

ПОСТРОЕНИЕ РАСПОЗНАЮЩИХ ПРАВИЛ И АЛГОРИТМА ОПРЕДЕЛЕНИЯ ПРИОРИТЕТНЫХ ЭЛЕМЕНТОВ ДОКУМЕНТА

Тишликов С.А.

Доцент, Гулистанского государственного университета

<https://doi.org/10.5281/zenodo.7317021>

Аннотация. При решении многих проблемных задач управления в информационных системах передаются на обработку большое количество документов, причем в десятках экземпляров, которые формируются в базах данных. Например, в системе управления кадрами промышленных предприятий малого и среднего бизнеса Сирдарьинского областного хокимията числится более 15 тысяч руководителей, на которых заводятся карточки учета кадров и дополнительные формы, связанные с учетом их деятельности.

Ключевые слова: процессах, множества решающих правил, обучающей, Алгоритм решения задачи

CONSTRUCTION OF RECOGNITION RULES AND AN ALGORITHM FOR DETERMINING THE PRIORITY ELEMENTS OF THE DOCUMENT

Abstract. When solving many problematic management problems in information systems, a large number of documents are transmitted for processing, and in dozens of copies that are generated in databases. For example, in the personnel management system of industrial enterprises of small and medium-sized businesses of the Sirdarya regional khokimiyat, there are more than 15 thousand managers who have personnel accounting cards and additional forms related to their activities.

Keywords: processes, sets of solving rules, training, Algorithm for solving the problem

ВВЕДЕНИЕ

Любой документ при хранении и размещении можно представить в виде матрицы сообщений, где строки описывают типы объекта и документа, а столбцы – показатели и реквизиты документов, которые несут информацию о количественных и качественных значениях.

Следует отметить, что когда используются документы одного типа при решении определенной задачи, то можно проследить число повторяющихся названий и значений показателей и реквизитов. Такая структурно-технологическая избыточность, содержащаяся в документе может быть использована при построении алгоритмов логического контроля достоверности информации и контроля по встроенным экспертным системам и базам данных.

МАТЕРИАЛЫ И МЕТОДЫ

Постановка задачи. Пусть на некотором множестве V^t ($t=1, k$) объектов (документов) $X_j^t = (x_{j_1}^t, \dots, x_{j_n}^t)$, ($i = 1, k, j = 1, m, i = 1, n$) определены подмножества V_1^t и V_2^t ($t = 1, k$) для элемента (строки либо столбца) C_t ($t=1, k$). Допустим, что V_1^t – подмножества годных документов, соответствующие образу V_1^t ($t = 1, k$), а V_2^t ($t=1, k$) – подмножества негодных документов на V^t ($t=1, k$) соответствующие образам V_2^t ($t = 1, k$).

Требуется в процессах автоматизированной обработки множества объектов на каждом элементе C_t построить для каждого элемента C задающее правило $R'(X)$, указывающее на принадлежность любого объекта

Решение задачи. Введем следующие ограничение и условие:

- обрабатываемые множества объектов однотипные. При этом объекты могут различаться характеристиками признаков; признаки разнотипные (количественные, качественные) [1];

- пороги для каждого признака соответствующего элемента задаются в виде системы интервалов, т. е.

$$C_1 : \left[\overset{\wedge_1}{\delta_{11}}, \overset{\vee_1}{\delta_{12}} \right], \left[\overset{\wedge_2}{\delta_{11}}, \overset{\vee_2}{\delta_{12}} \right], \dots, \left[\overset{\wedge_n}{\delta_{11}}, \overset{\vee_n}{\delta_{12}} \right];$$

$$C_2 : \left[\overset{\wedge_1}{\delta_{21}}, \overset{\vee_1}{\delta_{22}} \right], \left[\overset{\wedge_2}{\delta_{21}}, \overset{\vee_2}{\delta_{22}} \right], \dots, \left[\overset{\wedge_k}{\delta_{21}}, \overset{\vee_k}{\delta_{22}} \right];$$

.....

$$C_k : \left[\overset{\wedge_1}{\delta_{k1}}, \overset{\vee_1}{\delta_{k2}} \right], \left[\overset{\wedge_2}{\delta_{k1}}, \overset{\vee_2}{\delta_{k2}} \right], \dots, \left[\overset{\wedge_r}{\delta_{k1}}, \overset{\vee_r}{\delta_{k2}} \right],$$

где $\overset{\wedge_i}{\delta_{t1}}$ и $\overset{\vee_i}{\delta_{t2}}$ максимальное и минимальное значения для признака x^t_i соответствующего элемента $C_t(t=1,k)$;

в процессах обработки документа т.е. при переходе объектов от одного элемента к другому изменяется количество порогов

$$C_1 : \left[\overset{\wedge_1}{\delta_{t1}}, \overset{\vee_1}{\delta_{t2}} \right], \left[\overset{\wedge_2}{\delta_{t1}}, \overset{\vee_2}{\delta_{t2}} \right], \dots, \left[\overset{\wedge_n}{\delta_{t1}}, \overset{\vee_n}{\delta_{t2}} \right],$$

количество объектов m и количество признаков n , а также изменяются характеристики признаков $x^i_1, x^i_2, \dots, x^i_n$.

Предположим, что обрабатываемые объекты X^t_1, \dots, X^t_m поступают к t -му элементу C_t . Для признаков $x^t_{j_1}, \dots, x^t_{j_n}$ объектов $X(j=1,m)$ заданы пороги в виде системы интервалов

$$\left[\overset{\wedge_1}{\delta_{t1}}, \overset{\vee_1}{\delta_{t2}} \right], \left[\overset{\wedge_2}{\delta_{t1}}, \overset{\vee_2}{\delta_{t2}} \right], \dots, \left[\overset{\wedge_n}{\delta_{t1}}, \overset{\vee_n}{\delta_{t2}} \right],$$

Преобразуем признаки $x^t_{j_1}, \dots, x^t_{j_n}$ относительно соответствующих порогов

$$\left[\overset{\wedge_1}{\delta_{t1}}, \overset{\vee_1}{\delta_{t2}} \right], \dots, \left[\overset{\wedge_n}{\delta_{t1}}, \overset{\vee_n}{\delta_{t2}} \right] \quad \text{в виде}$$

$$\tilde{x}^t_{ji} = \begin{cases} 1, & \text{если } \overset{\wedge_i}{\delta_{t1}} \geq x^t_{ji} \geq \overset{\vee_i}{\delta_{t2}} \\ 0 - \text{во всех остальных случаях} \end{cases}, \quad (1)$$

где \tilde{x}_{ji}^t - преобразованное значение i -го признака j -го объекта, касающегося элемента S_t .

Применяя формулу (1) для признаков $X_{j_1}^t, \dots, X_{j_n}^t$, соответствующих выборку V^t , состоящую из нулей и единиц соответственно для каждого элемента S_t .

$$V^t = V_1^t U V_2^t = \begin{pmatrix} 1 & 0 & 1 \dots & 1 & 0 \\ 0 & 1 & 1 \dots & 1 & 1 \\ 1 & 0 & 0 \dots & 0 & 1 \\ \dots & \dots & \dots & & \dots \end{pmatrix} \quad (2)$$

Теперь используя признаки $x_{ji}^t \in V^t$, определяем так же, как и в [2] простые и сложные признаки относительно образа V_1^t или V_2^t . Признак называется простым признаком относительно образа V_1^t , если выполняются соотношения

$$\begin{cases} \forall X_j^t \in V_1^t \Rightarrow x_{ji}^t = 1, \\ \exists X_j^t \in V_2^t \Rightarrow x_{ji}^t = 0, \end{cases} \quad (3)$$

Либо

$$\begin{cases} \exists X_j^t \in V_1^t \Rightarrow x_{ji}^t = 1 \\ \forall X_j^t \in V_2^t \Rightarrow x_{ji}^t = 0 \end{cases} \quad (4)$$

Если для признака x_{ji}^t выполняется соотношение (3), то простой признак является простым признаком 1-го типа образа V_1^t если же выполняется соотношение (4), то простой признак является простым признаком 2-го типа образа V_1^t .

Подсистема признаков называется сложным признаком относительно образа V_1^t если выполняются соотношения

$$\begin{cases} \exists X_j^t \in V_1^t \Rightarrow x_{j_1}^t \vee \dots \vee x_{j_n}^t = 1 \\ \forall X_j^t \in V_2^t \Rightarrow x_{j_1}^t \vee \dots \vee x_{j_n}^t = 0 \end{cases} \quad (5)$$

либо

$$\begin{cases} \exists X_j^t \in V_1^t \Rightarrow x_{j_1}^t \wedge \dots \wedge x_{j_n}^t = 1 \\ \forall X_j^t \in V_2^t \Rightarrow x_{j_1}^t \wedge \dots \wedge x_{j_n}^t = 0 \end{cases} \quad (6)$$

Если для подсистемы признаков выполняются соотношения (5), то сложный признак является сложным признаком 1-го типа образа V_1^t , если же выполняются соотношения (6), то сложный признак является сложным признаком 2-го типа образа V_1^t .

РЕЗУЛЬТАТЫ

Допустим, что в процессе обучения объектов X_j^t множества V_t отобраны n' простых признаков 1-го типа образа V_1^t обеспечивающих разделение образов V_1^t и V_2^t . Тогда решающее правила имеет вид

$$R^t(X_j^t): \begin{cases} X_j^t \in V_1^t & \text{если} & \bigwedge_{i=1}^n x_{ji}^t = 1 \\ X_j^t \in V_2^t & \text{в противном случае} \end{cases} \quad (7)$$

Если для разделения образов V_1^t и V_2^t используется простые признаки 2-го типа образа V_1^t то решающее правило будет

$$R^t(X_j^t): \begin{cases} X_j^t \in V_1^t & \text{если} & \bigvee_{i=1}^n x_{ji}^t = 1 \\ X_j^t \in V_2^t & \text{в противном случае} \end{cases} \quad (8)$$

Теперь предположим что на множестве V^t в процессе обучения отобраны r сложных признаков 1-го типа образа V_1^t .

В этом случае решающее правило будет

$$R^t(X_j^t): \begin{cases} X_j^t \in V_1^t & \text{если} & \bigwedge_{k=1}^k \bigwedge_{i=1}^n x_{ji}^t = 1 \\ X_j^t \in V_2^t & \text{в противном случае} \end{cases} \quad (9)$$

В случае, когда в процессе обучения отобрано r сложных признаков 2-го типа образа V_1^t , решающее правило, обеспечивающих разделение образов V_1^t и V_2^t , имеет вид:

$$R^t(X_j^t): \begin{cases} X_j^t \in V_1^t & \text{если} & \bigvee_{k=1}^k \bigwedge_{i=1}^n x_{jki}^t = 1 \\ X_j^t \in V_2^t & \text{в противном случае} \end{cases} \quad (10)$$

ОБСУЖДЕНИЕ

Таким образом, используя обучающую выборку V^t для каждого элемента S_t можно построить множества решающих правил определяемых формулами (7÷10).

Далее, для того чтобы в процессе обучения найти распознающее правило $\tilde{R}_1^t(X)$ из множества решающих правил $R_1^t(X), \dots, X^t_p(X)$, нужно наложить на $R_1^t(X), \dots, P_1^t(X)$ дополнительные требования. Так же из в [3] введем количественные меры эффективности решающих правил.

В зависимости от типа признаков (простых или сложных) каждое решающее $R_\gamma^t(X_j)$ может ошибочно классифицировать объекты X_j^t , принадлежащие только определенным образом V_1^t или V_2^t . Так решающие правила, определяемые (7) и (9) не могут ошибочно классифицировать объекты $X_j^t (j=1, m')$ образа V_1^t , в то же время решающие правила.

Обозначим через S_γ^t - количества правильно классифицированных объектов на обучающей выборке V^t решающим правилом $R_\gamma^t(X)$.

Тогда величину

$$F_{\gamma}^t = \frac{S_{\gamma}^t}{m} \quad (11)$$

будем называть разделяющей силой решающего правила $R_{\gamma}^t(X)$, где $0 \leq F_{\gamma}^t \leq 1$.

Используя вычисленные значения разделяющей силы F_1^t, \dots, F_p^t соответственно для каждого решающего правила $R_1^t(X), \dots, R_p^t(X)$, можно определить наилучшее решающее правило $R_{\gamma}^t(X)$ среди $R_1^t(X), \dots, R_p^t(X)$ для элемента C_t в виде

$$R_{\gamma}^t(X) = \max_{1 \leq \gamma \leq p} (F_{\gamma}^t). \quad (12)$$

ВЫВОДЫ

Алгоритм решения задачи. Рассмотрим алгоритм, предназначенный для решения задачи определения приоритетных элементов документа.

Алгоритмом отбираются только такие решающие правила для элементов документа, разделяющая сила которых наиболее максимально. Алгоритм включает следующие основные этапы.

1. В оперативную память заносятся исходные данные в виде V_{mnl}^t , где m -количество объектов, n -количество признаков, l -количество классов, а также значения системы порогов $\left[\begin{matrix} \hat{\delta}_{t1} & \hat{\delta}_{t2} \\ \hat{\delta}_{t1} & \hat{\delta}_{t2} \end{matrix} \right]$, ($i=1, n$).

2. Преобразуются значения признаков относительно системы порогов $\left[\begin{matrix} \hat{\delta}_{t1} & \hat{\delta}_{t2} \\ \hat{\delta}_{t1} & \hat{\delta}_{t2} \end{matrix} \right]$ в виде (1).

3. Определяется преобразованная матрица, состоящая из нулей и единиц в виде (2).

4. Проверяется для бинарного признака соотношение (3) или (4). Если для признака выполняется (3) или (4), то x_{ji}^t является простым признаком и алгоритм переходит к шагу 5. В противном случае - к шагу 6.

5. Определяется тип простого признака x_{ji}^t соотношением (3) или (4).

6. Вызывается следующий $(I+1)$ признак, а затем повторяются шаги (4÷5). Алгоритм переходит к шагу 8 тогда, когда число признаков $n \geq 2$.

7. Составляется подсистема признаков $x_{j1}^t, \dots, x_{jn}^t$, для которой затем проверяется соотношение (5) или (6) и определяется тип сложного признака $x_{j1}^t, \dots, x_{jn}^t$ в виде (5) или (6).

8. Определяется решающее правило $R^t(X_j^t)$ в виде (7÷10).

9. Вычисляется разделяющая сила для каждого решающего правила $R^t(X_j^t)$ в виде (11).

10. Определяется наилучшее решающее правило $R_{\gamma}^t(X)$ из множества $R_{\gamma}^t(X)$

,..., $R_p^t(X)$ для каждого элемента C_i в виде (12).

11. Находится распознающей правила $\tilde{R}_\gamma^t(X)$ из множества $R_\gamma^t(X), \dots, R_p^k(X)$, которое определяет наиболее приоритетный элемент в виде (13).

12. Данный алгоритм реализован средствами Турбо -Паскаль для ПЭВМ и проверен на решении практической задачи с реальным массивом данных.

REFERENCES

1. Jumanov, I. I., Karshiev, K. B., & Tishlikov, S. (2019). A Examination of the Efficiency of Algorithms for Increasing the Reliability of Information on Criteria of Harness and the Cost of Processing Electronic Documents. *International Journal of Recent Technology and Engineering (IJRTE)*, 28(2).
2. Jumanov, I. I. (2019). Karshiev Kh. B., Tishlikov SA Examination of the Efficiency of Algorithms for Increasing the Reliability of Information on Criteria of Harness and the Cost of Processing Electronic Documents. *International Journal of Recent Technology and Engineering (IJRTE) ISSN, 2277-3878*.
3. Ахатов, А. Р., & Тишликов, С. А. (2016). Методы повышения достоверности передачи и обработки информации в системах электронного документооборота на основе нечеткой семантической гиперсети. *Проблемы вычислительной и прикладной математики*, (3), 10-19.
4. Тишликов, С. А., Жуманов, И. И., & Юлдошев, О. (2012). Контроль достоверности информации электронных документов на основе нейронечеткой семантической гиперсети. *«Химическая технология. Контроль и управление»*, (6), 38.
5. Тишликов, С. А. (2013). ОТНОШЕНИЯ АВТОМАТИЗИРОВАННОГО ДЕЛОПРОИЗВОДСТВА НА ОСНОВЕ СТРУКТУРНО-ТЕХНОЛОГИЧЕСКОЙ ИЗБЫТОЧНОСТИ. In *Современные инструментальные системы, информационные технологии и инновации* (pp. 247-252).
6. Ахатов, А. Р., & Тишликов, С. А. (2016). Методы повышения достоверности передачи и обработки информации в системах электронного документооборота на основе нечеткой семантической гиперсети. *Проблемы вычислительной и прикладной математики*, (3), 10-19.