

Методологические основы лингвистических измерений нечисловых характеристик сложных объектов

С. В. Прокопчина

Финансовый университет при Правительстве Российской Федерации

Аннотация. Статья посвящена актуальным вопросам использования нечисловой, неструктурированной информации при измерении характеристик и определении состояний сложных систем в условиях значительной информационной неопределенности. В такой ситуации для реализации измерительных и вычислительных процессов должны быть использованы методы и системы, ориентированные на указанную специфику такой информации. Подобная информация носит название лингвистической.

Поэтому измерения, реализуемые на ее основе, можно назвать лингвистическими. В статье рассмотрены методологические аспекты лингвистических измерений на основе методологии байесовских интеллектуальных технологий. Приведены примеры использования методов и средств лингвистических измерений в социально-экономической сфере региональной экономики.

Ключевые слова: лингвистические измерения, байесовский подход, интеллектуальные системы

I. ВВЕДЕНИЕ

Важную часть информации, особенно в условиях измерительных ситуаций с неопределенностью, представляют собой сведения в лингвистической форме: естественно-научные знания, наблюдения в форме описаний, аналогии из других предметных областей, лингвистические оценки экспертов, анкетные данные, а также решения, выводы и рекомендации.

Для обработки и принятия решений на основе подобной информации развиты специальные теории: лингвистических переменных [1], теории возможностей [2] и другие. Основой формализации такой информации в данных теориях является аппарат теории нечетких множеств, позволяющий отображать имеющуюся информацию в виде специальных функций – функций принадлежности $\mu(v|x)$, которым придается следующий смысл [4]:

$$\mu(v|x) \in \mu_{B^*}(v|x) \in [0; 1], \quad (1)$$

где $B^* = \{x; \mu(v|x)\}$ – нечеткое множество множества X_v , X_v – базовое множество или базовая шкала значений контролируемого свойства. Носителем B нечеткого множества B^* на такой шкале являются те элементы X_v , для которых значение функции принадлежности отлично от нуля.

Для формализации информации, представленной в лингвистической форме, как отмечалось выше, развита теория лингвистических переменных. Лингвистической переменной ставится в соответствие кортеж:

$$v = \langle \eta, \lambda(\eta), \Gamma_\lambda, \Gamma_v \rangle, \quad (2)$$

где η – название лингвистической переменной, $\lambda(\eta)$ – терм-множество вербальных значений переменной, причем каждое из них является нечеткой переменной с областью определения – синтаксическое правило (грамматика), порождающее нечеткую переменную v вербальных значений лингвистической переменной η , Γ_v – семантическое правило, которое сопоставляет v с $\mu(v|x)$; X_v .

II. ПОНЯТИЕ ЛИНГВИСТИЧЕСКИХ ИЗМЕРЕНИЙ И ПРИНЦИПЫ СОЗДАНИЯ ЛИНГВИСТИЧЕСКОЙ ШКАЛЫ НА ОСНОВЕ БИИ

Однако, в указанных выше работах не были реализованы принципы измерительного подхода, что не позволяло обеспечить метрологию результатов обработки лингвистической информации и совместить эту обработку с обработкой количественной информации на принципах единства измерений. Для реализации этого было необходимо использовать измерительный подход и разработать принципиальную основу совмещения и сопоставления количественных и лингвистических результатов обработки. Это было сделано в 90-х годах прошлого века на основе регулизирующего байесовского подхода (РБП) в рамках методологии байесовских интеллектуальных измерений (БИИ).

Методология и технологии БИИ позволяют производить обработку числовой и лингвистической информации отдельно, а также осуществлять совместную обработку. Существует класс задач, связанный с обработкой только лингвистической информации в виде формализованных и классифицированных знаний, высказываний, оценок, текстов, аудиоинформации. К этому широкому классу информационных задач относятся не только задачи анкетирования социальных опросов, обработки информации социальных сетей, обработки неструктурированной информации, но и, по сути, все практические управленческие, ситуационные и поведенческие задачи общества. Накопленные и непрерывно получаемые и возрастающие объемы информации, представленные в такой форме, многократно превышают объемы информации в количественной форме. Однако, методы обработки такой информации, используемые в настоящее время не позволяют контролировать качество получаемых решений, что значительно затрудняет использование их в информационных технологиях на практике и снижает эффективность как процессов обработки, так и получаемых решений.

Введение процессов метрологического обоснования в технологии обработки лингвистической информации

позволяет реализовать принципы измерительного подхода и, тем самым, не только обеспечить контроль и управление качеством получаемых решений, но и оптимизировать сами технологии обработки.

В связи с этим, целесообразно выделить измерительные процессы, связанные с обработкой неколичественной в отдельный вид измерений, который назовем *лингвистическими измерениями*.

Как известно атрибутами любого вида измерений являются измерительные шкалы, позволяющие реализовать процессы сравнения измеряемых свойств объектов с образцовыми свойствами этих объектов.

Поэтому, далее в этой статье будут рассмотрены методологические основы и технологии создания шкал лингвистических измерений. Такие шкалы позволяют структурировать неструктурированную информацию, тем самым делая ее пригодной для последующей вычислительной обработки.

Теоретической основой создания таких шкал являются теории нечетких множеств, лингвистических переменных и регуляризирующего байесовского подхода (РБП).

В условиях значительной информационной неопределенности такие лингвистические шкалы базируются на основе технологий регуляризирующего байесовского подхода и достаточно давно (начиная с 90-х годов прошлого столетия) используются для решения практических задач. Опыт их использования приведен, в таких работах как [1, 3–10] других.

На базе РБП построены также многомерные лингвистические шкалы, позволяющие производить измерения сложных свойств, представленных лингвистическими переменными. Далее будут приведены примеры таких шкал для практических задач.

В этой статье построение лингвистических шкал рассматривается как базовый этап реализации лингвистических измерений.

Создание подобных шкал связано с некоторыми преобразованиями и интерпретацией основных понятий теорий нечетких множеств и лингвистических переменных.

Из всего многообразия смысловой интерпретации функции $\mu(v|x)$ целесообразно выделить приведенную в [160], которая позволяет рассматривать значения функции принадлежности $\mu(v|(x; v))$ как условную вероятность реализации события $v = \lambda_i$ из терм-множества λ при условии реализации значения $x \in X_v$. Такая интерпретация, не фиксируя объективность характера условной вероятности (это может быть и субъективная вероятность, отражающая мнение специалиста), позволяет привести функцию принадлежности путем нормировки к виду условной плотности распределения вероятности $f(\lambda|X = x)$.

Если в выражении (2) придать входящим в него составляющим следующий смысл: η – переменная, характеризующая состояние или лингвистический результат БИИ; λ_η – терм-множество, описывающее лингвистические градации свойства ОИ; $X_v = \{h_s\}$ – носитель шкалы БИИ H_K ; Γ_λ – измерительное преобразование (алгоритм БИИ), Γ_v – алгоритм

интерпретации априорной лингвистической информации в соответствующее условное распределение $f(\tilde{v}|h_s|x_i)$, причем элемент терм-множества при этом считается как байесовская оценка неслучайным, случайной является реализация (оценка) \tilde{v} при условии недостатка априорной информации и привлечения выборочных данных для получения $h_s^{(\lambda)}$ – решения на лингвистической шкале, то можно применить РБП к синтезу лингвистической шкалы БИИ.

Такой подход позволяет синтезировать специальную, сопряженную с числовой через базовую шкалу, лингвистическую шкалу (типа ШДО), обладающую свойствами количественных шкал БИИ, например, метрическими, а также возможностями самообучения и развития в процессе измерений.

Базовая шкала результатов БИИ с носителем H_K , формируется как упорядоченное терм-множество $\lambda_i = \{\lambda_i\}$ в соответствии с условием

$$(\forall \lambda_i \in \lambda), (\forall \lambda_j \in \lambda), (i > j) \leftrightarrow (\exists x_i \in B_i), (\exists x_j \in B_j), (x_i > x_j). \quad (3)$$

III. МЕТРОЛОГИЧЕСКОЕ ОБОСНОВАНИЕ ШКАЛ ЛИНГВИСТИЧЕСКИХ ИЗМЕРЕНИЙ

Метрологические свойства результатов на лингвистической шкале определяются в соответствии с метрологическими свойствами базовой шкалы БИИ [7]. Так, точность определения результата на лингвистической шкале $\xi_{\text{доп}}$ можно определить, исходя из расстояния между соседними элементами терм-множества $H_{K\lambda}$, представляющего собой носитель шкалы. Это расстояние может быть вычислено с использованием принятой для базовой шкалы метрики в виде:

$$\rho(\lambda_i; \lambda_{i+1}) = \rho(h_i^{(\lambda)}; h_{i+1}^{(\lambda)}), \quad (4)$$

Точность же определения результатов на лингвистической шкале может быть определена по максимальному расстоянию между элементами $h_i^{(\lambda)}, h_{i+1}^{(\lambda)}; \forall (h_i^{(\lambda)}; h_{i+1}^{(\lambda)}) \in H_{K\lambda}$.

Используя методологию РБП, можно синтезировать алгоритм принятия решений БИИ на лингвистической шкале в соответствии *уравнением лингвистических измерений*:

$$\{h_i^{(\lambda)}|\lambda^{(o)}\} = \{\operatorname{argmin} r_{\lambda_i}[h_i^{(\lambda)}; \lambda^{(o)}|(h_s; x_i; Y_i; Y_i^{(\lambda)})]\}, \quad (5)$$

Аналогично теореме возможности реализации алгоритма БИИ [5], при заданных метрологических требованиях, можно записать условие, при котором для данного терм-множества и метрологических ($M_i^{(\lambda)}$) требований возможен синтез лингвистической шкалы:

$$\forall \lambda_i \in \lambda; \exists \rho: \rho(\lambda_i; \lambda_{i+1}) \leq \rho_{\text{доп}}^{(\lambda)}[\rho_{\text{доп}}(\xi_{\text{доп}}); M_i^{(\lambda)}], \quad (6)$$

В [5] сформулирована теорема о необходимых и достаточных условиях для построения лингвистической шкалы с заданными метрологическими характеристиками.

На основе методологии лингвистических измерений можно преобразовать формулу Байеса в следующем виде. При этом создается возможность применения

байесовской свертки для лингвистических переменных, что на основе лингвистических шкал с динамическими ограничениями позволяет использовать байесовский вывод для не количественной и неструктурированной информации.

$$P_k^{(\lambda)} = \sum_{s=1}^{I_z} \frac{P_k^{(\lambda a)} \cdot f(\lambda^{(o)} | \lambda_k | h_s | x_i | y_i; y_i^{(\lambda)})}{\sum_{j=k_0}^{k_K} P_j^{(\lambda a)} \cdot f(\lambda^{(o)} | \lambda_i | h_s | x_i | y_i; y_i^{(\lambda)})} \cdot P_s \quad (7)$$

Комплексы метрологических характеристик $\{MX\}_i^{(\lambda)}$, включающие показатели точности $\xi_i^{(\lambda)}$, надежности $V_i^{(\lambda)}$, уровней ошибок 1 $\alpha_i^{(\lambda)}$ и 2 $\beta_i^{(\lambda)}$ рода, достоверности $P_{\alpha_i}^{(\lambda)}$ могут быть определены для каждого элемента нечетких результатов лингвистических измерений и в целом для (14) в виде:

$$\{MX\}_i^{(\lambda)} = \{\xi_i^{(\lambda)}; V_i^{(\lambda)}; P_i^{(\lambda)}\} | \{MX\}_s; \quad (8)$$

$$V_i^{(\lambda)} = P_{\alpha_i}^{(\lambda)} \cdot \mathcal{D}_i^{(\lambda)}; \quad (9)$$

$$P_{\alpha_i}^{(\lambda)} = \int_{h_{\gamma_i}^{(\lambda)}}^{h_{\gamma_{i+1}}^{(\lambda)}} f(\tilde{h}_i^{(\lambda)} | H_{K\lambda} | h_i^{(\lambda)}) dh^{(\lambda)}; \quad (10)$$

$$\alpha_i^{(\lambda)} = 1 - P_i^{(\lambda)}; \mathcal{D}_i^{(\lambda)} = 1 - \beta_i^{(\lambda)}; \quad (11)$$

$$\beta_i^{(\lambda)} = \max_j \int_{h_{\gamma_i}^{(\lambda)}}^{h_{\gamma_{i+1}}^{(\lambda)}} f(\tilde{h}_j^{(\lambda)} | H_{K\lambda} | h_j^{(\lambda)}) dh^{(\lambda)}; \quad (12)$$

Апостериорная вероятность появления решений $\lambda_i \sim h_i^{(\lambda)}$ на лингвистической шкале БИИ определяется в соответствии с теоремой Байес и **модифицированной формулой Байеса для свертки лингвистических переменных в виде (7)**.

Применение лингвистической шкалы дает возможность получать результаты в виде выводов, лингвистических оценок состояния объекта измерения (ОИ) и рекомендаций по оптимальному управлению ОИ в измерительных задачах второго и третьего типов (измерительный контроль и измерительное управление), а также сопровождать их комплексами метрологических показателей, характеризующих их точность, надежность и достоверность.

Кроме того, важность привлечения такого типа шкал для получения решений на сопряженной шкале с числовой шкалой, как показано в [3, 5–10], обуславливается также и возможностью перехода от одной формы представления результата к другой, получения результата БИИ в числовой форме на одной из них и трансформации его в виде лингвистического решения на другую, что дает возможность использовать все формы представления информации для получения измерительных решений:

Следует отметить, что на основе методологии БИТ, разработана компьютерная платформа «Инфоаналитик», средствами которой создан ряд прикладных

информационных систем, в том числе и для реализации лингвистических измерений.

IV. ЗАКЛЮЧЕНИЕ

Методы и технологии лингвистических измерений являются весьма полезными при наличии неколичественной информации, представленной в разнообразных формах различных типов знаний. Эти технологии могут быть использованы как отдельный класс технологий, а также совместно с другими технологиями.

Практика применения интеллектуальных систем, построенных их основе, подтверждает их эффективность при решении различных прикладных задач в условиях значительной информационной неопределенности

Однако необходимо отметить, что полученные выводы являются условными при наличии данной неполной информации и ограничений на ее состав.

В моделях таких измерительных технологий, как правило, учитываются известные информационно-обеспеченные факторы, а также известные неучтенные информационно-необеспеченные факторы.

Естественно, что и при этих условиях модель является квазиадекватной, так имеется еще значительное число неучтенных неизвестных влияющих факторов, не вошедших в данную версию модели.

Однако радует тот факт, что данная модель принадлежит к типу моделей с динамическими ограничениями [3, 5–10], что позволяет ей развиваться при получении новой информации об имеющихся и неучтенных факторах сложных систем.

СПИСОК ЛИТЕРАТУРЫ

- [1] Заде Л. Понятие лингвистической переменной и его применения к принятию приближенных решений. М.: Мир, 1976. 163 с.
- [2] Дюбуа Д., Прад А. Теория возможностей. Приложение к представлению знаний в информатике. М.: Радио и связь, 1990. 288 с.
- [3] Prokopchina S.V. Bayesian intellectual measurements, M.: Ed. House "SCIENTIFIC LIBRARY", 2021. 495 p.
- [4] Нильсон Н. Искусственный интеллект. Методы поиска решений.- М.: Мир, 1973. 272 с.
- [5] Прокопчина С.В. Концепция байесовской интеллектуализации измерений в задачах мониторинга сложных объектов // Новости искусственного интеллекта. 1997, №3, с. 7-56.
- [6] Прокопчина С.В. Основы теории шкалирования в экономике. М.: Изд. Дом «НАУЧНАЯ БИБЛИОТЕКА», 2021. 295 с.
- [7] Прокопчина С.В. Новый тип нейросетей: байесовские измерительные нейросети (БИН) на базе методологии регуляризирующего байесовского подхода // Мягкие измерения и вычисления. М., 2018. №10.
- [8] Прокопчина С.В. Интеллектуальные сети ЖКХ на основе регуляризирующего байесовского подхода в задачах цифровой экономики // Мягкие измерения и вычисления. М., 2018, №5.
- [9] Прокопчина С.В. Моделирование и управление процессами цифровизации региональной экономики. М.: Изд. Дом «НАУЧНАЯ БИБЛИОТЕКА», 2021. 395 с.
- [10] Котляр Э.А., Прокопчина С.В. Интеллектуальный программный комплекс для мониторинга и управления горячим водоснабжением (ГВС) на основе регуляризирующего байесовского подхода. // Мягкие измерения и вычисления., 2022, №12.