New Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics
Volume II: Applications

Editors

Krassimir T. Atanassov
Władysław Homenda
Olgierd Hryniewicz
Janusz Kacprzyk
Maciej Krawczak
Zbigniew Nahorski
Eulalia Szmidt
Sławomir Zadrożny

SRI PAS IBS PAN

New Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. Volume II: Applications

New Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics Volume II: Applications

Systems Research Institute
Polish Academy of Sciences

New Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics
Volume II: Applications

Editors

Krassimir T. Atanassov
Władysław Homenda
Olgierd Hryniewicz
Janusz Kacprzyk
Maciej Krawczak
Zbigniew Nahorski
Eulalia Szmidt
Sławomir Zadrożny

IBS PAN SRI PAS

Systems Research Institute
Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland
www.ibspan.waw.pl

ISBN 83-894-7541-3

Dedicated to Professor Beloslav Riečan on his 75th anniversary

Dimension reduction of time series for clustering problem

Maciej Krawczak[1, 2] and Grażyna Szkatuła[1]

[1]Systems Research Institute, Polish Academy of Sciences,
Newelska 6, Warsaw, Poland
[2] Warsaw School of Information Technology,
Newelska 6, Warsaw, Poland
{krawczak, szkatulg}@ibspan.waw.pl

Abstract

In this paper we considered time series dimension reduction and clustering. The techniques of reduction of dimension of time series is based on the concept of upper and lower envelopes, aggregation of the envelopes and extracting essential attributes. Essential attributes were nominalized. The reduced representation of time series is characterized by nominal attributes. For such representation of time series we applied a definition of conditions domination within each pair of clusters. The developed hierarchical and agglomerative method is characterized both by high speed of computation as well as extremely good accuracy of clustering was applied to solve a clustering problem of time series data.

Keywords: time series, dimension reduction, nominal attributes, cluster analysis.

1 Introduction

Nowadays there is a problem of time series clustering. Time series data are characterized by huge number of objects and each object is characterized by a large number of attributes. Analysing such data very often we have to overcome the *curse of dimensionality* of the problem.

In this paper we will consider a time series reduction problem based on idea of envelopes developed by Krawczak and Szkatuła (2008) as well as a technique of essential attributes introduced by Krawczak and Szkatuła (2010a, 2010b).

New Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. Volume II: Applications *(K.T. Atanassov, W. Homenda, O. Hryniewicz, J.Kacprzyk, M. Krawczak, Z. Nahorski, E Szmidt, S. Zadrożny, Eds.), IBS PAN - SRI PAS, Warsaw, 2012.*

The values of the new reduced attributes were changed into intervals, and then the primary symbolic values were assigned to the ordinal intervals (Krawczak and Szkatuła, 2010c). In order to obtain the nominal representation of the considered time series we called each difference of two primary symbols by different letters of the alphabet.

This way we were able to reduce the high dimension of each object to representation characterized by a short sequence of nominal symbols.

The original time series data were remade and now each time series is represented by few nominal attributes instead of long string of numbers.

In order to cluster reduced dimension time series, represented by nominal attributes, we developed a new technique based on relation of dominance between clusters.

One can find several algorithms specialized to analysis of long chains of symbols. The algorithms found applications in text analysis or in bioinformatics (Apostolico *et al.*, 2002), Gionis and Mannila, 2003), (Lin *et al.,* 2007). These algorithms are based on some measure of distance between objects, e.g. Wang (2010).

Our algorithm to clustering nominal data is different than those known in the literature, additionally, it seems that efficiency of the new algorithm is also better.

However, the clustering algorithm has several features common with standard ones, for example our algorithm is hierarchical and agglomerative ("bottom-up"). The hierarchical clustering (defined by Johnson in 1967) starts with *N* single object clusters and ends reaching the prescribed number of clusters. This kind of algorithm allows finding the most similar pair of clusters and merge into a new single cluster.

In our algorithm we introduced a definition of the condition's dominance which allowed merging smaller clusters in order to get larger ones.

The algorithm was applied to solve a clustering problem of time series data available at the Irvine University of California (Alcock, Manolopoulos, 1999). The result of clustering confirmed the efficiency of the developed clustering algorithm.

2 Symbolic representation of data series

From the available database we selected 75 time series (objects) under the following assumption, namely 25 objects represented Class I, next 25 objects represented Class II and the rest 25 objects represented Class III. The affiliation of the chosen objects was not used during the clustering process.

In order to reduce time series dimensionality the following procedure was applied.

First, each object was normalized to have a mean equal to zero and a standard deviation equal to one.

In results we obtained the following data $\left[x_1(n), x_2(n), \ldots, x_{60}(n)\right]$, for $n = 1, 2, \ldots, 75$. It means that the number of considered object is $N = 75$, while the dimension of each object is $M = 60$.

Application of the approach developed by Krawczak and Szkatuła (2008, 2010a, 2010b) allows generating the *m*-step upper envelopes and/or *m*-step lower envelopes, where *m* denotes a number of sequent data values, under the assumptions that $m << M$, here it was taken $m = 4$.

Aggregation of the envelopes (Krawczak and Szkatuła, 2008, 2010a, 2010b) reduced $m = 4$ times dimension of each time series. In results we obtained the reduced form of the envelopes, now the dimension of each envelop (representing the object) is equal $\left\lfloor \frac{M}{m} \right\rfloor = 15$.

In order to get the further reduction of the objects representations the essential attributes were extracted. The heteroassociative neural network was applied to obtain *E* essentials attributes $\{b_j(n)\}_{j=1}^{j=E}\}$, where $E << \left\lfloor \frac{M}{m} \right\rfloor$, here it was assumed that $E = 5$.

In the next step, the new attributes are created on the base of the essential attributes as rearrangements of all differences of the essential attributes. This way we slightly enlarge dimensionality of the data series representation, but in the same time we provided in some sense the distances between the essential attributes.

Now each object is represented by 60 data points and the set $\{c_j(n)\}_{j=1}^{j=10}$ constitutes the new representation of e.g. lower envelopes of the time series. These new attributes of dimension $K = 10$ replaced the essential attributes.

In this step of procedure the real values of the attributes $\{c_j(n)\}_{j=1}^{j=10}\}$, $n = 1, 2, \ldots, 75$, are replaced by nominal values. The replacement is done in such a way that the ranges of the attributes are divided into some number of elements. Here the problem is calculated with the method called equal width interval discretization, which involves determining the domain of observed values of an attribute, and dividing this interval into equally size intervals. We obtained the nominal representation of the time series in the following form $\{a_j(n)\}_{j=1}^{j=10}\}$, $n = 1, 2, \ldots, 75$.

This way the data series $[x_1(n), x_2(n), \ldots, x_{60}(n)]$, for $n = 1, 2, \ldots, 75$, was replaced by nominal value of the attributes $\{a_1(n), a_2(n), \ldots, a_{10}(n)\}$, and can now be considered as data for clustering method for the nominal attributes.

After the application of the above described procedure we obtained the symbolic representation of the time objects in the following form:

There is a finite set of data objects $U = \{e^n\}$, $n = 1, 2, \ldots, N$. The objects are described in the form of conditions associated with the finite set of attributes $A = \{a_1, \ldots, a_K\}$. The set $V_{a_j} = \{v_{j,1}, v_{j,2}, \ldots, v_{j,L_j}\}$ is the domain of the attribute $a_j \in A$, $j = 1, \ldots, K$, where L_j - denotes number of values of the *j*-th attribute.

Each object $e^n \in U$ can be described in the form of conjunction of K elementary conditions in the following manner

$$e^n = (a_1 \in \{v_{1,t(1,n)}\}) \wedge \ldots \wedge (a_K \in \{v_{K,t(K,n)}\}) \tag{1}$$

where $v_{j,t(j,n)} \in V_{a_j}$ and $j = 1, \ldots, K$. The index $t(j,n)$ for $j \in \{1, 2, \ldots, K\}$ and $n \in \{1, 2, \ldots, N\}$ denotes that the attribute a_j takes value $v_{j,t(j,n)}$ in the object e^n.

For example, for the *j*-th attribute the set $V_{a_j} = \{v_{j,1}, v_{j,2}, \ldots, v_{j,L_j}\}$, using letters of the alphabet, can have the following symbolic form for $L_j = 9$

$$V_{a_j} = \{a, b, c, d, e, f, g, h, i\}.$$

An exemplary data object for a given $n \in [1, N]$ can be written as follows:

$$\begin{aligned} e^n = [&(a_1 \in \{b\}) \wedge (a_2 \in \{d\}) \wedge (a_3 \in \{f\}) \wedge (a_4 \in \{c\}) \wedge (a_5 \in \{e\}) \wedge \\ &(a_6 \in \{f\}) \wedge (a_7 \in \{c\}) \wedge (a_8 \in \{k\}) \wedge (a_9 \in \{a\}) \wedge (a_{10} \in \{g\})] \end{aligned}$$

3 Basic elements of the approach

The task of clustering can be formulated as follows: we want to splits the set of objects U into non-empty, disjoint subsets $\{C_1, C_2, ..., C_C\}$, $\bigcup_{g=1}^{C} C_g = U$, (called *clusters*) so that objects in the same cluster are similar in some sense. The set of clusters on U is denoted by *C(U)*. If a certain object belongs to a definite cluster then it could not be included in another cluster, by assumption. Basic elements of proposed method were introduced below.

Consider an attribute a_j, $j = 1,...,K$ and no empty sets $A_{j,t(j,k)}$ and $A_{j,t(j,n)}$, where $A_{j,t(j,k)} \subseteq V_{a_j}$, $A_{j,t(j,n)} \subseteq V_{a_j}$.

We say that the condition $(a_j \in A_{j,t(j,k)})$ *dominates* the condition $(a_j \in A_{j,t(j,n)})$ if the clause $A_{j,t(j,k)} \supseteq A_{j,t(j,n)}$ is satisfied, denoted by $(a_j \in A_{j,t(j,k)}) \succeq (a_j \in A_{j,t(j,n)})$.

Let us notice that condition $(a_j \in \{a, b, f\})$ dominates the condition $(a_j \in \{a, f\})$, i.e. $(a_j \in \{a, b, f\}) \succeq (a_j \in \{a, f\})$.

We assume that there is *lack of mutual dominance* two conditions $(a_j \in A_{j,t(j,k)})$ and $(a_j \in A_{j,t(j,n)})$ if the first condition does not dominates the second and the second condition does not dominates the first, denoted by $(a_j \in A_{j,t(j,k)}) \prec\succ (a_j \in A_{j,t(j,n)})$.

Let us notice that there is lack of mutual dominance of two conditions $(a_j \in \{a, b, f\})$ and $(a_j \in \{a, c\})$, i.e. $(a_j \in \{a, b, f\}) \prec\succ (a_j \in \{a, c\})$.

The *cluster* C_g can be expressed as follows:

$$(a_1 \in A_{1,t(1,g)}) \wedge ... \wedge (a_K \in A_{K,t(K,g)}) \tag{2}$$

where $A_{j,t(j,g)} \subseteq V_{a_j}$, for $j = 1,...,K$.

We say that the object $e^n = (a_1 \in \{v_{1,t(1,n)}\}) \wedge ... \wedge (a_K \in \{v_{K,t(K,n)}\})$ *belongs to the cluster* C_g if conditions:

$$\begin{gathered} (a_1 \in A_{1,t(1,g)}) \succeq (a_1 \in \{v_{1,t(1,n)}\}) \\ \cdots \\ (a_K \in A_{K,t(K,g)}) \succeq (a_K \in \{v_{K,t(K,n)}\}) \end{gathered} \tag{3}$$

are satisfied.

Let's consider two clusters: $C_{g_1} : (a_1 \in A_{1,t(1,g_1)}) \wedge ... \wedge (a_K \in A_{K,t(K,g_1)})$ and $C_{g_2} : (a_1 \in A_{1,t(1,g_2)}) \wedge ... \wedge (a_K \in A_{K,t(K,g_2)})$, for $A_{j,t(j,g_1)} \subseteq V_{a_j}$, $A_{j,t(j,g_2)} \in V_{a_j}$, $j = 1,...,K$.

We say that the cluster C_{g_1} and the cluster C_{g_2} are ω*-distinguishable* for the set of attributes $\{a_j : j \in I_k\}$, $card(I_k) = \omega$, if two conditions are satisfied:

1) $(a_j \in A_{j,t(j,g_1)}) \prec\succ (a_j \in A_{j,t(j,g_2)})$, $\forall j \in I_k$

(4)

2) $(a_j \in A_{j,t(j,g_1)}) \succeq (a_j \in A_{j,t(j,g_2)})$ or $(a_j \in A_{j,t(j,g_2)}) \succeq (a_j \in A_{j,t(j,g_1)})$,
$\forall j \in \{1,...,K\} \setminus I_k$.

We assume that the cluster $C_{g_1} : \{e^n :\ e^n \in U,\ n \in J_{g_1} \subseteq \{1,2,...,N\}\}$ and the cluster $C_{g_2} : \{e^n :\ e^n \in U,\ n \in J_{g_2} \subseteq \{1,2,...,N\}\}$ are ω-distinguishable for the set of attributes $\{a_j : j \in I_k\}$, $card(I_k) = \omega$.

The ω - *conditional action rule* is defined in the following manner:

$$\begin{aligned}
&\underset{j\in I_k}{\wedge}(A_{j,t(j,g_3)} := A_{j,t(j,g_1)} \cup A_{j,t(j,g_2)}) \text{ and} \\
&\underset{j\in\{1,2,..K\}\setminus I_k}{\wedge}(A_{j,t(j,g_3)} :=\ dom\ \{A_{j,t(j,g_1)}, A_{j,t(j,g_2)}\}) \qquad (5)\\
&\Rightarrow ((C_{g_1}, C_{g_2}) \to (C_{g_3}))
\end{aligned}$$

where *dom* - dominant condition.

In result a new cluster $C_{g_3} : (a_1 \in A_{1,t(1,g_3)}) \wedge ... \wedge (a_K \in A_{K,t(K,g_3)})$ contains the following objects $\{e^n :\ e^n \in U,\ n \in J_{g_1} \cup J_{g_2}\}$.

We proposed a hierarchical agglomerative approach to cluster nominal data. The bottom level of the structure has singular clusters while the top level contains one cluster with all objects. During iteration two clusters are heuristically selected. These selected clusters are then merged to form a new cluster.

Suppose we have a finite set of objects $U = \{e^n\}$, $n = 1, 2, \ldots, N$.

The objects are described in the form of conditions associated with the finite set of attributes. We want to splits the set of objects U into non-empty, disjoint subsets $\{C_1, C_2, ..., C_C\}$, $\bigcup_{g=1}^{C} C_g = U$.

Basic elements of proposed algorithm were introduced below.

Step 1. U – set of objects, K - number of attributes, C - waited number of clusters. Each object creates one-element cluster in the initial set of clusters *C(U)*, $card(C(U)) = N$, $\omega := 0$.

Step 2. From a pair of ω-distinguishable clusters we create new cluster. If $card(C(U)) = C$, go to Step 4; otherwise, if it exists pair of at the most ω-distinguishable clusters, repeat Step 2; otherwise, go to Step 3.

Step 3. $\omega := \omega + 1$; if $\omega \leq K$, repeat Step 2; otherwise, go to Step 4.

Step 4. STOP.

4 Experimental results

For the reduced representation of time series prepared according to the procedure described in Section 2 the proposed method introduced in Section 3 was applied in order to cluster the data set $U = \{e^n\}$, $n = 1, 2, \ldots, 75$, where the objects are described by ten nominal attributes $A = \{a_1, ..., a_{10}\}$, and the set $V_{a_j} = \{a, b, c, d, e, f, g, h, i, j\}$ is the domain of each attributes $a_j \in A$, $j = 1, ..., 10$. The values of the attributes of the objects are shown in Table 1.

Table 1

No.	$a_1(n)$	$a_2(n)$	$a_3(n)$	$a_4(n)$	$a_5(n)$	$a_6(n)$	$a_7(n)$	$a_8(n)$	$a_9(n)$	$a_{10}(n)$
1	*f*	*f*	*d*	*i*	*f*	*d*	*g*	*e*	*h*	*h*
...	...	...	...	...	...	...	...	...	...	...
40	*e*	*e*	*h*	*b*	*e*	*h*	*d*	*h*	*d*	*d*
...	...	...	...	...	...	...	...	...	...	...
48	*h*	*f*	*f*	*b*	*h*	*g*	*c*	*i*	*c*	*e*
...	...	...	...	...	...	...	...	...	...	...
55	*f*	*e*	*g*	*f*	*e*	*g*	*g*	*g*	*g*	*g*
...	...	...	...	...	...	...	...	...	...	...
75	*d*	*g*	*e*	*g*	*f*	*g*	*f*	*e*	*h*	*f*

Each object e^n is represented by ten elementary conditions in the following manner: $e^n = (a_1 \in \{v_{1,t(1,n)}\}) \wedge \ldots \wedge (a_1 \in \{v_{10,t(10,n)}\})$, where $v_{j,t(j,n)} \in V_{a_j}$ and $j = 1, ..., 10$, e.g. $e^1 = (a_1 \in \{f\}) \wedge (a_2 \in \{f\}) \wedge (a_3 \in \{d\}) \wedge \ldots \wedge (a_{10} \in \{h\})$.

Our goal is to partition the set of the objects U into three, non-empty, disjoint clusters $C(U)=\{C_{g_1},C_{g_2},C_{g_3}\}$, where $\bigcup_{i=1}^{3} C_{g_i}=U$, $C_{g_u}\cap C_{g_w}=\varnothing$, for $u,w\in\{1,2,3\}$, $u\neq w$.

The result of clustering the data set is shown in Table 2.

Table 2

C	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
C_{g_1}	*d, e, f*	*f, g, e*	*b, c, d, f, e*	*h, i, ,j*	*e, f*	*f, d, e*	*i, g, h*	*e, d, f*	*i, h*	*h, i*
C_{g_2}	*e, f, g, h*	*d, e, f*	*h, f*	*b, a, c*	*c, e, g, h*	*h, g*	*d, c*	*g, h, i*	*d, c, e*	*c, d, e*
C_{g_3}	*d, e, f*	*d, g, h, e*	*e, g, h*	*g, f, e, d*	*f, d, e*	*g*	*f, e, g*	*e, f, g, h*	*h, f, g*	*f, g*

Three clusters obtained in the space of two attributes: a_4 and a_8 is shown in Figure 1.

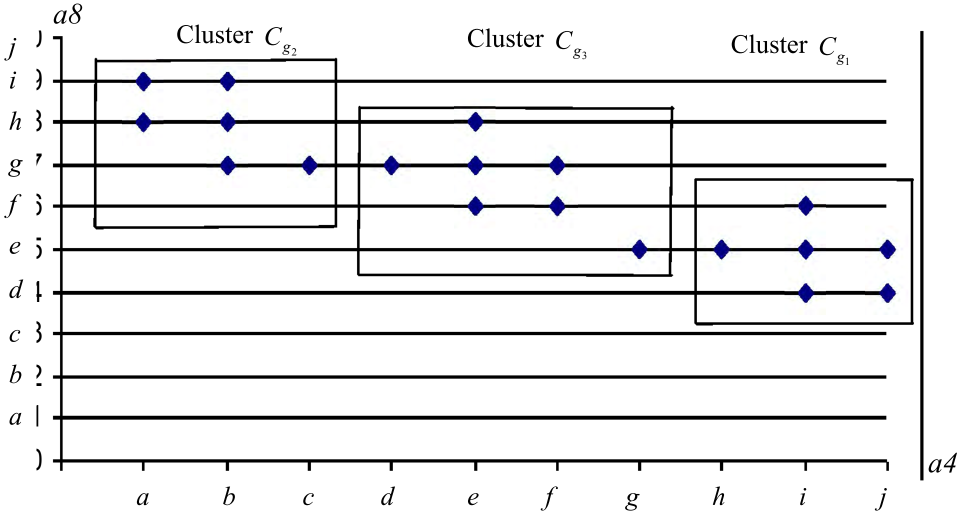

Figure 1: Three clusters obtained in the space of two attributes: a_4 and a_8

It is worth to notice that on the basis of Table 2 it is possible to create descriptions of clusters in the form of decision rules. These descriptions can be represented e.g. in the form of single-condition rules as follows

IF *certain condition is fulfilled,*
THEN *membership in a definite cluster takes place*.

In our case, the conditional part of the rules will contain the disjunction of conditions related to the subset of attributes selected for the description of the objects. The exemplary single-condition decision rules are shown below:

IF $(a_4 \in \{h,i,j\})$ THEN C_{g_1},
IF $(a_7 \in \{d,c\})$ THEN C_{g_2},
IF $(a_{10} \in \{f,g\})$ THEN C_{g_3}.

Consequently, the data series have been grouped into three clusters C_{g_1}, C_{g_2} and C_{g_3}. It must be emphasized that all 75 objects have been separated into three groups according to their affiliation to three different classes, so the clustering efficiency of the proposed methodology is 100%.

5 Conclusions

In this paper we considered time series data represented by a huge dimensionality. In order to reduce dimensionality of time series the new representation with nominal attributes of time series was obtained.

For data set described by nominal attributes we introduced and developed the algorithm based on the idea of dominations of conditions within each pair of cluster. Each cluster is described by a conjunction of conditions associated with attributes describing objects.

The solved example showed that the efficiency of the proposed methodology appeared effective in 100 per cent.

Acknowledgements

The research was partially supported by the Ministry of Science and Higher Education (Poland) under Grant Nr N N519 384936.

References

[1] Alcock, R. J., Manolopoulos, Y. (1999) Time-Series Similarity Queries Employing a Feature-Based Approach. 7th Hellenic Conference on Informatics, Ioannina, Greece.

[2] Apostolico R., Bock M. E., Lonardi S. (2002). Monotony of surprise in large-scale quest for unusual words. In: Proceedings of the 6th International conference on research in computational molecular biology, Washington, DC, April 18-21, 22-31.

[3] Gionis A., Mannila H. (2003). Finding recurrent sources in sequences. In: Proceedings of the 7th International conference on research in principles of database systems, Tucson, AZ, May 12-14, 249-256.

[4] Johnson S. C. (1967). Hierarchical Clustering Schemes, *Psychometrika*, 2:241-254.

[5] Krawczak M., Szkatuła G. (2008). On decisions rules application to time series classification. In: Atanassov K.T., Hryniewicz O., Kacprzyk J., Krawczak M., Nahorski Z., Szmidt E., Zadrożny S. (Eds.): Advances in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. Ac. Publ. House EXIT, Warsaw 2008, 95-104.

[6] Krawczak M., Szkatuła G. (2010a). On time series envelopes for classification problem. Developments of fuzzy sets, intuitionistic fuzzy sets, generalized nets, vol. II, 2010

[7] Krawczak M., Szkatuła G. (2010b). Time series envelopes for classification. In: Proceedings of the conference: 2010 IEEE International Conference on Intelligent Systems, London, UK, July 7-9 2010, 156-161.

[8] Krawczak M., Szkatuła G. (2010c). Redukcja wymiarowości szeregów czasowych. Studia i materiały Polskiego Stowarzyszenia Wiedzą, No. 31, 32-45.

[9] Lin J., Keogh E., Wei L., Lonardi S. (2007). Experiencing SAX: a Novel Symbolic Representation of Time Series. Data Min Knowledge Disc, 2, 15, 107–144.

[10] Nanopoulos A., Alcock R., & Manolopoulos Y. (2001). Feature-based Classification of Time-series Data. International Journal of Computer Research, 49-61.

[11] Wang B. (2010).A New Clustering Algorithm on Nominal Data Sets. *Proceedings of International MultiConference of Engineers and Computer Scientists 2010 IMECS 2010,* March 17-19, 2010, Hong Kong

[12] Wei L., Keogh E. (2006). Semi-Supervised Time Series Classification. In: *Proc. of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006),* 748 - 753, Philadelphia, PA, U.S.A., August 20-23, 2006.

The papers presented in this Volume 2 constitute a collection of contributions, both of a foundational and applied type, by both well-known experts and young researchers in various fields of broadly perceived intelligent systems.
It may be viewed as a result of fruitful discussions held during the Tenth International Workshop on Intuitionistic Fuzzy Sets and Generalized Nets (IWIFSGN-2011) organized in Warsaw on September 30, 2011 by the Systems Research Institute, Polish Academy of Sciences, in Warsaw, Poland, Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences in Sofia, Bulgaria, and WIT - Warsaw School of Information Technology in Warsaw, Poland, and co-organized by: the Matej Bel University, Banska Bystrica, Slovakia, Universidad Publica de Navarra, Pamplona, Spain, Universidade de Tras-Os-Montes e Alto Douro, Vila Real, Portugal, and the University of Westminster, Harrow, UK:

Http://www.ibspan.waw.pl/ifs2011

The consecutive International Workshops on Intuitionistic Fuzzy Sets and Generalized Nets (IWIFSGNs) have been meant to provide a forum for the presentation of new results and for scientific discussion on new developments in foundations and applications of intuitionistic fuzzy sets and generalized nets pioneered by Professor Krassimir T. Atanassov. Other topics related to broadly perceived representation and processing of uncertain and imprecise information and intelligent systems have also been included. The Tenth International Workshop on Intuitionistic Fuzzy Sets and Generalized Nets (IWIFSGN-2011) is a continuation of this undertaking, and provides many new ideas and results in the areas concerned.

We hope that a collection of main contributions presented at the Workshop, completed with many papers by leading experts who have not been able to participate, will provide a source of much needed information on recent trends in the topics considered.

ISBN-13 9788389475411

9 788389 475411