# AIR QUALITY CLASSIFICATION BY INTUITIONISTIC FUZZY RELATIONS AND KOHONEN'S SELF-ORGANIZING FEATURE MAPS

Vladimír Olej, Petr Hájek
Univerzita Pardubice, Fakulta ekonomicko-správní, Ústav systémového inženýrství a informatiky

***Abstract:*** *The paper presents a design of parameters for air quality classification of districts into classes according to their pollution. Therefore, the design of air quality classification is presented by intuitionistic fuzzy relations. Intuitionistic fuzzy relations and their composition were designed on the basis of Kohonen's self-organizing feature maps.*

***Keywords:*** *Air quality, Kohonen's self-organizing feature maps, K-means algorithm, intuitionistic fuzzy relations, classification.*

## 1. Introduction

The paper presents the parameters design for air quality modelling. Therefore, the data matrix **U** is designed where the vectors $\mathbf{p}_i^t$ characterize the districts $o_i^t \in O$. Next, the paper introduces basic notions of the intuitionistic fuzzy sets IFSs [1], intuitionistic fuzzy relations (IFRs), and Sanchez's approach [2], [3] for classification of the i-th district $o_i^t \in O$ to the j-th class $\omega_{i,j}^t \in \Omega$. The method of intuitionistic classification involves IFRs as defined in [4]. Further, the paper presents the basic concepts of Kohonen's self-organizing feature maps (KSOFMs) for the IFRs design. Moreover, the classification of the i-th district $o_i^t \in O$ to the j-th class $\omega_{i,j}^t \in \Omega$ presented in the paper assists state administration to evaluate air quality. The knowledge of notable experts in the field of air quality measuring gives support to the results of the classification.

## 2. Parameters Design for Air Quality Classification

Harmful substances in the air represent the parameters of air quality modelling. They are defined as the substances emitted into the external air or formed secondarily in the air which harmfully influent the environment directly, after their physical or chemical transformation or eventually in interaction with other substances. Except the harmful substances, other components also influence the overall air pollution. Both, the parameters concerning the harmful substances in the air and the meteorological parameters influence the air quality development. The design of parameters can be realized as presented in Tab. 1. Based on the presented facts, the following data matrix **P** can be created

$$\mathbf{P}=\begin{array}{c|cccccc:c}
 & p_1^t & \cdots & p_k^t & \cdots & p_m^t & \omega_{i,j}^t \\
\hline
o_1^t & p_{1,1}^t & \cdots & p_{1,k}^t & \cdots & p_{1,m}^t & \omega_{1,j}^t \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
o_i^t & p_{i,1}^t & \cdots & p_{i,k}^t & \cdots & p_{i,m}^t & \omega_{i,j}^t \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
o_n^t & p_{n,1}^t & \cdots & p_{n,k}^t & \cdots & p_{n,m}^t & \omega_{n,j}^t
\end{array},$$

where $o_i^t \in O$, $O=\{o_1^t, o_2^t, \ldots, o_i^t, \ldots, o_n^t\}$ are districts in time t, $p_k^t$ is the k-th parameter in time t, $t \in \{2002 \text{ till } 2007\}$, $p_{i,k}^t$ is the value of the parameter $p_k^t$ for the i-th object $o_i^t \in O$, $\omega_{i,j}^t$ is the j-th class assigned to the i-th object $o_i^t \in O$, $\mathbf{p}_i^t = (p_{i,1}^t, p_{i,2}^t, \dots, p_{i,k}^t, \dots, p_{i,m}^t)$ is the i-th pattern in time t, $P = \{p_1^t, p_2^t, \dots, p_k^t, \dots, p_m^t\}$ are parameters in time t.

Tab. 1: Parameters design for air quality modelling

| Parameters | |
|---|---|
| Harmful substances | $p_1 = SO_2$, $SO_2$ is sulphur dioxide. |
| | $p_2 = O_3$, $O_3$ is ozone. |
| | $p_3 = NO$, $NO_2$ ($NO_x$) are nitrogen oxides. |
| | $p_4 = CO$, CO is carbon monoxide. |
| | $p_5 = PM_{10}$, $PM_{10}$ is particulate matter (dust). |
| Meteorological | $p_6 = SW$, SW is the speed of wind. |
| | $p_7 = DW$, DW is the direction of |
| | $p_8 = T_3$, $T_3$ is the temperature 3 meters above the Earth's |
| | $p_9 = RH$, RH is relative air |
| | $p_{10} = AP$, AP is air pressure. |
| | $p_{11} = SR$, SR is solar radiation. |

## 3. Model Design for the Classification

Modelling air quality represents a classification problem. However, the descriptions of classes $\omega_{i,j}^t \in \Omega$ are known [5]. Therefore, it is suitable to realize the modelling of air quality by unsupervised methods. Data pre-processing is carried out by means of data standardization. Based on the analysis presented in [5], [6] the combination of the KSOFMs and K-means algorithm is a suitable unsupervised method for air quality modelling. The model for the classification of objects $o_i^t \in O$ into classes $\omega_{i,j}^t \in \Omega$ is presented in Fig. 1.

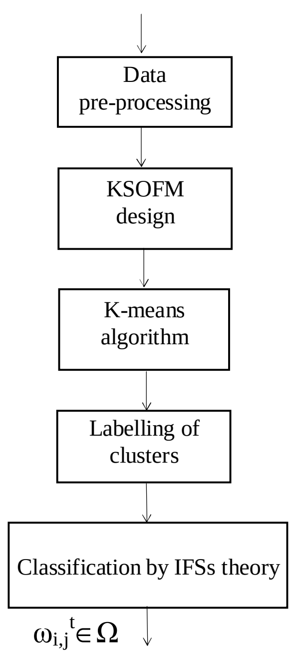



Fig. 1: The model for classification of objects $o_i^t \in O$ into classes $\omega_{i,j}^t \in \Omega$

The IFSs theory uses its results as the inputs. The theory of IFSs is well suited to dealing with vagueness. Recently, the IFSs have been used to intuitionistic classification models that can accommodate imprecise information. L. A. Zadeh defined fuzzy relation in [7]. E. Sanchez [2], [3] adopted Zadeh‘s max-min composition rule as an inference mechanism.

Let X and Y be two sets. Then the IFR R from X to Y (will be denoted $R(X \rightarrow Y)$) is an IFS of $(X \times Y)$ characterized by the membership function $\mu_R(x)$ and the non-membership function $\nu_R(x)$. If A is an IFS of X, then the max-min-max composition of the IFR $R(X \rightarrow Y)$ with A is an IFS B of Y (denoted by $B=R \circ A$) and is defined by the membership function [1], [8]

$$\mu_{R \circ A}(y) = \bigcup_x \, [\mu_A(x) \cup \mu_R(x,y)], \text{ and the non-membership function} \tag{1}$$

$$\nu_{R \circ A}(y) = \bigcup_x \, [\nu_A(x) \cup \nu_R(x,y)], \ \forall y \in Y, \text{ where } \bigcup = \max, \ \bigcup = \min. \tag{2}$$

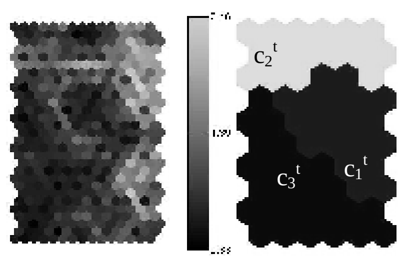
Let $Q(X \rightarrow Y)$ and $R(Y \rightarrow Z)$ be two IFRs. Then the max-min-max composition $T=R \circ Q$ is the IFR from $T(X \rightarrow Z)$, defined by the membership function [1], [8]

$$\mu_{R \circ Q}(x,z) = \bigcup_y \, [\mu_Q(x,y) \cup \mu_R(y,z)], \text{ and the non-membership function} \tag{3}$$

$$\nu_{R \circ Q}(x,z) = \bigcup_y \, [\nu_Q(x,y) \cup \nu_R(y,z)], \ \forall (x,z) \in (X \times Z) \text{ and } \forall y \in Y. \tag{4}$$

## 4. Determination of Intuitionistic Fuzzy Relations

Using the KSOFM as such can detect the data structure as presented in Fig. 2a. The **U**-matrix shows the square Euclidean distances d between representatives $\mathbf{w}_{i,j}$. The K-means algorithm can be applied to the adapted KSOFM in order to find clusters as demonstrated in Fig. 2b.




2a 2b

Fig. 2a: **U**-matrix of square Euclidean distances d

Fig. 2b: Clustering of the KSOFM by K-means algorithm

The K-means algorithm carries out the clustering of the adapted KSOFM as presented in Fig. 2b. The K-means algorithm belongs to the non-hierarchical algorithms of cluster analysis, where patterns $\mathbf{p}_1^t, \mathbf{p}_2^t, \dots, \mathbf{p}_i^t, \dots, \mathbf{p}_n^t$, n=60, are assigned to clusters $C=\{c_1^t, c_2^t, \dots, c_i^t, \dots, c_q^t\}$. The number of clusters q=3 is determined by indexes evaluating the quality of clustering [9]. A general title can be assigned to each of the clusters $c_i^t$ based on the districts dominating in the cluster. Green (residential) area, industrial area or traffic junction are samples of the titles. The following districts are analyzed in the city of Pardubice, the Czech Republic: bus stations of Cihelna (CI), Dubina (DU), Polabiny (PO), Rosice (RO), Rybitví (RY) and Srnojedy (SR), crossroads of Palacha-Pichlova (PP), Náměstí Republiky (NR) and Lázně Bohdaneč (LB), and the chemical factory of Paramo (PA).

Clustering process [5] is realized in two levels. In the first level, n objects are reduced to representatives $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_s$ by the KSOFM and the s representatives are clustered into q clusters. Clusters $C=\{c_1^t, c_2^t, \dots, c_i^t, \dots, c_q^t\}$ can be interpreted on the basis of parameters' values $\mathbf{p}_i^t=(p_{i,1}^t,p_{i,2}^t, \dots ,p_{i,k}^t, \dots ,p_{i,m}^t)$ for the s representatives of the KSOFM. The interpretation of parameters results from the air quality and dispersion conditions defined in [5]. The interpretation of clusters $C=\{c_1^t,c_2^t, \dots ,c_i^t, \dots ,c_q^t\}$ follows the values of all parameters $P=\{p_1^t,p_2^t, \dots ,p_k^t, \dots ,p_m^t\}$, (Fig. 3).

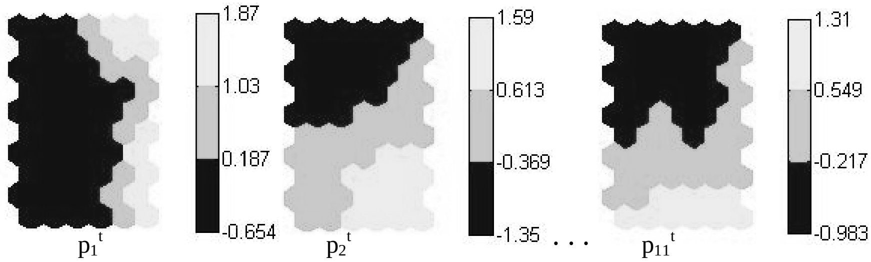



Fig. 3: The values of parameters $P=\{p_1^t,p_2^t, \dots ,p_k^t, \dots ,p_m^t\}$, m=11, for the representatives of the KSOFM

The interpretation of the clusters results in the assignment of the clusters to the classes $\omega_{i,j}^t \in \Omega$, Tab. 2. The clusters are labelled by the classes $\omega_{i,j}^t \in \Omega$, j=3, $\omega_{i,1}^t,\omega_{i,2}^t,\omega_{i,3}^t$ so that the class $\omega_{i,1}^t$ represents the least polluted air and the class $\omega_{i,3}^t$ represents the most polluted air. The intuitionistic fuzzy relation $R(P \to \Omega)$ is derived from the standardized values of parameters $P=\{p_1^t,p_2^t, \dots ,p_k^t, \dots ,p_m^t\}$, m=11, shown in Fig. 3. Then the proportion of high values (light grey colour) in the classes $\omega_{i,1}^t,\omega_{i,2}^t,\omega_{i,3}^t$ determines the membership functions $\mu_R(p_k^t,\omega_{i,j}^t)$, and the proportion of low values (black colour) determines the non-membership functions $\nu_R(p_k^t,\omega_{i,j}^t)$.

Tab. 2: Labelling of clusters by classes $\omega_{i,j}^t \in \Omega$, j=3, according to air quality

| Cluster | Parameters of harmful substances and dispersion conditions | $\omega_{i,j}^t$ |
|---|---|---|
| $c_1^t$ | Favourable air quality, slightly unfavourable dispersion conditions, acceptable health condition. | $\omega_{i,2}^t$ |
| $c_2^t$ | Poor air quality, unfavourable dispersion conditions, air-endangering health of the whole population. | $\omega_{i,3}^t$ |
| $c_3^t$ | Excellent air quality, slightly unfavourable dispersion conditions, healthy air. | $\omega_{i,1}^t$ |

Intuitionistic fuzzy relation $Q(O \to P)$ is derived from the values of parameters $P=\{p_1^t,p_2^t, \dots ,p_k^t, \dots ,p_m^t\}$ in the districts $o_i^t \in O$ during the time of monitoring. The mentioned fact is realized by the parameters of sigmoidal function.

## 5. Air Quality Classification

Next, we present an application of IFSs theory in Sanchez‘s [2], [3] approach for air quality classification. Air quality classification can be defined in this way:

Let $o_i^t \in O$ be the i-th district, $P=\{p_1^t,p_2^t, \dots ,p_k^t, \dots ,p_m^t\}$ be the parameters, and $\omega_{i,j}^t \in \Omega$ be the j-th class assigned to the i-th district $o_i^t \in O$. Then, we define intuitionistic air quality knowledge for classification as an IFR.

Let A be an IFS of the parameters P, and R be an IFR, $R(P \to \Omega)$, (Tab. 3 in appendix). Then the max-min-max composition B of the IFS A with the IFR $R(P \to \Omega)$ denoted by $B=A \circ R$ signifies the state of the districts $o_i^t \in O$ in terms of class as an IFS B of $\omega_{i,j}^t \in \Omega$ with the membership function given following way

$$\mu_B(\omega_{i,j}^t) = \bigvee_{p_k^t \in P} [\mu_A(p_k^t) \cup \mu_R(p_k^t,\omega_{i,j}^t)], \text{ and the non-membership function} \tag{5}$$

$$\nu_B(\omega_{i,j}^t) = \bigwedge_{p_k^t \in P} [\nu_A(p_k^t) \cup \nu_R(p_k^t,\omega_{i,j}^t)],\ \forall \omega_{i,j}^t \in \Omega. \tag{6}$$

If the state of a given district $o_i^t \in O$ is described in terms of the IFS A of the parameters $P=\{p_1^t, p_2^t, \ldots, p_k^t, \ldots, p_m^t\}$, then district $o_i^t \in O$ is assumed to be assigned to classes in terms of IFS B of $\omega_{i,j}^t \in \Omega$, through an IFR R of intuitionistic air quality knowledge from P to Ω, $R(P \rightarrow \Omega)$. Next, let be given n districts $o_i^t \in O$, i=1,2, ... ,n and let R be an IFR $R(P \rightarrow \Omega)$, then an IFR Q can be constructed from the set of districts $o_i^t \in O$ to the set of parameters P, $Q(O \rightarrow P)$, (Tab. 4, Tab. 5 in appendix). The composition T of IFRs R and Q, $T=R \circ Q$, describes the state of the district $o_i^t \in O$ in terms of the classes $\omega_{i,j}^t \in \Omega$ as an IFR from O to Ω, $T(O \rightarrow \Omega)$ given by the membership function

$$\mu_T(o_i^t,\omega_{i,j}^t) = \bigvee_{p_k^t \in P} [\mu_Q(o_i^t,p_k^t) \cup \mu_R(p_k^t,\omega_{i,j}^t)], \text{ and the non-membership function} \tag{7}$$

$$\nu_T(o_i^t,\omega_{i,j}^t) = \bigwedge_{p_k^t \in P} [\nu_Q(o_i^t,p_k^t) \cup \nu_R(p_k^t,\omega_{i,j}^t)],\ \forall o_i^t \in O \text{ and } \omega_{i,j}^t \in \Omega. \tag{8}$$

The composition T of IFRs R and Q and association index $\xi_T = \mu_T(o_i^t,\omega_{i,j}^t) - \nu_T(o_i^t,\omega_{i,j}^t) \times \pi_T(o_i^t,\omega_{i,j}^t)$ are presented in Tab. 6 in appendix. Association index $\xi_T$ [10] emphasizes high values of the membership function $\mu_T(o_i^t,\omega_{i,j}^t)$ (association) and reduces low values of the non-membership function $\nu_T(o_i^t,\omega_{i,j}^t)$ (non-association). Based on the analysis of the association index $\xi_T$ it is possible to classify the districts of Pardubice (DU,SR) to the class $\omega_{i,1}^t$. This class represents excellent air quality, slightly unfavourable dispersion conditions and healthy air with respect to the parameters of harmful substances and dispersion conditions. The districts (DU,SR) can be labelled as green (residential) areas. On the contrary, the districts of Pardubice (PP,RY,NR) are classified to the class $\omega_{i,3}^t$. This class represents poor air quality, unfavourable dispersion conditions and air endangering health of the whole population. The districts (PP,RY,NR) can be labelled as traffic junctions and decanting plants. The districts of Pardubice (PA,CI,PO,RO,LB) represent favourable air quality, slightly unfavourable dispersion conditions and acceptable health condition.

## 6. Conclusion

The intuitionistic fuzzy relation $R(P \rightarrow \Omega)$ and the intuitionistic fuzzy relation $Q(O \rightarrow P)$ were designed based on the presented results and the knowledge of experts in the field of air quality measuring (by the mobile monitoring system HORIBA). After the realization of intuitionistic fuzzy relations composition and the calculation of association index $\xi_T$, the districts $o_i^t \in O$ (bus stations, traffic junctions, and chemical factory) can be classified into the classes $\omega_{i,j}^t \in \Omega$. The gained results represent the recommendations for the state administration of the city of Pardubice in the field of air quality development.

This work was supported by the scientific research project of Ministry of Environment, the Czech Republic under Grant No: SP/4i2/60/07 with title Indicators for Valuation and Modelling of Interactions among Environment, Economics and Social Relations.

## References:


[1] ATANASSOV, K. Intuitionistic fuzzy sets. Fuzzy Sets and Systems, 20, 1986, pp.87-96.

[2] SANCHEZ, E. Solutions in composite fuzzy relation equation, application to medical diagnosis in Brouwerian logic. Gupta, M.M., Saridis, G.N., Gaines, B.R., Eds., Fuzzy Automata and Decision Process, Elsevier, North-Holland, 1977.

[3] SANCHEZ, E. Resolution of composition fuzzy relation equations. Inform. Control, 30, 1976, pp.38-48.

[4] BISWAS, R. Intuitionistic fuzzy relations. BUSEFAL, 70, 1997, pp.22-29.

[5] OLEJ, V., HÁJEK, P. Air quality modelling by Kohonen's self-organizing feature maps and intuitionistic fuzzy sets. A.P. Pobil, Eds., In Proc. of the 12th IASTED Int. Conference on Artificial Intelligence and Soft Computing, ASC 2008, 2008, Palma de Mallorca, Spain, ACTA Press, Calgary, Alberta, Canada, pp.22-27.

[6] KOHONEN, T. Self-organizing maps. 3rd edition, New York, Springer-Verlag Berlin Heidelberg, 2001.

[7] ZADEH, L.A. Fuzzy sets. Inform. and Control, 8, 1965, pp.338-353.

[8] ATANASSOV, K. Intuitionistic fuzzy sets. Springer-Verlag Berlin Heidelberg, 1999.

[9] STEIN, B., MEYER ZU EISSEN, S., WISSBROCK, F. On cluster validity and the information need of users. In Proc. of the Int. Conf. on Artificial Intelligence and Applications (AIA 03), Benalmádena, Spain, 2003, pp.216-221.

[10] DE, S.K., BISWAS, R., ROY, A.R. An application of intuitionistic fuzzy sets in medical diagnosis. Fuzzy Sets and Systems, 117, 2001, pp.209-213.


## Appendix

Tab. 3: Intuitionistic fuzzy relation R(P $\to\Omega$)

| R | $\omega_{i,1}^t$ | | $\omega_{i,2}^t$ | | $\omega_{i,3}^t$ | |
|---|---|---|---|---|---|---|
| | $\mu$ | $\nu$ | $\mu$ | $\nu$ | $\mu$ | $\nu$ |
| $p_1$ | 0.25 | 0.50 | 0.00 | 0.91 | 0.38 | 0.46 |
| $p_2$ | 0.47 | 0.14 | 0.00 | 1.00 | 0.00 | 0.92 |
| . . . | | | | | | |
| $p_{11}$ | 0.78 | 0.22 | 0.00 | 1.00 | 0.00 | 1.00 |

Tab. 4: Intuitionistic fuzzy relation Q(O $\to$P), the membership function $\mu$

| Q | $p_1$ | $p_2$ | | $p_{11}$ |
|---|---|---|---|---|
| DU | 0.15 | 0.70 | | 0.55 |
| PP | 0.39 | 0.03 | | 0.22 |
| PA | 0.64 | 0.24 | | 0.37 |
| CI | 0.34 | 0.59 | | 0.45 |
| PO | 0.17 | 0.52 | . . . | 0.47 |
| RO | 0.25 | 0.92 | | 0.82 |
| RY | 0.29 | 0.87 | | 0.51 |
| SR | 0.27 | 0.64 | | 0.77 |
| NR | 0.40 | 0.01 | | 0.06 |
| LB | 0.13 | 0.72 | | 0.78 |

Tab. 5: Intuitionistic fuzzy relation Q(O $\to$P), the non-membership function $\nu$

| Q | $p_1$ | $p_2$ | | $p_{11}$ |
|---|---|---|---|---|
| DU | 0.75 | 0.00 | | 0.06 |
| PP | 0.35 | 0.76 | | 0.56 |
| PA | 0.00 | 0.00 | | 0.00 |
| CI | 0.50 | 0.09 | | 0.28 |
| PO | 0.65 | 0.19 | . . . | 0.14 |
| RO | 0.48 | 0.00 | | 0.00 |
| RY | 0.50 | 0.11 | | 0.36 |
| SR | 0.58 | 0.00 | | 0.00 |
| NR | 0.24 | 0.98 | | 0.91 |
| LB | 0.81 | 0.14 | | 0.00 |

Tab.6: Composition relations T=R $\circ$ Q and association index $\xi_T$

| T | $\omega_{i,1}^t$ | | $\omega_{i,2}^t$ | | $\omega_{i,3}^t$ | | $\xi_T$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\mu$ | $\nu$ | $\mu$ | $\nu$ | $\mu$ | $\nu$ | | $\omega_{i,1}^t$ | $\omega_{i,2}^t$ | $\omega_{i,3}^t$ |
| DU | 0.55 | 0.14 | 0.27 | 0.36 | 0.46 | 0.31 | DU | 0.51 | 0.14 | 0.39 |
| PP | 0.36 | 0.34 | 0.27 | 0.27 | 1.00 | 0.00 | PP | 0.26 | 0.15 | 1.00 |
| . . . | | | | | | | | | | |
| LB | 0.78 | 0.14 | 0.82 | 0.00 | 0.36 | 0.31 | LB | 0.77 | 0.82 | 0.25 |

**Kontaktní adresa:**

prof. Ing. Vladimír Olej, CSc., Ing. Petr Hájek, Ph.D.
Ústav systémového inženýrství a informatiky
Fakulta ekonomicko-správní
Univerzita Pardubice
Studentská 84, 532 10 Pardubice
email: vladimir.olej@upce.cz, petr.hajek@upce.cz
tel.: +420 466 036 004