex with Renewable Energy Sources based on Fuzzy Models. // International Russian Automation Conference (RusAutoCon), Sochi, 2019, 1-6 pp.