Journal of Computer Science 8 (8): 1285-1293, 2012
ISSN 1549-3636

Arabic Named Entity Recognition Using Artificial Neural Network

Naji F. Mohammed and Nazlia Omar
School of Computer Science, Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

Abstract: Problem statement: Named Entity Recognition (NER) is a task to identify proper names as well as temporal and numeric expressions, in an open-domain text. The NER task can help to improve the performance of various Natural Language Processing (NLP) applications such as Information Extraction (IE), Information Retrieval (IR) and Question Answering (QA) tasks. This study discusses on the Named Entity Recognition of Arabic (NERA). The motivation is due to the lack of resources for Arabic named entities and to enhance the accuracy that has been reached in previous NERA systems. **Approach:** This system is designed based on neural network approach. The main task of neural network approach is to automatically learn to recognize component patterns and make intelligent decisions based on available data and it can also be applied to classify new information within large databases. The use of machine learning approach to classify NER from Arabic text based on neural network technique is proposed. Neural network approach has performed successfully in many areas of artificial intelligence. The system involves three stages: the first stage is pre-processing that cleans the collected data, the second involves converting Arabic letters to Roman alphabets and the final stage applies neural network to classify the collected data. **Results:** The accuracy of the system is 92 %. The system is compared with decision tree using the same data. The results showed that the neural network approach achieved better than decision tree. **Conclusion:** These results prove that our technique is capable to recognize named entities of Arabic texts.

Key words: Arabic, Natural language processing, named entity recognition, neural network approach, Information Extraction (IE), artificial intelligence, Question Answering (QA), Arabic script

INTRODUCTION

The use of Named Entity Recognition (NER) concept has emerged as an important approach in natural language processing environments. NER systems are essential particularly when identifying proper names in open-domain texts. NER is crucial and can assist in improving the performance of Natural Language Processing (NLP) applications. For instance, when executing tasks related to handling massive amounts of information, NER systems could help in Information Extraction (IE), Information Retrieval (IR) and Question Answering (QA) tasks (Grover *et al*., 2008). Named Entity Recognition (NER) is synonymously well-known as NE extraction, NE detection or NE identification. It is regarded as one of the most important sub tasks in the process of Information Extraction. However, it is also significant in the recognition and classification of defined named entities from large text, or in general context of news-wires (Maynaed *et al*., 2008). The main goal of Named Entity Recognition (NER) task is the attempt to increase performance accuracy with regard to the identification and extraction of named entities. The most significant named entities in Arabic script include organizations (companies, government organizations, committees), persons, locations (cities, countries, rivers) dates and time expressions and monetary amounts (percent, money, weight) from open-domain Texts (Elsebai *et al*., 2009). Recently, there has been a sudden increase in NER task research for Arabic language. Many researchers have worked on this problem in diverse languages using different approaches. However, the available studies have

Corresponding Author: Nazlia Omar, School of Computer Science, Faculty of Information Science and Technology, University Kebangsaan Malaysia, 43600 UKM Bangi,, Selangor, Malaysia Tel: 60-3-89216733 Fax: 603-8921 6732

suggested that there are very few researchers who seem to be interested in Named Entity Recognition (NER) for Arabic texts. There is scarcity of resources for Arabic named entities and a lack of accuracy in previous NERA systems. These factors have created a negative effect among potential researchers particularly in the Arabic natural language processing field. In this work we use machine learning approach to classify NER from Arabic text based on neural network technique. The principal task of machine learning approach is to automatically learn from previous data. Furthermore, the classification techniques can be employed to organize new information within large database.

The Arabic language: The Arabic language is one of the most important languages spoken in the world today .It has the unique features and some of the inherent similarities between Arabic and other Semitic languages (Albared *et al*., 2009). The styles of the Arabic language are Classical Arabic, Modern Standard Arabic and Colloquial Arabic. The Arabic language is characterized by a wealthy vocabulary and a complex morphology. The named entities of Arabic language have the single and composite types of the Arab personal names (Benajiba, 2009).

Challenges: There exist two major challenges posed by focusing on Arabic NER. There are as follows:

- Absence of capital letters in the orthography: The Arabic language is different from Latin languages. Other languages have a signal in the orthography and that is capitalization of the first letter. The capital letter is used to point to a word or a string of words as named entities. The Arabic language has no such unique signals that lead to the recognition of NEs. Arabic does not deal with capital letters. Hence, it is harder to detect the NEs (Elsebai, 2009)
- The Arabic language is highly inflectional: Most single words in Arabic language have more than one affixes such as: Word = prefix (es) + lemma + suffix (es). The prefixes can be articles, prepositions, or conjunctions. Also, the suffixes can be objects or personal/possessive anaphora (AbdelRahman *et al*., 2010). For example, the Arabic word “وبحسناتهم” is intended to mean in English “and by their virtues”. In order to tackle this problem, it is needed to perform a segmentation of each word (tokenization) as a pre-processing step. It helps particularly for the NER task to overcome two major difficulties: (i) make the NEs appear always in the same form (which lowers the number of unseen NEs); (ii) reduce the number of surface forms of the contexts in which the NEs appear.

Related work: Research activities in the area of Arabic NER are a recent development. Abdul-Hamid and Darwish (2010) conducted a study where they applied a basic set of features that could strongly identify NER for Arabic without the common requirement for morphological or syntactic analysis or gazetteers. The same process was applied to word series features and word extent. The recommended set of features gave a better result of a 9 point F-measure improvement for recognizing persons. Elsebai *et al*. (2009) designed a system to improve and execute Arabic named entity recognition for persons’ names. The model that was chosen employed a rule based approach which in turn makes use of the Bulkwater Arabic Morphological Analyzer (BAMA). The results achieved an F-measure of 89% which was better than the two results produced by Persons Named Entity Recognition of Arabic (PNERA) system in which its first result with gazetteers achieved about 87.5% of F-Measure and the second without gazetteers is about 75% of F-Measure. Shaalan and Raza (2008) used a rule based approach to develop a Named Entity Recognition System for Arabic (NERA). The performance results achieved an F-measure of 87.7% for the person, 85.9% for location and 83.15% for organization. Benajiba *et al*. (2008a) used SVM to investigate the influence of using different sets of features that were both language independent and language dependent for Arabic NER. Their system got the highest performance of an F1 score of 82.71. Benajiba *et al*. (2008b) applied SVM and CRF to check the effect of using diverse sets of features for Arabic NER. Their system indicated a performance 83.5 % of F1 measure of Automatic Content Extraction (ACE2003) and Broadcast News data ACE 2004 and ACE 2005 data sets respectively. Yassine and Rosso (2008) attempted to improve NERA system by using the Conditional Random Fields (CRF) method. It was apparently evident that CRF obtained good results when combined with appropriate features. On the other hand, it obtained less desirable results when it used the same features individually. When combining "all" the features, the following results were obtained: 86.90% for precision, 72.77% for recall and 79.21 for F-measure.

Benajiba *et al*. (2007) presented an NER system constructed for Arabic texts based on Maximum Entropy (ME). They built their own training and test corpora (ANERcorp) and gazetteers (ANERgazet) to evaluate and present their system. When they used ANERsys (without using ANERgazet) on the ANERcorp test they got a precision result of 62.72%, recall 47.58% and f-measure of 54.11%. Unlike when they applied ANERsys (using ANERgazet) on the ANERcorp test, they achieved 63.21% precision, 49.04% recall and 55.23% F-measure. The related work an attempted to present outline the remarkable works that have been done in named entity recognition for Arabic text so far. As the aim of this review to improve the accuracy that obtained from past research by introduce neural network technique in the literature.

MATERIALS AND METHODS

The classes of named entity: A Named Entity in Arabic language exists in many types. Table 1 illustrates some of the proper names and each one with an example:

Neural Network: The main task of neural network approach is to automatically learn to recognize component patterns and make intelligent decisions based on available data and it can also be applied to classify new information within large databases. The data for training and testing have been taken from experiments (Sathyabalan *et al*., 2009).

Neural Network has several advantages that make it feasible and therefore encourage its implementation in different system applications (Gupta, 2006; Ahmadi *et al*., 2008). The advantages are as follows:

- Adaptive learning: An application with a neural network component has the capability to learn how to perform tasks based on training data or initial experience
- It also has the ability to derive meaning from complex or imprecise data. Hence it can be used to recognize and extract useful patterns that cannot be noticed by humans or other computer based techniques
- Self-Organization: It can organize and precisely make representations of the information that is gathered during or through a learning process
- There are different structural designs that need diverse types of algorithms and that could be potentially difficult for an ordinary system yet comparatively uncomplicated for an artificial neural network based system

The artificial Neural Network (ANN): It is one of AI important techniques. It is considered to be a familiar approach to machine learning whereby an ANN can develop the performance and learning abilities of an intelligent system. Figure 1 shows the architecture of a typical Artificial Neural Network.

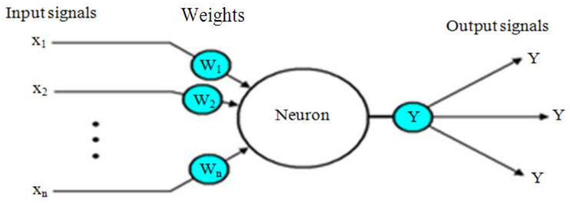

Fig. 1: The neuron as a simple computing element

Table 1: Proper names with example

العلم اسم (Named entity)	Example
الاسم (Person)	احمد (Ahmed)
المكان (Location)	كوالالمبور (Kuala Lumpur)
الشركه (Company)	مايكروسوفت (Microsoft)
التاريخ (Date)	الجمعه (Friday)
الوقت (Time)	12:30 pm
السعر (Price)	5$00
القياس(Measurement)	5 كيلوميتر (5Kilometers)
الهاتف رقم (Phone Number)	0173482402
للكتاب الدولي الرقم (ISBN)	435678 (international standard book number)

Back-Propagation Net (BPN): Back-Propagation Net (BPN) is one of most important classifier elements of the ANN. It is a feed-forward neural network which classifies layers with Log-sigmoid activation functions. The classifier of BPN propagates errors backwards during the learning process. Whenever there is a difference between actual and desired output patterns that indicates an error. In order to calculate an error and reduce it, the weights must be adjusted accordingly (Suresh *et al*., 2005; Helmy and El-Taweel, 2010) as Fig. 2.

Learning by ANN: The most famous learning algorithms are updated by back-propagation. To reduce the difference between the actual and desired outputs of back-propagation small adjustments must be made in the weights, as outlined in the next steps (Ramlall, 2010):

- Back-propagation has initial weights (random), normally in range [0.5, 0.5]
- Update to get the output agree with the training sets
- Compute the error as expected output minus actual output. Error e = Y expected -Y actual
- Adjust the weights to decrease the error

The architecture of The NERA model: The architecture of the system is shown in Fig. 3. The process includes four stages. These are data collection, Pre-Processing and Romanization and application of ANN classification.

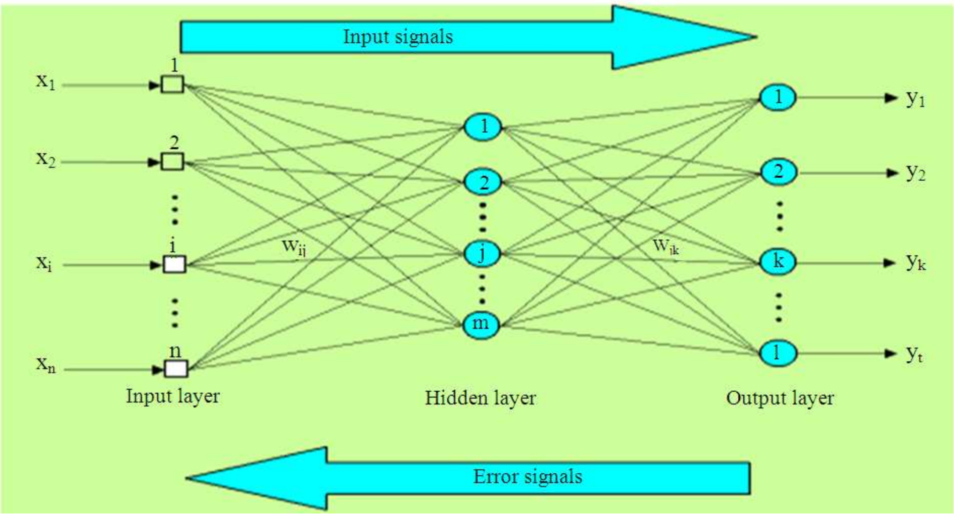

Fig. 2: Three-layer back-propagation neural networks

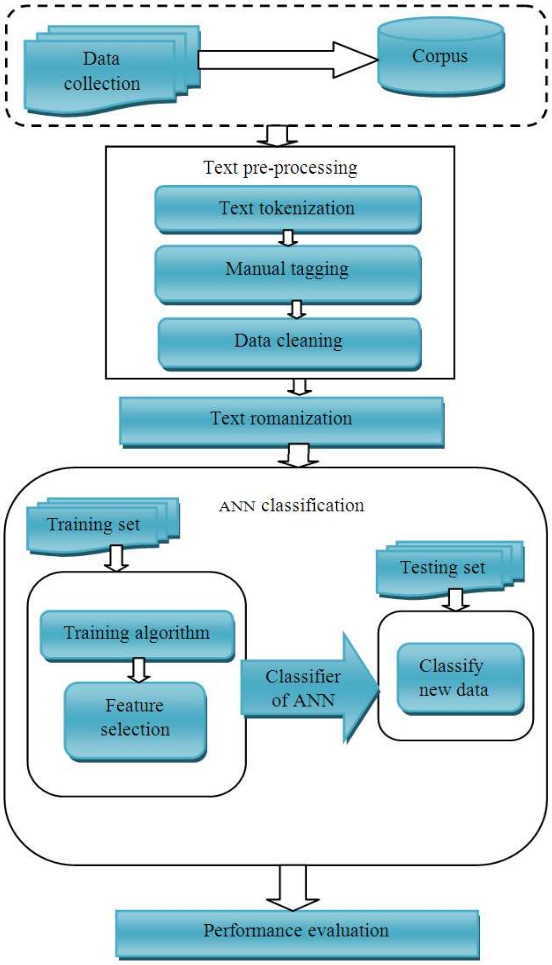

Fig. 3: General Structure of NERA model

Table 2: List of tokens

Tokens
[قام]
["]
[منظمه]
[اجنبيه]
[ل]
[سياسات]
[العامة]
[في]
[شتاء]
[عام]
[2008]
[ب]
[دراسه]
[الانتخابات]
["]

Phase1. Data Collection: ANERcorp was developed by Benajiba *et al*. (2007) and some other additional contributions have been added to it. The data is collected manually from diverse web sources. It includes 150 Kbytes and contains many articles of Modern Standard Arabic (MSA) from diverse web resources like Aljazeera web site 35%, Raya16%, Arabic. wikipedia 7%, other websites 24% and studies or magazines which account for 18%. All the articles preferred are from different types of web resources and various newsstudys in order to get a more generalized corpus. The ANERcorp is corpus that has two corpora. One corpus is for training and another one is for testing.

Phase2. Pre-processing: 1. Text Tokenization: The text tokenization is a module that is employed to divide the text into simple tokens such as number, punctuation, symbols and words. Here the corpus contains more than150 tokens.

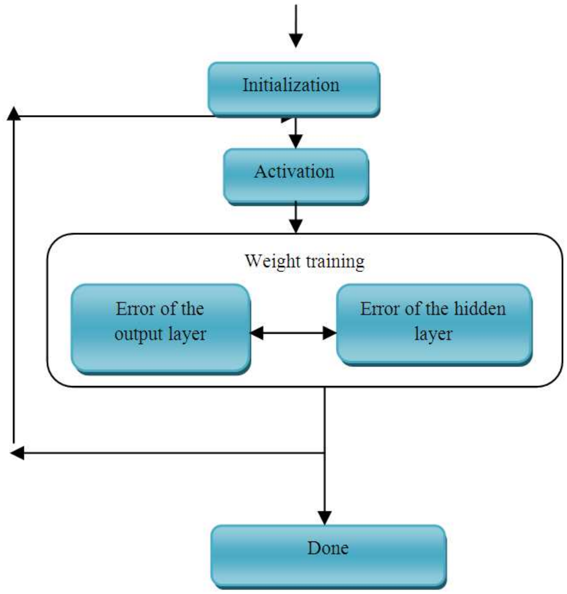

Fig. 4: A simplified flow of back propagation

Table 3: An extract of an IOB tagged corpus

Arabic	English	Tag
مع	With	O
امين	Security	O
عام	General	O
الامم	Nation	B-ORG
المتحده	United	I-ORG
كوفى	Kofi	B-PERS
عنان	Annan	I-PERS
فى	In	O
فينا	Vienna	B-LOC
.	.	O

In this phase every sentence is split into tokens .For example, Table 2 shows the list of tokens produced from this sentence:
بدراسه 2008عام شتاء فى العامه لسياسات اجنبيه منظمه"قام
."الانتخابات (The foreign organization for public policies in the winter of 2008 studying elections).

Manual Tagging (IOB): At this stage, the Inside-Outside-Beginning (IOB) tagging scheme by Benajiba *et al*. (2008b) is involved. The corpus has the same classes which have been used in the CoNLL2002 conference for (person, location, organization and Misc). Table 3 shows the tagging scheme in the IOB of ANERcorp for English translation of the 'With Security Union Nation United Kofi Anan in Vienna.'

Data cleaning: This is an important step in data processing for NER for Arabic language. Since the data collected from various web sources is not clean, it must undergo this process. In order to make the data suitable for use into the proposed model, the data must be cleaned by removing words from the English language, removing symbols like quota, full stops, parentheses and, question marks or removing numbers and empty spaces.

Phase 3Text Romanization: Text Romanization is the process of converting the text of non-Latin into Latin letters. This helps a person who speaks a foreign language and does not know the original alphabet. He or she simply reads the sounds of the language (Yousif, 2007). Most Arabic NLP researchers prefer to use the Buckwalter transliteration (a simple one-to-one map from Arabic letters to Roman letters). The programs for converting Arabic letters to Roman alphabets are designed to make Arabic texts more accessible and easy to handle in the proposed system.

Phase4. ANN Classification: Training algorithm: Back- Propagation Net (BPN) is one of most important classifier elements of the ANN. It is a feed-forward neural network which classifies layers with Log-sigmoid activation functions. The classifier of BPN propagates errors backwards during the learning process, whenever there is a difference between actual and desired output patterns that indicates an error. In order to calculate the error and reduce it, the weights must be adjusted accordingly. The back-propagation training algorithm can be modified and improved by adjusting the weights of the inputs with supervised learning (Jones, 2008).

Figure 4 shows a simplified flow of back propagation that explains the steps of training algorithm.

The steps are as follows: Step1.Initialisation: Determine all weights initialization of the network to random numbers also threshold value θ, normally in small range as [-0.5, 0.5].

$$\left(-\frac{2.4}{F_i},+\frac{2.4}{F_i}\right)$$

where is the total number of inputs of neuron i in the network. The weight initialization is done on a neuron-by-neuron basis.

Step2.Activation: To Activate the BPN by using inputs x1(p), x2(p),…, xn(p) and desired outputs yd,1(p), yd,2(p),…, yd,n(p).

Compute the actual outputs of the neurons in the hidden layer:

$$y_j(p) = \text{sigmoid}\left[\sum_{i=1}^{n} x_i(p)\cdot w_{ij}(p) - \theta_j\right]$$

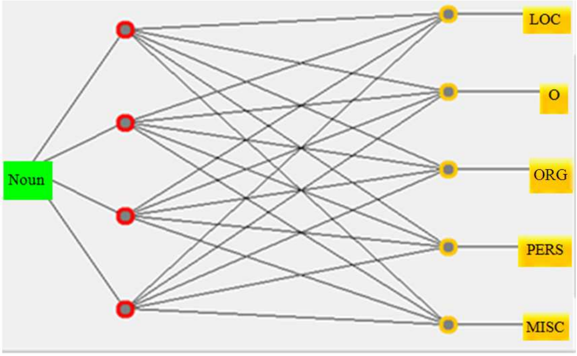

Fig. 5: An example for the neural network output

where, sigmoid is the sigmoid activation function, n is the number of inputs of neuron j in the hidden layer and θ is threshold value to compares the weighted sum of the input signals with value threshold θ that value given.

Compute the actual outputs of the neurons in the output layer:

$$y_k(p) = \text{sigmoid}\left[\sum_{j=1}^{m} x_{jk}(p) \cdot w_{jk}(p) - \theta_k\right]$$

where m is the number of inputs of neuron k in the output layer.

Step 3. Weight training: Update the weights in the BPN spreading backward the errors that related with output neurons. Calculate the error gradient for the neurons in the output layer and the hidden layer, as follows:

- For the output: $w_{jk}(p+1) = w_{jk}(p) + \Delta w_{jk}(p)$,

Where $\Delta w_{jk}(p) = \alpha \cdot y_j(p) \cdot \delta_k(p)$

- For the hidden layer $w_{ij}(p+1) = w_{ij}(p) + \Delta w_{ij}(p)$,

where $\Delta w_{ij}(p) = \alpha \cdot x_i(p) \cdot \delta_j(p)$.

Step 4: Iteration: Increase the iteration p by one, then go back to Step 2 and repeat the process until the selected error criterion is satisfied

Feature Selection: The most challenging aspect of any machine learning approach is deciding on the optimal feature sets. By selecting the appropriate set of features, a good of classification can be obtained as follows:

Person:

- السيد-السيده-الأستاذ-الأستاذه”/“Mr.-Mrs.” as good person name indicators habitually come before names.
- بن -الأول-الثاني-...“ - ”/“the son of -the first-the second,..” may often occur as parts of person names.
- حفظه الله-سدد خطاه -رعاه الله- المغفور له“ - ”/“Bless him” may be frequently occur before or after person names

Location:

- ميدان -شارع ”/ “street-square”
- العاصمة “ - /The Capital”
- جزيره - جبل- ميدان- بحر“ - ”/“Island-Mountain-Square-Sea” may occur as parts of location names.

Organization:

- A s parts of organization names, the words, such as,”صحيفه – منظمه – حزب – شرآه – وكاله – إتحاد – معرض“

The Fig. 5 shows an example for how the final result would be after all the processes are completed. The neural network categorizes the token as one of the named entity types which are: LOC, ORG, PERS or MISC. If the token does not belong to one of these types, the neural network will categorize it as ‘O’ which refers to others.

RESULTS AND DISCUSSION

The evolution methods: We have used the f- measure below for evaluation:

$$F = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} * \text{recall}}$$

where, precision is the percentage of NEs found by the system and which are correct. It can be expressed as:

$$\text{precision} = \frac{\text{number of correct named entries found by the sySTem}}{\text{number of named entries found by the system}}$$

And recall is the percentage of NEs existing in the corpus and which were found by the system. It can be expressed as:

$$\text{Re call} = \frac{\text{number of correct named entries found by the system}}{\text{total number of NEs}}$$

Summary of results:

Corpus: The corpus used in the evaluation contains many articles of Modern Standard Arabic (MSA) which were collected from diverse web sources. The corpus includes more than 150 tokens. In addition, percentage of NEs per class in the evaluation corpus is 39% for PERSon, 30.4% for LOCation, 20.6% for ORGanization, 10% for MISCellaneous class as shown in Fig. 6.

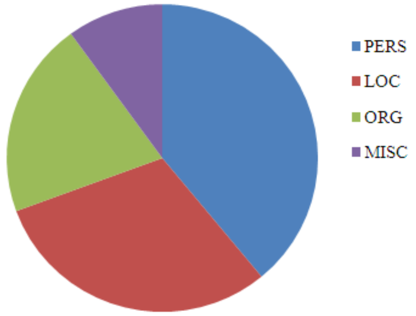

Fig. 6: Percentages of NEs per class in the evaluation corpus.

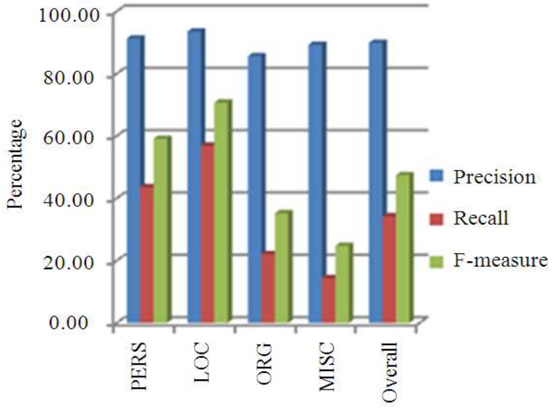

Fig. 7: Comparison between the measurements by using Neural Network

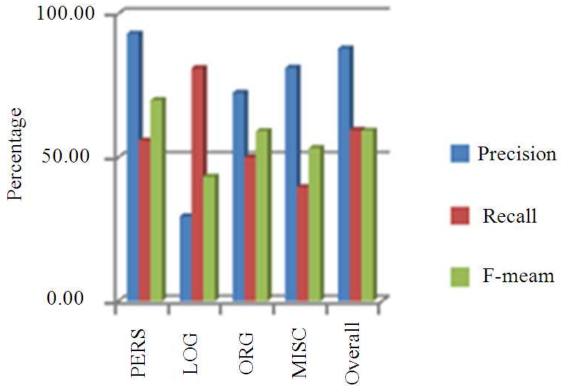

Fig. 8: Comparison between measurements of Decision Tree

Experimental results: The corpus was split into 90% for the training set and remaining set is for testing where training set represents the input values for the classification model of ANN. Moreover, the corpus represents the data entries in this model. The Decision Tree was chosen in the comparison between the proposed models. The aim of the experiments presented here is to evaluate the performance of neural network compared with other machine learning approach. The result of ANN models was compared with the result of Decision Tree with using the same dataset. Decision tree as mentioned above is well known as a common classification technique because it can handle huge data size (Cai, 2008) Decision tree provides directly the categorization of NE types and due to this fact, the decision tree classifier is a straight method by humans (Paliouras *et al*., 2000). The comparison involved decision tree to ensure its ability in contrast with neural network for NERA task. . The results obtained after executing all the experiments are as follows:

First phase of experiments of Neural Network (ANN): The Table 4 shows the accuracy in terms of the precision, recall and F-measure for each class of named entity in Arabic by applying neural network.

Figure 7 shows the comparison between the three standard evaluation measures for classes of Named Entity by applying Neural Network. The aim of this comparison is to show the different in percentage of each measure in each class. It is clear that the precision measure in each class gave higher percentage over the other two measures.

First Phase of Experiments of Decision Tree (DT): On the other hand, Table 5 shows the accuracy in terms of precision, recall and F-measure for each class of named entity in Arabic by using Decision Tree.

The Fig. 8 illustrates the average of the measurements for each class of named entity by using Decision Tree. Consequently, it can ensure that the precision measure gives higher percentage among the other measures.

Third Phase of Experiments of Comparison between (ANN) and (DT): In this comparison, the corpus was used in different sizes to examine the relationship between the sizes of the corpus and the accuracy. In the first comparison, 10k byte from the corpus was used for training and testing and the overall accuracy for precision measurement was taken in each techniques. Table 6 summarizes the results between the two techniques. The results showed that the neural network approach performs better than Decision Tree. Table 6 summary of results between the two techniques Furthermore, the results showed that the neural network approach obtain higher accuracy than Decision Tree as illustrated in the curve in Fig. 9.

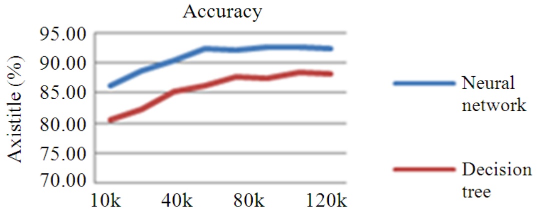

Fig. 9: Comparison between the accuracies

Table 4: The measures for each class by neural network

NE	Recall	Precision	F-measure
PERS	55.90%	93.10%	69.90%
LOC	81.10%	29.50%	43.30%
ORG	50%	72.50%	59.20%
MISC	39.70%	81.20%	53.30%

Table 5: The measures for each class by Decision Tree

NE	Recall	Precision	F-measure
PERS	43.60%	91.30%	59.10%
LOC	56.90%	93.50%	70.70%
ORG	22.10%	85.60%	35.20%
MISC	14.40%	89.30%	24.80%

Table 6: Summary of results between the two techniques

Size	Neural network	Decision tree
10k	86.20%	80.40%
20k	88.80%	82.05%
40k	90.47%	85.15%
60k	90.50%	86%
80k	90.67%	87.23%
100k	90.27%	87.45%
120k	90.61%	88.29%
Over All data	92.36%	87.93%

It is clear that ANN give better performance in extracting NER for Arabic. Neural network produced 92% accuracy and Decision Tree gave 87% accuracy. One promising possibility that could be extended from our technique is that it achieves the main goal of dealing with NERA problem.

CONCLUSION

The study has presented an attempt to develop the named entity recognition model for Arabic language using neural network technique. The aim of this model is to improve the precision of NER in Arabic language introduced by different approaches in the literature. ANN has been adopted in this research. The corpus was split into 90% for the training set and the remaining set is used for testing where the training set represents the input values for the classification model of ANN. The typical evaluation measures in the IE (Sitter *et al.*, 2004) i.e., Precision, Recall and F-measures have been used in the evaluation of the proposed model. The experiment performed is compared with Decision Tree approach using the same test set. The result showed that the neural network approach overcomes the Decision Tree in its performance and in terms of accuracy. The neural network achieves 92% while decision Tree gained 87% for precision measurement.

REFERENCES

AbdelRahman, S., M. Elarnaoty, M. Magdy and A. Fahmy, 2010. Integrated machine learning techniques for Arabic named entity recognition. IJCSI Int. J. Comput. Sci., 7: 27-36.

Abdul-Hamid, A. and K. Darwish, 2010. Simplified feature set for Arabic named entity recognition. Proceedings of the 2010 Named Entities Workshop, (NEWS’ 10), ACM Press, USA., pp: 110-115.

Ahmadi, N., R.K. Moghadas and A. Lavaei, 2008. Dynamic analysis of structures using neural networks. Am. J. Applied Sci., 5: 1251-1256. DOI: 10.3844/ajassp.2008.1251.1256

Albared, M., N. Omar and A.B.M.J. Aziz, 2009. Arabic part of speech disambiguation: A survey. Proceedings of the 5th International Conference on Rough Set and Knowledge Technology, (RSKT’ 09), pp: 517-517.

Benajiba, Y., 2009. Arabic named entity recognition. Ph. D. Thesis, University of Valencia, Spain.

Benajiba, Y., M. Diab and P. Rosso, 2008a. Arabic named entity recognition: An SVM-based approach. Proceedings of the International Arab Conference on Information Technology, Dec. 16-18, Sfax, Tunisia.

Benajiba, Y., M. Diab and P. Rosso, 2008b. Arabic named entity recognition using optimized feature sets. Proceedings of the Conference on Empirical Methods in Natural Language Processing, (EMNLP’ 08), ACM Press, USA., pp: 284-293.

Benajiba, Y., P. Rosso and J.M. BenediRuiz, 2007. ANERsys: An Arabic named entity recognition system based on maximum entropy. Comput. Linguistics Intell. Text Process., 4394: 143-153. DOI: 10.1007/978-3-540-70939-8_13

Cai, J., 2008. Decision Tree Pruning Using Expert Knowledge. 1st Edn., VDM Publishing, USA., ISBN: 9783836491556, pp: 236.

Elsebai, A., 2009. A rules based system for named entity recognition in modern standard Arabic. Ph.D. Thesis, University of Sanford, U.K.

Elsebai, A., F. Meziane and F.Z. Belkredim, 2009. A rule based persons names Arabic extraction system. Commun. IBIMA, 11: 53-59.

Grover, C., E. Klein, C. Manning, K. Markert and M. Nissim, 2008. Machine learning of entity recognizers for modular retargetable natural language processing.

Gupta, C., 2006. Implementation of back propagation algorithm (of neural networks) in VHDL. Ph.D. Thesis, DSpace at Thapar University (TU).

Helmy, A.K. and G.S. El-Taweel, 2010. Neural network change detection model for satellite images using textural and spectral characteristics. Am. J. Eng. Applied Sci., 3: 604-610. DOI: 10.3844/ajeassp.2010.604.610

Jones, M.T., 2008. Artificial Intelligence: A Systems Approach. 1st Edn., Jones and Bartlett Publishers, Sudbury, ISBN-10: 9780763773373, pp: 500.

Maynaed, D., Y. Li and W. Peters, 2008. NLP techniques for term extraction and ontology population. Proceedings of the 2008 Conference on Ontology Learning and Population: Bridging the Gap Between Text and Knowledge, (BGTK' 08), ACM Press, Amsterdam, pp: 107-127.

Paliouras, G., V. Karkaletsis, G. Petasis and C. Spyropoulos, 2000. Learning decision trees for named-entity recognition and classification. Proceedings of the 14th European Conference on Artificial Intelligence (ECAI), Aug. 20-25, Berlin, Germany.

Ramlall, I., 2010. Artificial intelligence: Neural networks simplified. Int. Res. J. Finance Econ.

Sathyabalan, P., V. Selladurai and P. Sakthivel, 2009. ANN based prediction of effect of reinforcements on abrasive wear loss and hardness in a hybrid MMC. Am. J. Eng. Applied Sci., 2: 50-53. 10.3844/ajeassp.2009.50.53

Shaalan, K. and H. Raza, 2008. Arabic named entity recognition from diverse text types. Adv. Natl. Language Process., 5221: 440-451. DOI: 10.1007/978-3-540-85287-2_42

Sitter, B., T. Bathen, B. Hagen, C. Arentz and F.E. Skjeldestad *et al*., 2004. Cervical cancer tissue characterized by high-resolution magic angle spinning MR spectroscopy. MAGMA, 16: 174-181.

Suresh, S., S. Omkar and V. Mani, 2005. Parallel implementation of back-propagation algorithm in networks of workstations. IEEE Trans. Parallel Distributed Syst., 16: 24-34. DOI: 10.1109/TPDS.2005.11

Yassine, B. and P. Rosso, 2008. Arabic named entity recognition using conditional random fields. Universidad Politecnica de Valencia.

Yousif, J., 2007. Automatic part of speech tagger for Arabic language using neural network. Ph.D. Thesis, University of UKM.