Vol 9, Issue 6, 2018
ISSN NO:0377-9254

Usage of Fuzzy Rule and SOM Based Model to Identify a Handwritten Chemical Symbol or Structures

RAVI REGULAGADDA, P SRINIVAS, SUNIL BOLLAM
Dept of IT, MRIET, Hyderabad

***Abstract*— The basic components of chemical expressions and its corresponding reactions are chemical symbols and structures. To recognize a handwritten or printed chemical expression, proper identification of the chemical symbols and structures are important.**

This paper has reviewed the existing algorithms and models used for identifying the organic chemical structures. The objective of this paper is to find out chemical structures and symbols which are in a handwritten format and proposed model is based on fuzzy image segmentation technique.

Keywords— fuzzy sets, fuzzy rule, Self Organizing Map (SOM), chemical expression, chemical structure, image segmentation.

I. INTRODUCTION

In the field of chemistry the chemical reactions are expressed on paper as a chemical expression in various domains. While recognizing such chemical expression first identification of chemical structures and the symbol is important as they are part of chemical expression. According to the Tom Y. Ouyang [1] the organic reactions widely contain the symbols & diagrams to represent the chemical structure. Symbol identification is one of the main topics of Graphics Recognition, which have been an intensive research work in the last decades. Tabbone, S, Wendling L [2] covers technical symbol recognition. Belongie, S. et al. [3] speaks about handwritten symbol recognition. Symbol indexing and spotting is represented by Rusiñol, M. et al. in [4]. However, in the field of character recognition, According to Yang Zhang et al. [5] identifying a chemical symbol & structure in chemical expression is still a challenging problem. Most of the work is done on printed a document or a document created using digital ink, very limited work has reported on the offline handwritten chemical reactions. N. Hewahi et al. [6] state- most researchers use Template matching, Syntactic Structural Approach, Statistical Approach & Neural Networks technique of pattern recognition. Fuzzy logic techniques of pattern recognition for identification of handwritten chemical symbols have not been explored.

Fuzzy logic is a form of reasoning that is approximate not fixed. Fuzzy logic is used to handle the partial true conditions. It is used to deal with uncertain & variable conditions. In symbol recognition feature extraction is the important part Ref. [8] shows that Fuzzy logic is used for extracting the feature from handwritten symbol. To understand which part of the symbol is curved or horizontal, fuzzy information is better than the binary information. I.P. Morns and S.S. Dlay [9] prove that Fuzzy logic based feature extraction gives better understanding about what is to be present in symbol.

II. RELATED WORK

Zhang, Shi & Yang [11] have proposed an HMM (Hidden Markov Model) based method for identifying an online handwritten chemical symbol. This method used 11 dimensional normalized local feature techniques to recognize the chemical symbol & structures. Some organic reaction contains complex structures to recognize such complex chemical structures. Ouyang T.Y and Davis, R [10] used a method which is combination of spatial context & domain knowledge. The chemical reactions are classified in two parts, inorganic reaction which contains ring structures and organic reactions contains organic ring structures. To classify non-ring structures and organic ring structure a double stage classifier has been used in [11]. The first stage of this classifier, an SVM method is used to separate the non-ring structure & organic structure, the second stage used HMM method for fine recognition. This classifier is used for

online handwritten chemical symbols & structures.
A Novel approach [12] had presented for separation of chemical symbol from online handwritten chemical formulas. This approach adopts Freeman chain code to separate organic ring. The proposed method consists of four steps as 1) Detection of organic ring based on Freeman code 2)Analysis of pen trace to locate the connected point 3) Determination of the type of connection 4)Separation of the two connected chemical symbols and removal of the ligature if it exists. It only recognizes either organic ring or inorganic symbol, not able to recognize the chemical symbol as entire chemical compound.

A rule based approach for recognizing chemical structures is presented by Noureddin M. Sadawi [17]. In this model the principal recognition steps for molecule diagrams is a strict rule based system. It provides rules to identify the main components atoms and bonds as well as to resolve possible ambiguities. Jungkap Park and Gus R Rosania [18] explain the technique for identifying and extracting the chemical structures from the digital raster images. They developed a framework called as ChemReader, which accepts a raster image as input from chemical document and convert these analog images to digital image format. To identify and extract the chemical structure this framework uses the following steps :- pre-processing, separation of lines and character, line detection algorithm to identify chemical bonds, bond type identification, ring structure identification, chemical spell checker. This model efficiently extracts the chemical structures from the digital documents but not tested for handwritten document which contains chemical structures.

The previous discussion about recent approaches shows that most of approaches are proposed for identifying online handwritten chemical structures and symbols. Although little work for offline handwritten chemical symbol identification has been done. An SVM based classifier has proposed for recognition of offline handwritten benzene structure [23]. In which a comparison of two classifiers based on SVM and Logistic Regression has done to recognize a handwritten benzene structure from handwritten chemical expression. The model proposed by Ouyang T.Y. [10] reports not so good performance in case of messy sketches due to bad handwriting and template based component. The model proposed by Lin Zhao [12] is used for offline handwritten recognition, but it only separates organic ring from the entire chemical formula and recognize the inorganic symbols.

A Fuzzy logic approach has been used to recognize handwritten mathematical symbols. In their approach fuzzy rules are used for feature extraction, symbol classification and to access spatial relationship [13]. The hybrid approach has been proposed [14] to recognize handwritten symbol based on fuzzy logic have presented by combining model based and discriminative classifier. It uses a combination of fuzzy rule based recognizer and self organizing map recognizer technique to recognize a handwritten symbol. The average fuzzy direction technique has been introduced to recognize a Chinese handwritten character [15]. M. Hanmandlu *et al*. [16] presents a fuzzy model to recognize handwritten Hindi numerals. This approach uses an optimized strategy called as “foraging model of E.coli bacteria”. A stroke based Neuro-fuzzy system is presented [19] for identification of handwritten Chinese characters. Stroke extraction, feature extraction and recognition are the three main components of this system.

III. PROBLEM DEFINITION

It has been observed that the existing fuzzy logic approach is successfully used for identification of handwritten symbols and characters in various fields. Specifically to identify the handwritten Chinese characters, fuzzy logic approaches have reported good results. The Chinese characters are represented by complex structures, similarly organic chemical symbols also has complex structures. Here, this paper has been proposed, that fuzzy logic approach would be used to identify handwritten chemical symbols and structures based on fuzzy rule base and self organizing map (SOM).

IV. PROPOSED MODEL

The propose model based is based on fuzzy ruled base and SOM for identification of chemical symbols & structures and this model has six stages, namely:-Input scanned image, Pre-Processing, Fuzzy rule and SOM based feature extraction, Image segmentation, Classification and Representation.
The Fig.1 depicts the flow of above six stages of model along with the prototyped image database of handwritten chemical symbols. This proposed model is divided in two parts the first part consists of first three stages deals with the preprocessing and feature extraction for both prototyped image and unknown image of chemical symbols. The second part which consists next three stages deals with recognition of unknown chemical symbols by comparing the similarities in a number of segments between unknown chemical symbol and known prototyped class of chemical symbol or structure.

Input scanned Images

1. This stage of model deals with the accepting images of hand wrote chemical expressions of chemical reactions for identification and various images of handwritten chemical symbols and structures for prototyping purposes. These images are created by collecting the handwritten chemical reaction and

chemical symbols from various peoples and scanned them through the scanner.

Pre-Processing

This stage of the model further subdivided into three steps, namely: noise removal, image conversion, edge detection.

Noise removal

Noise removal technique is used to remove the unwanted handwritten shapes & structures which are not part of the chemical reaction. This noise is outcome of bad handwriting of the user. To get the best result for identification, noise has to be removed in the initial step. Noise removal can be done by using certain standard filtering like Linear filtering, Median filtering and Adaptive filtering. Apart from this filtering, fuzzy based noise removal techniques can be used.

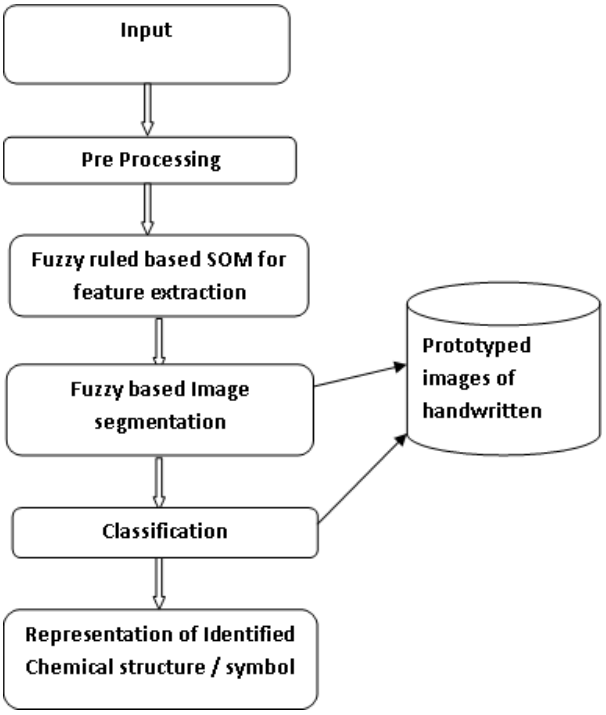

Fig.1 Proposed Model: Identification of Chemical Structures and Symbols

Image conversion

Once the noise has been removed from the image, then this image has converted from the RGB format to binary image format. This format is the most popular format for edge detection, feature extraction and segmentation. To convert an RGB image into binary image, intensive based thresholding technique is used. This technique separates background and object based on the intensity value of the pixel in the image. The concept used in this technique is simple, a parameter Θ called the intensity threshold of a pixel is chosen and applied to an image a[m,n] as follows:-

If (a[m,n] < Θ a[m,n] =1 = object
Else a[m,n]=0= background

This technique assumes that the high intensity value represents an object and low intensity represent the background and convert the RGB image to binary.

Edge Detection

Edge detection has been carried out to detect the presence of organic symbol and structure in the chemical reactions. Edge detection has been done with the help of Canny's (1986) algorithm [20].

Fuzzy rule and SOM based feature extraction

Feature extraction to be used for chemical symbol recognition. It is possible to use fuzzy logic approach for feature extraction. This paper has identified four types of features which commonly present in chemical symbols and structures likely are Vertical Line (V-Line), Horizontal Lines (H-Line), Slant Lines (S-Lines) used to represent chemical bonds, O (circle) shape is used to represent a ring in chemical structure. This paper has a belief that by identifying these four features, this model is able to find presences of organic symbol and structures.
This model identifies two types of fuzzy rules, namely low level fuzzy rule and high level fuzzy rule. High level rule defines the properties the input must if it belong to particular features. Low level rule access the extent to which these properties are presented. General form of high level rule is as follows:-

Feature (Z) ← Property_1 (Z) ∩ Property_2 (Z) ∩……∩ Property_k (Z)

Each Property_1 or Property_k is a simple property such as something, straightness, slanting nature, connectivity of line. A single representative value form Property_1 (Z) to Property_k (Z) to be calculated using t-norm or Yager [21] intersection operator. A low level fuzzy rule is used to determine the membership values of fuzzy sets Property_1 to Property_k By using these two fuzzy rules a feature extraction has to be done to find the four identified types of feature which help in recognition & classification.

A self-organizing map (SOM) is a type of artificial neural network (ANN) that is trained using unsupervised learning to produce a low-dimensional (typically two-dimensional), discretized representation of the input space of the training samples, and called a map**.** The SOM is the most used network in pattern recognition for feature extraction. SOM is used for converting the raw image data into more meaningful representation and identifying the variant and invariant aspects in handwritten symbol.

Image segmentation

A particular region to be found where entire chemical symbols and structures has drawn, for that image segmentation has to be used. To partition the image into multiple region or set of pixel, image segmentation will helpful. This model proposes use of the fuzzy thresholding technique for image segmentation. The problem may be involved in thresholding is to identify a threshold value T and segmenting the image into different region. For that a fuzzy thresholding has been used which involved in partitioning the image into different fuzzy sets representing the different region of images. This model has to identify a membership function associates with each region. The result of the segmentation will be the set of different region containing the chemical symbols or structures.

Classification

The objectives of this stage have to label and create different classes for different chemical symbols & structures while processing the prototyped images. This stage will also used to identify a chemical symbol or structure from unknown image. While identifying the input image, the different segments (region) which are extracted in previous stage will be used. It compares these segments to find the similarities from the classes present in the prototyped images. If it found similarities then it labelled that segment, according to the label of chemical symbols or structures present in the prototyped images.

Representation

In this stage the identified chemical symbol or structure to be represented in a computer understandable format like SMILES proposed by David Weininger [22]. Input for this stage is labelled symbol or structure from previous stage.

V. CONCLUSION

This paper has proposed a theoretical fuzzy rule and SOM based model for identification of handwritten chemical symbols and structures. This model has been used fuzzy rules to identify the different feature which helps to improve the performance of the model. The image segmentation could be done by using fuzzy thresholding instead of traditional segmentation technique. The outcome of the proposed model would be a chemical symbol or structure in a computer understandable format.

VI. LIMITATIONS AND FUTURE WORK

This proposed model will be used to identify organic chemical symbols or structures. This model will be able to identify the symbols or structures which are present in the prototyped image data base. This model required the input images is in particular fixed size.

Much of our future work involves practical implementation of the proposed theoretical model and converts this model into an experimental model by using some image processing tools. Also this model has to be tested on real life scenarios.

REFERENCES

[1]. Tom Y. Ouyang , Randall Davis ChemInk: A Natural Real-Time Recognition System for Chemical Drawings IUI'11, February 13–16, 2011, Palo Alto, California, USA.

[2]. Tabbone, S., Wendling, L.: Technical symbols recognition using the two-dimensional radon transform. In Proceedings of the 16th International Conference on Pattern Recognition, vol. 3, pp. 200–203 (2002).

[3]. Belongie, S., Malik, J., Puzicha, J.: Shapematching and object recognition using shape contexts. IEEE Trans. Pattern Anal. Mach.Intell. 24(4), 509–522 (2002)

[4]. Rusiñol, M., Borràs, A., Lladós, J.: Relational indexing of vectorial primitives for symbol spotting in line-drawing images. Pattern Recogni. Lett. 31(3), 188–201 (2010).

[5]. Yang Zhang, Guangshun Shi, Jufeng Yang "HMM-based Online Recognition of Handwritten Chemical Symbols" Proceedings of the International Conference on Document Analysis and Recognition (ICDAR), 2009.99

[6]. N. Hewahi, M.Al Nono, M. Nasar, M. Hamed, H. Hamed, "Chemical Ring Handwritten Recognition Based on Neural Networks", Ubiquitous Computing And Communication Journal 3, no. 3, 2008.

[7]. L.A. Zadeh "Outline of a new approach to the analysis of complex systems and decision processes", Man and Computer , pp.130-165, 1972.

[8]. N.R. Gomes and Lee Ling ," Feature extraction based on fuzzy set theory for handwritten recognition",ICDAR'01 pp.655-659, 2001

[9]. I.P. Morns and S.S. Dlay ," The DSFPN: A New Nural Network and Circuit Simulation for Optical Character Recognition", IEEE Transactions on signal Processing Vol 51,N0.12 pp. 3198-3209,2003

[10]. Ouyang T.Y., Davis R., "Recognition of Hand Drawn Chemical Diagrams", In Proc. of the National Conf. on Artificial Intelligence, 2007, pp.846-851.

[11]. Yang Zhang, Guangshun Shi, Kai Wang "A SVM-HMM Based Online Classifier for Handwritten Chemical Symbols" Proceedings of the International Conference on Pattern Recognition (ICPR), 2010 Pages 1888-1891.

[12]. Lin Zhao, Hu Yan, Guangshun Shi, Jufeng Yang "Segmentation of Connected Symbols in Online Handwritten Chemical Formulas" Proceedings of the International Conference on System Science, Engineering Design and Manufacturing Informatization (ICSEM), 2010 Pages 278-281.

[13]. J. Fitzgerald, F. Geiselbrechtinger, T. Kechad "Mathpad: A Fuzzy Logic-Based Recognition System for Handwritten Mathematics", ICDAR Sep. 2007 PP. 694-698.

[14]. Alex Cronin, John A. Fitzgerald, Tahar Kechadi, "A Hybrid Recogniser for Handwritten Symbols Based on Fuzzy Logic and Self-Organizing Maps", ICTAI 2006, pp. 693-700.

[15]. Weiping ZHU , Wei LIU ,Zhuqing HUANG, "Average Fuzzy Direction Based Handwritten Chinese Characters Recognition Approach", IEEE WKDD 2008, pp. 42-47.

[16]. M. Hanmandlu, A.V. Nath, A.C. Mishra and V.K. Madasu, "Fuzzy Model Based Recognition of Handwritten Hindi Numerals using Bacterial Foraging", ICIS 2007, pp. 309-314.

[17]. Noureddin M. Sadawi, Alan P. Sexton, and Volker Sorge "Chemical Structure Recognition: A Rule Based Approach" Proc. SPIE 8297, 82970E (2012)

[18]. Jungkap Park, Gus R Rosania, Kerby A Shedden, Mandee Nguyen, Naesung Lyu and Kazuhiro Saitou: Automated extraction of chemical structure information from digital raster images, Chemistry center Journal 2009 3:4.

[19]. Jue-Wenlin, Shie-Jue Lee, and Hsin-Tai Yang, "A Stroke-based neuro-fuzzy system for handwritten Chinese character recognition", Applied Artificial Intelligence: An International Journal, 15:6, pp.561-586

[20]. Canny, John,"A Computational Approach to Edge Detection",IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol- PAMI 8,No.6,pp 679-698,1986.

[21]. R. R. Yager. On a general class of fuzzy connectives. Fuzzy Sets and Systems, 4(1):235.242, 1980.

[22]. David Weininger. Smiles: a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci., 28(1):31–36, 1988.

[23]. Mapari, Shrikant, and Ajaykumar Dani. "Recognition of Handwritten Benzene Structure with Support Vector Machine and Logistic Regression a Comparative Study." The International Symposium on Intelligent Systems Technologies and Applications. Springer International Publishing, 2016.