Arabic Text Classification Using Linear Discriminant Analysis

Fawaz Al-Anzi and Dia Abuzeina

April 18, 2018

Arabic Text Classification Using Linear Discriminant Analysis

Fawaz S. Al-Anzi
Department of Computer Engineering
Kuwait University
Kuwait City, Kuwait
fawaz.alanzi@ku.edu.kw

Dia AbuZeina
Department of Computer Engineering
Kuwait University
Kuwait City, Kuwait
abuzeina@ku.edu.kw

***Abstract*—Linear Discriminant Analysis (LDA) is a dimensionality reduction technique that is widely used in patter recognition applications. LDA aims at generating effective feature vectors by reducing the dimensions of the original data (e.g. bag-of-words representation) into a low dimensional space. Hence, LDA is a convenient method for text classification that generally characterized by high dimensional feature vectors. In this paper, we empirically investigated two LDA based methods for Arabic text classification. The first method based on computing the generalized eigenvectors of the ratio (inverse within-class and between-class) scatters, the second method include linear classification functions that assume equal population covariance matrices (i.e. pooled sample covariance matrix). We used a textual data collection that contains 1,750 documents belong to five categories. The testing set contains 250 documents belong to five categories (50 documents for each category). The experimental results show that the linear classification functions method outperforms the eigenvalue decomposition method.**

Keywords—Arabic, Text, Classification, Linear discriminant analysis, Fisher.

I. Introduction

Linear discriminant analysis (LDA), also known as Fisher discriminant analysis, is a dimensionality reduction method that is widely used in supervised learning pattern recognition. LDA is a projection a high dimensional feature vectors into a low dimensional space in order for better handling of large number of attributes such as in face recognition , text classification, or any application involving high-dimensional data. Hence, LDA seeks for reasonable objects features representation for effective use by various statistical classification methods. Dimensionality reduction, furthermore, can help to cope with the curse-of-dimensionality problem. Nevertheless, minimizing feature vectors should always preserve the essential information of the original data without sacrificing the classification performance. In addition to LDA, there are other popular dimensionality reduction methods such as singular value decomposition (SVD), principal component analysis (PCA), etc.

Large textual feature vectors generally characterize text classification problems. In particular, vector space model (VSM) is a widely used technique that based on the entire vocabulary to represent a single document. Hence, employing features reduction is, indeed, the key to the performance of the classifiers. In fact, utilizing dimensionality reduction techniques is extremely important especially for textual applications that have high-dimensional data which sometimes beyond the hardware capabilities. Text classification is generally about ten thousands dimensions in common; therefore, a dimensionality reduction technique can be though as data compression technique since it projects the original data into a smaller space. For example, It indicated in [1] that using dimensionality reduction significantly improve the performance. Reference [2] indicated that dimensionality reduction is the process of finding a suitable lower-dimensional space for several reasons such as: exploring high-dimensional data to discover structure that lead to the formation of statistical hypotheses, visualizing the data using 2-D or 3-D, and analyzing the data using statistical methods such as clustering or classification. Dimensionality reduction typically employed in very high-dimensional space domains such as data visualization, data mining, and information retrieval (IR) [3]. Even though using dimension reduction is extremely important in many applications, however, [4] listed some of constrains such as the complexity of the algorithms for data reduction, predictive/descriptive accuracy (relevant features), and representation of the Data-Mining model (simplicity of representation). The concept of dimensionality reduction is something like what Fig. 1 presented. It shows a transformation process of textual features to generate a reduced dimensional space; the reduced feature vectors are arbitrary chosen to be three dimensions.

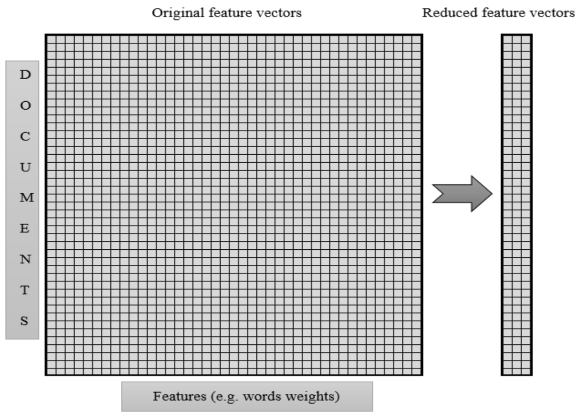

Fig. 1. Atransformation process to generate a lower dimensitonal space.

There are many algorithms to implement dimensionality reduction, intuitively; no single algorithm can be best suited for all applications. The algorithm's selection based on the reduced time, improved accuracy, and simplified representation. The LDA initially introduced for two-class classification. Later, the extension of two-class known as multiple discriminant analysis for multiclass applications. In this paper, we discuss two LDA based methods known as generalized eigenvectors of the ratio (inverse within-class and between-class) scatters and linear classification functions where each class has its linear classification rule.

We have organized the rest of this paper as follows. In the next section, we present some of the Arabic text classification challenges. In section 3, we present the literate review followed by the linear discriminant analysis in section 4. We present the proposed method in section 5 followed by the experimental results in section 6. Finally, we conclude in section 7.

II. TEXT CLASSIFICATION CHALLENGES

Text classification has a number of challenges such as noise and mixed contents documents. Noise are the words that have no discriminative power such as small words prepositions {من, الى, في, عن}. In general, such noise listed as stop words set for discarding before classification process. However, the true challenge in text classification is the mixed contents documents that have words belong to different categories. In this case, the classification process might give less than optimal performance (i.e increasing of misclassification rate). For example, Fig. 2 is an article that belong to the Religion category; however, it has some words that belong to other categories.

An example of Arabic article
في امسية احتفالية واجواء طغى عليها الطابع الرمضاني وتخللتها العديد من الفقرات والجوائز القيمة اقام البنك الاهلي الكويتي غبقته السنوية وذلك يوم الثلاثاء الموافق يونيو بقاعة الراية التابعة لفندق ماريوت كورت يارد شهدت الامسية التي ضمت اكثر من موظف من مختلف الادارات العديد من الفقرات والفعاليات الممتعة والجوائز القيمة التي فاز بها الموظفون في البداية رحب الرئيس التنفيذي للبنك الاهلي الكويتي ميشال العقاد بالموظفين موجها لهم كلمة بمناسبة حلول شهر رمضان الكريم وشكر الموظفين على ما يبذلونه من جهد لتقديم افضل الخدمات المصرفية متمنيا لهم الاستمرار في العطاء والعمل الجاد تضمنت الامسية العديد من الفقرات والسحوبات التي منحت الموظفين فرصة الحصول على جوائز قيمة من بينها هواتف سامسونغ وشاشات سامسونغ وتذاكر سفر الى لندن وباريس واسطنبول ودبي بالاضافة الى الجائزة الكبرى وهي سيارة جي ام سي تيرين الجديدة والتي كانت من نصيب الفائز اسرار احمد خان وقد اضاف لأجواء الامسية وجود اثنين من فناني الكاريكاتير وهما السيد محمد القحطاني والسيد محمد خليل اللذان امتعا الحضور برسوماتهم الكاريكاتيرية المميزة للموظفين طوال الامسية كما استمتع الحضور بالموسيقى على انغام الفرقة الكويتية وذلك وسط اجواء رمضانية مبهجة
The translation using Google translator
In the evening festive atmosphere overshadowed by the character of Ramadan and punctuated by several paragraphs and prizes hosted Ahli Bank of Kuwait Annual Gbakth and on Tuesday, June Hall of the Hotel Courtyard Marriott banner saw the evening, which included most of the employees from the various departments of several paragraphs and events fun and valuable prizes that won by staff at first Chief Executive of the Bank of Kuwait Ahli Michel Accad welcomed staff directed them to a speech marking the holy month of Ramadan and thanked the staff for their efforts to provide the best banking services and wished them to continue in the tender and hard work included the evening several paragraphs and raffles that gave employees a chance get valuable prizes including Samsung and screens, Samsung and tickets to travel to London, Paris, Istanbul and Dubai phones in addition to the grand prize, a GMC Terrain's new car, which went to the winner of the secrets Ahmed Khan has been added to the atmosphere of the evening and the presence of two artists, cartoonists, namely Mr. Mohammad Al-Qahtani Mr. Mohamed Khalil, who Amtaa attendance Brsumaathm cartoons special staff throughout the evening as attendees enjoy music on the music of the Kuwaiti band amid the atmosphere of Ramadan exhilarating

Fig. 2. An Arabic article with its translation using Google translator.

Examples of particular categories' words include some related to Technology such as (سامسونغ=Samsung, شاشات=screens, هواتف=Phones), Economy (البنك=bank, الرئيس التنفيذي=chief executive), Tourism (لندن=London, باريس=Paris, دبي=Dubai, فندق = hotel), Sports (الفائز = winner, جائزة = prize), Art (فنان = artist, بالموسيقى = music). In this article, only one word founded that directly related to the Religion category that is (رمضان= Ramadan). However, this word might be a person name. In addition, other words are neither stop swords nor categories' words such as the word (اسرار) which mean (secrets), but it used as a person name. The Arabic article in Fig. 2 translated to English using Google translator in [5]. The document intentionally has no numbers, commas, full stops, English characters, or other symbols such as :{à ® é \ ² … è ² ™ ¯ > × − ô é ü ÷ ¹ × ‘ ° © ½ ¾ ë <> ' [ ] ” _ ; ؛ -- · = % ’ * , / : ! « ? » ؟} as the document has to only include Arabic letters in the preprocessing stage.

Intuitively, since the discriminate words play a crucial role to classify a document, the variety of document's contents might be a source of errors. The previous example shows that text classification requires methods for choosing high quality discriminative words especially when the corpus documents significantly have such content diversity. For more details of Arabic challenge, the reader can refer to reference [6].

III. LITERATURE REVIEW

The literature shows that LDA used in different pattern recognition applications specially face recognition. However, few studies found to employ LDA for text classification. In this literature, we demonstrate some of LDA based research in different classification domains. Reference [7] presented LDA based for multiclass text categorization. Reference [8] presented LDA based method for document classification. The work in [9] investigated several generalized LDA algorithms and compared their performances in text classification and face recognition. Reference [10] employed LDA for detection and confirmation of counterfeits in documents of legal importance. Reference [11] demonstrated a comparison between LAD and PCA using a face database. Reference [12] introduced a Gabor representation and LDA for face recognition. Reference [13] proposed a dual-space LDA approach that found to be more stable to use all discriminative information compared with the standard LDA approach in high dimensional face recognition. Reference [14] proposed kernel discriminant analysis, which deals with the nonlinearity of the face patterns' distribution as well as effectively solving the "small sample size" problem.

A comprehensive comparison between 1D-LDA (standard) and 2D-LDA (matrix-based) presented in [15]. The work in [16] presented that 2D-LDA (based on image matrix features) has many advantages over other methods such as 2D-PCA. Reference [17] employed LDA for assessing the geographical origin and the year of production for olive oils from various Mediterranean areas. Reference [18] used LDA for forensic classification of ballpoint pen inks using high performance liquid chromatography and infrared spectroscopy. Reference [19] demonstrated a method to avoid the LDA singularity problem by an intermediate dimension reduction stage using

PCA before LDA. Reference [20] employed LDA gender classification.

IV. Linear Discriminant Analysis

LDA is dimensionality reduction technique that projects N data vectors that belong to c different classes into a (c-1) dimensional space in such way that the ratio of between group scatter S_b and within group scatter S_w is maximized [21]. Hence, the classes' centroids in the transformed space spread out as much as possible. Fig. 3 shows the projection of some points (two dimensions) into a new line (z) defined by the set of weight parameters. The discriminant function yields the scores on z is given by the following form: $z = w'x = w1 \times x1 + w2 \times x2 + \ldots + wk \times xp$, where p is total number of dimensions in each feature vector.

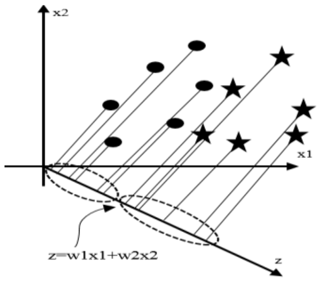

Fig. 3. A projection of some points into a new space.

Hence, the first step is to find a set of weight values that maximize the ratio of the between-class scatter to within-class scatter using the training set samples. The eigenvalue decomposition of the between-class (Sb) to within-class (Sw) scatter matrices is the key towards finding the weight values (eigenvectors). Getting the lower dimensions achieved by discarding some eigenvectors that has smallest corresponding eigenvalues. Table I shows the formulas used to extract the weights that are used to generate the transformed values in the reduced space.

TABLE I. Extracting Weights Formulas

<table>
<tr><td>within-class (Sw) scatter</td><td>between-class (Sb) scatter</td></tr>
<tr><td>$\sum_{classes\ c} P_c \times Cov_c$</td><td>Sb=covariance (training data set)-Sw</td></tr>
<tr><td colspan="2">Eigenvalues and Eigenvectors = eigenvalue decomposing ($\frac{Sb}{Sw}$)</td></tr>
</table>

Once finding the eigenvectors (i.e the weights) that correspond to some of the maximum eigenvalues, the transformed values of the new space are generated by multiplying the weights with the the original data. For classification, the discriminant function z used to classify a new data point. That is, the new sample is transformed to the new space for classification purpose. The classification process performed using a distance measure such as Euclidean distance or cosine similarity measure. The LDA based classification algorithm summarized as follows:

1) *Finding weights through Eigenvalues decomposition*
   *= inverse (within-class) * between-class covariance*
   = Eigenvalues decomposition of $Sw^{-1}Sb$
2) *The original data vectors are transformed into the new dimensional subspace*
   *= Selected maximum Eigenvectors * Original data points*
3) *Find the the transformed score of the testing data point*
   *= Eigenvectors * a new data point*
4) *The transformed score compared with classes' centroids using any distance measure*

To simplify the projection process, we consider a small data set that contains 15 samples belong to three classes. The following Fig. 4 shows some information of this set. In this example, we demonstrate how to represent 3-dimenstinal feature vectors into 1-dimenstional feature vectors using LDA.

<table>
<tr><td colspan="2">Class 1 (5 instances)</td><td colspan="2">Class 2 (5 instances)</td><td colspan="2">Class 3 (5 instances)</td></tr>
<tr><td colspan="2">-5,-8,-6 ➔1
-7,-6,-8➔1
-8,-6,-7➔1
-6,-4,-5➔1
-5,-8,-4➔1</td><td colspan="2">1,2,3➔2
3,2,1➔2
2,1,5➔2
2,4,3➔2
1,4,3➔2</td><td colspan="2">10,14,11➔3
12,11,14➔3
11,9,12➔3
14,10,11➔3
12,10,9➔3</td></tr>
<tr><td colspan="2">Apriori probability=5/15</td><td colspan="2">Apriori probability=5/15</td><td colspan="2">Apriori probability=5/15</td></tr>
<tr><td colspan="2">The covariance of class 1 (Cov1)
[[0.56 -0.36 0.50]
[-0.36 0.93 -0.16]
[0.50 -0.16 0.83]]</td><td colspan="2">The covariance of class 2 (Cov2)
[[0.23 -0.11 -0.16]
[-0.11 0.60 -0.16]
[-0.16 -0.16 0.66]]</td><td colspan="2">The covariance of class 3 (Cov3)
[[0.73 -0.51 -0.05]
[-0.51 1.23 0.03]
[-0.05 0.033 1.10]]</td></tr>
<tr><td colspan="3">Within-class scatter: Sw = 5/15×Cov1+5/15×Cov2+5/15×Cov3
[[1.53 -1.00 0.28]
[-1.00 2.76 -0.30]
[0.28 -0.30 2.60]]</td><td colspan="3">The covariance matrix of the training data points is (C):
[[59.40 54.33 56.10]
[54.33 55.23 53.21]
[56.10 53.21 56.31]]</td></tr>
<tr><td colspan="3">Between-class scatter (Sb=C-Sw):
[[57.87 55.33 55.81]
[55.33 52.47 53.51]
[55.81 53.51 53.71]]</td><td colspan="3">Eigenvalues and Eigenvectors:
[-92.75 0.0079 -7.0]
[[0.78 -0.79 0.053]
[-0.25 -0.56 0.68]
[-0.55 -0.24 -0.72]]</td></tr>
<tr><td colspan="3">After Sorting the Eigenvalues and the Eigenvectors:
[0.0079 -7.0 -92.75]
[[-0.79 0.053 0.78]
[-0.56 0.68 -0.25]
[-0.24 -0.72 -0.55]]</td><td colspan="3">The final weight values is the first column of the Eigenvectors since we need just 1 dimension:
[[-0.79044526]
[-0.56090417]
[-0.24613575]]</td></tr>
<tr><td colspan="6">The transformed value of the 15 training instances:
[[9.91], [10.86], [11.41], [8.21], [9.42], [-2.65], [-3.73], [-3.37], [-4.56], [-3.77], [-18.46], [-19.10], [-16.69], [-19.38], [-17.30]]</td></tr>
<tr><td colspan="6">The centroids of the new data representation: class 1, class 2, class 3
[[9.96736178], [-3.61955955], [-18.19096662]]</td></tr>
<tr><td colspan="6">The testing set : [[-3,-3,-3],[2,2,2],[10,10,10]]
The transformed values: [[4.79], [-3.19], [-15.97]]
Clearly, the points belong to class 1, class 2, and class 3, respectively.</td></tr>
</table>

Fig. 4. Classification using Eigenvalue Decomposition

Fig. 5 shows the original data in three-dimensional space. The goal of LDA is to have these pointes represent using less than three features, one or two, of course without losing the discriminative quality for each class.

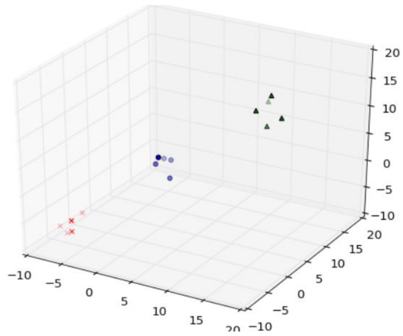

Fig. 5. Three class data set in the orginal space.

Fig. 6 shows a chart that represent how LDA transforms 3-dimensional input feature vectors into a reduced 1-dimensional feature vectors with preserving the differences among the classes' centroids. Even the points clearly separated into three different classes; however, some overlaps might appear in complex applications. In Fig. 6, the new transformed values used for both coordinates x, y, and z. For example, the value 9.91 is used for x, y, and z coordinates. Naturally, the new instance assigned to the closest centroid in the projected space.

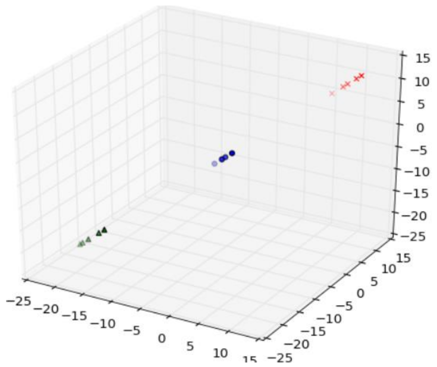

Fig. 6. The Data in the transformed space of three classes.

Although a single discriminant function z can separate samples into several classes, however, multiple discriminant analysis used to construct a separate discriminant function for each class. That is, instead of solving for the eigenvalue/eigenvector, linear classification functions can be used for classification. Linear classification functions estimate the common population covariance matrix by a pooled sample covariance matrix known as (S_{pl}):

$$\mathbf{S}_{pl} = \frac{1}{N-k}\sum_{i=1}^{k}(n_i - 1)\mathbf{S}_i$$

Where (Si) is the covariance matrix of the class i, k is the total number of the classes, N is the total number of training instances in all classes, and n_i is the number of instances in a particular class. For classification using linear classification functions, assign the new instance (y) to the group for which Li(y) is maximum as follows [22]:

$$L_i(\mathbf{y}) = \overline{\mathbf{y}}_i'\mathbf{S}_{pl}^{-1}\mathbf{y} - \tfrac{1}{2}\overline{\mathbf{y}}_i'\mathbf{S}_{pl}^{-1}\overline{\mathbf{y}}_i, \qquad i = 1, 2, \ldots, k$$

Where ($\overline{\mathbf{y}}_i$) is the mean of the class i. Fig. 7 shows how to utilize linear classification functions for the previous simple example.

Class 1 (5 instances)	Class 2 (5 instances)	Class 3 (5 instances)
-5,-8,-6 ➔1 -7,-6,-8➔1 -8,-6,-7➔1 -6,-4,-5➔1 -5,-8,-4➔1	1,2,3➔2 3,2,1➔2 2,1,5➔2 2,4,3➔2 1,4,3➔2	10,14,11➔3 12,11,14➔3 11,9,12➔3 14,10,11➔3 12,10,9➔3
mean(class 1) [-6.20 -6.40 -6.00]	mean(class 2) [1.80 2.60 3.00]	mean(class 3) [11.80 10.80 11.40]
S_{pl} = [[1.53 -1.00 0.28] [-1.00 2.76 -0.30] [0.28 -0.30 2.60]]	S_{pl}^{-1} = [[0.86 0.30 -0.05] [0.30 0.47 0.02] [-0.05 0.02 0.39]]	Testing set -3 -3 -3 2 2 2 10 10 10
The testing set : [[-3,-3,-3],[2,2,2],[10,10,10]] The scores of [-3,-3,-3] are { -1.70, -21.48, -224.24} The maximum is the first value that means it belong to class 1. The scores of [2 ,2 ,2] are { -72.38, 4.24, -95.29} The maximum is the first value that means it belong to class 2. The scores of [10 ,10, 10] are { -185.46, 45.41, 111.02} The maximum is the third value that means it belong to class 3.		

Fig. 7. Classification using Linear Classification Functions

V. THE PROPOSED METHOD

To investigate the proposed method, we prepared an Arabic text corpus that contains 1,750 documents for training and 250 documents for testing. The training set contains 929,205 words with 80,156 unique words. The collected documents belong to five categories as shown in Table II. The corpus was prepared with help by Alqabas newspaper in Kuwait [23]. Table II shows the statistical information of the corpus used.

TABLE II. THE CORPUS INFORMATION

The training data collection				
#	Category	Number of documents	Number of words	Number of unique words
1	Economy	350	225,659	31,942
2	Health	350	151,912	25,835
3	Education	350	224,078	35,294
4	Sports	350	135,473	24,056
5	Tourism	350	192,083	28,218
	Total	**1,750**	**929,205**	**80,156***
The testing data collection				
1	Economy	50	22,031	6,959
2	Health	50	29,722	9,386
3	Education	50	35,338	8,961
4	Sports	50	15,168	5,482
5	Tourism	50	20,268	6,794
	Total	**250**	**122,527**	**24,496***

* It is not algebraic summation since the common words not counted.

The preprocessing include deleting all characters that are out of the Arabic alphabetic characters. It also include deleting numbers, commas, full stops, and all other symbols. A normalization process also performed to change some Arabic characters such as (أ→ا) and (إ→ا). The stoplist declared that contains all common words in the five categories of training documents. The proposed method summarized in the following algorithm:

1) *Stoplist declared as the common words of the training categories documents.*
2) *Document Frequency (DF) feature is set.*

3) *Training and testing feature vectors generated using VSM.*
4) *LDA used to generate the transformed feature vectors of training and testing sets.*
5) *The Euclidian distance used for classification using categories centroids of the transformed values.*

VI. Experimental Results

This section presents the experimental results of the proposed method. The stoplist contains 1,846 words. In addition to stoplist, we discarded all one or two characters since they have no effect in classification process, an example of single character is the shorthand of doctor (Dr ➔ د). Different DF values used in the experiments as indicates in table III. The DF aims at discarding any word that appears in less than or equal DF threshold. Hence, VSM considered all words appear in more than DF threshold to create 1,750 feature vectors of the training set. Each dimension of a created VSM feature vector cantinas the total number of a particular word's occurrences in the document. Of course, the VSM was used to create the testing set feature vectors based on the dictionary prepared using the training data set. The total number of features of each feature vector specified according to the choice of DF threshold. Therefore, the total number of words' dictionary based on the selected DF threshold. Table III shows the proposed method performance using both eigenvalue decomposing and linear classification functions. The results shows that text classification using linear classification functions outperforms the eigenvalue decomposing.

TABLE III. The Results Using eigenvalue decomposition and Linear Classification Functions

#	*DF*	*# of original dimensions*	*Eigenvalue Accuracy (%)*	*Linear functions Accuracy (%)*
1	64	335	82.0	83.2
2	65	315	81.2	84.4
3	66	305	80.8	82.4

Table IV shows the accuracies of each class (DF=65) using both eigenvalue decomposition and linear classification functions. Four LDA dimensions used with the results presented in Table IV.

TABLE IV. The Accuracy of each Class

#	*Category*	*Eigenvalue Accuracy (%)*	*Linear functions Accuracy (%)*
1	Economy	82	88
2	Health	86	86
3	Education	100	98
4	Sports	66	72
5	Tourism	72	78
	Average	**81.2**	**84.4**

We repeated the experiments for different LDA dimensions in the transformed space with the following values: one, two, and three dimensions. The reported accuracies (with DF=65) are 46.0%, 63.2%, 74.4%, respectively. We then repeated the experiments using DF equal to 40. This DF gives 923 features; the accuracy scored 70.0% using both eigenvalue decomposition and linear classification functions method (only four dimensions used).

VII. Conclusion

This paper discussed the LDA dimensionality reduction method for Arabic text classification. The Euclidean distance measure used for classification. Even LDA is not widely used for text classification; however, the results reveal that LDA is an option for this task. The results demonstrate how to employ eigenvalue decomposition and linear classification functions to reduce high dimensional feature vector into very few dimensions with an acceptable results. With document frequency (DF=65), the LDA reduces the dimensions form 315 to 4 with 84.4% accuracy for five categories. However, we emphasize that this study is to explore the LDA possibilities for text classification as, we believe, other classification method might give better performance such as SVD along with cosine similarity measure.

References

[1] Marsland, Stephen. Machine learning: an algorithmic perspective. CRC press, 2015.

[2] Martinez, Wendy L., et al. Exploratory data analysis with MATLAB. CRC Press, 2010.

[3] Theodoridis, S. and K. Koutroumbas (2008). Pattern Recognition, Fourth Edition, Academic Press.2010.

[4] Kantardzic, Mehmed. Data mining: concepts, models, methods, and algorithms. John Wiley & Sons, 2011.

[5] Avialable: https://translate.google.com/

[6] Al-Anzi, Fawaz S., and Dia AbuZeina. "Stemming impact on Arabic text categorization performance: A survey." 2015 5th International Conference on Information & Communication Technology and Accessibility (ICTA). IEEE, 2015.

[7] Li, Ming, and Baozong Yuan. "2D-LDA: A statistical linear discriminant analysis for image matrix." Pattern Recognition Letters 26.5 (2005): 527-532.

[8] Torkkola, Kari. "Linear discriminant analysis in document classification."IEEE ICDM Workshop on Text Mining. 2001.

[9] Park, Cheong Hee, and Haesun Park. "A comparison of generalized linear discriminant analysis algorithms." Pattern Recognition 41.3 (2008): 1083-1097.

[10] Silva, Carolina Santos, et al. "Classification of blue pen ink using infrared spectroscopy and linear discriminant analysis." Microchemical Journal 109 (2013): 122-127.

[11] Martínez, Aleix M., and Avinash C. Kak. "Pca versus lda." IEEE transactions on pattern analysis and machine intelligence 23.2 (2001): 228-233.

[12] Liu, Chengjun, and Harry Wechsler. "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition."*IEEE Transactions on Image processing* 11.4 (2002): 467-476.

[13] Wang, Xiaogang, and Xiaoou Tang. "Dual-space linear discriminant analysis for face recognition." Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. Vol. 2. IEEE, 2004.

[14] Lu, Juwei, Konstantinos N. Plataniotis, and Anastasios N. Venetsanopoulos. "Face recognition using kernel direct discriminant analysis algorithms." IEEE Transactions on Neural Networks 14.1 (2003): 117-126.

[15] Zheng, Wei-Shi, Jian-Huang Lai, and Stan Z. Li. "1D-LDA vs. 2D-LDA: When is vector-based linear discriminant analysis better than matrix-based?."Pattern Recognition 41.7 (2008): 2156-2172.

[16] Li, Tao, Shenghuo Zhu, and Mitsunori Ogihara. "Using discriminant analysis for multi-class classification: an experimental

investigation." Knowledge and information systems 10.4 (2006): 453-472.

[17] Rezzi, Serge, et al. "Classification of olive oils using high throughput flow 1 H NMR fingerprinting with principal component analysis, linear discriminant analysis and probabilistic neural networks." Analytica Chimica Acta 552.1 (2005): 13-24.

[18] Kher, Ashwini, et al. "Forensic classification of ballpoint pen inks using high performance liquid chromatography and infrared spectroscopy with principal components analysis and linear discriminant analysis." *Vibrational spectroscopy* 40.2 (2006): 270-277.

[19] Ye, Jieping, Ravi Janardan, and Qi Li. "Two-dimensional linear discriminant analysis." Advances in neural information processing systems. 2004.

[20] Jain, Amit, and Jeffrey Huang. "Integrating independent components and linear discriminant analysis for gender classification." Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on. IEEE, 2004.

[21] Duda, Richard O., and Peter E. Hart. Pattern classification and scene analysis. Vol. 3. New York: Wiley, 1973.

[22] Rencher, Alvin C. Methods of multivariate analysis. Vol. 492. John Wiley & Sons, 2003.

[23] Avialable : http://www.alqabas.com.kw/Default.aspx