On the associative memory utilization in English- Arabic natural language processing

Ahmad Abdul Qadir AlRababah *

Faculty of Computing and Information Technology in Rabigh, King Abdulaziz University, Rabigh 21911, Saudi Arabia

A B S T R A C T

This paper discusses denotational means of lexicon dictionary by using the Associative Memory (AM), which decreased the memory location used to store different types of information. The AM flexibility retrieves the information in both sides (Arabic or English). The AM is a memory to store and retrieve the information using the associative patterns. The AM works as candidates for information and to recognize the errors caused by transmitted and wrong computed when store and retrieve the information. In other side the lexicon dictionary is supporting the artificial Intelligent Applications, especially the machine and instant translation.

1. Introduction

Neural Networks are very powerful candidates for managing and processing information systems, because one of their main characteristics is considered the dynamic performance shows constant stage that act as basing of fascination toward which nearby positions progress in time (Pustejovsky and Stubbs, 2012). This time progression of an array of neuron like components toward these stability themes can be construed as the development of an inadequate shape toward the accurate shape.

For the determination of assessment in a numerical computer, a preferred set of information called a memory is elicited when the accurate address of memory is given. This is named Address Addressable Memory (AAM) (Pustejovsky and Stubbs, 2012; Petrov, 2013; Shahabi and Rahmati, 2009). In disparity, as associative memory is capable to retrieve a full memory set of information, when it is excited with a sufficiently large slice of that memoirs information named a key. Accordingly a full set of memory information is re-called a Content Addressable Memory (CAM) (Kessentini et al., 2009; El-Beltagy and Rafea, 2009).

Data association is the process of development replicating in a more compact form. People compress data to save space in memory and decrease data transmission time (Slimane et al., 2009; Juravsky et al., 2014). That compression ratio is measurement of how well a particular piece of data is compressed.

The Associative Memory (AM) did certainly accomplish as an influential system discovery and it is one of the important forms of human erudition and to represent the information in memory location with little involved to give a human rumored and retrieves information (lane, 2010; Juravsky et al., 2014; Hassanein et al., 2011). For example, when we see someone and we need to know this person by using the outline shape (pattern recognition) and when we have some match with our store information, then we retrieve his information like color, name, tall, employee etc. store and retrieve information is called associative memory (Kessentini et al., 2009).

2. Content addressable memories (CAM)

One of the most impressive and attractive features of neural networks is the simplicity with which they can be automated as content addressable or associative memories. Content addressable memories are particularly exciting from a computer science the point of view as of the following features:

- Information is deposited and repossesses from memory identifying the stored code, which is addressed in memory.
- The associative memory is effective to identify the shapes which may use loud input, unfinished input and even fault input.

Consider that P pairs (X, Y) of associative pattern have been learned by a network, where X are the

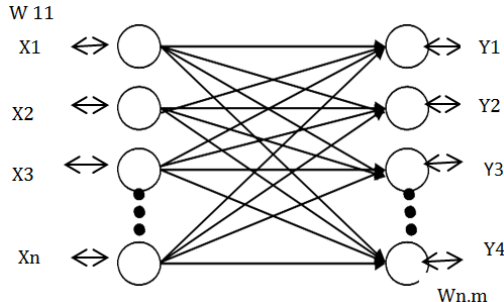input vectors and Y are the associative patterns as shows in Fig. 1.

Fig. 1: A feed forward associative Network

As a measure of associative network enactment, it is a capacity C. The capacity is the maximum number of shapes that can be deposited and retrieves the information with least average of errors (Atallah and Omar, 2009). There are several types of AM. The humblest type of AM is the linear associated. Such networks are of the feed accelerative type where the recommence vector Y is calculated from the input vector X in a single feed frontward procedure. A second type of associative network is the recurring feedback type, which has the output from one or more elements linked back to unit inputs (Al-Muhtaseb et al., 2009). These networks recall the output pattern after a cyclical recursive calculation where the output is varied with the input till the network converts constant (Chang et al., 2009) For example we consider a single layer networks with *n* input vectors and *m* associative vectors (Kandel et al., 2012), where net *j* computed as follows (Eq. 1):

$$net_j = \sum_{ij}^{W} X_{ji} = 1 = W_i.X \qquad (1)$$

where X are the input vectors and W_i is the I-th row of weight matrix W

Deliberate that a single layer network with weight matrix W, the next step is calculated the weight matrix to attain effective storage and retrieve the associative patterns.

3. Hebb's simple learning rule

Is a model for a number of associative memory ANNs. If the input to a collection of neurons is X and the output excitations are y = y(x), then the result weight matrix W_{ij} is given by Eq. 2

$$W_{ij} = \alpha x_i y_j \qquad (2)$$

where (α) is regularizing constant. The weight matrix (W) can be written in matrix representation as follows Eq. 3:

$$W = \alpha y(x)^T \qquad (3)$$

The weight matrix W in Eq. 3 stores only a single pattern pair (Xp, yp) the p-th pattern should be written in Eq. 3 as Eq. 4:

$$W^P = \alpha y^P (x^P)^T \qquad (4)$$

To indicate that the p-th pattern couple has been deposited, the remaining total weight matrix is calculated by addition the specific shape matrices as follows (Eq. 5):

$$\sum_{p=1} W^P. \qquad (5)$$

We consider an example to illustrate the above concepts. Assume that we have two pairs of patterns, (X1, Y1) and (X2, Y2) defined as follows:

X1 = (101001)   Y1= (1010)
X2 = (101110)   Y2 = (1011).

The first step in encoding these vectors into a connection matrix W, is to transform them into a bipolar form by changing (1 →1), (0 →-1):

X1 = (1-11-1-11)   Y1 = (1-11-1)
X2 = (1-1111-1)   Y2 = (1111)

The next step is computed the individual matrices Wp (Juravsky et al., 2014; Bikel and Zitouni, 2012) by using Eq. 4

$$W_1 = \begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \quad W_2 = \begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$$

Now we compute the weight matrix W by using Eq. 5

$$W = \begin{bmatrix} 2 & -2 & 1 & 0 \\ -2 & 1 & -2 & 0 \\ 2 & -2 & 1 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & -2 \end{bmatrix}$$

How we can be recovered the associative patterns **y**? To recover the associative pattern we multiply the input vector (for example X1) by the weight matrix *(W* 1*)* and the result is Net j, as follows: Net y1 = (6, – 6, 6, – 6) and the bipolar , vector is y1 as: y1 = (1, –1, 1, –1).

Data Compression is the approach that permits big files to be reduced so they take less area of storage in memory and decreases transmission time. Code words = {0, 1}; Code Length: Table 1.

2*3+3*3+4*3+5*3+6*3+7*3+8*3+5*3 = 120

Table 1: Data compression

Source message	Code	Source message	Code
a	00	B	001
c	011	D	011
e	100	F	101
g	110	Space	111

An image file or few minutes of sound file can take up huge area of memory. Similarly large text files needs lots of memory space. Large files not only occupy lots of memory, but they can also be very slow in transmission.

Records compression is frequently denoted to by way of coding, where coding is precise universal term surrounding several superior demonstrations of records, which gratifies a certain requirement. Information theory is demarcated to be the learning of effective coding and its significance in the formula of speediness of communication and possibility of mistake.

Records compression can be noticed as a division of information theory. A humble classification is that is comprises converting a series of characters in various demonstration in to novel series (of bits) which hold several information but whose size is as lesser as probable.

Records compression has essential solicitation in the zones of data communication and data stowage. Numerous data processing presentations entail storing of big capacities of data, besides the amount of such solicitations is frequently growing as the use of computer encompasses to new punishments. At the same time, the propagation of computer communication networks is causing in huge transmission of data over communication relations.

Compacting data to be deposited or transferred or bargain storage and/or over communication charge. Once the volume of data to be transferred is compact, the conclusion is that of aggregate the size of the communication networks. Likewise, squeezing a folder to a half of its original dimension is correspondent to doubling the volume of stowage average.

It can then become achievable to store the data at a greater, therefore quicker, level of the stowage pyramid and decrease the capacity on input/output networks of the computer system.

Data density is of importance in corporate data treating, both as of the price saving it proposals and since of a huge size of data employed in several business solicitations.

The kinds of resident severance existing in business data files comprise shots of zero in numeric fides, structure of spaces in alphanumeric field, and fields, which are existing in various accounts and null in others. Null destruction may be consummate over the usage of incidence bits. Alternative technique adventures circumstances in which a restricted set of characteristic values occurs. Lexicon replacement requires substituting alphanumeric illustrations of information such as bank account type; assure policy, sex, month, etc. By rare bits compulsory to extant the limited quantity of probable characteristic values.

A code: Is a mapping of source message into code words or a message from a source alphabet to code alphabet?

Source Message: Words from the source alphabet and they are two elementary parts into which the series to be characterized is divided. These elementary parts may be solitary symbols from the foundation alphabet or they may be strings of symbols (Table 2).

Code words: Words of the alphabet code and its determination are taken as {0, 1}.

Codes categorized as:

- **Block-Block:** indicate that the source messages and code words are of stable size.
- **Variable-Variable:** Map variable-length source message into variable-length code words.

Example = if we have the following string character:

aa bbb cccc ddddd eeeeee fffffff gggggggg:

We want to code the block-block code and variable-variable code. Source message will be = {a, b, c, d, e, f, g, space}.

Table 2: Block-block code

Source message	Code	Source message	Code
Aa	0	bbb	1
Ccc	10	ddddd	11
Eeeeee	100	gggggggg	110
Fffffff	101	space	111

Variable-Variable
Code Length:

1*1+1*1+2*1+2*1+3*1+3*1+3*1+3*5 = 30

A code is separate if every word code is illustrious from all others.

A distinct code is exclusively decodable: if each word code is distinguishable when engrossed in a series of code word.

A uniquely decodable is a prefix of (prefix free code) if it has prefix property, which requires that no code words are a proper prefix of any other code words.

Coding: The procedure of transmuting is a foundation collaborative (sequence of message) into a code message.

Encoder: The process that constructs and use the charting to transmute the source collaborative.

Decoder: Perform the inverse operation, restoring the coded message to its original form as shown in Fig. 2.
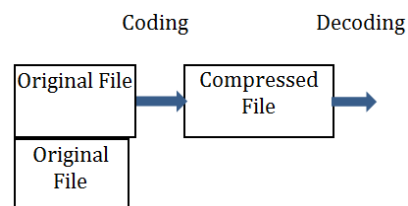

Fig. 2: Coding-decoding process

4. Results and discussion

This research discusses the representation of Lexicon that represents two dimensional arrays. The

first dimension contains the word and the other contains its meaning, by using the associative memory, which gives us the ability to connect the information and their characteristics.

The AM is able to store and retrieve the information in an effective and simply way by computing the weight matrix, which is considered as a monitoring function to recognize the associative patterns.

The using of AM to represent the Lexicon dictionary is decreased the used memory location, because the associative patterns are represented in memory by littlie size, for example suppose that we have length of word (16 bits), this word is storing as associative pattern with size (10 bits) or less than, which helps to reduce huge memory area.

Note that each location must have a unique address and can hold a unique data element.

The input information translates in a suitable form that depends on usage coding, like the ASCII Code.

The representation of different words, that has different lengths, produces different vectors lengths. This leads to a new problem, which must be solved. Also, it will increase the storage size in memory, because every character represents as a vector and the total vectors become as the results. The final connections between these vectors required the computing of weight matrix for every pattern W, by using Eq. 4. The total weight matrix must be computed from equal vectors. Therefore, that required addition to new cells in order to equal the vectors. Moreover, this becomes over flow storing, leads to slow retrieving. To explain the above concepts, consider the following two different words and their meanings:

Go (yd<u>h</u>hb)
(dftr) - Copybook

By using the ASCII Code for coding the English words, and the ASM0449 System to code the Arabic words. We will have the following vectors shown in Table 3.

Table 3: ASCII and ASM0449code

G	1000111	<u>h</u>	1110011
O	1001111	y	0101011
C	100011	d<u>h</u>	0000101
P	1010000	b	0001001
Y	1011001	d	1111001
B	1000010	f	1000011
K	1001011	t	0101001
		r	1000101

The user inputs the information by using the keyboard or any other input units, then the information stored in memory as it is clear in Table 3. The address represents the method or technique in which the computer refers to that cell or location, while the contents represent the data stored in that particular memory location.

We compute W1 for the first word and its meaning as matrix (14*28) and W2 For the second word and its meaning as matrix (56*28). In order to compute the total weight matrix (W), we must have equal matrices. This research discusses the operation of storing and retrieving the Lexicon dictionary by using the addresses to reduce the amount of storage Table 4. The link between the word and its meaning is represented as vectors decrease in memory.

The representations of address as a vector decrease the storage size, because the address contains in maximum 16 bits. That produces particular matrices W2 in maximum (16*16).

Consider that we have the following terms and particular matrices contain their meaning and their locations as following shown in Table 3.

Table 4: Meaning-location linkage

Word	Location	Meaning	Location
Able	1001101	mqtdr	10101
Poet	1100101	s<u>h</u>aʿr	11011
Level	1110011	mstwy	10011

Then we have the following connections

X1 = (1001101)   y1 = (10101)
X2 = (1100101)   y2 = (11011)
X3 = (1110011)   y3 = (10011)

By using the meaning of bipolar vectors the above vectors can be written:

X1 = (1-1-111-11)   y1= (1-11-11)
X2 = (11-1-11-11)   y2 = (11-111)
X3 = (111-1-111)   y3 = (1-1-111)

Now we can compute the total weight matrix W by using Eq. 5 as follows:

$$\begin{bmatrix} 3 & -1 & -1 & 1 & 3 \\ 1 & 1 & -3 & 3 & 1 \\ -1 & -1 & -1 & 1 & -1 \\ -1 & -1 & 3 & -3 & -1 \\ 1 & 1 & 1 & -1 & 1 \\ -1 & -1 & -1 & 1 & -1 \\ 3 & -1 & -1 & 1 & 3 \end{bmatrix}$$

Finally if we need to retrieve the meaning of the third word (Level), we must compute the net j by using Eq. 1 as follows:

$net_j = (5 - 3 - 11\ 11\ 5).$

Moreover, if we write this vector in bipolar form, we have

$net_j = (1 - 1 - 1\ 11).$

5. Conclusion

The suggested method for representing the lexical semantic dictionary would help much in retrieving words in both English and Arabic for it depends on an associative memory. By contrast, the present-day approach stars with the English word so as to find out its' meaning in Arabic.

The suggested method is highly competent and more economical as regard with the amount of storage because we are trying to store the accompanied patterns with almost half of their size or less than that. This provides a greater storage area because of the vast number of words and their meanings that are going to be stored. The more words stored are the more competent and useful the dictionary will be.

As a result, an active system will appear for natural language processing. It also reduces the time required for search and retrieval because the weight function will be counted only once and the required pattern could be reached directly –furthermore, it makes it possible to add and modernize the dictionary data easily by re-calculating the weight function by installing the new directions into the computer. These will represent the words that are added or modernized – they also allow to process words weather regular or distorted. This is so because the computer will deal with the specific direction of the words rather than the words itself.

Acknowledgement

References

Al-Muhtaseb HA, Mahmoud SA, and Qahwahi RS (2009). A novel minimal script for Arabic text recognition databases and benchmarks. International Journal of Circuits, Systems and Signal Processing, 3(3): 145-153.

Atallah AS and Omar K (2009). A comparative study between methods of arabic baseline detection. In the International Conference on Electrical Engineering and Informatics, IEEE, Selangor, Malaysia, 1: 73-77. https://doi.org/10.1109/ICEEI.2009.5254814

Bikel D and Zitouni I (2012). Multilingual natural language processing applications: from theory to practice, IBM Press. Indianapolis, USA.

Chang Y, Chen D, Zhang Y, and Yang J (2009). An image-based automatic Arabic translation system. Pattern Recognition, 42(9): 2127-2134.

El-Beltagy SR and Rafea A (2009). KP-Miner: A keyphrase extraction system for English and Arabic documents. Information Systems, 34(1): 132-144.

Hassanein AT and Abdou K, and El Seoud DA (2011). The concise Arabic-English Lexicon of verbs in context. The American Univesity in Cairo Press, Cairo, Egypt.

Juravsky D and Martin JH (2014). Speech and language processing. Pearson Education, Upper Saddle River, USA

Kandel ER, Schwartz JH, Jessell TM, Siegelbaum SA, Hudspeth AJ (2012). Principles of neural science. McGraw-Hill Education, New York, USA.

Kessentini Y, Paquet T, and Hamadou AB (2009). A multi-lingual recognition system for Arabic and Latin handwriting. In the 10th International Conference on Document Analysis and Recognition, IEEE, Barcelona, Spain: 1196-1200. https://doi.org/10.1109/ICDAR.2009.55

lane EW (2010). An Arabic-English lexicon: Derived from the best and the most copious eastern sources. Nabu Press, Charleston, USA.

Petrov AA (2013). Associative memory-based reasoning: A computational model of analogy-making in a decentralized multi-agent cognitive. LAMBERT Academic Publishing, Saarbrücken, Germany.

Pustejovsky J and Stubbs A (2012). Natural language annotation for machine learning. O'Reilly Media, Sebastopol, USA.

Shahabi F and Rahmati M (2009). A new method for writer identification of handwritten Farsi documents. In the 10th International Conference on Document Analysis and Recognition, IEEE, Barcelona, Spain: 426-430. https://doi.org/10.1109/ICDAR.2009.290

Slimane F, Ingold R, Kanoun S, Alimi AM, and Hennebert J (2009). Database and evaluation protocols for Arabic printed text recognition. Internal Research Report (DIUF), University of Fribourg, Fribourg, Switzerland. Available online at: http://diuf.unifr.ch/diva/APTI/publications.html