Paweł FIDEREK, Radosław WAJMAN, Jacek KUCHARSKI

Institute of Applied Computer Science, Lodz University of Technology

Fuzzy clustering based algorithm for determination of the two-phase gas-liquid flows similarity level

***Abstract**. In the article there is a description of the FCM based algorithm for the determination of the similarity level of the current investigated by the capacitance process tomography methods to the previously prepared pattern flows. Additionally readers can find a description of raw tomographic data collection method used, preparation of the most significant features vector routine and basic theoretical issues.*

***Streszczenie.** Artykuł zawiera opis algorytmu opartego na klasyfikacji rozmytej mającego na celu wyznaczenie wartości podobieństwa badanych za pomocą metod pojemnościowej tomografii przemysłowej przepływów dwufazowych do wcześniej przygotowanych przepływów wzorcowych. W artykule można znaleźć opis wykorzystanych metod akwizycji danych, przygotowania wektora cech znaczących oraz podstawowych zagadnień teoretycznych. (**Algorytm do wyznaczania stopnia podobieństwa przepływów dwufazowych typu gaz ciecz oparty na klasyfikacji rozmytej**).*

Słowa kluczowe: klasyfikacja rozmyta, surowe dane tomograficzne, przepływy dwufazowe gaz-ciecz, logika rozmyta.
Keywords: fuzzy c-means, raw tomographic data, two phase gas-liquid flows, fuzzy logic.

doi:10.12915/pe.2014.02.14

Introduction

The importance of the two phase gas-liquid flows can be easily noticed in various fields of the industry, for example in power plant facilities – as a tool of cooling system checking, in bioengineering facilities – as a tool of various processes investigation etc. Rapid development of knowledge about this kind of flow regimes allows developing new, quick methods of their identification and control. Nowadays there is a significant trend in two phase gas-liquid flow regimes identification – image reconstruction and identification [8].

Despite of common usage of such methods and technological advancement of devices used in such identification processes, there is still a problem with the long time needed to identification of the flow regimes. In this article the authors presents a new method of two phase gas-liquid flows analysis based on a raw tomographic data (described in [1, 2]) and a fuzzy algorithms without using the image reconstruction.

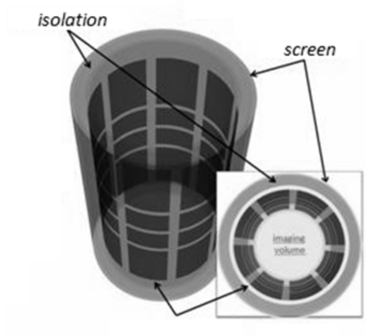
The raw tomographic data used in studies came from electrical capacitance tomography (ECT) method, and were obtained from semi industrial research facility located in the authors institute.

Research setup

During the research the authors used 3D ECT system described in [8]. Two most important modules of that system are:

32 electrodes capacitance sensors - The sensors used in the research were constructed on the basis of the technique of multiple winding a synthetic fabric fiber with an epoxy resin laminate [4]. The process of determining the optimal structure of the sensor was conducted by both theoretical and experimentally. The theoretical analysis was applied to the standard rules of determining the capacitance distribution within the capacitors on the basis of the physical phenomena occurring in the electric field. However, the analysis of between electrodes capacitance, collected using Agilent E4980A LCR meter, made it possible to effectively identify the areas of the sensor interior with a lower sensitivity and adopt a strategy for the sensor interior capacitance electrodes structure, characterized by uniform sensitivity measurement throughout the whole scan.

ECT tomograph – this device is the most important part in all installations and setups of that kind. The tomograph is responsible for raw tomographic data acquisition. Raw tomographic data are the between sensor electrodes capacitance measurements. During the research the raw data was normalized before the analysis - the desirability of this procedure was described in [6].



Pic.1. The schema of 32 electrodes capacitance sensor used in described research.

The raw data obtained by the tomograph are aggregated in the form of measurement vectors – the measurement frames. Each frame contains 496 between electrodes capacitance measurements, used in research tomograph device allow to collect 12 measurement frames per second.

Before the measurement frames could be aggregated into a analyze matrix (analyze matrix consist of the series of frames collected during one type of flow. The range of the matrix depends on the frame acquisition time) there is a need of their values validation. All of the frames, where there is even one wrong measurement (wrong means that the measurement value has out of range value – overflow) are dropped and not taking into account during the analysis.

Features describing the investigated and pattern flows are calculated from each row of the analyze matrix.

Table 1. The schema of the analyze matrix

F_1M_1	F_2M_1	…	F_nM_1
F_1M_2	F_2M_2		F_nM_2
…	…	…	…
F_1M_{i-1}	F_2M_{i-1}	…	F_nM_{i-1}
F_1M_i	F_2M_i	…	F_nM_i

where: F_1M_1 – first measurement inside the first frame, F_nM_i – i-th measurement inside n-th frame.

Introduction to fuzzy logic

Fuzzy logic is a multi-valued logic and a generalization of the classical two-valued approach. It was proposed by Lotfi Zadeh and is closely related to his theory of fuzzy sets [10]. The fuzzy logic, in opposite to standard two-valued logic, between the state of 0 (false) and the state of 1 (true) extends a number of intermediate values that determine the degree of membership of the element to the investigated collections (clusters).

Fuzzy logic is proved useful in engineering applications where the classical logic of classifying (only by the criterion of true and false) cannot effectively deal with many ambiguities and contradictions. There is a lot of applications [7, 8], including electronic control systems (machines, vehicles etc.), data mining tasks, or in the construction of expert systems where the fuzzy logic and fuzzy logic based algorithms are in use. The best example of fuzzy logic application is the vehicle ABS system presented by Mitsubishi Company in 1992.

Fuzzy c-means

FCM (fuzzy c-means) classifier is the most popular classifier based on fuzzy logic. This algorithm classification mechanism allows to the classification of one object to more than one cluster (group) with varying degrees of belonging to them – instead of standard classic 2 valued logic based classification where an object belongs to the cluster or not. A characteristic feature of the algorithm is that the shape of each cluster in the feature space is the same, and depends on the accepted standards. Operation of the algorithm is based on the minimization criterion:

(1) $$J(X;U,V)=\sum_{i=1}^{c}\sum_{k=1}^{M}(\mu_{ik})^m \| X_k - V_i \|_A^2$$

where: U is a membership matrix [10, 3], and V is a matrix representing the centers of clusters determined by the algorithm.

Fuzzy classifier algorithm can be divided into the following steps:

-Randomly initialize the membership matrix U:

(2) $$U=|\mu_{ik}|$$

where:

(3) $$\sum_{i=1}^{c}\mu_{ij}\ ,\forall_{j}=1,2,...,n$$

-Determination of the cluster centers

(4) $$c_i=\frac{\sum_{j=1}^{n}\mu_{ij}^m X_j}{\sum_{j=1}^{n}\mu_{ij}^m}$$

where:

(5) $$\mu_i=\frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{m-1}}}$$

where: d_{ij} - Position of the center of the j-th cluster in the i-th iteration;

-Determination of a new membership matrix U (including new cluster centers)

-Checking the algorithm stop condition - For the FCM the most common stop conditions are the convergence of the matrix membership U and the number of iterations.

Before the algorithm start user needs to define a distance measure, which will be measured distance between an object and clusters centers (for example the Euclidean distance – as used in presented research), the number of desired clusters and the fuzzyfication parameter *m*. The m parameter defines the amount of classification results fuzzyfication and its value should be as follows:

(6) $$m>1$$

The value of *m* parameter close to one will make the classification results to be similar to the one obtained using crisp classifiers (KNN, HCM), while the value of the m parameter is too high, degrees of objects membership to a particular class will be close to the inverse of their number. In practice, [3] the most common value of the m parameter is equal to 2.

Features calculation

In the word literature from feature extraction and calculation field the authors found that there are some features that are commonly used in the various kind of raw data description process – namely the statistical features.

In the presented research the authors used 4 most common (in the literature [5, 9]) statistical features for each of test set (a 180-piece set of corresponding measurement values in the test ECT frames sequence – 180 is a set of frames collected in the time of 15 seconds and with the speed of 12 frames per second). Knowing that every ECT frame holds 496 measurements, which with usage of 4 statistical features for every set gives1984 features describing a given sequence of test ECT frames.

The arithmetical mean is the first statistical feature used in the research, which can be used to determine a similarity between tested set of frames and a predefined set representing some flow pattern.

(7) $$\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}$$

where: n is the number of values in the test set (in the described study it is 180) and x_i is the i-th value of the collection.

The second statistical feature selected was the standard deviation. In statistics and probability theory, standard deviation shows how much variation or dispersion exists from the average (mean), or expected value. A low standard deviation indicates that the data points tend to be very close to the mean; high standard deviation indicates that the data points are spread out over a large range of values.

(8) $$\sigma=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$$

Third feature used to describe investigated sets is the skewness of a set. In probability theory and statistics, skewness is a measure of the extent to which a probability distribution of a real-valued random variable "leans" to one side of the mean. The skewness value can be positive, negative or even undefined.

(9) $$g_1=\frac{m_3}{m_2 * \sqrt{m_2}}$$

where:

(10) $$m_2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$

(11) $$m_3 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{n-1}$$

The last feature used to the tested sets description is kurtosis. It is used to describe if the set distribution is slender flattened. In other words the kurtosis of the set is used to determine the concentration of the variable around the mean value:

(12) $$g_2 = \frac{m_4}{m_2^2} - 3$$

where:

(13) $$m_4 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^4}{n}$$

Obtained results

Practical results are presented for the example of the 60 mm diameter pipe placed in the horizontal section of the installation [8]. Obtained results were validated by the expert with great experience in two phase gas-liquid flows field. First, the authors prepared measurement data for 9 two phase flows (all obtained with the same installation condition). Taking into the consideration the expert opinion, the authors separated 3 patterns measurements for each of 3 observed flow types:

-Slug flow;
-Layer flow;
-Foam flow;

Table 2. Example of the significant features calculation for one flow type

Slug pattern flow1	***Slug pattern flow2***	***Slug pat tern flow4***
0,92	***0,97***	***1,11***
1,45	1,15	1,03
1,03	**0.92**	***1,05***
…	…	…
1,35	1,09	1,24

where features considered as a similar are bolded

Table 3. Example of the most significant features for each type of flow

Slug flow	Layer flow	Foam flow
0,9732931	**0,809204**	**1,265042**
1,156781	1,74779	1,184039
	1,065585	1,195555
0,9795476	0,8923343	1,188933
0,0089927	***0,03218058***	***0,0278025***
…	…	…
1,036832	1,062161	1,188965
0,9574885	***0,7492933***	***1,269663***
1,093333	1,047041	1,196884
0,9817279	0,8536359	1,169193

where features considered as a different are bolded

Table 4. Example similarity determination results:

1st class	2nd class	3rd class	Determined flow type
0,001307145	0,9973721	0,0013207	slug
0,0148088	0,9718004	0,0133908	slug
0,007913345	0,9837972	0,0082895	slug
0,005334495	0,0078185	0,986847	layer
0,004842961	0,0087146	0,9864424	layer
2,07E-06	3,37E-06	0,9999946	layer
0,7386973	0,1993449	0,0619578	layer
0,9989628	0,0006552	0,0003819	foam
0,9483826	0,0294117	0,0222057	foam

Following calculations shows the procedure of determination of most significant features vector and the results of FCM algorithm. The necessity of most important features determination is caused by the nature of investigated phenomena. Most of the features describing one flow have no connection with corresponding features in another same type flow. In order to accurate classification results the features needs to be filtered. Most significant feature is the feature similar inside one flow type and different between different flow types. All values of features similarity or difference where experimentally determined during the research. 2 values are similar when the difference between them is lower or equal than 10% and 2 values are different when the difference between them is more than 30% to 50%.

As it could be noticed in table 4, every pattern flow has high value of similarity inside one type of flow and very low value of similarity between different flow types.

By the similarity level value the authors mean the membership value calculated by the FCM algorithm. Used in the research version of the algorithm returns the information about the object classification (flow classification) in range (0,1) (0-the object is not a member of the investigated cluster, 1-the object is a member of only the investigated cluster both of those values are unreachable due to the algorithm nature).

Obtained membership value can be interpreted as value of the level of similarity (where 0 – completely different from the flow placed in the cluster center, 1- exactly the same as a flow placed in the cluster center) where the flow placed in the cluster center is the perfect example of one two phase gas-liquid flow and its position in the feature space is determined in the iterations of the FCM algorithm.

Method application

Presented method and its results have various applications. Two of the applications are their usage in fuzzy inference about two phase gas-liquid flows regimes systems and installation of two phase gas-liquid flows fuzzy control systems. Both of these projects are currently being implemented by the authors. The simplest way to use presented results is to connect them with context information about the two phase gas-liquid flows – namely gas flow rate and liquid flow rate. Those information can be merged in the fuzzy reasoning rules [10].

As an example of such rule we can present the rule that determines the layer flow (one of many rules):

if
investigated flow is very similar to a layer flow (presented method results)
and
gas flow rate is medium
and
liquid flow rate is medium
then
investigated flow is a layer flow

Corresponding fuzzy rules can be also used in the installation control process, for example in the situation when installation user wants to obtain waved flow when currently investigated flow is a layer flow:

if
investigated flow is very similar to a layer flow (presented method results)
and
gas flow rate is medium
and
liquid flow rate is medium
and
flow regime should be a waved flow
then
increase the gas flow rate slightly

All used fuzzy values as slightly, medium or very similar needs to be previously determined in the inference algorithm learning process. In this moment the authors are able to perform an inference only for few flow types, but obtained results are very promising and conform to the presented fuzzy rules.

Conclusions

Presented in this paper method allows to obtain the value of similarity level of two phase gas-liquid flow regimes using only normalized raw tomographic data. The speed and the accuracy of the obtained results allow to use presented method in the applications where the speed of the algorithms is crucial – to example in the on-line two phase flows installation monitoring system.

The most important task before the method usage is appropriate input data preparation. Properly selected coefficients of the most significant features provides to obtaining very accurate results, even for flows which are difficult to recognize even for the expert.

This work is sponsored by National Science Centre as a part of the research project no 2011/01/D/ST6/07209.

REFERENCES

[1] Banasiak R., Soleimani M., Shape based reconstruction of experimental data in 3D electrical capacitance tomography, *NDT & E International*, Volume 43, Issue (2010), 241-249
[2] Banasiak R., Ye Z., Soleimani M., Improving three-dimensional electrical capacitance tomography imaging using approximation error model theory, *Jurnal of Electromagnetics*, Waves and Applications, Vol. 26 (2012), 411-421
[3] Bezdek J.C., Ehrlich R., Full W., FCM: The fuzzy c-means clustering algorithm, *Computers and Geosciences,* Volume 10, Issue 2-3 (1984), 191-203
[4] Brandisky K., Sankowski D., Banasiak R., Dolapchiev I., ECT sensor optimization based on RSM and GA, *COMPEL - The International Journal for Computation and Mathematics in Electrical and Electronic Engineering, Volume 31, Issue 3 (2012), 858-869*
[5] Burges, C.J.C., A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery*, Volume 2, Issue 2 (1998), 121-167
[6] Chaniecki Z., Algorytmy przetwarzania i analizy danych pomiarowych elektrycznej tomografii pojemnościowej w diagnostyce wybranych procesów przemysłowych, *PhD thesis* (2006)
[7] Jaworski T., Kucharski J., An algorithm for reconstruction of temperature distribution on rotating cylinder surface from a thermal camera video stream, *Przegląd Elektrotechniczny, Volume 89, Issue 2 A (2013), 91-94*
[8] Wajman R , Fiderek P, Fidos H., Jaworski T., Nowakowski J., Sankowski D., Banasiak R., Metrological evaluation of a 3D electrical capacitance tomography measurement system for two-phase flow fraction determination, *Measurement Science and Technology,* Volume 24, Issue 6 (2013), art. no. 065302
[9] Xu R., Wunsch II D., Survey of clustering algorithms, *IEEE Transactions on Neural Networks,* Volume 16, Issue 3 (2005), 645-678
[10] Zadeh L.A., Fuzzy sets, *Information and Control, Volume 8, Issue 3 (1965), 338-353*

***Autorzy**: mgr inż. Paweł Fiderek, Politechnika Łódzka, Instytut Informatyki Stosowanej, 90-924, Łódź, E-mail: p.fiderek@kis.p.lodz.pl; dr inż. Radosław Wajman, Politechnika Łódzka, Instytut Informatyki Stosowanej, 90-924, Łódź, E-mail: rwajman@kis.p.lodz.pl; dr hab. inż. Jacek Kucharski, prof. PŁ., Politechnika Łódzka, Instytut Informatyki Stosowanej, 90-924, Łódź, E-mail: jkuchars@kis.p.lodz.pl.*