PRAJÑĀ - Journal of Pure and Applied Sciences Vol. 21: 54 - 59 (2013)
ISSN 0975 - 2595

RESEARCH DIRECTIONS IN NEW ARTIFICIAL INTELLIGENCE: A CASE OF NEURO-FUZZY SYSTEM FOR WEB MINING

Priti Srinivas Sajja*
Department of Computer Science, Sardar Patel University, Vallabh Vidyanagar

ABSTRACT

Aim of this article is to introduce constituents of new artificial intelligence and their applications in various domains. This article defines field of new artificial intelligence and its ubiquitous nature to solve variety of problems in different domains. To demonstrate the utility of the new artificial intelligence, the possible research applications are enlisted in this article. These examples are categorized in two classes, namely pure research applications and applied new artificial intelligence techniques. The pure applications are more generic in nature and contribute to the field by invention of new techniques, models and structures. Applied class examples also contributes some innovative techniques, however they are less generic in nature. As an example, article also presents a case of neuro-fuzzy system for web mining. The architecture and design of the system is presented here. The sample design includes fuzzy user profile structure, code snippet to generate a multilayer neural network and structure of the training data set.

INTRODUCTION

New artificial intelligence is a field dedicated to handle complex problems using nature inspired computing methods. The field encompasses consortium of nature or bio inspired techniques. The prime goal of such field is to solve problems, which are very difficult to handle in highly effective manner by the traditional methods. The field is also known as soft computing technology. New artificial intelligence techniques are directly inspired from natural techniques such as evolutionary computation, neural computation, ecological computation, quantum computation, complex self-adaptive system and other fields inspired by natural systems. The field offers unique characteristics and advantages of self-adaptive, self-organizing and self-learning in problem solving. According to Zadeh [1] "In contrast to traditional hard computing, soft computing exploits the tolerance for imprecision, uncertainty and partial truth to achieve tractability, robustness, low solution-cost and better rapport with reality". Because of its ability to solve complex problems that are difficult for traditional computational methods, natural inspired computation is widely used in many fields, such as simulation to human expertise, innovative techniques, natural evolution, model free learning, goal driven, intensive numerical computations, dealing with partial and incomplete information, fault tolerance, machine learning, optimal design, optimal control, network security and creative design.

The major components of the new artificial intelligence techniques are fuzzy logic, artificial neural network and genetic algorithms. To solve high level of complex, unstructured and innovative problems in human like intelligent manner, these major techniques may participate with each other resulting hybrid techniques such as neuro-fuzzy, neuro-genetic and neuro-genetic-fuzzy techniques.

RESEARCH POSSIBILITIES

Scope of the new artificial intelligence techniques is nearly unlimited and ubiquitous. Application of intelligence enhances any business and improves problem solving. This makes the application of the new artificial intelligence techniques ambient. Researchers can contribute in the concepts and techniques of the field or they can identify novel mechanisms, algorithms and/or data structures related to these techniques which can enhance the techniques themselves. Contribution to the pure concepts, techniques and standards may be known as pure research contribution in the area. The applied research, on the other hand, considers different domain and applies modified or specially designed techniques which can enhance the domain techniques as well as contributes in the concepts also. It is to be noted that application of the existing techniques in a domain is not considered as a research. Table 1 describes a few research possibilities in the pure as well as applied field for three major techniques of the new artificial intelligence namely fuzzy logic, artificial neural network and genetic algorithms along with hybrid techniques.

*Corresponding author: *priti@pritisajja.info*

Pure Research Applications in the field of New Artificial Intelligence	Applied Research Applications in the field of New Artificial Intelligence
Fuzzy Logic (FL)	
• Innovative fuzzification and defuzzification methods • Fuzzy arithmetic and Boolean operations such as Fuzzy ANDing and ORing • Multi valued logic such as five value logic, seven value logic and their membership functions, fuzzification and defuzzification methods. • Fuzzy queries from distributed database • Fuzzy inference techniques, forward and backward inference and hybrid inference engine for fuzzy logic based systems • Fuzzy knowledge representation such as fuzzy rule incorporation in script, frame and semantic net • Type 2 fuzzy functions and type reducer applications along with complete generic model for type 2 fuzzy rule based system • 3 dimensional fuzzy logic systems • Tools to develop and manage fuzzy logic based system such as MATLAB (http://www.mathworks.in) and Scilab (http://www.scilab.org/)	• Fuzzy expert system models in medical diagnosis, portfolio management, risk assessment, aptitude testing and fault finding etc. [2] • Fuzzy logic application for various education measures and learning such as eLearning, mLearning (mobile) and pLearning (personalized/customized) [3] • Fuzzy logic as an interface and natural language for special audience such as handicapped, children and farmers • Fuzzy ontology for specific domain such as online shopping and ontology mining [4] • Novell standards for fuzzy ontology mapping, merging and integration • Type 2 fuzzy logic application in software project management and cost monitoring • Fuzzy programming editors and compilers • Image processing, compressing and improving techniques enhanced with fuzzy logic • Fuzzy sensors in domestic applications, nanotechnology and embedded systems
Artificial Neural Network (ANN)	
• Innovative neural network models with suitable learning algorithm • Heuristics for back propagation learning in order to decide error on which back propagation is done • Library to automatically develop neural network • Innovative activation functions • Automatic generation of validation sets • Hybrid learning algorithms • Self organizing maps and automatic wrapper generation • Automatic evolution of ANN structures through genetic algorithms	• Application of learning algorithms • Neural network for forecasting and prediction based on heuristic for a domain such as predicting demand in market • All type of classifications (eg. Protein), trends suggestion (eg. Fashion) and advisory systems • Pattern matching with neural network [5] • Learning in real time interactive gaming • Support vector machine for watermarking • Intrusion detection • Robot control
Genetic Algorithm (GA)	
• Novell fitness functions • New and generalized genetic operators • New algorithm on evolution of desired schema such as modification on building block • Novel encoding strategy • Determining penalty for application dependent fitness function.	• GA for aptitude testing and advisory system • Self evolving rules • Software quality metrics and automatic test case generation using GA • Idea generation system for business process reengineering and total quality management
Hybrid Systems	
• Neurons with fuzzy activation function and its network with learning algorithm • Co-operative ANN with FL or GA such as fine tuning ANN output with the FL • Neuro-fuzzy-genetic system structures • Soft computing library with visual development support • Self evolving rule based with genetic fuzzy functions	• Application of FL based fitness functions for product promotion • Fuzzy interface to enhance human computer interaction in e-commerce, aptitude system, product promotion, trust based network etc. [6] • Genetic fuzzy operators for given domain such as identifying emotional intelligence of a student [7] • Web mining using neuro-fuzzy system[8]

Table 1: Pure and applied research possibilities in the area

DESIGN OF NEURO-FUZZY SYSTEM FOR WEB MINING

The web mining refers to effective extraction of required information from the Web using web log mining; web page classification is one of the techniques of mining the Web in order to categorize web pages into some predefined categories. Data mining, machine learning and other artificial intelligent techniques are used for such web page classification. Work done by Cohen W et al.[9], Freitag, D. and McCallum [10], Kushmerick, N [11], etc. illustrates use of partly automated machine learning techniques for structured data extraction from web pages. Work illustrated in Chang and Lui[12], Liu et al.[13], Crescenzi et al.[14], Arasu and Garcia-Molina[15], Embley et al.[16], Lerman et al.[17], Pinto et al.[18], etc. demonstrates the automatic wrapper generation for web mining. The work presented in Bharanipriya et al[19], Carullo et al [20], Deepa, S. and Hariharan, S[21], Gemmell et al[22], Jayanthi, S.K. and Prema, S.[23], Poonkuzhali et al[24], Qi, X. and Davison, B.D.[25] describe applications and state of the art surveys in the field of web mining and web page classification.

These solutions mine the Web or classify web pages into some rigid categories. Further, it lacks user friendly interface. The case presented here uses a neuro-fuzzy approach to effectively classify web pages in vague categories. The fuzzy mechanism of the system is used to enable customized mining of the web content in user friendly manner allowing vague input and storing user's interest in fuzzy form in a fuzzy user profile. The neural network component of the system aids in automatic learning of users interest and mines the web content

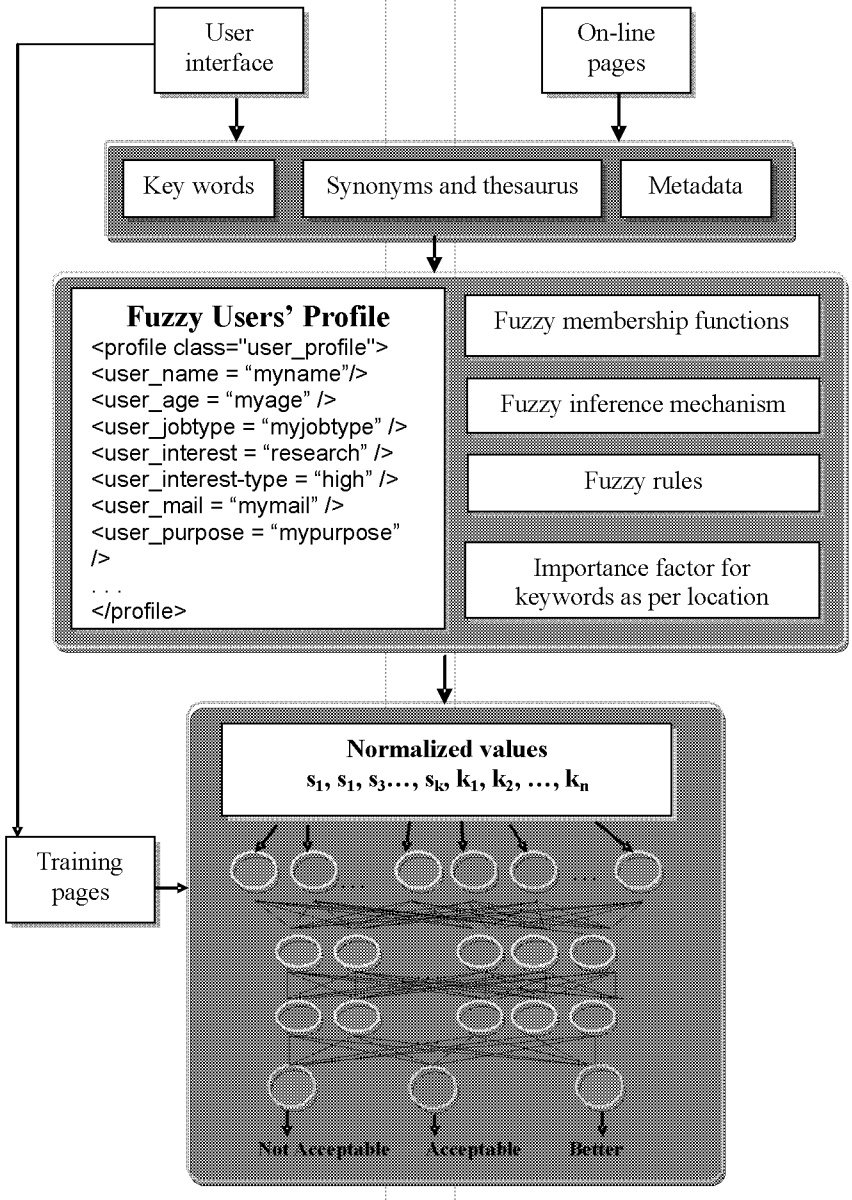

Figure 1: Architecture of the neuro-fuzzy system

automatically. Figure 1 presents a brief introduction of methodology and architecture of the system.

As the objective is to classify web pages in customized manner as per the user's nature, style and level; users' information must be stored into the system. According to the architecture shown in the Figure 1, users' information is stored in fuzzy manner in form of fuzzy user profile repositories. The fuzzy user profile repositories stores following information:

- personal information such as age, gender, job type and location
- educational qualification
- log (history) of web pages
- interests in research
- interest in problem solving
- interest in education and training
- personal preferences
- etc.

Some of the above parameters, such as interest in research are vague (fuzzy) in nature. Fuzzy variables are used to represent values of such fuzzy parameters. In case of fuzzy parameter 'interest in research' possible values are 'High', 'Moderate' and 'Low'. Along with every fuzzy variable an appropriate fuzzy membership function is used. Such membership function is used to map crisp (non-fuzzy) values into corresponding fuzzy values and vice versa.

The format of the user profile [26] can be given as follows.

```
<profile class="user_profile">
<user_name = “myname”/>
<user_age = “myage” />
<user_jobtype = “myjobtype” />
<user_interest = “research” />
<user_interest-type = “high” />
<user_mail = “mymail” />
<user_purpose = “mypurpose” />
. . .
</profile>
```

Once information about users is stored into the repository, users need to provide keywords to search web pages. The keywords provided by users are preprocessed for noise clearance, if any. The thesaurus of system may also provide matching aliases and synonyms. These keywords total frequency within the web page is calculated. These frequencies along with the location of the keyword are multiplied with application specific importance factor of the keyword as per its location, which is used to normalize the input to the base neural network.

The neural network component follows multi layer back propagation learning mechanism. The neural network is trained in offline manner with dataset according to the target users' requirements. The normalized values from the fuzzy component are input to the neural network input layer. There are two hidden layers which has the variation of sigmoid function. The final layer has three neurons, as the web pages are intended to be classified into three categories such as 'Not Acceptable', 'Acceptable' and 'Better'. Number of categories can be enhanced on need.

For example, to categories web pages for students of an educational institute, it is required to check that harmful or objectionable content must not be there on the web pages. Web page having phrases such as- 'film, 'games' or 'video' may not be advised for the presentation to the students. The system should categorize the web page in 'Not Acceptable' class. For this, first following information can be collected from the web page.

Keywords	Frequency	Location of the term
Video	0	Page title
Legal age	2	Page content
Game	0	Page title
Name of social network site	1	Header text
Film	1	Page content
Fun link	0	Title word list
……	….	……

If the objectionable keyword appears either at important location (such as page title) of the web page or its frequency is very high, then chance of the web page to fall in 'Not Acceptable' category is high.

As stated earlier, the base neural network requires to be trained. To prepare training data set, first target users are asked to study selected web pages and provide feedback regarding its classification. Such data forms the training set of the base neural network. Once the network is trained according to the users expectations, it is ready to classify the web page into given categories using keywords provided. For different target audiences, different training data sets are prepared in order to train the base neural network.

The base neural network can be generated using code in an object oriented environment as shown in Table 2.

```
public class NeuralNetwork
  {  protected Layer[] layers;
     protected int ni;
     protected LearningAlgorithm la;
     public int N_Inputs
     { get { return ni; }
     }
    public int N_Outputs
     { { return layers[N_Layers - 1].N_Neurons; }      }
public int N_Layers
     { get { return layers.Length; } }
public LearningAlgorithm LearningAlg
     { { return la; }
       set { la = (value != null) ? value : la; } }
public Layer this[int n]
     { { return layers[n]; }}
public NeuralNetwork(int inputs, int[] layers_desc,
ActivationFunction n_act, LearningAlgorithm learn)
     {    if (layers_desc.Length < 1)
          throw new Exception("PERCEPTRON :
cannot build perceptron with this input ");
          if (inputs < 1)
          throw new Exception("PERCEPTRON :
cannot build perceptron, it must have at least 1
input");
          la = learn;
          ni = inputs;
        layers = new Layer[layers_desc.Length];
        layers[0] = new Layer(layers _desc[0], ni);
for (int i = 1; i < layers_desc.Length; i++)
        layers[i] = new Layer(layers_desc[i],
layers_desc[i - 1], n_act);       }
public void randomizeWeight()
     { foreach (Layer l in layers)
          l.randomizeWeight(); }
     public void randomizeThreshold()
     { foreach (Layer l in layers)
          l.randomizeThreshold();  }
     public void randomizeAll()
     { foreach (Layer l in layers)
          l.randomizeAll();  }
     public void
setActivationFunction(ActivationFunction f)
     { foreach (Layer l in layers)
          l.setActivationFunction(f);  }
     { foreach (Layer l in layers)
          l.setRandomizationInterval(min, max); }
     public float[] Output(float[] input)
     { (input.Length != ni)
          throw new Exception("PERCEPTRON : Wrong
input vector size, unable to compute output value");
      float[] result;
        result = layers[0].Output(input);
        for (int i = 1; i < N_Layers; i++)
          result = layers[i].Output(result);
        return result;
     }
}  }
```

Table 2: Code snippet to generate the base artificial neural network

The neural network has an ability to learn from data sets; however, it is not able to provide explanation and justification of the classification it makes. In future, this system can be enhanced to fine tuned output of the neural network to provide necessary explanation and justification.

CONCLUSION

The new AI methods are ubiquitous in nature and have high applicability on number of domains. There are a large number of opportunities to contribute in pure computer science research using the new AI methods. There are some applications, where the new AI methods can contribute towards identification of novel mechanisms for the domain. This article describes more than 50 research ideas, applications and projects in the field of the new artificial intelligence, among which an experiment case for web page classification using generic neuro-fuzzy system is discussed in detail. One may use the neuro-fuzzy system for other applications such as product and employee classification with minor changes. This application is a hybrid application example, which uses neural network and fuzzy logic in co-operative manner to achieve dual advantages of both the approaches. From the enlisted applications and ideas, any idea can be picked up as a full fledge research or project work. The finished product may be considered as commercial or free tool for seeking patent or may be used to uplift society and industry.

REFERENCES

[1] L. A. Zadeh. (1994) Soft computing and fuzzy logic. Software, IEEE, 11(6): pp 48-56.

[2] Sajja, P.S. and Akerkar, R.A. (2010) *Advanced knowledge-based systems: Models, applications and research trends*, TMRF,Kolhapur, India, pp 1-11.

[3] Sajja, P.S. (2011) Personalized content representation through hybridization of mobile agent and interface agent, in Susmit Bagchi (Ed.), Ubiquitous Multimedia and Mobile Agents: Models and Implementations, Chapter 4, IGI Global Book Publishing, Hershey, PA, USA, pp 85-112.

[4] Sajja, P.S. and Akerkar, R.A. (2012) Mining sentiment using conversation ontology, in Patricia Ordóñez de Pablos, Héctor Oscar Nigro, Robert Tennyson, Sandra E. González Císaro and

Waldemar Karwowski (Eds.), Advancing Information Management through Semantic Web Concepts and Ontologies, Chapter 16, IGI Global Book Publishing, Hershey, PA, USA pp 302-315.

[5] Sajja, P.S. and Akerkar, R.A. (2013) Bio-inspired models and the semantic web, in Xin-She Yang, Zhihua Cui, Renbin Xiao, Amir Hossein Gandomi and Mehmet Karamanoglu (Eds.), Swarm Intelligence and Bio-Inspired Computation: Theory and Applications, Chapter 12, Elsevier, Waltham, MA, USA, pp 273-294.

[6] Akerkar R.A. and Sajja, P.S. (2009) *Knowledge-based systems*, Jones & Bartlett Publishers, Sudbury, MA, USA, pp 129-152.

[7] Mankad, K.B. and Sajja, P.S. (2012) Measuring human intelligence by applying soft computing techniques: A genetic fuzzy approach, in Naidenova XA & Shagalov VL (Eds.), Diagnostic Test Approaches to Machine Learning and Commonsense Reasoning Systems, Chapter 6, IGI Global Book Publishing, Hershey, PA, USA, pp 130-155.

[8] Sajja, P.S. and Akerkar, R.A. (2012) *Intelligent technologies for web applications*, CRC Press (Taylor & Francis Group), Boka Raton, FL, USA.

[9] Cohen, W., Hurst, M. and Jensen, L. (2002) A flexible learning system for wrapping tables and lists in HTML documents. In: Proceedings of the Eleventh International World Wide Web Conference (WWW), Honolulu, Hawaii, USA.

[10] Freitag, D. and McCallum, A. (2000) Information extraction with HMM structures learned by stochastic optimization. In: Seventeenth National Conference on Artificial Intelligence (AAAI), Austin, Texas, USA.

[11] Kushmerick, N. (2000) Wrapper induction: efficiency and expressiveness. Artificial Intelligence. 118(1--2): pp 15-68.

[12] Chang, C.H. and Lui, S.L. (2001) IEPAD: information extraction based on pattern discovery. In: Proceedings of the Tenth International World Wide Web Conference (WWW), Hong Kong.

[13] Liu, B., Grossman, R. and Zhai, Y. (2003) Mining data records in web pages. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA.

[14] Crescenzi, V., Mecca, G. and Merialdo, P. (2001) Roadrunner: towards automatic data extraction from large web sites. In: 27th International Conference on Very Large Databases (VLDB), Roma, Italy.

[15] Arasu, A. and Garcia-Molina, H. (2003) Extracting structured data from web pages. In: Proceedings of the International Conference on Management of Data (SIGMOD), San Diego, California.

[16] Embley, D., Jiang, Y. and Ng, Y. (1999) Record-boundary discovery in web documents. In: Proceedings of the International Conference on Management of Data (SIGMOD), Philadelphia, Pennsylvania, USA.

[17] Lerman, K., Getoor L., Minton, S. and Knoblock, C. (2004) Using the structure of web sites for automatic segmentation of tables. In: Proceedings of the International Conference on Management of Data (SIGMOD), Paris, France.

[18] Pinto, D., McCallum, A., Wei, X. and Bruce, W. (2003) Table extraction using conditional random fields. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), Toronto, Canada.

[19] Bharanipriya, V. and Kamakshi Prasad, V. (2011) Web content mining tools: a comparative study. International Journal of Information Technology and Knowledge Management. 4(1), pp 211-215.

[20] Carullo, M., Binaghi, E. and Gallo, I. (2009) An online document clustering technique for short web contents. Pattern Recognition Letters. 30(10), pp 870-876.

[21] Deepa, S. and Hariharan, S. (2012) Web content mining: issues and challenges. The IUP Journal of Information Technology. 8(3), pp 64-72.

[22] Gemmell, J., Shepitsen, A., Mobasher, B. and Burke, R. (2008) Personalization in folksonomies based on tag clustering. In: Proceedings of the 6th Workshop on Intelligent Techniques for Web Personalization and Recommender Systems (AAAI), Chicago, USA.

[23] Jayanthi, S.K. and Prema, S. (2011) Word sense disambiguation in web content mining using Brill's tagger technique. International Journal of Computer and Electrical Engineering. 3(3), pp 358-362.

[24] Poonkuzhali, G., Kishore Kumar, R., Sudhakar, P., Uma, G.V. and Sarukesi, K. (2012) Relevance ranking and evaluation of search results through web content mining. In: Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS), Hong Kong.

[25] Qi, X. and Davison, B.D. (2007) Web page classification: features and algorithms. In: Technical Report LU-CSE-07-010. Computer Science and Engineering, Lehigh University.

[26] Sajja, P.S. (2013) Intelligent web content filtering through neuro-fuzzy approach, International Journal of Data Mining and Emerging Technologies, Vol. 3, No. 1, pp 33-39.