

ПРИКЛАДНЫЕ АСПЕКТЫ МАТЕМАТИКИ И ИНФОРМАТИКИ

УДК
658.512.011

МЕТОДЫ ИСПОЛЬЗОВАНИЯ СВОЙСТВ И ХАРАКТЕРИСТИК ОБУЧАЮЩИХ ЭТАЛОННЫХ НАБОРОВ ДЛЯ ОБЕСПЕЧЕНИЯ ДОСТОВЕРНОСТИ ЭЛЕКТРОННЫХ ДОКУМЕНТОВ

Акмал Рустамович Ахатов

д.т.н., доцент

akmalar@rambler.ru

г. Самарканд (Узбекистан)

Файзулло Махмадиярович Назаров

ассистент

fayzullo-samsu@mail.ru

г. Самарканд (Узбекистан)

Эштемиров Бунёд Шерали угли

студент

bunyodamaliy@bk.ru

г. Самарканд (Узбекистан)

Самаркандский государственный
университет

Самаркандский государственный
университет

Самаркандский государственный
университет

Аннотация. Рассмотрена проблема проектирования практических приложений, связанных с построением систем электронного документооборота (СЭД) на основе методов интеллектуального анализа данных и аппарата мягких вычислений, а также формирования нелинейных принципов нечетких выводов, в результате чего создается эффективный инструментарий для решения прикладных задач различных проблемных областей. Сложный характер динамических процессов, происходящих на этапах передачи и обработки электронных документов (ЭД), неоднозначность, неполнота исходной информации выделены в качестве аргументов разработки методов и алгоритмов оптимизации повышения достоверности информации на основе новых современных технологий. Сформулирована задача повышения достоверности данных в системах электронного документооборота на основе использования свойств и характеристик обучающих наборов-эталонов. Предложены принципы построения методов повышения достоверности электронных документов (ЭД) на базисе информационной избыточности различной природы. Разработаны модели эталонной проверки достоверности информации ЭД на основе формирования и использования наборов обучающих данных, в которых отражаются закономерности, свойства, особенности сильных вариаций, случайных всплесков, нестационарности информации, что позволяет обнаруживать и корректировать различного рода искажения, возникающие на этапах ввода, передачи и обработки данных. Выработаны правила контроля достоверности обработки данных для поддержания и регулирования значения выходного (расчетного) параметра управления на требуемом уровне для минимизации среднеквадратического

отклонения значений контролируемого элемента ЭД от порогового, который заложен в наборе-эталоне обучения. Разработан алгоритм, реализующий процедуры разбиения общего интервала значений элементов ЭД на непересекающиеся подмножества, выделения и формирования эффективного набора-эталона. Предложена методика оценки эффективности стратегий контроля достоверности информации ЭД.

Ключевые слова: электронный документ, достоверность информации, обучающий набор, свойства, характеристики, набор-эталон, адаптивное правило, разрешенное и запрещенное подмножество, стратегия контроля, оценка эффективности,

Актуальность темы. Одним из значимых исследований по разработке современных методов оптимизации обработки данных за счет повышения достоверности информации в системах электронного документооборота (СЭД) предприятий и организаций, которые характеризуются нестационарными параметрами переходных процессов на этапах сбора, передачи, хранения и обработки электронных документов (ЭД) является создание технологий интеллектуального анализа данных (ИАД) слабоструктурированных и нестационарных областей знаний [Jumanov I.I., Akhatov A.R., 2010: 71]. Эффективным механизмом обеспечения достоверности информации ЭД на сегодняшний день может выступать формирование и использование наборов обучающих данных, в которых отражаются закономерности, свойства, особенности сильных вариаций, случайных всплесков, нестационарности информации, что позволяет обнаруживать и корректировать различного рода искажения, возникающие на этапах ввода, передачи и обработки данных [Жуманов И.И., Ахатов А.Р., 2011: 42].

В связи с этим, настоящая работа посвящена исследованию и разработке методов оптимизации передачи и обработки данных за счет обеспечения их достоверности на основе эталонного обучения, в котором обобщены свойства статистических и динамических моделей идентификации, анализа и обработки информации.

Принципы повышения достоверности ЭД на основе эталонного обучения. В предлагаемом подходе к повышению достоверности ЭД при предварительной обработке данных множество значений элементов ЭД разбивается на подмножества разрешенных значений, в интервале которых данные считаются достоверными и запрещенных значений, в интервале которых данные считаются искаженными, т.е. недостоверными. Принципом повышения достоверности данных является формирование и использование эталонного обучающего набора, из которого предполагается выделение характеристик, особенностей, полезных сведений и знаний об ЭД [Воронцов К.В., 2004: 7].

Математически задачу повышения достоверности передачи и обработки ЭД на основе эталонного обучающего набора можно сформулировать следующим образом. Пусть множество $E_0 = \{S_1, \dots, S_m\}$ значений элемента ЭД разделяется на l

непересекающихся подмножеств K_1, \dots, K_l , $E_0 = \bigcup_{i=1}^l K_i$. Каждая компонента ЭД

$S_i \in E_0$ описывается n достоверными и недостоверными элементами из последовательности исходных данных $X(n) = (x_1, \dots, x_n)$. Для последовательности элементов задаются некоторые пороги δ ($\delta > 0$) в интервальной шкале и $n - \delta$ в номинальной шкале. Последовательность элементов разделяется на подмножества разрешенных и запрещенных значений, т.е. из множества элементов ЭД

представляются пары (S_u, S_v) , где $S_u \in K_i$ - достоверный элемент, $S_v \in K_j$ - недостоверный элемент, $u \neq v$, $i = 1, \dots, l$, $j = 1, \dots, m$. Важным моментом исследования является определение взаимосвязанности элементов S_u и S_v по некоторому отношению R , которое представляется различными критериями, в частности по критерию важности, информативности, эквивалентности, взаимной эквивалентности и других [Загоруйко Н.Г., Кутненко О.А. и др., 2014: 892].

Выделение и формирование набора-эталона. Задается $R_S = \rho(S_u, S_v)$ - расстояние от подмножества элементов S_u до ближайшего подмножества S_v , которое определяется по порогу δ и по Евклидовой метрике. Формируется E_0^* - подмножество взаимосвязанных элементов. С учетом этого оценивается взвешенное расстояние между контролируемыми элементами S и S' с подмножествами S_u, S_v и S'_u, S'_v

$$d(S, S') = \frac{\rho(S, S')}{R_S} + \frac{\rho(S, S')}{R_{S'}}.$$

Проверяется принадлежность элемента $S' \in K_t$ к одной из последовательности групп K_1, \dots, K_l по условию если $d(S, S') = \min_{S_u \in E_0} d(S_u, S')$ и по результатам проверки выделяется E_{ed}^* - набор-эталон и E_k^* - контрольный набор, $E_0^* = E_{ed}^* \cup E_k^*$.

Набор-эталон далее используется для выделения и формирования подмножества E_0^* искаженных элементов. Формируется рациональный набор обучающих данных E_0^* по принципу последовательного исключения недостоверных элементов и такая процедура поиска начинается с $E_k^* = \emptyset$. Определяются кандидаты на редукцию набора-эталона E_{ed} по минимальному значению R_S на основе отношения $\{R_S\}_{S \in \Gamma_u}$ по группе Γ_u , $u = 1, \dots, \delta$. Производится упорядочение групп взаимосвязанных элементов как $|\Gamma_1| \geq \dots \geq |\Gamma_\delta|$. Полученные результаты позволяют решить задачу оптимизации контроля и оценки достоверности передачи и обработки данных на основе набора-эталона и элементов мягких вычислений.

Оптимизация контроля достоверности обработки данных на основе набора-эталона. В предположении, что входной элемент x представляется в виде временного ряда, динамика элемента ЭД в переходных процессах задается

дифференциальным уравнением $\frac{dx}{dt} = uk - v$, где k - коэффициент динамической характеристики элемента; u - параметр идентификации элемента в переходном процессе; v - параметр внешних воздействий и помех.

Исходя из этого функционал качества контроля достоверности представляется в виде [Воронцов К.В., 2014:12]

$$Q = \int_0^T (x(t) - \delta)^2 dt \rightarrow \min$$

Принципом контроля достоверности обработки данных является поддержание и регулирование значения выходного (расчетного) параметра u на требуемом уровне для минимизации среднеквадратического отклонения значений контролируемого элемента ЭД от порогового, который заложен в наборе-эталоне обучения. В соответствии с этим,

расчетный выход сравнивается со значениями характеристик набора-эталона, выявляются разности, значения которых формируются бинарной последовательностью 0 и 1, где «0» - соответствует недостоверному элементу, а «1» - достоверному. Отметим, что оптимизация контроля достоверности по традиционной постановке решения задачи, как правило, основывается на задании условия Коши $x(0) = 0$. В данном случае для оптимизации известный принцип максимума Понтрягина неприменим, так как при достижении $x(t)$ порогового уровня δ могут существовать бесконечное количество переключений. В связи с этим, предлагается эвристическое правило контроля достоверности обработки данных на основе «мягких вычислений» в следующем виде:

- если $x = \text{«меньше } \delta\text{»}$ то $u = \text{«около 1»}$;
- если $x = \delta$ то $u = \text{«около } v/k\text{»}$;
- если $x = \text{«больше } \delta\text{»}$ то $u = \text{«около 0»}$.

Элемент x представляется в реальном режиме времени с определенным интервалом значений. Если в принятом интервале значение контролируемого элемента превышает уровень порога δ , то считается, что значение расчетного элемента недостоверно. Заключительный этап исследования состоит в определении оценки эффективности E_0 для стратегий контроля достоверности обработки данных на основе предложенных правил по принципу применения набора-эталона. Для этого анализируется следующее соотношение $E_0 = \lambda D_s + (1 - \lambda) T_s$, где D_s , T_s – соответственно емкости подмножества разрешенных и запрещенных значений элементов; $\lambda \in [0,1]$ - значения вероятностной функции оценки принадлежности элемента к подмножествам D_s и T_s . Установлено, что эффективность алгоритма контроля достоверности обработки данных на основе набора-эталона в первую очередь зависит от степени принадлежности элемента к границам подмножеств D_s и T_s , стратегии расчета параметра u , регулированием значений которого повышается либо уменьшается значение E_0 .

Список литературы

1. Воронцов К.В. (2004) Обзор современных методов по проблеме качества обучения алгоритмов // Таврический вестник информатики и математики. №1. С. 5-25.
2. Жуманов И.И., Ахатов А.Р. (2011) Интеллектуальный контроль достоверности текстовой информации на основе учета свойств и распределений данных // «Химическая технология. Контроль и управление» ТГТУ. Ташкент. № 1 (37). С.41-47.
3. Загоруйко Н.Г., Кутненко О.А., Зырянов А.О. (2014) Обучение распознаванию образов без переобучения // Машинное обучение и анализ данных, том 1. №7 С.891-901.
4. Jumanov I.I., Akhatov A.R. (2010) The control of information transfer reliability in intellectual control systems on the basis of statistical redundancy // Sixth World Conference on Intellectual Systems for Industrial Automation, Uzbekistan, TSTU. Tashkent. p. 70-75

METHODS OF USING THE PROPERTIES AND CHARACTERISTICS OF TRAINING ETHALONS SETS FOR ENSURING THE RELIABILITY OF ELECTRONIC DOCUMENTS

A.R. Akhatov | Samarkand State University

Dr. Sci. (Engineering), Associate Professor akmalar@rambler.ru Samarkand (Uzbekistan)	(Samarkand, Uzbekistan)
F.M. Nazarov assistant fayzullo-samsu@mail.ru Samarkand (Uzbekistan)	Samarkand State University (Samarkand, Uzbekistan)
B.Sh. Eshtemirov student bunyodamaliy@bk.ru Samarkand (Uzbekistan)	Samarkand State University (Samarkand, Uzbekistan)

Abstract. The problem of designing practical applications are considered with relation in the construction of electronic document management (EDM) systems based on methods of data mining and soft computing, as well as the formation of nonlinear principles of fuzzy inferences, which results allow to create the effective tool for solving applied problems of various problem areas. Complex nature of the dynamic processes occurring at the stages of transmission and processing of electronic documents (ED), ambiguity, incompleteness of the initial information are identified as arguments for developing the methods and algorithms for optimizing the reliability of information based on new modern technologies. The problem of increasing the reliability of data in electronic document management systems is formulated on the basis of using the properties and characteristics of learning sets of standards. The principles are proposed for constructing the methods of increasing the reliability of electronic documents (ED) on the basis of information redundancy of various nature. The models of the standard verification of ED information reliability are developed on the basis of formation and use of training data sets in which regularities, properties, features of strong variations, random bursts, nonstationary information are reflected, and it allows to detect and correct various distortions arising at the stages of input, transmission and processing data. Rules for controlling the reliability of data processing to maintain and control the value of the output (calculated) control parameter at the required level are developed to minimize the standard deviation of the values of monitored ED element from the threshold value, which is laid down in the training standard set. The developed algorithm implements the procedures for partitioning a common interval of values for ED elements into disjoint subsets, separating and forming an effective reference set. The methodology for evaluating the effectiveness of control strategies for the reliability of ED information is proposed.

Keywords: electronic document, information authenticity, training set, property, characteristics, set-standard, adaptive rule, permitted and forbidden subset, control strategy, estimation of efficiency.

References

1. Vorontcov K.V. (2004) Obzor sovremenny`kh metodov po probleme kachestva obucheniia algoritmov [Review of current methods on the quality of learning algorithms] Taurian Bulletin of Informatics and mathematics. №1, pp. 5-25.
2. Zhumanov I.I., Ahatov A.R. (2011) Intelletal`ny`i` kontrol` dostovernosti tekstovoi` informatcii na osnove ucheta svoi`stv i raspredelenii` danny`kh [Intelligent control of reliability of text information based on properties and data distributions] Taurian Bulletin of Informatics and mathematics. №1, pp. 41-47.
3. Zagorui`ko N.G., Kutnenko O.A., Zy`rianov A.O. (2014) Obuchenie raspoznavaniuu obrazov bez pereobucheniia [Learning to recognize images without retraining] Machine

learning and data analysis. №7. Т. 1, pp. 891-901.

4. Jumanov I.I., Akhatov A.R. (2010) The control of information transfer reliability in intellectual control systems on the basis of statistical redundancy // Sixth World Conference on Intellectual Systems for Industrial Automation, Uzbekistan, TSTU. Tashkent. p. 70-75.

УДК 513.7 | **СТЕПЕННЫЕ ПРЕОБРАЗОВАНИЯ ДЛЯ СИСТЕМ НЕЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ**

Ахмаджон Солеев | Самаркандский государственный
д.ф.-м.н., профессор | университет, Узбекистан
asoleev@yandex.ru
Самарканд, Узбекистан

Аннотация. В работе рассматриваются степенные преобразования, которые преобразуют нелинейные алгебраические уравнения или системы нелинейных алгебраических уравнений в уравнения или системы алгебраических уравнений с меньшей переменной. Эти преобразования делают геометрию показателей степеней более содержательной. Показаны изменения геометрических конструкций после указанных преобразований. Приведены понятия размерности системы алгебраических уравнений на основе построения многогранников Ньютона и нормальных конусов его граней. Построения многогранников Ньютона и нормальных конусов его граней рассматривались в работах А.Д.Брюно (1980), А.Солеева (1982), А.Г.Хованского (1983). Рассмотрены некоторые существенные свойства степенных преобразований. Доказано, что если размерность системы равна $d < n$, то существует унимодулярная матрица α такая, что степенным преобразованием с матрицей α и подходящими сокращениями эта система приводится к системе уравнений относительно d переменных. Также предложен алгоритм нахождения унимодулярных матриц степенных преобразований. В конце приведены конкретные примеры.

Ключевые слова: Степенные преобразования, системы нелинейных алгебраических уравнений, размерность системы, линейные пространства, многогранники Ньютона, нормальные конусы, унимодулярные матрицы.

Рассмотрим преобразования системы

$$f_i(X) \equiv \sum_{j=1}^{k_i} a_{ij} X^{Q_j} = 0, \quad i = 1, \dots, m, \quad (1)$$

которым в пространстве R^n показателей степеней Q соответствуют параллельные переносы (для каждого f_i свой) и аффинные преобразования (одно для всех f_i).

Эти преобразования делают геометрию показателей степеней [Брюно, Солеев, 1991: 89] более содержательной и дают способ для решения определённых систем.

Отыскание тех решений системы (1), у которых одна из координат тождественно равна нулю, будем считать решенной задачей, так как она приводится к задаче аналогичной исходной, но с меньшей размерностью. Поэтому будем искать решения системы (1) для которых ни одна координата не равна нулю тождественно. Для таких решений в каждом из уравнений системы (1) можно производить сокращение на любое