ISSN No. 231-2705 | DOI: 10.51244/IJRSI | Volume X Issue II February 2023

Polynomial Networks Model for Arabic Text Summarization

Mohammed Salem Binwahlan

Information Technology Department, College of Applied Science, Seiyun University

Received: 24 January 2023; Accepted: 08 February 2023; Published: 21 March 2023

Abstract- Online sources enable users to get their information needs. But, finding the relevant information, in such sources, became a big challenge and time consumption due to the massive size of data those sources contain. Automatic text summarization is an important facility to overcome such a problem. To this end, many text summarization algorithms have been proposed based on different techniques and different methodologies. Text features are the main entries in text summarization, where each feature plays a different role for showing the most important content. This study introduces the polynomial networks (PN) for Arabic text summarization problem. The role of the polynomial networks (PN) is to compute optimal weights, through the training process of PN classifier, where these weights were used to adjust the text features scores. Adjusting the text features scores creates a fair dealing with those features according to their importance and plays an important role in the differentiation between higher and less important ones. The proposed model produces a summary of an original document through classifying each sentence as summary sentence or non-summary sentence. Six summarizers (Naïve Bayes, AQBTSS, Gen–Summ, LSA–Summ, Sakhr1 and Baseline–1) were used as benchmarks. The proposed model and benchmarks were evaluated using the same dataset (EASC – the Essex Arabic Summaries Corpus). The results shew that the proposed model defeats the all six summarizers. In addition, the rate error results of both the proposed model (PN classifier) and Naïve Bayes (NB classifier), it is a clear that the proposed model (PN classifier) works better. In general, the proposed model provides a good enhancement indicating that the polynomial networks (PN) are a promising technique for text summarization problem.

Keywords- Automatic text summarization, polynomial networks, sentence similarity, term frequency, text feature.

I. Introduction

Online sources enable users to get their information needs. But, finding the relevant information, in such sources, became a big challenge and time consumption due to the massive size of data those sources contain. For this reason, it is believed that automatic text summarization is, the process of scanning a full text for discovering its parts bearing the most important meaning and presenting those parts in a limited size space, an important facility to overcome such a problem. The requirement of including the most informative parts in that limited size space (which is called a summary) addresses a big challenge. Such challenge forces the researchers in the area of text summarization to deal with it in two directions, the first one is how to determine the most important parts of a full text and second one is how to control the inclusion of those parts in the limited size space (the summary). A summary of the full text content helps readers to make a decision to read the whole document or not. Reading the summary, instead of the full text, can save the time and effort. To this end, many text summarization algorithms have been proposed based on different techniques and different methodologies. Those proposed algorithms were classified into two main categories, extractive and abstractive (Mani, 2001). Extractive algorithms insert the most important parts of the original document, without changing the structure of those parts (simple copy), into the final summary. Similar to extractive algorithms, abstractive algorithms insert the most important parts of the original document into the final summary, but after editing the structure of those parts (perform paraphrasing). And this makes the abstractive algorithms more complicated than extractive algorithms.

The cornerstone of automatic text summarization systems is those approaches which dates back to the 1950s and 1960s [Luhn, 1958; Edmundson, 1969]. Such approaches depend on a linear combination of shallow features of text units to calculate the score of these units [Luhn, 1958; Edmundson, 1969; Baxendale, 1958]. Luhn (1958) proposed that the word significance is determined by frequency of its occurrence and the significance of sentence is determined by the relative position of its words. A combination of these two measurements determines the significance factor of a sentence. The highest score sentences are chosen as summary sentences “auto-abstract” whereas the sentences are reordered based on their significance order. Edmundson (1969) presented summarization system to generate extracts in which four features are used: word frequency, positional importance, cue words, and title or heading words. Each sentence is scored by the weights of the four features. Each feature is given a weight manually. The advantages of these approaches are simplicity and efficiency. In Baxendale's study (1958), a sentence is selected as a candidate for the summary based on its position. The sentence appearing in the beginning and the end of the paragraph has been given more significance. Zechner (1996) presented a pure statistical abstract-based system employing only tf*idf weight to score the text sentences. The system is a neutral of domain knowledge and text characteristics. Although automatic text summarization has gained researchers’ attention since Luhn’s work (Luhn, 1958) but the topmost work on it started from the year 2000 (Binwahlan, 2015). Narayan et al. (2018) criticized a number of text summarization methods (Cheng and Lapata, 2016; Nallapati et al., 2017; Narayan et al., 2017; Yasunaga et al., 2017) being having two weaknesses: lacking of a ranking-based objective, which learn to order the candidate sentences according to their importance and another weakness is no convergence between the evaluation criterion, used by those works, and the learning objective. The researchers proposed a reinforcement learning model based text summarization method, to overcome above mentioned weaknesses through optimization process for evaluation metric and learning to get a suitable sentence ranking for final summary creation. The proposed method composes of three main parts; the first one is Convolutional Neural Networks (CNNs) based sentence encoder, second one is Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) based document encoder and third one is sentence extractor based on RNNs + LSTM for summary creation). The objective function consists of a mix of maximum-likelihood cross-entropy loss and policy gradient RL based rewards.

The features are the main entries in text summarization [Luhn, 1958; Edmundson, 1969], each feature plays a different role for showing the most important content. Extraction of the most important content is affected by the features selected. So, the feature selection received much attention by many approaches [Luhn, 1958; Edmundson, 1969; Baxendale, 1958; Zechner, 1996; Binwahlan et. al.,2009b; Aristoteles et. al., 2012]. Therefore, this work focuses on those more complicated features rather than simple features.

Unlike English text, which has received much attention of the researchers in this field, Arabic text is still lake to such serious investigations. This reason gave the author of this paper strong motivation to participate in a pushing Arabic language into the concern domain of automatic text summarization researchers by introducing the polynomial networks (PN) [Campbell, et. Al., 2001; AL-Tahrawi & Abu Zitar, 2008; Al-Tahrawi & Al-Khatib, 2015] for Arabic text summarization problem. To the best of my knowledge, the polynomial networks (PN) are investigated in a problem of text summarization, in general, for the first time in the literature. The role of the polynomial networks (PN) is to compute optimal weights, through the training process of PN classifier, where these weights were used to adjust the text features scores. Adjusting the text features scores using some weights [Binwahlan et. al.,2009b; Aristoteles et. al., 2012] creates a fair dealing with those features according to their importance and played an important role in the differentiation between the features having high importance and those having low importance. Most features used in the proposed model are of different structures and more complicated features; because such features can guide to the most important content of any document (Binwahlan, 2009b). In addition, most of those features are considered to identify the relation between each sentence and the document containing it. Many works on text summarization considered concepts (or key phrase) as key entry to discover most important text units in a document [Imam, et. Al., 2013a; Imam, et al. 2013b; D'Avanzo, et. al, 2004; Hirao, et. al., 2004]. The proposed model is similar to these methods in terms of considering the importance of multi-word concepts, but it is more similar to (Hirao, et. al., 2004). The proposed model creates the multi-word concepts as bi-grams (two words), it divides the sentence into a number of bi-grams (each bi-gram shares its first word with the its precedent bi-gram (if any) and shares its second word with the its subsequent bi-gram (if any)). Hirao et. al. (2004) divides the sentence into a number of uni-grams, bi-grams, tri-grams, and there are no shared words between any two successive divisions. Another complicated feature used by the proposed model, and consider more important, is log entropy which was used by Gulcin et. al. (2010) for Turkish Texts. Another feature that can reflect the relation between each sentence and the document containing it is a sentence similarity to its document [Ko & Seo, 2008; Kim & Hwang, 2001].

II. Arabic Text Summarization

Recently, Arabic language gained an attention of automatic text summarization researchers. In this section, a number of works, which have been done on Arabic text, are reviewed in general. Where, the first attempt to push Arabic language into the concern domain of automatic text summarization researchers was in 2004 when Douzidia and Lapalme (2004) proposed an extractive approach based summarizer, called "Lakhas, that generates very short (headline) summary. The scoring mechanism considers four features: sentence position in the document, number of subject terms (i.e., words that appear in the headline) in the sentence, number of "indicative words" in the document, and the tf.idf value of each word in the sentence. "Lakhas" participated in the DUC-2004 task 3 of generating very short summaries (75-bytes). Conroy *et. al.* (2006) introduced a summarization system called CLASSY, in which, to score a sentence, oracle score is used, which enable to determine the number of terms shared between abstract and a sentence. That score led to propose a new score called "approximate oracle", the new score determines whether a sentence is included in the summary or not. A new Traveling Salesperson (TSP) formulation was used to order sentences in the summary. CLASSY achieved a high performance in the DUC 2006 competition. El-Haj *et. al.* (2009) presented two Arabic summarization systems based on vector space model (VSM) [Salton *et. al.*, 1975]: Arabic query-based text summarization system (AQBTSS) and Arabic concept-based text summarization system (ACBTSS). In both system, term frequency (TF) and inverse document frequency (IDF) are used as weighting scheme to determine the relevance degree which is represented in the VSM. The problem with those two systems is that they require a user participation in the summarization process. Elhaj (2012) proposed two generic extractive single-document summarizers for Arabic, based on the idea of [El-Haj *et. al.*, 2009], called "Gen–Summ" and "LSA–Summ". In these two systems, the user participation in the summarization process, as in [El-Haj *et. al.*, 2009], was removed. Gen–Summ is special query-based summarizer; because it uses the document's first sentence as a query. In Gen–Summ, VSM [Salton *et. al.*,

ISSN No. 231-2705 | DOI: 10.51244/IJRSI | Volume X Issue II February 2023

1975] is used as in [El-Haj *et. al.*, 2009]. The difference between LSA–Summ and Gen–Summ is that LSA–Summ used latent semantic analysis (LSA) [Deerwester et. al, 1990] instead of the VSM.

Imam, et. Al., (2013) presented an Arabic text query based, single document summarizer using knowledge base. The proposed summarizer consists of two modules. The first module is for building the knowledge, in which the multi-word concepts are extracted. The second module is for summarization; it starts with a user's query expansion. Then the second module is finalized by the summarization of a document. Each sentence receives score of 1 or 0.5 for each exact matching between each of its words with the original and the expanded query respectively. The current study related to Imam, et. al.'s study **(**2013) in terms of the multi-word concepts (with difference of creation and usage) and the evaluation Corpus. Ibrahim *et. al.* (2013) proposed Arabic text summarizer based on rhetorical structure theory (RST) and vector space model (VSM). The most important paragraphs in the document is identified using RST based on functional and semantic criteria. Then, those most important paragraphs are represented, and ordered based on their cosine similarity score, in VSM. El-Haj and Rayson (2013) proposed a single-document and multi-document summarizers, in which a summary is generated by selecting those sentences with high scores. The documents sentences are scored in three steps. Firstly, word frequency lists from the corpus are produced. Secondly, the log likelihood score for each word in the word frequency lists is calculated. Thirdly, the sentence score is calculated by summing up the log likelihood scores of its words. Al Qassem et al. (2019) presented fuzzy logic based Arabic text summarization method, where a number of text features are extracted for each sentence. Among those features, Tf-IDF which is assigned only to extracted nouns the sentences. The role of fuzzy logic is to score the sentences based on input features. Elgamal et al. (2019) conducted a comparison between two text summarization methods, which are clustering algorithm using an adaptive latent semantic analysis LSA model-based and the second one is deep learning using deep learning method with Restricted Boltzmann Machine RBM method. The researchers reported that the second method achieved a good performance better than first one. Lagrini and Redjimi (2021) stated, through their review for previous Arabic text summarization works, that all those works ignored text semantic and rhetorical relations, therefore, the generated summary shows low coherence and weak degree of readability. To conquer this gap, the researchers presented a method based on a rhetorical analysis and a statistical method. The method creates two summaries, initial summary which is created using rhetorical relations and final summary which is created based on a statistical module.

The rest of the paper is organized as follows: related work on Arabic text summarization using classification technique is presented in Section 2, Polynomial Networks (PNs) is presented in Section 3, Pre-Processing and Feature Extraction are presented in Section 4, text summarization using PNs is presented in section 5, experimental design, experimental results and discussion, conclusion and future work, are presented in sections 6, 7, 8, respectively.

III. Related Work

Polynomial Networks (PNs) based classifier, as machine learning technique, demonstrated its ability as a good classifier in a text classification field, and could defeat a number of well-known text classifiers in literature such as SVM, KNN, NB, Logistic Regression (LR) and Radial Basis Function Networks (RBF) [AL-Tahrawi & Abu Zitar, 2008; AL-Tahrawi, 2014; AL-Tahrawi, 2015; Al-Tahrawi & Al-Khatib, 2015]. To the best of my knowledge, PNs based classifier is never used for text summarization problem. While other machine learning approaches [Kupiec, et. al., 1995; Lin & Hovy, 1997; Lin, 1999; Conroy & O'leary, 2001; Osborne, 2002; Svore & Burges, 2007; Fattah & Ren, 2008; Alotaiby & Alkharashi, 2012] have been investigated in a such problem and proven their ability in improving the summarization performance.

A number of classification based studies, especially those for Arabic text summarization, are presented. Imam *et. al.* (2013b) modified their work in (Imam, et. Al., 2013a), they trained the decision tree algorithm (C4.5) on a set of features extracted from the original documents. The summary is generated by including those sentences selected by the C4.5 trained model. For the testing purpose, the Essex Arabic Summaries Corpus (EASC, 2013) 2013 was used.

A number of researchers dealt with the issue of text features scores weighting in their works [Binwahlan et. al., 2009b; Aristoteles et. al., 2012; Jain et al., 2022]. Binwahlan et al. (2009b) proposed Particle Swarm Optimization (PSO) algorithm based model, the researchers trained the model to find adjusting weights that can be used to show an importance of each text feature. Aristoteles et. al. (2012) trained a genetic algorithm model using eleven text features to achieve right weights for those features, where the model recommends the most effective mixture of those feature, that can produce a good a document summary. Jain et al. (2022) proposed Real Coded Genetic Algorithm based method, which accomplishes features weighting process based on different factors like strings of features, chromosomes selection, and reproduction operators. This aimed to assign each feature to its worth weight, and scoring the sentences fairly. So, the highest scored sentences form the final summary.

IV. Polynomial Networks (PNS)

A Polynomial network is a powerful supervised nonlinear classifier, in which a squared activation works as the usual activation function works in neural network. The functions of a polynomial network are capable of approximating any continuous multivariate function from a collection of input–output data [Liu, 2006; Shanableh & Assaleh, 2010; Blondel et. al., 2017]. A long time ago in

ISSN No. 231-2705 | DOI: 10.51244/IJRSI | Volume X Issue II February 2023

the literature, Polynomial Network (PN) have been recognized as a supervised machine learning algorithm for classification problem [Fukunaga, 1990; Campbell, et. Al., 2001; Assaleh & Al-Rousan, 2005; Liu, 2006; AL-Tahrawi & Abu Zitar, 2008; AL-Tahrawi, 2014; AL-Tahrawi, 2015; Al-Tahrawi, 2013]. In this research, the polynomial networks (PN) are introduced to a problem of text summarization.

4.1 The polynomial networks (PN) structure

The polynomial networks (PN) model, as described in [Campbell, et. Al., 2001; AL-Tahrawi & Abu Zitar, 2008; Al-Tahrawi & Al-Khatib, 2015; Shanableh & Assaleh, 2010], consists of two levels. A vector of inputs (x_1, x_2, x_3 …, x_v) is feed to the first level of network, the terms of input vector are used to build a group of K-degree order monomial basis functions p(x). for example, for two inputs vector (x_1, x_2) and a second order network, we get the following second order polynomial basis function:

$$p(x) = [1\ x_1\ x_2\ x_1^2 x_1 x_2\ x_2^2]\ Eq(2)$$

For any K-degree order polynomial, the basis terms of p(x) take the following form:

$$\prod_{j=1}^{N} x_j^{kj}\ , where\ k_j \geq 0\ and\ 0 \leq \sum_{j=1}^{N} kj \leq K\ Eq(1)$$

The outputs of the first level of the polynomial networks (PN) are used as inputs for second level. All these inputs are combined to appear in a form of a score $w^t p(x)$, where w is referred as the classification model. The inner product, as a score, is calculated for each input vector, x_i, and each class j. The final output results in the calculation of average of the total score over all feature vectors [Campbell, et. Al., 2001; AL-Tahrawi & Abu Zitar, 2008; Al-Tahrawi & Al-Khatib, 2015], as bellows:

$$S_j = \frac{1}{np}\sum_{i=1}^{np} w^{op} p(x_i)\ \ Eq(3)$$

Where np is a number of polynomial expansions of Sj.

For identification or verification of unknown inputs, the final output is used.

The proposed model produces a summary of an original document through classifying each sentence as summary sentence or non-summary sentence using Eq(3). It exploits the optimal weight, obtained from classifier learning, and second degree order basis functions, as in Eq(1), of six text features, such features can guide to the most important content of any document (Binwahlan, 2009b).

V. Pre-Processing and Feature Extraction

The proposed model exploits six text features, (average TF-ISF(ATI), sentence length(SL), sentence position(SP), sentence similarity to document (SSD), sentence concepts(SC), and log entropy(LE)) (Binwahlan, 2015), as input feature vectors to polynomial network, producing second degree order basis functions polynomial expansions of those feature vectors. A preprocessing of the original document, like breaking the input document into a list of sentences, stemming and removing stop words, is done first, and then those features, as described below, are extracted:

5.1 Sentence concepts (SC) feature

The extraction of this feature is similar to [Binwahlan, 2015; Hirao, et. al., 2004]. Multi-word concepts are crated as bi-grams (two words), where the sentence is divided into a number of bi-grams (each bi-gram shares its first word with the its precedent bi-gram (if any) and shares its second word with the its subsequent bi-gram (if any)). Whereas Hirao *et. al.* [2004] divides the sentence into a number of uni-grams, bi-grams, tri-grams, and there are no shared words between any two successive divisions.

After extraction of each sentences concepts, every sentence is scored based on the following equation:

$$SC(s_j) = \left[\left[\sum_{i=1}^{nc(s_j)} \sum_{k=1}^{m} \sum_{q=1}^{nc(s_k)} c_i \times c_q \right] \Big/ MAX\_SC \right] | \ j \neq k \qquad Eq.1$$

Where *SC* is a sentence concept score, *nc* is a number of concepts in a sentence, *m* is a number of sentences in the document and *MAX_SC* is a maximum.

ISSN No. 231-2705 | DOI: 10.51244/IJRSI | Volume X Issue II February 2023

5.2 Log entropy (LE) feature

Log entropy is another complicated feature was used by *Gulcin et. al. (*2010), for Turkish Texts, and Binwahlan (2015).

$$sum = \sum_i p(i,j)\log_2(p(i,j)) \qquad Eq.2$$

$$global(i) = 1 + \frac{sum}{\log_2(m)} \qquad Eq.3$$

$$local(i,j) = \log_2(1+f(i,j)) \qquad Eq.4$$

$$EL(s_j) = global \times local \qquad Eq.5$$

where p(i,j) is the probability of word i that is appeared in sentence j, f(i,j) is the number of times word i appeared in sentence j, and m is the number of sentences in the document.

5.3 Average TF-ISF(ATI) feature

Because the term frequency is considered as one of the key entries to find the important sentences in a document is [Zechner, 1996; Binwahlan, 2015]. An average Tf-Isf are used to evaluate each document sentence by average Tf-Isf weights summation of its words. Tf.Idf method [Salton, 1989] is modified to calculate Tf-Isf weight Eq.6 (Binwahlan, 2015), where the document parameter is replaced by the sentence parameter and document collection parameter is replaced by the document parameter:

$$TI(t_i) = Tf\_Isf(t_i) = tf(t_i) + \log(TNT(D)/TON(t_i, D)) \qquad Eq.6$$

where tf(ti) is the term frequency of ith word in the sentence, TNT(D) is total number of terms in the document, and TON(ti, D) is total number of ti occurrences in the document

$$ATI(s) = \left(\frac{\sum_{i=1}^{n} TI(t_i, s)}{TNT(D)} \right) \qquad Eq.7$$

Where ATI(s) is average Tf-Isf score of sentence s, n is a number terms in sentence s, and TI(ti,s) is Tf-Isf of term i in sentence s.

5.4 Sentence similarity to document (SSD) feature

To determine the relation between each sentence and the document containing it, a sentence similarity to its document [Binwahlan, 2015; Gulcin, et al., 2010; Ko & Seo, 2008] is used.

$$SSD(s_j) = \sum_{k=1, k\neq j}^{m} sim(s_j, s_k) \qquad Eq.8$$

Where m is a number of sentences in the document (D)

To calculate the sentence similarity between two sentences sj and sk, cosine similarity measure as in Eq.9 is used [Binwahlan, 2015; Saggion & Gaizauskas, 2004]:

$$sim(s_j, s_k) = \frac{\sum_{i=1}^{n} (w_i, s_j) \bullet (w_i, s_k)}{\sqrt{\sum_{i=1}^{n} (w_i, s_j)^2} \cdot \sqrt{\sum_{i=1}^{n} (w_i, s_k)^2}} \qquad Eq.9$$

Where wi is Tf-Isf (TI) of term ti in the sentence si or sj, n is a number terms in sentence s.

ISSN No. 231-2705 | DOI: 10.51244/IJRSI | Volume X Issue II February 2023

5.5 Sentence length (SL) feature

More important meaning, contained in any document, is possible hold in longer sentences of that document. The total number of a sentence words determines its length [Binwahlan, 2015; Nobata & Sekine, 2004].

$$SL(s_i) = TNT(s_j) / \max\_SL \qquad Eq.10$$

Where TNT(Sj) is a total number of words in sj and max_SL is the max sentence length in the document.

5.6 Sentence position (SP) feature

The opening sentence of a paragraph is more important than other paragraph sentences in the original document [Mani, 2001; Binwahlan, 2015; Lin & Hovy, 1997; Hovy & Lin, 1999; Alfonseca, et. al., 2004]. A score of 1 is given to each paragraph opening sentence and score of 0 is given to the other sentences in the same paragraph.

$$SP(S_j) = \begin{cases} 1\ if\ paragraph\ starting\ sentence \\ 0\ otherwise \end{cases} \quad Eq.11$$

VI. Text Summarization Using PNS

The proposed model is defined as a combination of weighted second degree order basis functions polynomial expansions of the document sentence feature vectors as in Eq(3). Where the values of those polynomial expansions are adjusted using the optimal weights resulting in the training of the PN classifier. Therefore, the first part in this model is for training the PN classifier, EASC data set [EASC, 2013], an Arabic natural language resources containing 153 Arabic articles cover different topics, is used as training and testing data. The second part in this model is for testing the proposed model. For a summarization task, all sentences of in document are scored, using Eq(3), and ranked in a descending order. Then the top n sentences are selected as summary, where n is equal to the predefined summary length.

6.1 The training procedure

As a first step, in the training phase, is calculating the text features for each summary sentence in the training data set. Then, features vectors, for each summary sentence in the training set, are formed. Next, a second degree order basis function is constructed for feature vector of all summary sentences in the training set using Eq (1). Using the formed second degree order basis functions, we construct a polynomial expansion matrix for each summary sentence, called Mi, as in Eq (4). So for all summary sentences, a matrix M is constructed as in Eq (5).

$$M_i = [p(x_{i,1})p(x_{i,2})p(x_{i,3}) \dots p(x_{i,Ni})] \quad Eq(4)$$

$$M = [M_1\ M_2\ M_3 \dots\ M_{nc}] \qquad Eq(5)$$

$$w_i^{opt} = arg_w min\|Mw - O_i\| \quad Eq(6)$$

A number of researcher from different fields [Campbell, et. Al., 2001; AL-Tahrawi & Abu Zitar, 2008; AL-Tahrawi, 2015] trained PN to approximate the ideal output using the mean-squared error criterion as in Eq. (6), but in our experiment, we trained PN to approximate the ideal output using a modified version of Eq. (6) as follows:

$$Ss_i = \frac{\sum_{j=1}^{n}\sum_{z=1}^{k}|Mw_{jz} - O_{jz}|^2}{n * k}\ Eq\ (7)$$

Where *Mw* is a sentence matrix consisting of *n* polynomial basis functions, Mw_{jz} represents the z^{th} basis term of the j^{th} polynomial basis functions, *k* is a number of basis terms of the polynomial basis function, *n* is a number of polynomial basis functions, Ss_i is a score of i^{th} summary sentence in a matrix *M*, and *O* is ideal outputs consists of ones. So, a vector, called $Vss(Ss_1\ Ss_2\ Ss_3,..\ Ss_{ns})$ is formed for the scores (Ss) of all summary sentences. By the end of each training run, on any summary sentences, the polynomial expansion of only one summary sentence is selected, from the matrix M, as the best. This selection decision is taken based on the value of that sentence score in the vector *Vss,* where the minimum score is a target as follows:

$$Best\_pn_d = \arg min_v\ for\ v = 1,2, \dots, ns\ Eq(8)$$

Where *ns* is a number a summary sentences and *d* is a number of a document summary in the training set. the selected polynomial expansions of the best sentences (one best sentence for each summary in the training set) were arranged in a matrix, called M_{BestPn}. The final optimal weight, w^{op} is calculated as average of all basis terms of the selected polynomial expansions in the matrix, M_{BestPn}

6.2 The testing procedure

In this experiment, leave-one-out cross validation was performed to measure the performance of the trained classifier in Arabic text summarization task. Based on that validation, the training process was run a number of times equal to a number of documents in the whole data set. Where in each run, one document is reserved for testing, as unseen document, and remaining documents are used for training the classifier. So, firstly, features weights Vectors for sentences of all documents in data set are created. Next, a second degree order basis function is constructed for feature vector of all unseen document sentences as in Eq (1). Using the formed second degree order basis functions, we construct a polynomial expansion matrix for each document sentence, called TM_i, as in Eq (4). So for all unseen document sentences, a matrix TM is constructed as in Eq (5). Next every basis term of the polynomial expansions in TM is weighted using the corresponding weight in w^{op}, the final score of any unseen document sentence is calculated as in Eq (3). The scored sentences are ranked in a descending order. Then the top n sentences are selected as summary, where n is equal to the predefined summary length. In this study, we used 50% of the total number of the document sentences as summary length. For evaluating the generated summaries, we use the ROUGE [Lin, 2004].

The classification error rate was considered as performance measure for the trained classifier as well. Each document sentence, $s_{j,}$ in the data set has a label as "SS" (summary sentence) or "NS" (non-summary sentence), that label was determined by human experts. We referred to both labels, SS and NS, as $HL(s_j)$. The trained classifier recognizes each document sentence, s_j in the test set, with label $CL(s_{j,})$ which is SS or NS, using Eq(3), where human label of a sentence s_j is unknown. Then, if the label, $HL(s_j)$, assigned to the current sentence s_j, by human, is "SS" and the trained classifier assigned a label, $CL(s_{j,})$, to that sentence as "NS", it is counted as classification error:

$$Error_j = \begin{cases} 1\ iff\ HL(s_j) = \mathrm{SS}\ and\ CL(s_j) \neq N\mathrm{S} & Eq(9) \\ 0\ otherwise & \end{cases}$$

The error rate is calculated as average of the total error numbers the trained classifier recorded on the whole testing set as follows:

$$Error_{rate} = {\sum_{j=1}^{n} error_j}/_{n} \quad Eq(10)$$

Where *n* is a number of unlabeled sentences which checked by the trained classifier.

VII. Experimental design

For the evaluation of the proposed model, The EASC (2013) is used, it is an Arabic natural language resources. It contains 153 Arabic articles cover different topics (art & music, education, environment, finance, health, politics, religion, science & technology, sports, and tourism) and 765 human-generated extractive summaries of those articles, five human summaries for each original document. These summaries were generated using Mechanical Turk [EASC, 2013]. This study follows the same strategy explained in [Binwahlan, 2015; Salton *et. al.*, 1975] for reproducing three types of human summaries (referred as level 3, level 2 and level all) based on the five human summaries of each original document. The summary of the first type (Level 3) contains all sentences appeared in at least three of the five human summaries. The summary of the second type (Level 2) contains all sentences appeared in at least two of the five human summaries. Finally, the summary of the third type (Level all) contains all sentences appeared in at least one of the five human summaries.

Summaries of length 50% is created by the proposed model. To evaluate those summaries, ROUGE (Recall-Oriented Understudy for Gisting Evaluation) toolkit [Lin, 2004] is used. ROUGE compares a system generated summary against a human generated summary to measure the quality of the system summary. ROUGE is the main metric in the DUC text summarization evaluations. It has different variants. In the experiment of this study, ROUGE-N (N = 2) is used. The reason for selecting this measure is that measure works well for single document summarization [Lin, 2004]. One of ROUGE settings is to determine a number of words to be selected from a summary being evaluated, in this study, that setting called "ROUGE-cut", where each summary is evaluated based on ROUGE-cut 100. The reason behind the determination of summaries lengths as 50%, is what reported in [Salton *et. al.*, 1975], that the length of human summary is not same for three human summary types (level all, level 2 and level 3), it falls in a range between five words and 515 words with an average of 114 words (five sentences) per summary. the average of words per the summary of, the first type (level all), the second type (level 2), the third type (level 3) is 250 words, 175 words, 98 words respectively. The summary length was also limited in text analysis conference (TAC) to be in a range from 240 to 250 words. six summarizers (Naïve Bayes, AQBTSS, Gen–Summ, LSA–Summ, Sakhr1 and Baseline–1) all those summarizers have been evaluated in [Salton *et. al.*, 1975], using the same dataset (EASC – the Essex Arabic Summaries Corpus [EASC, 2013], except the first one, which has been evaluated in this study using the same dataset. The proposed model is also evaluated using the same dataset, therefore, those six summarizers are used as benchmarks, to compare their performances with the performance of the proposed method. In addition, the error rate based performances of both Naïve Bayes and the proposed model are compared.

VIII. Experimental results and discussion

The proposed model and the benchmarks are used to create a summary for each document in the document set used in this study. Each system created a good summary compared to the reference (human) summary. The results are drawn in Figure 1. The performance of the proposed model is compared with the benchmarks based on ROUGE-2 measure only, because all the benchmarks have been evaluated in [Binwahlan, 2015; Salton *et. al.*, 1975] using that measure. Based on average of ROUGE-2 measure scores, it is noticed that the proposed model defeats the all six summarizers. In addition, Table 1 shows the rate error results of both the proposed model (PN classifier) and Naïve Bayes (NB classifier), it is a clear that the proposed model (PN classifier) works better.

Based on the evaluation results shown in Figure 1, a big difference can be realized between the performance of, the proposed model and the six benchmarks, where the proposed model outperforms all benchmarks on average scores. Where the summaries created by the proposed model and benchmarks were compared with the human summary of first type "level all"). But, the results evaluation using the human summary of second type "level 2"), it can be noticed that the results of the proposed model are better on average than the results of the six benchmarks.

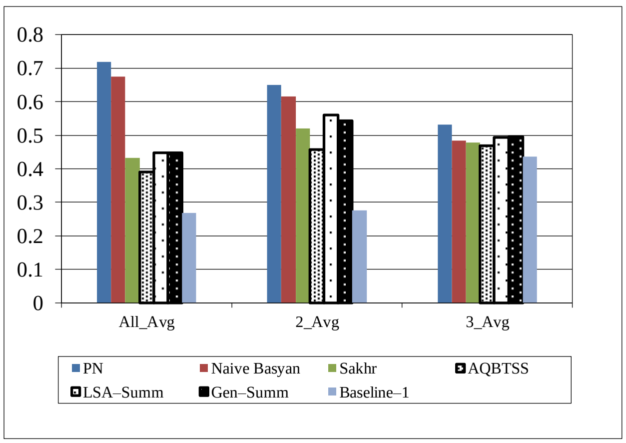

Figure 1: Comparison of the proposed model (summary length 50%, ROUGE-cut: 100) with Naïve Bayes, Sakhr, AQBTSS, LSA–Summ, Gen–Summ and Baseline–1 based on average of, ROUGE -2 measure of the three levels (Level all, Level 2, and Level 3). (no stemmer).

Table 1: Performance of PN and NB

	PN	NB
Average error	0.07389	0.17291
Max error	0.47368	0.68421
Min error	0.0	0.0

The same thing can be said on the results of, the proposed model and the six benchmarks (where the summaries created by the proposed model and benchmarks were compared with the human summary of third type "level 3"), it can be seen that the results of the proposed model defeats all benchmarks On average of recall and precision.

In general, the proposed model provides a good enhancement indicating that the polynomial networks (PN) are a promising technique for text summarization problem. We could infer two observations from the experimental results. First, as mentioned above that three types of human summaries (referred as level 3, level 2 and level all) based on the five human summaries of each original document, the summary of the first type (Level 3) contains all sentences appeared in at least three of the five human summaries. The summary of the second type (Level 2) contains all sentences appeared in at least two of the five human summaries.

ISSN No. 231-2705 | DOI: 10.51244/IJRSI | Volume X Issue II February 2023

Finally, the summary of the third type (Level all) contains all sentences appeared in at least one of the five human summaries. From experimental results, it a clear that the agreement among those humans on selection of specific sentences to form the summary sentence is low, this supported by the results drawn in Figure 1 and Table 1, where a high performance was achieved with summaries of "level all" more than the performance of other levels "level 2 and level 3" summaries, and the performance gained with summaries of "level 2" better than the performance of "level 3" summaries. Second observation, that the adjusting of the text features scores by some weights, like those obtained from the training of PN classifier, could lead to improve the performance. This is a clear, when we look at the results of the proposed model.

IX. Conclusion

In this paper, we introduced the polynomial networks (PN) for Arabic text summarization problem, where the polynomial networks (PN) are investigated in this problem for the first time in the literature. The role of the polynomial networks (PN) is to compute optimal weights, through the training process of PN classifier, where these weights were used to adjust the text features scores for creating a fair dealing with the text features according to their importance and played an important role in the differentiation between higher and less important features. The outstanding performance of the proposed model among all benchmarks, used in this study, gives an indication that the polynomial networks (PN) are a promising technique for text summarization problem. An inclusion of redundant sentences into a summary consumes its length and prevent important information from being considered as summary sentences, so for future work, we are planning, to improve our model performance by adding an annotator module for filtering such redundant sentences, and apply the proposed model for multi document summarization problem.

References

1. Mani, I. (2001). Automatic Summarization. (1st ed.). Amsterdam: John Benjamins Publishing Company.
2. Luhn, H. P. (1958). The Automatic Creation of Literature Abstracts. IBM Journal of Research and Development. 2(92), 159-165.
3. Edmundson, H. P. (1969). New Methods in Automatic Extracting. Journal of the Association for Computing Machinery. 16(2), 264-285.
4. Baxendale, P. (1958). Machine-made Index for Technical Literature - an Experiment. IBM Journal of Research Development. 2(4), 354-361.
5. Zechner, K. (1996). Fast Generation of Abstracts from General Domain Text Corpora by Extracting Relevant Sentences. In Proceedings of the 16th International Conference on Computational Linguistics. 986–989, Copenhagen, Denmark.
6. Binwahlan, M. S. (2015). Extractive Summarization Method for Arabic Text – ESMAT. International Journal of Computer Trends and Technology. 21(2), pp. 103-109. (Impact Factor: 1.517).
7. Binwahlan, M. S., Salim, N., & Suanmali, L. (2009b). Swarm based features selection for text summarization. IJCSNS International Journal of Computer Science and Network Security, 9(1), 175–179.
8. Aristoteles, Herdiyeni Y., Ridha, A., & Adisantoso J. (2012). Text Feature Weighting for Summarization of Document in Bahasa Indonesia Using Genetic Algorithm. IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 1, May 2012. ISSN (Online): 1694-0814. www.IJCSI.org
9. Campbell, W.M., & Assaleh, K.T., Broun, C.C. (2001). A novel algorithm for training polynomial networks. In: Int NAISO Symp Information Science Innovations ISI'2001, Dubai, UAE, March 2001. doi:http://dx.doi.org/10.1.1.28.5119.
10. AL-Tahrawi, M.M., & Abu Zitar, R. (2008). Polynomial networks versus other techniques in text categorization. Int. J. Pattern Recognition. Artif. Intell. (IJPRAI) 22 (2), 295–322. http://dx.doi.org/10.1142/S0218001408006247.
11. Al-Tahrawi, M.M., & Al-Khatib, S.N. (2015). Arabic text classification using Polynomial Networks. King Saud Univ. J. King Saud Univ. – Comput. Inf. Sci. 27, 437–449.
12. Imam, I, Nounou, N., Hamouda, A., & Abdul Khalek, H. A. (2013a). Query Based Arabic Text Summarization. International Journal of Computer Science and Technology. 4(2), 2013, Pp. 35-39
13. Imam, I., Hamouda, A., & Abdul Khalek H, A. (2013b). An Ontology-based Summarization System for Arabic Documents. International Journal of Computer Applications Volume 74– No.17, 2013, pp.0975 – 8887
14. D'Avanzo E., Magnini B., & Valli A. (2004). Keyphrase Extraction for Summarization Purposes: The LAKE System at DUC2004. In Proceedings of the 4th Document Understanding Conferences. DUC.
15. Hirao, T., Suzuki, J., Isozaki, H. & Maeda, E. (Hirao, et. al., 2004). NTT's Multiple Document Summarization System for DUC 2004. In Proceedings of the 4th Document Understanding Conferences. DUC.
16. Gulcin, M., Ilyas O., Cicekli F., & Alpaslan N. (2010). Text Summarization of Turkish Texts using Latent Semantic Analysis. Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), pages 869–876, Beijing, August 2010
17. Ko Y., & Seo J. (2008). An effective sentence-extraction technique using contextual information and statistical approaches for text summarization. Pattern Recognition Letters 29 (2008) 1366–1371

18. Kim, J., & Hwang, D. (2001). Korean text summarization using an Aggregation Similarity. In: Proc. 5th Internat. Workshop Information Retrieval with Asian Languages, pp. 111–118.
19. Douzidia, F. & Lapalme, G. (2004). Lakhas, an Arabic Summarising System. In Proceedings of the 4th Document Understanding Conferences, pages 128–135. DUC.
20. Conroy, J. Schlesinger, O'Leary, J. D. & Goldstein, J. (2006). Back to Basics: CLASSY 2006. In Proceedings of the 6th Document Understanding Conferences. DUC.
21. El-Haj, M., Kruschwitz, U. & Fox, C. (2009). Experimenting with Automatic Text Summarization for Arabic. In Zygmunt Vetulani, editor, 4th Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, LTC'09, "Lecture Notes in Artificial Intelligence", pages 490–499, Poznan, Poland, 2009. Springer.
22. Salton, G., Wong A., & Yang, S. (1975). A Vector Space Model for Automatic Indexing. Communications of the ACM, vol. 18, no. 11, (pp. 613–620).
23. Elhaj M. (2012). Multi-document Arabic Text Summarisation. PhD thesis, 2012, University of Essex
24. Deerwester, Dumais, S., Furnas, G. Landauer, T. & Harshman, R. (1990). Indexing by Latent Semantic Analysis. Journal of the American Society for Information Science, 41(6):391–407, 1990.
25. EASC, Essex Arabic Summaries Corpus (2013). [Online] Available: http//:privatewww.essex.ac.uk/~melhaj/easc.htm, (14-01-2013.)
26. Ibrahim, A., Elghazaly, T., & Gheith, M. (2013). A Novel Arabic Text Summarization Model Based on Rhetorical Structure Theory and Vector Space Model. International Journal of Computational Linguistics and Natural Language Processing. Vol 2 Issue 8 August 2013.
27. El-Haj, M. & Rayson, P. (2013). Using a Keyness Metric for Single and Multi Document Summarisation Proceedings of the MultiLing 2013 Workshop on Multilingual Multi-document Summarization, pages 64–71, Sofia, Bulgaria, August 9 2013. C 2013 Association for Computational Linguistics
28. AL-Tahrawi, M.M. (2014). The significance of low frequent terms in text classification. Int. J. Intell. Syst. 29 (5), 389–406. http://dx.doi.org/10.1002/int.21643.
29. AL-Tahrawi, M.M. (2015). Class-based aggressive feature selection for polynomial networks text classifiers – an empirical study. U.P.B. Sci. Bull. Ser. C 77 (2), 93–110, ISSN: 2286-3540.
30. Kupiec, J., Pedersen, J. & Chen, F. (1995). A trainable document summarizer. In Proceedings of the ACM. SIGIR conference. July 1995. New York, USA, 68-73.
31. Lin, C. Y. & Hovy, E. (1997). Identifying topics by position. In Proceedings of the Fifth conference on Applied natural language processing. March. San Francisco, CA, USA, 283-290, 1997.
32. Lin, C. Y. (1999). Training a selection function for extraction. In Proceedings of the Eighteenth Annual International ACM Conference on Information and Knowledge Management (CIKM). 2-6 Nov. 1999. Kansas City, Kansas, 55-62.
33. Conroy, J. M. & O'leary, D. P. (2001). Text summarization via hidden markov models. Proceedings of SIGIR '01. 9-12 September 2001. New Orleans, Louisiana, USA, 406-407.
34. Osborne, M. (2002). Using maximum entropy for sentence extraction. Proceedings of the ACL'02 Workshop on Automatic Summarization. July 2002. Morristown, NJ, USA, 1-8.
35. Svore, K., Vanderwende, L. & Burges, C. (2007). Enhancing single document summarization by combining RankNet and third-party sources. Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. June 2007. Prague: Association for Computational Linguistics, 448–457.
36. Fattah, M. A. & Ren, F. (2008). GA, MR, FFNN, PNN and GMM based models for automatic text summarization. Computer Speech and Language. 2008. 23(1), 126-144.
37. Alotaiby, F., Foda, S. & Alkharashi, I. (2012). New approaches to automatic headline generation for Arabic documents. Journal of Engineering and Computer Innovations Vol. 3(1), pp. 11-25, February 2012
38. Liu, C.L. (2006). Polynomial Network Classifier with Discriminative Feature Extraction, Joint IAPR International Workshops, SSPR 2006 and SPR 2006, Hong-Kong. doi:http://dx.doi.org/10.1007/11815921_80.
39. Shanableh, T. & Assaleh, K. (2010). Feature Modeling Using Polynomial Classifiers and Stepwise Regression. Neurocomputing, Vol. 73, No. 10-12, 2010, pp. 1752-1759. doi:10.1016/j.neucom.2009.11.045.
40. Blondel, M., Niculae, V., Otsuka, T., & Ueda, N. (2017). Multi-output Polynomial Networks and Factorization Machines. In Advances in Neural Information Processing Systems 30. 2017.
41. Assaleh, K., & Al-Rousan, M. (2005). A new method for Arabic sign language recognition. In: EURASIP J Appl Signal Processing. Hindawi Publishing Corporation, New York, pp. 2136–2145.
42. Fukunaga, K., (1990). Introduction to Statistical Pattern Recognition. Academic Press.
43. Al-Tahrawi, M.M. (2013). The role of rare terms in enhancing the performance of polynomial networks based text categorization. J. Intell. Learn. Syst. Appl. 5, 84–89. http://dx.doi.org/10.4236/jilsa.2013.52009.

ISSN No. 231-2705 | DOI: 10.51244/IJRSI | Volume X Issue II February 2023

44. Salton, G. (1989). Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer. Addison-Wesley Publishing Company.
45. Saggion, H., & Gaizauskas, R. (2004). Multi-document summarization by cluster/profile relevance and redundancy removal. In Proceedings of the 4th Document Understanding Conferences. DUC.
46. Nobata, C. & Sekine, S. (2004). CRL/NYU Summarization System at DUC-2004. In Proceedings of the 4th Document Understanding Conferences. DUC.
47. Lin, C. Y. & Hovy, E. (1997). Identifying Topics by Position. In Proceedings of the Fifth conference on applied natural language processing. March. San Francisco, CA, USA, 283-290.
48. Hovy, E. H. & Lin C-Y. (1999). Automated Text Summarization in SUMMARIST. In Mani I. and Maybury M. (Eds.). Advances in Automated Text Summarization. (pp. 81–94). Cambridge: MIT Press.
49. Alfonseca, E., Guirao, J. M., & Moreno-Sandoval, A. (2004). Description of the UAM system for generating very short summaries at DUC-2004. In Proceedings of the 4th Document Understanding Conferences. DUC.
50. Lin, C. Y. (2004). Rouge: A Package for Automatic Evaluation of Summaries. Proceedings of the Workshop on Text Summarization Branches Out, 42nd Annual Meeting of the Association for Computational Linguistics. 25–26 July. Barcelona, Spain, 74-81.
51. Cheng, J. & Lapata, M. (2016). Neural summarization by extracting sentences and words. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany, pages 484–494.
52. Nallapati, R., Zhai, F., & Zhou, B. (2017). SummaRuNNer: A recurrent neural network based sequence model for extractive summarization of documents. In Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, California USA, pages 3075–3081.
53. Narayan, S., Papasarantopoulos, N., Cohen, S. B., & Lapata, M. (2017). Neural extractive summarization with side information. CoRR abs/1704.04530.
54. Yasunaga, M., Zhang, R., Meelu, K., Pareek, A., Srinivasan, K., & Radev, D. (2017). Graph-based neural multi-document summarization. In Proceedings of the 21st Conference on Computational Natural Language Learning. Vancouver, Canada, pages 452–462.
55. Al Qassem, L., Wang, D., Barada, H., Al-Rubaie, A., & Almoosa, N. (2019). Automatic Arabic text summarization based on fuzzy logic, in Proc. 3rd Int. Conf. Natural Lang. Speech Process., 2019, pp. 42–48.
56. Elgamal, M., Hamada, S., Aboelezz, R., & Abou-Kreisha, M. (2019). Better Results in Automatic Arabic Text Summarization System Using Deep Learning based RBM than by Using Clustering Algorithm based LSA. International Journal of Scientific & Engineering Research Volume 10, Issue 8, August-2019. ISSN 2229-5518
57. Lagrini and Redjimi (2021). A New Approach for Arabic Text Summarization. 4th International Conference on Natural Language and Speech Processing, Trento, Italy, November 12-13, 2021. Pp: 247—256.
58. Jain, A.; Arora, A.; Morato, J.; Yadav, D.; Kumar, K.V. (2022). Automatic Text Summarization for Hindi Using Real Coded Genetic Algorithm. Appl. Sci. 2022, 12, 6584. https://doi.org/10.3390/app12136584