Impact Factor:	ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
	ISI (Dubai, UAE) = 0.829	РИНЦ (Russia) = 0.234	PIF (India) = 1.940
	GIF (Australia) = 0.564	ESJI (KZ) = 1.042	IBI (India) = 4.260
	JIF = 1.500	SJIF (Morocco) = 2.031	

SOI: 1.1/TAS DOI: 10.15863/TAS

International Scientific Journal

Theoretical & Applied Science

p-ISSN: 2308-4944 (print) **e-ISSN:** 2409-0085 (online)

Year: 2017 **Issue:** 02 **Volume:** 46

Published: 28.02.2017 **http://T-Science.org**

Ivan Borisovich Krylov
Head of Information Technology
Research Library, applicant
Orenburg State University,
Russia, Orenburg
krilovib@mail.ru

SECTION 4. Computer science, computer engineering and automation.

CALCULATION ALGORITHM OF TOTAL NUMBER OF QUOTATIONS UNDER THE CONDITIONS OF NONCOMPLETE DETERMINATION

Abstract: *It was reviewed the problem of total number of quotations calculation on the base of data from RSCI and SCOPUS under the conditions of noncomlete determination. It was offered the calculation algorithm of total number of quotations the base of which makes mathematical apparatus of fuzzy decision trees. Fuzzy decision tree was learnt on learning data set. The developed algorithm was researched on testing data set that allowed make conclusions about its usage for calculation of total number of quotations under the conditions of noncomplete determination.*

Key words: *scientometrical indexes, noncomplete determination conditions, total number of quotations calculation, fuzzy decision trees.*

Language: *English*

Citation: Krylov IB (2017) CALCULATION ALGORITHM OF TOTAL NUMBER OF QUOTATIONS UNDER THE CONDITIONS OF NONCOMPLETE DETERMINATION. ISJ Theoretical & Applied Science, 02 (46): 209-214.

Soi: http://s-o-i.org/1.1/TAS-02-46-36 ***Doi***: https://dx.doi.org/10.15863/TAS.2017.02.46.36

Introduction

At the moment none of the existing quotation systems individually displays the full picture about the authors' publication activity: not all list of publications and citations comes into account. One of the possible solutions of the problem is the building of aggregative list of publications on the base of data from different quotation systems and also the calculation of total number of quotations for each found publication. The special actuality this problem gets under the conditions of noncomplete determination specified by limited access to the quotation systems. One of the most popular foreign quotation systems is Scopus [1] and Russian is RSCI [2].

In this case under the conditions of noncomplete determination is understood the lack of possibility to associate quotations in SCOPUS system with the list of quotations in RSCI for each publication. In foreign quotation systems it's provided only the number of quotations on each chosen publication. There is the problem of calculation of total number of quotations for each publication. As the result it's getting impossible to calculate the main scientometrical indexes (citation index, Hirsch index [3] and others).

The aim of this work is the development and experimental researches of calculation algorithm of the total number of quotations on the base of data from RSCI and SCOPUS quotation systems under the conditions of noncomplete determination.

Materials and Methods

To calculate the total number of quotations for each publication the data used received at the stage of building of aggregative list of publications [4] precisely the number of quotations in RSCI, number of quotations in SCOPUS and also the number of found duplicated publications in SCOPUS system.

The base of calculation algorithm of total number of quotations under the conditions of noncomplete determination makes mathematical apparatus of fuzzy decision trees [5]. This mathematical apparatus combine decision trees and fuzzy logic advantages: allows operate quality characteristics of the subject; used in situations when it's difficult to classify the subject exactly according to any attribute позволяет; provides training on comparable small data set.

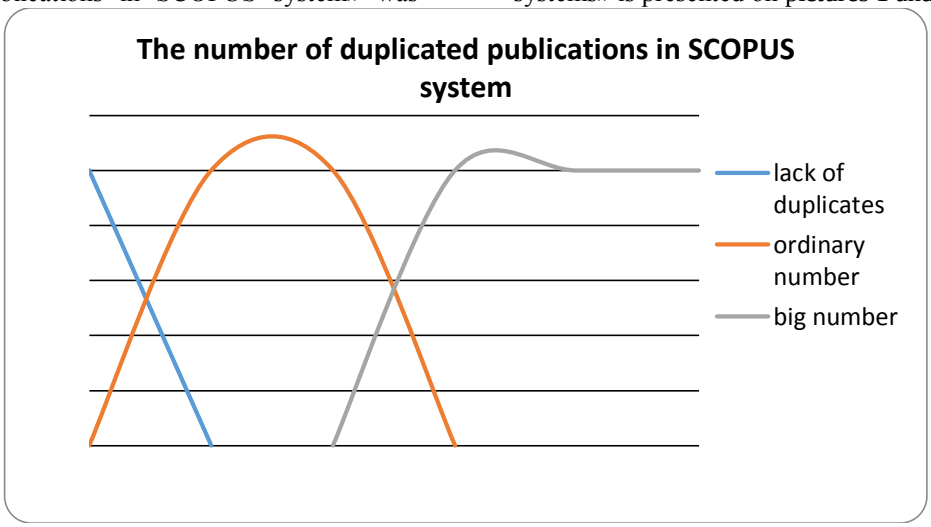
By building fuzzy decision tree for each attribute were selected some linguistic variables and defined examples membership degree. Instead of

number of examples for each knot fuzzy decision tree groups their membership degrees.

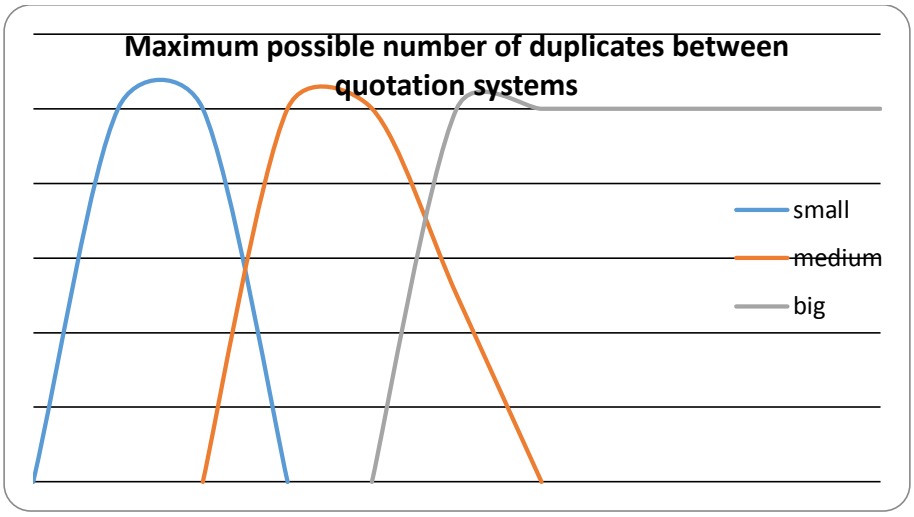
There were selected 2 target classes: «small portion of intersectional quotations» (negative result), «big portion of intersectional quotations» (positive result).

As attributes by which decision tree was built were selected the following: «the number of duplicated publications in SCOPUS system», «maximum possible number of duplicates between quotation systems». The attribute «the number of duplicated publications in SCOPUS system» was given by linguistic variable with the following term-set of meanings: «lack of duplicates», «ordinary number», «big number». The attribute «maximum possible number of duplicates between quotation systems» was given by linguistic variable with the following term-set of meanings: «small», «medium», «big».

The type of membership function [6, 7] for term-sets on attributes «the number of duplicated publications in SCOPUS system» and «maximum possible number of duplicates between quotation systems» is presented on pictures 1 and 2.



Picture 1 – Membership functions for term-sets on attribute «the number of duplicated publications in SCOPUS system».



Picture 2 – Membership functions for term-sets on attribute «maximum possible number of duplicates between quotation systems».

Numeric value of attribute «maximum possible number of duplicates between quotation systems» is calculated in the following way:

$$x_i^1 = \min(K_i^{SCOPUS}; K_i^{RISC}),$$

where K_i^{SCOPUS} - the number of quotations in SCOPUS system for the current publication, K_i^{RISC} - the number of quotations in RSCI system for the current publication.

For construction of fuzzy decision tree is used the algorithm consisting of several stages.

At the first stage of algorithm work general entropy is calculated.

In the following stage are calculated coefficients P for each possible node. The calculation of coefficients P for each node N is accomplished in the following way [8]:

$$P_i^N = \sum_{S^N} \min(\mu_N(D_j), \mu_i(D_j)),$$

where $\mu_N(D_j)$ – membership level of the training example D_j to the node *N*, $\mu_i(D_j)$ – membership level of training example toward objective value *i*, S^N - the variety of all examples.

Coefficient defining main characteristics of the node *N* is calculated in the following way:

$$P^N = \sum_i P_i^N ,$$

In the following stage is calculated the entropy that estimates the average number of information to determine the object class from the set P^N:

$$E(S^N) = -\sum_i \frac{P_i^N}{P^N} \cdot \log_2 \frac{P_i^N}{P^N}.$$

Then the entropy for each attribute individually is calculated:

$$E(S^N, A) = \sum_j \frac{P^{N|j}}{P^N} \cdot E(S^{N|j}),$$

where *N*|*j* – child of node *N*.

Then the information gain on each attribute is calculated:

$$G(S^N, A) = E(S^N) - E(S^N, A).$$

Finally as root attribute is chosen the attribute with the maximum information gain.

Then node *N* is devided into subnodes *N*|*j*. Membership level of each example D^k for node *N*|*j* is calculated from node *N*:

$$\mu_{N|j}(e_k) = \min(\mu_{N|j}(D^k); \mu_{N|j}(D^k, a_j)),$$

where $\mu_{N|j}(D^k, a_j)$ demonstrates the membership level D^k to the attribute a_j. In case if none of the examples belongs to the node *N*|*j*, this node is deleted.

The algorithm work continues until all the attributes are used or all the examples aren't classified.

The membership to the target class for the new recording is found in the following way:

$$\delta_j = \frac{\sum_l \sum_k P_k^l \cdot \mu_l(D_j) \cdot x_k}{\sum_l (\mu_l(D_j) \cdot \sum_k P_k^l},$$

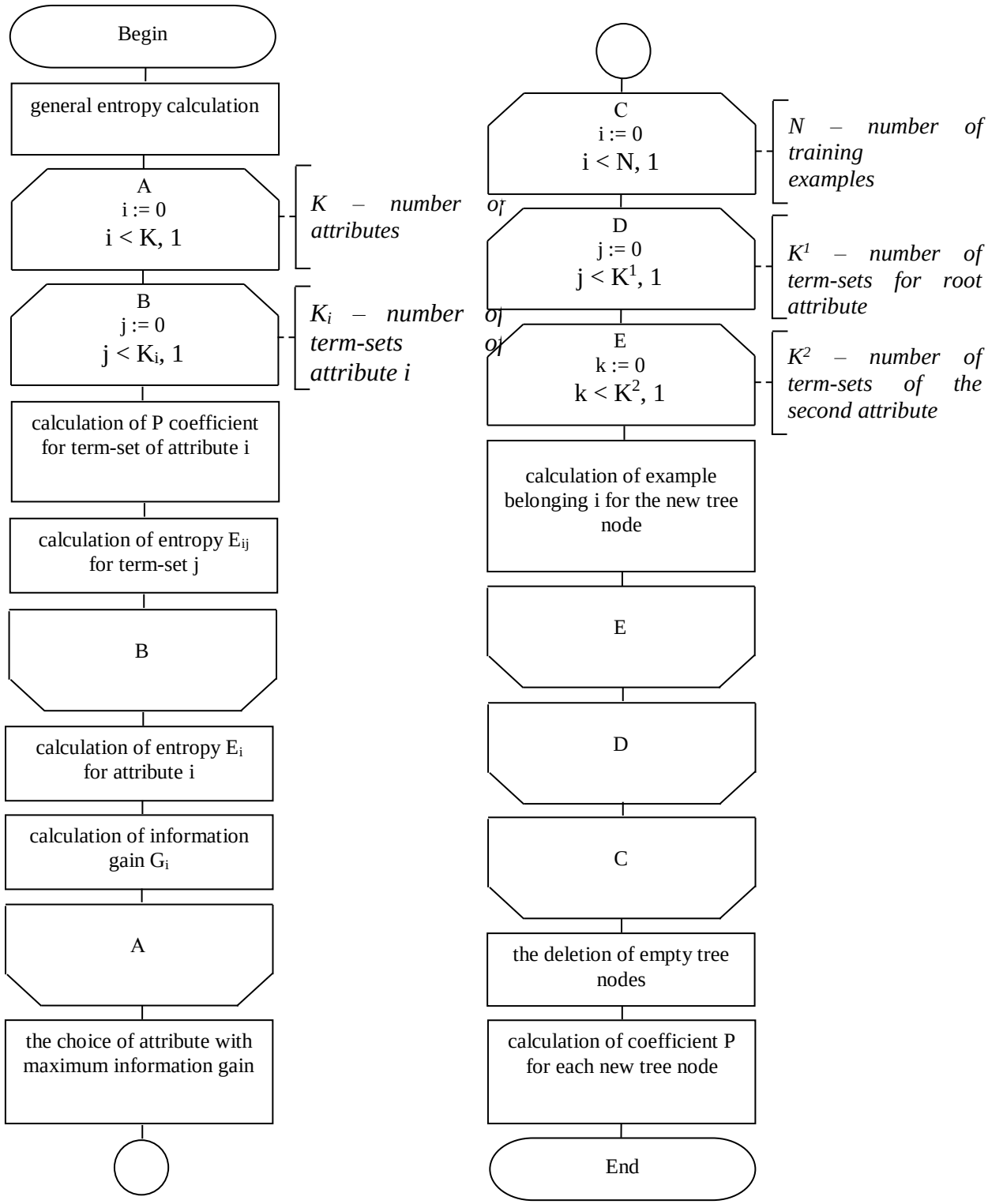
where P_k^l - coefficient displaying the correlation of examples *l* for target class values *k*, $\mu_l(D_j)$ - membership level of example *j* to the node *l*, x_k - membership of the target class value *k* to the positive result.

The calculation of total number of quotations is made in the following way:

$$S_i = K_i^{SCOPUS} + K_i^{RISC} - \min(K_i^{SCOPUS}; K_i^{RISC}) \cdot \delta_i$$

.

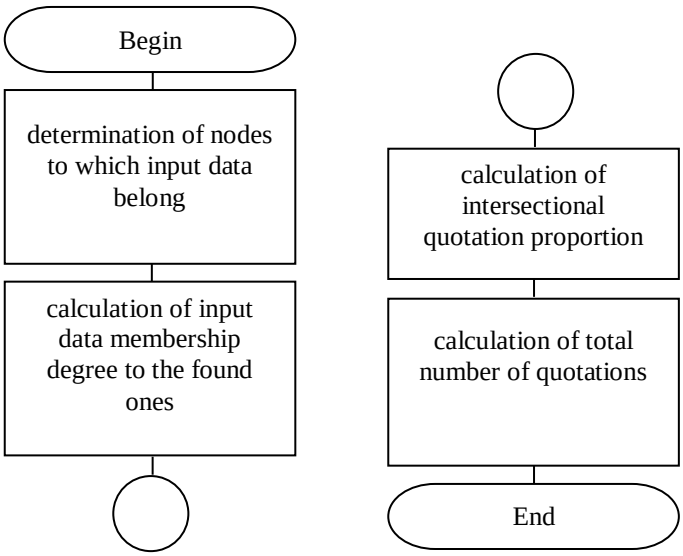
The algorithm of building of fuzzy decision tree [9, 10] on the base of training set is presented on picture 3.



Picture 3 – Algorithm of building of fuzzy decision tree.

The generalized chart of calculating algorithm of total number of quotations by means of built decision tree is presented on picture 4.

.

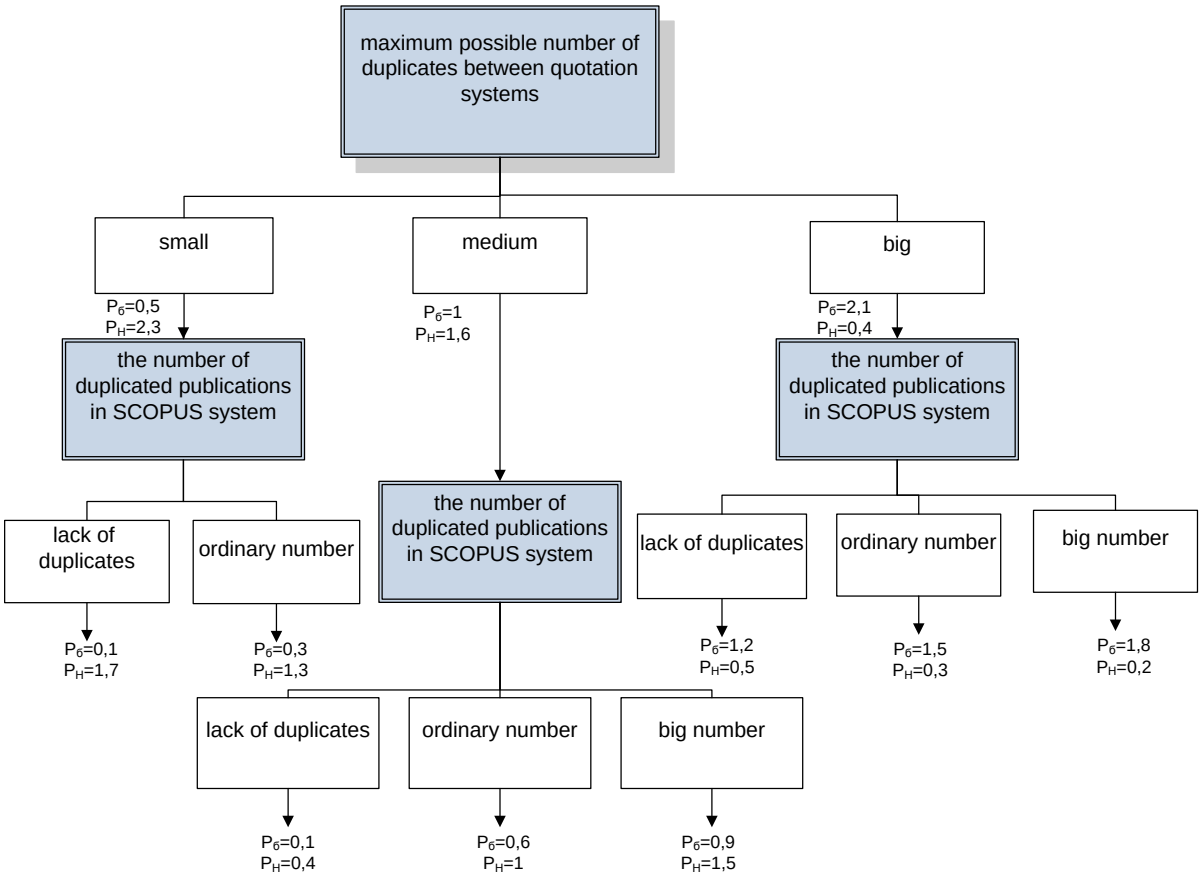


Picture 4 – Generalized chart of calculating algorithm of total number of quotations.

For analyses of algorithm work results on calculation of total number of quotations for each publication at first stage was carried out the building of fuzzy decision tree on the base of training set.

Training set was formed on the base of publications list of 5 leading authors of Orenburg State University having fairly large number of duplicated publications in RSCI and SCOPUS quotation systems.

Constructed fuzzy decision tree is presented on picture 5.



Picture 5 – Constructed fuzzy decision tree.

At the next stage trained tree was tested. The test set was formed on the base of list of publications of 3 leading authors of Orenburg State University that aren't included in training set and having fairly large number of duplicated publications in RSCI and SCOPUS systems.

Conclusion

As result of algorithm work in training set were received the following results: for more than 82% publications were noticed the coincidence of total number of quotations calculated by means of developed and real algorithm; for 18% of publications calculated total number of quotations by means of developed algorithm differed slightly from real value.

Thus it may be concluded that it's possible to use the developed algorithm for calculating of total number of quotations on the base of data from RSCI and SCOPUS quotation systems.

References:

1. (2017) Scopus: citation database of peer-reviewed literature. Available: http://www.scopus.com/ (Accessed: 03.02.2017).
2. (2017) eLIBRARY.RU: nauchnaja jelektronnaja biblioteka. Available: http://elibrary.ru (Accessed: 03.02.2017).
3. Hirsch JE (2005) «An index to quantify an individual's scientific research output». Proceedings of the National Academy of Sciences. N. 102, p. 16569-16572.
4. Krylov IB, Boldyrev PA (2015) Several characteristics of existing automated systems according to survey of russian scientists publishing activity. Theoretical & Applied Science. N. 5 (25), p. 6–9.
5. Janikow CZ (1998) Fuzzy Decision Trees: Issues and Methods. IEEE transactions on systems, man, and cybernetics – part b: Cybernetics. V. 28, N. 1, p. 1-14.
6. Shtovba SD (2017) Vvedenie v teoriju nechetkih mnozhestv i nechetkuju logiku. 11. Available: http://matlab.exponenta.ru/fuzzylogic/book1/2.php (Accessed: 03.02.2017).
7. Rutkovskaja D, Pilin'skij M, Rutkovskij L. (2006) Nejronnye seti, geneticheskie algoritmy i nechetkie sistemy. Gorjachaja linija-Telekom. 452 p.
8. Jusupova NI, Ahmetova JF, Bogdanova DR (2013) Klassifikacija klientov na osnove nechjotkoj informacii. Vestnik UGATU. N. 5 (58), p. 93-100.
9. Chernov VG (2010) Nechjotkie derev'ja reshenij (nechjotkie pozicionnye igry). Informacionno-upravljajushhie sistemy. N. 5, p. 8-14.
10. Kul'chin JN, Kim J, Notkin BS, Ljuhter AB (2014) Postroenie algoritma nechjotkogo dereva reshenij na osnove jeksperimental'nyh dannyh pri obrabotke signalov RVOIS. Informatika i sistemy upravlenija. N. 3 (41), p. 103-114.