Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Arabic Text Classification based on Analogical Proportions

Myriam Bounhas[1] | Bilel Elayeb*[2] | Amina Chouigui[2] | Amir Hussain[3] | Erik Cambria[4]

[1] LARODEC Research Laboratory, ISGT, Tunis University, Tunisia
[2] RIADI Research Laboratory, ENSI, Manouba University, Tunisia
[3] School of Computing, Edinburgh Napier University, Edinburgh EH10 5DT, UK
[4] School of Computer Science and Engineering, Nanyang Technological University, Singapore

Correspondence
*Bilel Elayeb Email: bilel.elayeb@riadi.rnu.tn

Present Address
RIADI Research Laboratory, ENSI, Manouba University,Tunisia

Summary

Text classification is the process of labeling a given set of text documents with predefined classes or categories. Existing Arabic text classifiers are either applying classic Machine Learning algorithms such as k-NN and SVM or using modern Deep Learning techniques. The former are assessed using small text collections and their accuracy is still subject to improvement while the latter are efficient to classify big data collections and show a limited competency to classify small corpora mainly with a large number of categories.This paper proposes a new approach to Arabic text classification able to treat small and large data collections while improving the classification rates of existing classifiers. We first demonstrate the ability of analogical proportions (AP) (statements of the form "x is to y as z is to t"), which have recently been shown to be effective in classifying "structured" data, to classify "unstructured" text documents requiring preprocessing. We design an analogical model to express the relationship between text documents and their real categories. Then based on this principle, we develop two new analogical Arabic text classifiers. These latter rely on the idea that the category of a new document can be predicted from the categories of three others, in the training set, in case the four documents build together a "valid" analogical proportion on all or on a large number of components extracted from each of them. The two proposed classifiers (denoted AATC1 and AATC2) differ mainly in terms of the keywords extracted for classification. To evaluate the proposed classifiers, we perform an extensive experimental study using five Arabic text collections with small or large sizes, namely ANT (Arabic News Texts) v2.1 and v1.1, BBC-Arabic, CNN-Arabic and AlKhaleej-2004. We also compare analogical classifiers to both classical ML-based and Deep Learning-based classifiers. Results show that AATC2 has the best average accuracy (78.78%) over all other classifiers and the best average precision (0.77) ranked first before AATC1 (0.73), NB (0.73) and SVM (0.72) for ANT corpus v2.1. Besides, AATC1 shows the best average precisions (0.88) and (0.92) respectively for BBC-Arabic corpus and AlKhaleej-2004, and the best average accuracy (85.64%) for CNN-Arabic over all other classifiers. Results highlight the interest in analogical proportions for text classification. In particular, the proposed analogical classifiers significantly outperform some existing Arabic classifiers, and in many cases they perform similarly as the robust SVM classifier.

KEYWORDS:
Explainable Artificial Intelligence Arabic text classification Analogical proportions Analogical inference Feature extraction and selection Text representation.

1 | INTRODUCTION

Given the huge amount of textual documents easily accessible to the user, extracting useful information from those "unstructured" data is seen as a hard task requiring a complex step of data processing and management. This may justify why there has been more interest in developing efficient tools for automatic text mining in recent years. Text classification techniques are among the relevant fields in text mining (see for example 1 for a recent work on text classification). The basic process of automatic text classification consists to assign the most relevant single or multiple classes to a given set of text documents based on a set of classification features. It has to be noted that text classification is involved in many application domains such as question answering 2, sentiment classification 3 4, spam filtering 5, genre classification 6, dialects identification 7 and sentiment analysis 8, 9, 10, 11, 12.

Most of the existing works on text classification are designed for English and other language's text collections. In the recent few years, Arabic language became the sixth official language stated by United Nations 13 and the fourth most used language on the Internet. Nevertheless, only a few Arabic text classifiers exist due to the lack of availability of a large collection in the Arabic language 14.

Existing Arabic text classifiers in the literature are mainly of two types (see 15 for a recent review on Arabic text classification): i) Arabic text classifiers directly applying classic Machine Learning (ML) algorithms such as k-Nearest Neighbor (k-NN), Decision Trees (DT), Naive Bayes (NB) and Support Vector Machine (SVM), etc. We cite for examples these recent works 16, 17, 18, 19, 20. Results in these works show that SVM is among the most accurate ML algorithms to classify Arabic texts 15. Nevertheless, three main challenges are still noted with these classic ML classifiers (as we will show in the related works study): first of all, most of these classifiers are assessed using small/medium-sized Arabic text collections. Second, most of those corpora are not freely available 15 for the community for reevaluation. Finally, the classification rates reported for these Arabic classifiers can be significantly improved 21. ii) Arabic text classifiers applying Deep Learning (DL) techniques, recently developed. DL approaches are either used to enhance the text representation process 22 or utilized as a text classification technique 23, 24, 14. Globally, DL techniques are efficient in Arabic text categorization using big text collections. However, small corpora as well as text collections having a large number of categories are still one of the main challenges of existing DL-based Arabic text classifiers 14.

Taking into consideration the complexity of the Arabic language from one side and the reduced number of new Arabic text classifiers (not directly using classic ML algorithms), on the other side is our main motivation in this paper to explore new approaches for Arabic text classification. Our main interest in this research work is two-fold: on the first hand, we aim to explore and develop a new text classification technique that performs well when dealing with small or large text collections (which is one of the main shortcomings of previous works). On the second hand, inspired by the recent achievements of analogical proportions in classifying *structured* data 25, 26, 27, our second aim is to investigate the performance of such proportions to classify *unstructured* text data collections.

During the last two decades, significant improvements in terms of complexity and accuracy have been registered in several application domains of analogical proportions 28, 29. They have demonstrated their efficiency to model morphological linguistic analysis 30 in addition to document-query matching in domain-specific Information Retrieval (IR) 31 and text summarization 32. Moreover, analogical proportions have recently shown their efficiency in the classification of *structured* data, applied first to Boolean data 33, 34 and then extended to numerical data 35, 26, 27, 25, in dataset enlargement 36 and preference learning 37. See also 38 for recent works on logical proportions-based classification behind analogy.

In this paper, our intuition is based on the idea that it may exist some kind of analogical link between text documents and their corresponding categories as in the case of classic analogical classification 26, 25. If such a relationship could be defined appropriately in the first stage, and if a sample set of documents whose categories are known is given, the category of a new document can be predicted on the basis of the analogical relationship between this document being classified (in the testing set) and the three other pairs (document, class) in the training set. To the best of our knowledge, these kinds of analogical relationships with other documents and their corresponding classes in the collection have never been applied in the classification of texts for any language including Arabic.

Analogical reasoning is a kind of human thinking that relies on comparisons with past experiences to solve any newly faced problem 39. For example, the children's reasoning when faced with unknown situations is naturally based on the analogy with known cases. Analogical learning systems aim at simulating the human ability to learn by analogy. The analogical reasoning performed by these systems is a part of research efforts in Explainable Artificial Intelligence (XAI). Unlike other existing learning approaches such as SVM or DL-based techniques, which may be seen in some cases as "black boxes", analogical learners have shown their ability to explain and justify their results 40, 41 since they are considered as a kind of natural human reasoning. Analogical learners are not only able to compete with other previous classifiers in terms of prediction rates 25, but also they prove their capability to justify/explain why such prediction is made for any item. In particular, exploring the change/non-change in the set of features that is responsible for the change/non-change in the class label is considered as an effort to make analogical learners explainable 40, 41. For example, if we are able to extract the subset of features whose change leads to a change in the prediction, it is easy to conclude that the subset of remaining features is irrelevant for decision-making.

The concept of analogical proportions is very close to analogical reasoning since it aims to construct a matching between known pairs of cases in order to approximate new unknown cases. Analogical proportions, known as statements of the form "x is to y as z is to t", are formally represented by $x : y :: z : t$. This relationship states that "x differs from y as z differs from t", as well as "y differs from x as t differs from z" 28 or similarly the pair *(x, y)* is analogous to the pair *(z, t)* 42. In the context of classification, the analogical inference is founded on the following hypothesis: if four objects x, y, z, t form a valid analogical proportion between their feature values, this will still continue to hold between their class labels. And if the class labels of objects x, y and z are known while the label of t is unknown, this states the basic to guess the class of t based on the class labels of the triple (x, y, z).

The paper is structured as follows. Section 2 presents the main objectives of this work. In Section 3, we review and discuss the most recent existing Arabic text classifiers. A brief background on analogical proportions as well as in the Boolean and nominal settings is presented in Section 4. Section 5 states first the fundamental principle of the proposed analogical proportions-based approach for Arabic text classification. Then, the basic procedure for the two designed classifiers is detailed. In Section 6, we provide the experimental results of analogical Arabic text classifiers and we compare them to classic ML algorithm-based approaches such as k-NN and SVM as well as to DL-based techniques for classification. Finally, Section 7 summarizes our main contributions and suggests some perspectives for future works.

2 | RESEARCH OBJECTIVES

In this paper, our main objective is to investigate if analogical proportions are $still$ efficient to classify *unstructured* text data in the same way as for the general context of *structured* nominal or numerical data 26.

First, we establish the basis for analogical text classification and we design an analogical model to represent the relationship between text documents and their corresponding categories. This model is based on the assumption that similar documents, having common relevant features, may be tagged with the same category while very dissimilar documents should have different categories. From an analogical point of view, this can be stated as a matter of comparisons between input text documents and the corresponding set of relevant keywords useful to identify their suitable categories.

Second, based on this formalized model, we develop two new Analogical Arabic Text Classifiers (we denote AATC1 and AATC2). In AATC1, the set of classification features (keywords) is extracted from the document being classified, while in AATC2 the set of keywords is extracted from the whole set of documents in each category.

Finally, we aim to assess the efficiency of the two proposed classifiers with regard to other classification techniques. We first compare them to classic ML classifiers (such as SVM, NB, DT and k-NN) that we tested on two Arabic collections: the small “BBC-Arabic" and the large corpus ANT v2.1. Then we also compare them to some existing Arabic text classifiers either applying classic or modern algorithms like DL and using different data collections: “CNN-Arabic", “AlKhaleej-2004" and ANT corpus v1.1 43.

The main contributions of this work are summarized in the following:

- Despite the achievement of analogical proportions in different application domains, as far as we know analogical proportions have never been applied to text classification for any language. We design an analogical model enabling to predict the category for a new document being classified.
- We develop two new analogical Arabic text classifiers applying the above principle.
- The performance of the proposed classifiers is assessed using a variety of small and large benchmarks.
- We also discuss and compare the efficiency of AATC with regard to some recent classic ML classifiers as well as to some existing DL-based Arabic text classifiers.

3 | RELATED WORK

In this section, we review the most recent Arabic text classifiers that appeared in the last few years. Most of them are using classic ML algorithms, while some others are based on DL techniques. We also review some other recent classifiers that are using different algorithms and techniques. A complete review of the state-of-the-art Arabic text classification can be found in 44, 45, 46, 47.

3.1 | Machine Learning algorithms for Arabic text classification

Some related works focus on feature selection or text representation as a preprocessing step, while others investigate mainly text classification.

Feature selection consists to extract the most relevant features to use in model construction. Hassanein and Nour (2019) 48 have recently suggested the use of a clustering technique to develop a feature selection method, useful for Arabic text classification. The model computes first a degree of correlation between every feature and an input class, aiming to select the set of highly correlated features using a cut-off threshold. In order to reduce the number of the best-selected features, it also computes intra-correlation scores between those features. Then, a set of logical operations (AND, OR) has been applied to perform the process of features' values fusion by considering their semantics, nature and structure. Finally, a features clustering process is performed using the Cosine rules for clusters' fusion.

Hamdan (2019) 16 has used some feature selection techniques to test ML algorithms for text classification such as NB, SVM, Artificial Neural Networks (ANN), k-NN, C4.5 Decision Tree and Rocchio classifier. The author has used the Information Gain (IG), Gain Ratio (GR) and Chi-square methods for feature selection and a weighted vector for text representation.

Text representation and classification using a bi-gram alphabet approach has been proposed and tested by Elghannam (2021) 49 to overcome the need for language-dependent tools, as well as the high dimensions of vector space. The process of building feature terms, useful for text classification, applies a bi-gram alphabet technique. Thus, the dimensions of vector space for large corpus have been significantly reduced by using constant feature terms rather than document vocabularies, and without any need for a Natural Language Processing (NLP) tool.

Moreover, Elnahas et al. (2020) 18 have studied the impact of feature selection techniques on the efficiency of Arabic text classifiers. They have proposed an approach using semantic fusion and multiple-words (SF-MW) and they have compared it to some existing feature selection approaches. The authors have shown some improvements in the previous feature selection methods, especially by using: (i) a combination of Chi-Square and TF-IDF feature selection; (ii) Gini Index and TF-IDF feature selection; (iii) Information Gain and TF-IDF feature selection.

Still in the context of feature selection, Basabain et al. (2023) 50 demonstrate the impact of using hidden label information to extract enhanced feature representation of an input text. The proposed approach prove its efficiency to improve classification performance especially when input texts are scarce/short such as in tweets for example.

Another recent Chi-square feature selection technique (denoted ImpCHI) has been used by Bahassine et al. (2020) 19 to decrease the large number of features and then improve the performance of Arabic text categorization. The proposed ImpCHI method is also compared to three classic feature selection assessments: information gain (IG), Chi-square and mutual information (MI). Besides, two ML classifiers: DT and SVM are also tested to evaluate the efficiency of the ImpCHI technique using the CNN Arabic collection (5070 documents and six categories). Results highlight that SVM classifier combined with ImpCHI technique overcomes other combinations in terms of recall, precision and F_1.

On the other hand, to treat text representation, El-Alami and El Alaoui (2016) 22 have developed a deep RBM (Restricted Boltzmann Machines) auto-encoder, as an efficient deep neural network approach 51 used to enhance the efficiency of the proposed Arabic text classifier. The authors have used word-count vectors as input to the proposed deep stacked auto-encoder considered as a deep architecture to generate a high-level abstraction presentation. Besides, short-reproduced codes are applied to consider better implicit semantics and to reduce the dimension of representation space. The experimental scenario involves three ML classifiers: Naïve Bayes (NB), Support Vector Machine (SVM) and Decision Tree (DT). Moreover, Khoja and Light stemmers are used to test these classifiers on CNN Arabic news collection. Light stemmer and 128-code vectors have achieved the best text classification results.

In the general context of Arabic text classification, a Linear Discriminant Analysis (LDA) technique is used by AbuZeina and Al-Anzi (2018) 52 for multiclass Arabic text classification. The authors have investigated two multiclass LDA-based techniques: (i) linear classification functions (a linear classification rule for each class), and (ii) generalized eigenvectors of the ratio (inverse within-class and between-class scatters). An Arabic corpus of 2000 documents and five classes, issued from Alqabas newspaper in Kuwait, has been used to test the proposed LDA-based approach. Results confirm that the accuracy of the linear function slightly outperforms the Eigenvalue accuracy. Besides, the efficiency of the semantic loss LDA-based technique is almost similar to the semantic-rich SVD (singular value decomposition).

To classify the documents of modern Arabic poetry, Abdullah et al. (2019) 17 have proposed an approach using ML algorithms such as NB, SVM, and Linear Support Vector for classification (LSVC). The authors have applied a words' occurrence process for feature selection and a Boolean vector model for text representation.

Kanan et al. (2020) 53 have proposed and experimented an approach combining some existing ML classifiers with an enhanced P-Stemmer, which contributes to enhance the efficiency of Arabic text classification. The P-Stemmer is an extended version of the existing Larkey's light stemmer. The improved P-Stemmer is combined with some traditional classifiers such as NB, SVM, Random Forest, K-Star and k-NN.

Finally, Chowdhury et al. (2020) 20 have explored the effect of diversifying the example set on the efficiency of Arabic text classifiers. The authors have first developed two Arabic text classification benchmarks (ASND and AITD). In this work, the authors have evaluated four different classifiers (SVM, mBERT, AraBERT and QARiB) in terms of macro F_1 and Precision.

3.2 | Deep Learning Techniques for Arabic text classification

Recently, DL approaches have been applied to overcome some Arabic language challenges, especially in the fields of sentiment analysis 54, 11, 55 and emotion classification at Task-1 of SemEval-2018 56, 57, 58. Interested readers can find a complete review of Arabic sentiment analysis research works in the recent survey/review in 59, 60, 61. Besides, DL techniques for Arabic NLP are surveyed in 62. In the following, we only give an overview of the most recent Arabic text classifiers based on DL techniques.

A deep neural network LSTM-RNN (Long Short-Term Memory Recurrent Neural Network) approach has been proposed by Alwehaibi and Roy (2018) 23 to assess the efficiency of several pre-trained Word Embedding (WE) vectors in Arabic text categorization. The proposed DL model has been tested for three classes (negative, positive and neutral) using an Arabic collection generated from Twitter (AraSenTi). Experimental results highlight important enhancement, in terms of accuracy, in Arabic text categorization.

Galal et al. (2019) 63 have used the known efficient Convolution Neural Network (CNN) technique to classify Arabic texts. During the preprocessing step, the authors have proposed their $Gstem$ method based on extra Arabic letters and word embedding distances to group similar Arabic terms with the same root. The goal is to reduce the number of distinct terms before the classification step. Besides, the Word2Vec technique has been used to learn word vectors and convert the input text to a vector space. The authors have applied the $Gstem$ method combined with CNN to perform Arabic text classification.

The SATCDM Arabic text classifier of Alhawarat and Aseeri (2020) 64 is based on a DL model using CNN and word embedding. This classifier uses the efficient multi-kernel CNN architecture and applies the skip-gram word embedding model enriched with sub-word information. The terms of a text are represented as a real-valued vector in a vector space, while feature selection applies a fast-text model.

Elnagar et al. (2020) 14 have developed nine DL models to perform single and multi-label Arabic text classification. For this purpose, the authors have built two large collections, namely NADiA and SANAD useful respectively for multi and single-label categorization tasks without any need for the preprocessing steps. Besides, the authors have also studied the influence of using Word2Vec embedding models to enhance Arabic text classification rates.

The effect of stemming and word embedding on DL-based Arabic text classification has been investigated by Almuzaini and Azmi (2020) 24. The Word2Vec technique is used for word representation of Arabic text documents. The latter are classified using seven DL methods. The ANT corpus v1.1 and SPA collection are used to evaluate these DL techniques in terms of F_1. Results highlight that stem-based methods outperform root-based approaches.

Moreover, Chaturvedi et al. 65 have applied a Deep Learning approach for domain adaptation of sentiment analysis across different languages. They developed a Genetic Programming model that enables to classify sentences of variable lengths. Results show that the proposed classifier outperforms baseline approaches in terms of prediction accuracy.

More recently, Mohamed et al. (2023) 66 have developed a DL approach applying Gated Recurrent Units and features extraction to conduct Aspect-Based Sentiment Analysis. The proposed approach is trained on a hotel's Arabic reviews to address aspect extraction and polarity classification in a sentiment analysis task.

3.3 | Other diverse related classifiers

In this subsection, we first refer to some diverse Arabic text classifiers not using classic or modern techniques cited above. For example, Al-Radaideh and Al-Abrat (2019) 67 have suggested an Arabic text classifier using term weighting and multiple reducts. The authors have used the reduct concept of rough set theory for feature selection, to classify Arabic texts based on some classification rules and using a reduced number of terms. In this approach, the authors have used a decision table for document representation, in which the rows represent the documents and the columns represent terms. The authors have implemented a multiple minimal reduct extraction algorithm to enhance the Quick reduct algorithm. The proposed rough set classifier is based on classification rules generated via the multiple reducts.

Moreover, Hawalah (2019) 68 has proposed an improved Arabic topic-discovery architecture (EATA) based on semantic ontology to classify Arabic texts. The goal of the proposed semantic clustering mechanism (SCM) is to enhance the process of Arabic text classification. The feature selection is based on TF-IDF scores, while a vector space model, based on TF-IDF, has been used for text representation.

More recently, Al-Taani and Al-Sayadi (2020) 69 have developed an Arabic text classifier that they denote SVD-FCM. This latter combines a singular value decomposition (SVD) technique for feature extraction with Fuzzy c-means algorithms (FCM) for text classification. While a document-term matrix model is utilized for text representation.

Finally, since Arabic text classification is a particular case of the general context of classification, it may benefit from the achievements of other classification methods in various fields. In the following, we restrict our interest to the most recent classification approaches.

In particular, various classification techniques have been recently developed in the context of the medical field. For example, Vulli et al. (2022) 70 have developed a DL classification task based on DenseNet with 169 layers that has been used for breast cancer metastasis prediction. Moreover, a DL-enhanced Elman neural network (IENN) technique has been suggested by Kumar et al. (2022) 71 to classify sensitive and non-sensitive data in an Internet of Medical Things (IoMT) architecture. Mandal et al. (2021) 72 have recently performed disease classification using a tri-stage wrapper-filter feature selection method. The authors have exploited four filter methods (Xvariance, Mutual Information, Chi-Square and Relief) for feature selection in the first stage. Highly correlated features have been detected using the Pearson correlation technique in the second stage. The best subset of optimal features has been obtained by applying an XGBoost classification method in the first and second stages. While a whale optimization algorithm has been used in the last stage to reduce the feature dimension and to reach better accuracy. A dynamic ensemble classifier-based model has been used by Juraev et al. (2022) 73 to predict the mortality of neonate patients. Finally, El-Sappagh et al. (2022) 74 have applied a heterogeneous ensemble classifier to automatically detect the progression of Alzheimer's diseases.

In the context of IoT, a multi-task classification technique based on the DL model (LSTM) has been proposed by Ali et al. (2022) 75 for IoT malware detection. The authors have used a feature selection technique at the modality and feature levels to improve the model performance through the best-selected features and modalities. Besides, Bhattacharya et al. (2022) 76 have proposed a DL-based classification technique (CNN and LSTM) using an ensemble measurement for smartphone sensor-based Human Activity Recognition (HAR). The proposed model for biomedical measurement has achieved good accuracy when assessed on different benchmark datasets.

4 | BACKGROUND ON ANALOGICAL PROPORTIONS

A numerical proportion expresses a relationship between four values a, b, c and d where (a, b) and (c, d) form the two pairs of the relationship. At the origin, a standard numerical proportion is defined as an equality statement between the two ratios $\frac{a}{b} = \frac{c}{d}$. This statement expresses also that the differences between the two items of each pair are equal (i.e. $a - b = c - d$).

Drawing a parallel from numerical proportions, an analogical proportion extends the same principle from numbers to symbolic objects. Analogical proportions represent a relationship between 4 objects $A, B, C, D \in X$, where X is the domain of A, B, C and D. This relationship is usually denoted $A : B :: C : D$ and reads "A is to B as C is to D". Both links "is to" and "as" are highly related to the domain X 77.

4.1 | Analogical proportions and Boolean attributes

Let us assume that we have four objects A, B, C and D that can be represented by their Boolean values $a, b, c, d \in \{0, 1\}$. To state whether A, B, C and D are in analogy or not, we assess if the analogical proportion linking their corresponding values, i.e: $a : b :: c : d$ holds true or not which is commonly represented as $a : b :: c : d$ if it is valid.

In the Boolean setting, analogical proportions satisfy the following axioms (see 78 for further details):

- **Symmetry:** if $a : b :: c : d$ then $c : d :: a : b$
- **Central permutation:** If $a : b :: c : d$ then $a : c :: b : d$
- Based on the central permutation property, we can derive two following implications: if $a : b :: a : x$ then $x = b$ and if $a : a :: b : x$ then $x = b$
- **Transitivity:** if $a : b :: c : d \wedge c : d :: e : f$ then $a : b :: e : f$

The examples $(0, 1, 0, 1)$ and $(1, 0, 1, 0)$ are in analogical proportion, proven by reflexivity axiom, and so are $(0, 0, 0, 0)$ and $(1, 1, 1, 1)$ since we may have $a = b$, the two remaining possible assignments are $(0, 0, 1, 1)$ and $(1, 1, 0, 0)$ obtained from central permutation. In total we have 6 possible arrangements for the 4 values of A, B, C and D that build a valid analogical proportion out of 16 (2^4) as shown in Table 1.

In the Boolean setting, we are interested to know how two objects are similar and how they are different using Boolean indicators. Given a pair (a, b) of Boolean variables, we have four indicators that correspond to this pair 79:

- *Positive similarity/ negative similarity:* $a \wedge b$ and $\neg a \wedge \neg b$. Logical value $a \wedge b$ (resp. $\neg a \wedge \neg b$) is true iff only both a and b are true (resp. false).
- *Dissimilarity:* $a \wedge \neg b$ and $\neg a \wedge b$, which are true iff only one of the values a or b is true and the other is false.

In 78, the author defines an analogical proportion as a conjunction of two equivalences that exploits only dissimilarities indicators defined as:

$$a : b :: c : d = (a \wedge \neg b \equiv c \wedge \neg d) \wedge (\neg a \wedge b \equiv \neg c \wedge d) \tag{1}$$

This relationship can be read: *"what is true for a and not for b is exactly what is true for c and not for d, and vice versa"*. This meaning perfectly fits with the statement "a differs from b as c differs from d and vice versa".

4.2 | Extension to nominal setting

The above definition for analogical proportion has been extended to nominal setting in 26. Let t, v be two distinct values, $t, v \in D$, and D is a given finite set. An analogical proportion always holds only for the three following patterns: (t, t, t, t), (t, t, v, v) and (t, v, t, v). All the remaining possible arrangements contradict the logic of analogical proportions. Especially the patterns (t, t, t, v), (t, t, v, t), (t, v, t, t), (v, t, t, t), and (t, v, v, t) are invalid because we cannot assume that "t is to t as t is to v" for $t \neq v$ as an example.

4.3 | Analogical equation for Boolean vectors

Analogical proportions have been also extended to deal with Boolean vectors in $\{0,1\}^n$. Let $\vartheta \in \{0,1\}^n$ be a given set of vectors and $\vec{x} \in \vartheta$ is a vector of n Boolean attributes represented as: $\vec{x} = (x_1, \cdots, x_n)$. Now if we consider four vectors $\vec{a}, \vec{b}, \vec{c}$ and $\vec{d} \in \vartheta$, the analogical proportion between these four vectors $\vec{a} : \vec{b} :: \vec{c} : \vec{d}$ is defined as:

$$\vec{a} : \vec{b} :: \vec{c} : \vec{d} \text{ iff } \forall i \in \{0,1\},\ a_i : b_i :: c_i : d_i \tag{2}$$

This relationship expresses that four Boolean vectors are in a valid analogical proportion iff this proportion holds true componentwise between all their individual attributes.

4.4 | Analogical inference

In 33 and 26, the authors suggest an analogical inference process for binary classification problems applied to Boolean datasets. Based on the continuity principle, they assume that "if the analogical equation holds componentwise for all features of 4 Boolean instances, this analogical equation should still hold for their classes". If we consider that we have 4 Boolean examples $\vec{a}, \vec{b}, \vec{c}$ and $\vec{d}$, the first three examples are in the sample set with known classes $class(\vec{a}), class(\vec{b}), class(\vec{c})$ and the last one is with an unknown class (to be predicted). The analogical inference is defined as:

$$\frac{\forall i \in [1,n], a_i : b_i :: c_i : d_i}{class(\vec{a}) : class(\vec{b}) :: class(\vec{c}) : class(\vec{d})} \tag{3}$$

In this inference process to classify the new instance $\vec{d}$, we should first solve the equation $class(\vec{a}) : class(\vec{b}) :: class(\vec{c}) : x$ and then we assign to $\vec{d}$ its solution. To solve such analogical equation on classes, an extrapolation mechanism is useful in which a suitable fourth item $x = class(\vec{d})$ is computed in order to complete the proportion $class(\vec{a}) : class(\vec{b}) :: class(\vec{c}) : x = 1$ (this can be seen as another view of the well-known "rule of three"; i.e: $\frac{A}{B} = \frac{C}{x}$ in the numerical case). In Boolean classification, it is clear that the class equation is not always solvable since the triple $class(\vec{a}), class(\vec{b}), class(\vec{c})$ may take $2^3 = 8$ values, while $class(\vec{a}) : class(\vec{b}) :: class(\vec{c}) : x$ is valid only for 6 distinct 4-tuples. For example, the equations $1 : 0 :: 0 : x = 1$ and $0 : 1 :: 1 : x = 1$ have no solution.

In the most general case of multiple-classification (which represents the context of Arabic text classification) for which more than two distinct classes may exist, $class(\vec{a}) : class(\vec{b}) :: class(\vec{c}) : x$ is valid only for 3 distinct patterns of the triple $(class(\vec{a}), class(\vec{b}), class(\vec{c}))$ as shown in Table 2 (where t and v are two distinct values).

Note that other patterns like ($t, v, v, x = 1$) have no solution.

To illustrate the analogical inference for text classification, let us consider 4 text documents, the three first ones are in the training set (for which the class label is known) and the last one is to be classified. We also assume that we have four terms: t_1=*"Goal"*, t_2=*"Inflation"*, t_3=*"Competition"* and t_4=*"Winner"* extracted from text documents. For simplicity, these terms are considered as representing the main features used for classification.

In Table 3, we show the terms-documents matching as well as the class label of the first three classified documents. As can be seen in this table, the analogical proportion $Doc_1(t_i) : Doc_2(t_i) :: Doc_3(t_i) : Doc_x(t_i)$ is $valid$ for each term t_i s.t: $i = 1, ..., 4$ (see vertical columns). Observing a valid analogical proportion on each feature helps to infer the class of Doc_x in a second stage. In light of the class label of the three first documents, the analogical inference on the class $Economy : Sport :: Economy : x$ yields to predict $x = Sport$ and thus Doc_x is classified as a Sport document.

4.5 | Analogical classification: basic principle

Analogical classification relies on the idea of selecting triples $(\vec{a}, \vec{b}, \vec{c})$ of items in the training set that build a valid analogical proportion with the item $\vec{d}$ to be classified. In 26, the authors suggest a brute force method that looks for all triples $(\vec{a}, \vec{b}, \vec{c})$ in the sample set for which the class equation $class(\vec{a}) : class(\vec{b}) :: class(\vec{c}) : x$ is solvable (has solution *l*). The next step consists to compute a score $P(\vec{a}, \vec{b}, \vec{c}, \vec{d})$ for each of these triples as the average of the truth values obtained in a componentwise manner for each individual attribute. Finally, assign to $\vec{d}$ the class label having the highest cumulative score P.

The authors 26 have also developed an optimized algorithm of this brute force approach in which they restrict the search of triples $(\vec{a}, \vec{b}, \vec{c})$ to the ones such that $\vec{c}$ is one of the k-nearest neighbours of $\vec{d}$.

This procedure can be summarized as follows:

1. Select all triples $(\vec{a}, \vec{b}, \vec{c})$ in the training set s.t: $\vec{c} \in N_k(\vec{d})$.
2. Solve the class equation $class(\vec{a}) : class(\vec{b}) :: class(\vec{c}) : x$.

3. If the class equation on classes is solvable (with a solution l), increment the score ($score(l)$) with $P(\vec{a}, \vec{b}, \vec{c}, \vec{d})$ as $score(l) += P(\vec{a}, \vec{b}, \vec{c}, \vec{d})$.

4. Assign to $\vec{d}$ the class label with the best score as $class(\vec{d}) = argmax_l(credit(l)))$

This algorithm, applied first to Boolean or nominal attributes, has been also extended to numerical settings using appropriate extension of the analogical proportions on $[0, 1]$ domain. 80 and 81 provide further explanations of the implementation of the Analogical Proportion-based classifier. The first paper treats the case of Boolean and nominal data while the second focuses on numerical data.

5 | ANALOGICAL PROPORTIONS FOR ARABIC TEXT CLASSIFICATION

As in the general context of classification, text categorization consists to assign a category, among a set of predefined categories, to a given text document. For example, classifying a media announcement as a political, cultural or sport document. Contrary to the context of classification of structured data, where each example to be classified is assumed to be represented by a set of fixed features defined for all considered examples, unstructured text documents do not have necessarily any predefined structure. This raises the first challenge for text classification since it requires a preprocessing phase aiming to expose a text document as a set of predefined terms which represent the basic features of the document. Then the problem returns to apply any classical classification technique to predict the category for this document.

In this paper, we suggest using analogical proportions for text classification. We first set the fundamental concepts of analogical text classification and show how analogical proportions may serve to design new text classifiers then we present two different approaches based on this principle.

5.1 | Fundamentals of Arabic text classification

Let's first formalize the text classification task in a more formal way. For this purpose, let us consider a set C of K predefined categories of documents $C = \{c_1, ..., c_K\}$ and a collection of documents $D = \{(d_1, c_1), (d_2, c_2), ..., (d_n, c_n)\}$ where $c_i \in C$ represents the category of each document d_i.

In the following, we suggest representing the "document-category" relationship more formally. For this purpose, we assume that documents are indexed by a set of relevant keywords extracted in a preprocessing step. We exploit a Term Frequency-based approach 48 for feature selection. Then, a document-term matrix model is used to perform text representation 69: in such a model, each text is represented as a matrix that contains the terms (keywords) and their corresponding frequencies. Let $W = \{w_1, w_2, ..., w_m\}$ be the set of relevant keywords for a given document. To represent the importance of a particular keyword w_j for a document d_i, let us consider d_i^j a Boolean predicate indicating if the given keyword w_j (the index j in each predicate refers to the keyword w_j) is relevant/irrelevant for a document as in 32 i.e: $d_i^j = 1$ (resp. $d_i^j = 0$) means that the keyword w_j is relevant (resp. irrelevant) to represent the document d_i. This meaning can easily be extended to a multi-valued setting by allowing more graded values for d_i^j, i.e: d_i^j may be considered as a value in $[0, 1]$ instead of 0 or 1. For example, $d_i^j = 0.8$ indicates a high relevance of the term w_j for document d_i while $d_i^j = 0.1$ assumes that w_j is almost irrelevant for d_i. In this paper, we are limited to deal with the Boolean representation of documents.

Now consider a triple of documents $t = (d_1, d_2, d_3) \in D^3$ whose corresponding categories are known and a new document $d_x \notin D$ for which we aim to predict the category c_x. The analogical proportion for a given keyword w_j, is defined as:

$$d_1^j : d_2^j :: d_3^j : d_x^j \tag{4}$$

states that, according to the keyword w_j, document d_1 differs from d_2 in the same way as document d_3 differs from d_x. For example, the proportion $0 : 1 :: 0 : 1$ ($d_1^j = 0$, $d_2^j = 1$, $d_3^j = 0$ and $d_x^j = 1$) informs that w_j, is being relevant for d_2 but not for d_1, in the same way as being relevant for d_x but not for d_3.

As said before, the analogical inference may be useful to represent the "document-category" relationship which may help to infer the category for a new document to be classified. This can be stated as follows: if document d_1 differs from document d_2 as document d_3 differs from document d_x, this similarity/dissimilarity between documents will still apply to categories, i.e: c_1 differs from c_2 as c_3 differs from c_x. Considering all keywords, the above analogical inference given in Section 4 (see Equation (3)) can be adjusted as follows for text classification:

$$\frac{\forall j \in [1, m], d_1^j : d_2^j :: d_3^j : d_x^j}{c_1 : c_2 :: c_3 : c_x} \tag{5}$$

Then the classification task returns to solve the analogical equation on classes as introduced in Table 2 and finally assign to d_x the solution of this equation.

5.2 | Analogical proportions-based Arabic text classification approach

In this paper, based on the above analogical inference we develop two different Analogical Arabic Text Classifiers that we denote AATC1 and AATC2. The main idea of these classifiers is to find triples of documents $(d_1, d_2, d_3) \in D^3$ that build a valid analogical proportion with the document d_x on all or on a large number of keywords w_j and whose class equation is solvable. The two classifiers differ basically in two regards: first, in AATC1, the keywords used for classification are extracted from the document d_x for which we aim to predict the class c_x while in AATC2 these keywords are extracted from the whole class set.

Second, following this process for AATC2, some keywords $w_j \in W$ are not necessarily relevant for d_x ($d_x^j = 0$); however, they are still useful for classification as they are relevant for the whole class. To explain better this second difference, Table 4 summarizes the different situations where proportion P4 is exploited differently for each classifier.

As can be seen in Table 4, only valid proportions P1, P2 and P3 are useful for prediction in AATC1 since all useful keywords are extracted from d_x and therefore d_x^j must be equal to 1. while in AATC2 P1, P2, P3 as well as P4 are relevant ($d_x^j = 0$ or $d_x^j = 1$).

Let us first focus on the classification process of AATC1: for each triple of documents $t = (d_1, d_2, d_3) \in D^3$, we compute an analogical score p_t as the average of analogical proportions valuations obtained componentwise for each keyword w_j and we consider only those that are "relevant" for d_x (with patterns P1, P2 or P3) and making a valid analogical proportion with d_x. We also weight those individual analogical proportions with the frequency of this keyword in document d_x. Considering keyword frequency in analogical scoring may contribute to raise the effect of frequent keywords that are making an analogical proportion with the considered triple t. In AATC1, the analogical score p_t (see Equation (6)) of a given triple $t = (d_1, d_2, d_3)$ is computed as:

$$p_t = \frac{\sum_{j=1}^{m}((d_1^j : d_2^j :: d_3^j : d_x^j) * Frequency(w_j))}{m} \tag{6}$$

Now for AATC2, since keywords are extracted from each category c_k, even proportion P4 (for which keywords j are not relevant for d_x but relevant for c_k) serves for classification. In this context, we estimate two complementary credits for each triple t: p_t^+ (see Equation (7)) accumulates analogical scores for keywords that are relevant for both document d_x and category c_k (using proportions P1, P2 and P3) and p_t^- (see Equation (8)) accumulates analogical scores for keywords that are relevant for category c_k but not for d_x (using proportion P4). By this last negative credit, we aim to penalize document categories whose most keywords are not relevant for d_x:

$$p_t^+ = \frac{\sum_{j=1}^{m}(d_1^j : d_2^j :: d_3^j : d_x^j \wedge d_x^j)}{m} \tag{7}$$

$$p_t^- = \frac{\sum_{j=1}^{m}(d_1^j : d_2^j :: d_3^j : d_x^j \wedge \neg d_x^j)}{m} \tag{8}$$

Finally, we aggregate the two credits to estimate the final analogical score p_t (see Equation (9)) for this triple t as:

$$p_t = p_t^+ - p_t^- \tag{9}$$

Except the two previously introduced distinctions, AATC1 and AATC2 can be described by the following general process:

1. Extract the set of keywords to represent a document or a whole class using the Term Frequency-based approach 48. These keywords are then used to calculate analogical scores.
2. Partition the training set of documents into sets D_{c_k} of documents belonging to the same category c_k.
3. For each category c_k, look at each triple $t = (d_1, d_2, d_3)$ in the training set D_{c_k}.
4. Compute an analogical score p_t for each triple $t = (d_1, d_2, d_3)$ in $D_{c_k}^3$ using Equation (6) or (9).
5. Accumulate the $score(c_k)$ with p_t as $score(c_k) += p_t$.
6. Select the category having the highest score as $c_x = argmax_{c_k}(score(c_k))$.

Note that a triple $t = (d_1, d_2, d_3)$ may build valid analogical proportion with d_x (Equations 6 or 9) $only$ on a small subset of terms w_j which reduces its classification power. Thus the previous process can be optimized by only considering the N best triples in each category, those having the highest scores p_t (N is a tuned threshold). For this purpose, triples are sorted decreasingly according to their scores and then only the N top triples having the best scores are aggregated by averaging their scores for the final decision.

Algorithm 1 Analogical Arabic Text Classification (AATC1)

Input: $D = \{(d_1, c_1), ..., (d_n, c_n)\}$: a training set of classified documents,

$C = \{c_1, ..., c_K\}$: a set of documents' categories,

a test document $d_x \notin D$ whose c_x is unknown,

N: a threshold.

Output: Identify the correct category c_x for d_x.

Generate the set of keywords from d_x: $W_x = \{w_1, ..., w_m\}$

Partition D into sets $D_{c_k} = \{d_i,$ s.t: $category(d_i) = c_k\}$

for each $c_k \in C$ **do** score(c_k) = 0 **end for**

for each class $c_k \in C$ **do**

  $Best_{p_t} = \emptyset$

  for each triple t = $(d_1, d_2, d_3) \in D^3_{c_k}$ **do**

    Compute p_t

    $Best_{p_t}$.add(p_t)

  end for

  $Best_{p_t}$.sort()

  for N max $p_t \in Best_{p_t}$ **do**

    score(c_k) = score(c_k) + p_t

  end for

  score(c_k) = $\frac{score(c_k)}{N}$

end for

$c_x = argmax_{c_k}(score(c_k))$

5.3 | Implementation

The following algorithms 1 and 2 implement respectively the two previously described classifiers AATC1 and AATC2.

5.4 | Illustrative example

Let us consider a triple of documents $t = (d_1, d_2, d_3)$ in the training set and a new document d_x to be classified. In this example, we aim to show the process of score computation for each triple in the training set using the two proposed classifiers (accumulating these individual scores for each triple and for each category will help to select the best category for d_x as described above). In this process, we first consider 10 keywords ($w_1, w_2, ..., w_{10}$) extracted from a given category c_k of documents.

Table 5 presents the truth values of each predicate used to compute p_t^+ and p_t^- for this triple $t = (d_1, d_2, d_3)$ in AATC2.

Applying Equation 4 in each row of this table enables computing an individual analogical score for each term w_j as shown in column 6. Then averaging these scores over all rows leads to compute p_t^+ = (5/10) = 0.5, p_t^- = (2/10) = 0.2 and the final score for this tripe p_t = 0.3. This relatively low score shows that the triple t is not so relevant for classification. This may seem intuitive since 30% (3 out of 10) of the keywords from this category are not building a valid analogical proportion with d_x (see last three rows for which $d_1^j : d_2^j :: d_3^j : d_x^j = 0$ in Table 5) and for two more keywords there is no agreement between the category and the document d_x (especially w_2 and w_7), even though there is a total agreement on 70% of the remaining keywords (those whose $d_x^j = 1$).

Now if we consider AATC1, as said before, only keywords extracted from d_x are considered in this calculation. For this reason, w_2, w_7 and w_{10} are discarded from Table 6 and replaced by new terms w_{11}, w_{12} and w_{13} whose $d_x^j = 1$. For simplicity, in this example, we assume that all keywords have a frequency equal to 1 in Equation 6. In the same way as in AATC2, averaging individual analogical scores over all rows leads to compute $p_t = (6/10) = 0.6$ which is relatively higher than the above score for the same triple in AATC2. We may expect that looking only at

Algorithm 2 Analogical Arabic Text Classification (AATC2)

Input: $D = \{(d_1, c_1), ..., (d_n, c_n)\}$: a training set of classified documents,

$C = \{c_1, ..., c_K\}$: a set of documents' categories,

a test document $d_x \notin D$ whose c_x is unknown,

N: a threshold.

Output: Identify the correct category c_x for d_x.

Generate the set of keywords for each class c_k: $W_{c_k} = \{w_1, ..., w_m\}$

Partition D into sets $D_{c_k} = \{d_i,$ s.t: $category(d_i) = c_k\}$

for each $c_k \in C$ **do** score(c_k) = 0 **end for**

for each class $c_k \in C$ **do**

  $Best_{p_t} = \emptyset$

  for each triple t = $(d_1, d_2, d_3) \in D^3_{c_k}$ **do**

    Compute p_t^+, p_t^-

    $p_t = p_t^+ - p_t^-$

    $Best_{p_t}$.add(p_t)

  end for

  $Best_{p_t}$.sort()

  for N max $p_t \in Best_{p_t}$ **do**

    score(c_k) = score(c_k) + p_t

  end for

  score(c_k) = $\frac{score(c_k)}{N}$

end for

$c_x = argmax_{c_k}(score(c_k))$

keywords extracted from the document to be classified in the analogical inference may help to converge faster towards the best triples useful for classification in each category.

5.5 | Sample of document classification

To show the basic process of the proposed classifiers, we consider a concrete sample of a document extracted from the CNN source of "ANT corpus v2.1" presented in Figure 1. This dataset will be presented with more details in the next Section. To classify this document, we start by removing the stop words from the input text and stem it using the light stemmer 82. Then, we calculate the classes' scores as explained before and finally select the class with the highest score. As said before, we only select the N top triples (those having the highest values of p_t) for scoring, $N = 30$ in this example.

Table 7 shows the obtained class scores for this document using respectively Algorithms AATC1 and AATC2.

Comparing scores for the two algorithms in Table 7, we can see that although both classifiers succeed to guess the correct category "Sport" for this document, AATC2 seems to have little higher scores for all classes. However, AATC1 seems more accurate to distinguish between other categories (AATC2 confuses between categories "Economy" and "InternationalNews"). This is probably due to the use of word frequency in AATC1 which raises the effect of the most frequent words and lowers the effect of those that are less frequent in this document.

<DOC>
<DOCNO>CN-sport-380-20180826</DOCNO>
<URL>
http://arabic.cnn.com/sport/article/2018/08/26/mohammed-salah-egyptian-football-association
</URL>
<SRC>CNN</SRC>
<CAT>sport</CAT>
<TITLE>
محمد صلاح ينتقد الاتحاد المصري لكرة القدم: لا أدري لماذا كل هذا؟
</TITLE>
<TIME>Sun, 26 Aug 2018 19:37:32 +0000</TIME>
<AUTHOR></AUTHOR>
<ABSTRACT>
انتقد اللاعب المصري محمد صلاح، نجم فريق ليفربول الإنجليزي، الأحد، الاتحاد المصري لكرة القدم، متهما الاتحاد بأنه لا يسعى إلى حل مشاكل لاعبيه.
</ABSTRACT>
<TEXT>
دبي، الإمارات العربية المتحدة ـ انتقد اللاعب المصري محمد صلاح، نجم فريق ليفربول الإنجليزي، الأحد، الاتحاد المصري لكرة القدم، متهما الاتحاد بأنه لا يسعى إلى حل مشاكل لاعبيه.
وقال محمد صلاح، في تغريدة عبر حسابه على تويتر، إنه من الطبيعي أن أي اتحاد كرة يسعى لحل مشاكل لاعبيه حتى يوفروا له الراحة.. لكن في الحقيقة ما أراه عكس ذلك تمامًا.. ليس من الطبيعي أن يتم تجاهل رسائلي ورسائل المحامي الخاص بي... لا أدري لماذا كل هذا؟ أليس لديكم الوقت الكافي للرد علينا؟!
ولم يذكر صلاح بوضوح سبب تغريدته عن الاتحاد المصري، ولكن كانت قد اندلعت أزمة قبل كأس العالم بين اللاعب والاتحاد بسبب استخدام صوره في الدعاية من قبل الشركة الراعية للمنتخب المصري، مما سبب له أزمة بسبب تعاقده مع شركة أخرى.
منح نجم المنتخب المصري، محمد صلاح، فريقه الإنجليزي، ليفربول، الفوز الثالث على التوالي في الدوري الإنجليزي الممتاز، السبت، بإحرازه هدف الانتصار على فريق برايتون آند هوف ألبيون.
</TEXT>
</DOC>

<DOC>
<DOCNO>CN-sport-380-20180826</DOCNO>
<URL>
http://arabic.cnn.com/sport/article/2018/08/26/mohammed-salah-egyptian-football-association
</URL>
<SRC>CNN</SRC>
<CAT>sport</CAT>
<TITLE>
Mohamed Salah criticizes the Egyptian Football Federation: I don't know the reason behind all this
</TITLE>
<TIME>Sun, 26 Aug 2018 19:37:32 +0000</TIME>
<AUTHOR></AUTHOR>
<ABSTRACT>
The Egyptian player and Liverpool star, Mohamed Salah, criticized the Egyptian Football Federation on Sunday, accusing Al-Ittihad of not seeking to solve the problems of its players.
</ABSTRACT>
<TEXT>
Dubai, United Arab Emirates - The Egyptian player and Liverpool star, Mohamed Salah, criticized the Egyptian Football Federation on Sunday, accusing Al-Ittihad of not seeking to solve the problems of its players.
Mohamed Salah said, in a tweet via his Twitter account, that it is natural that any football association seeks to solve the problems of its players in order to provide them with comfort, but in reality what I see is the exact opposite ... it is not normal that my letters and the letters of my lawyer are being ignored. ... I don´t know the reason behind all this? Do you not have enough time to respond to us ?!
Salah did not clearly state the cause of his tweet about the Egyptian federation. However, a crisis had erupted before the World Cup between the player and the federation due to the use of his photos in advertising by the Egyptian team´s sponsoring company, which caused him a crisis because of his contract with another company.
The Egyptian national football star, Mohamed Salah, awarded his English team, Liverpool, the third consecutive victory in the English Premier League, Saturday, by scoring the goal of victory over the Brighton & Hove Albion team.
</TEXT>
</DOC>

Figure 1 Document sample and its English translation extracted from CNN source of "ANT corpus v2.1"

5.6 | Discussion and comparison with previous analogy-based classification approaches and with some related ML algorithms

The extrapolation mechanism, founded on the idea that a fourth item can be predicted/computed from the three others, if the 4 items a, b, c and d build a valid analogical proportion, has been recently applied with success in different prediction methods. For example, this principle has been exploited for predicting user preference between pairs of choices by extrapolating triples or pairs of choices in the training set 83, 37.

In the context of classification, a variety of algorithms has been developed to predict the class for a new example $\vec{d}$ by extrapolating the class of each triple of feature vectors ($\vec{a}$, $\vec{b}$, $\vec{c}$), with known labels, in the training set 33, 34, 35, 26. More recently, 27 have studied the effectiveness of a particular case of analogical proportions, denoted continuous analogical proportions, for classification in the case of numerical data. Such proportions use pairs ($\vec{a}$, $\vec{c}$) from the training set (rather than triples) to interpolate the class label for the new example $\vec{d}$.

The proposed analogical text classifiers seem close to the AP-classifier in 26 since both algorithms exploit triples of examples for prediction. However, AATC classifiers differ from previous AP-classifier 26 on more than one side:

In this work, our main objective is to develop first a theoretical framework to model text classification through analogical proportions. This model enables to treat $unstructured$ data by extracting relevant features from a document collection. The proposed analogical model might be generalized for predicting different features in other textual contexts such as in sentiment analysis or user preferences learning. Contrary to AATC, classic AP-classifiers 26 can only deal with $structured$ data in which examples are represented as a fixed and predefined set of features with known characteristics. Second, in AATC we are limited to deal with the Boolean view of analogical proportions (introduced before in subsection 4.1) since documents are represented by Boolean predicates expressing the existence/non-existence of a keyword in each document while AP-classifier 26 is a generic algorithm that can be applied to any type of data: Boolean, nominal or numerical data and exploits the Boolean, nominal and multi-valued setting of analogical proportions.

Due to the use of triples of items from the training set for classification, analogical classifiers still have a cubic complexity in general. In an attempt to reduce the computational burden, the algorithm in 26 constrains item $\vec{c}$ to be one of the nearest neighbors of the item $\vec{d}$ to be classified. Thus the complexity turns to quadratic instead of cubic. In AATC no such restriction is applied i.e.: all triples of documents (d_1, d_2, d_3) in the training set are evaluated. However, we apply another type of optimization restricting the use of only those triples leading to the "best" analogical score for final scoring. This kind of selection has not been used in AP-classifier 26 since all triples ($\vec{a}$, $\vec{b}$, $\vec{c}$) (such that $\vec{c}$ is among the nearest neighbors of $\vec{d}$) are involved in the final classification decision. Finding the optimal set of triples necessary to achieve the best classification rates is still an open question 26 and requires further investigation in the future.

We also establish the links between analogical text classifiers and some classic ML classification models, in particular, the k-NN and Naïve Bayes classifiers since they are seen as the most simple predictors in the literature while being reasonably accurate when applied to various datasets. A complete study of the links and differences between analogical proportions-based classifiers and the k-NN classifier can be found in 26.

Although analogical classifiers and k-NN are both considered as lazy classification methods since they require no training time and take rather more classification time to process the example set, they are quite different in their principle. We summarize in the following the main benefits of AATC when compared to the classical k-NN:

- Exploiting a larger number of examples from the training set for classification clearly raises the complexity of AATC if compared to k-NN. In return, this helps to avoid much misclassification due to the restriction to the close neighborhood (in k-NN) and thus improves the classification rates.

- As introduced in Section 5.2, AATC makes a complex score calculation for each triple and then sums these elementary scores for each class label. Thus, the analogical text classifiers estimate a sort of distance from the document to be classified to each triple of documents in the example set. While classic k-NN simply applies a vote on the class labels of the nearest neighbors without taking into consideration the distance. This extra computational effort made by analogical text classifiers justifies why are significantly better than a classic k-NN (as we will see in the experimental section 6.4).

On the other hand, note that the proposed analogical text classifier is entirely different from Naïve Bayes classifier for at least two reasons.

On the first hand, the Bayesian classifier computes the conditional probability for each predefined feature from the training set and then aggregates these probabilities by multiplication to proceed with the final decision. In AATC, as we are dealing with Boolean valuations of analogical proportion, we aggregate features' analogical scores by summation (as illustrated in the previous example) instead of multiplication. One more difference, in AATC2, we use two kinds of scores p_t^+ and p_t^- to distinguish two kinds of triples: those that are in total agreement with d_x and those that are in disagreement.

More importantly, in AATC, to decide which category fits better to a given document d_x, each triple of documents (d_1, d_2, d_3) in the training set, helps to predict an individual decision/score. Those computed scores are then averaged to get the final decision. In Bayesian classifier, conditional probabilities are estimated by computing the $conformity$ of a given feature w.r.t to the whole class by looking at the whole training set contrary to our approach which is likely to have a more local view of the data. What matters in AATC is the extent to which a given document (being classified) builds valid analogical proportions on all or a maximum set of features as well as with a maximum number of triples of documents belonging to a given class. If we reconsider the example in Table 6, it is clear that the keywords w_{12} or w_{13} for example, do not help for any prediction as the analogical proportions $0 : 1 :: 1 : 1$ or $1 : 0 :: 0 : 1$ are not valid. Such features (existing in some training documents and not existing in others) may serve to compute conditional probabilities in a classic Bayesian classifier. Finally Naïve Bayes requires more memory space to store the prior conditional probabilities of different features while AATC has no prior estimates thus not much requiring in terms of memory space.

6 | EXPERIMENTAL RESULTS AND COMPARISON

In this section, we conduct a series of experimentations to assess the efficiency of the proposed analogical classifiers. We first present the two data collections used for the main experiments: “ANT corpus v2.1" and “BBC-Arabic corpus". The three other tested data collections are presented later in subsection 6.4.3 for a comparative study with other existing Arabic classifiers. Then, we introduce the testing strategy, the classic ML classifiers that we have tested for comparison and the assessment metrics. Finally, we present the experimental results and provide a comparative study with state-of-the-art Arabic text classifiers.

6.1 | Data collections

To evaluate our proposed classifiers, we have used the datasets ANT corpus v2.1 (Arabic News Texts Corpus) 43 and BBC-Arabic corpus 84. We provide in the following a brief description for each of them.

6.1.1 | ANT corpus v2.1

The ANT corpus v2.1 is a freely available online data collection[1] having different versions 85, 86, 87. It includes data from five news sources: AlArabiya[2], BBC[3], CNN[4], France24[5] and SkyNews[6]. The collected articles are classified into 6 categories: “Culture” (ثقافة), “Economy” (اقتصاد), “InternationalNews” (العالم), “MiddleEast” (الشرق الأوسط), “Sport” (رياضة) and “Technology” (تكنولوجيا). Table 8 presents the number of documents in each category for each source.

6.1.2 | BBC-Arabic corpus

BBC-Arabic corpus includes 4763 Arabic texts divided into 7 categories: “Divers" (منوعات), “Economy" (اقتصاد و اعمال), “InternationalNews" (اخبار العالم), “Journal" (عرض الصحف), “MiddleEast" (اخبار الشرق الأوسط), “Sport" (رياضة) and “Technology" (علوم و تكنولوجيا). Table 9 summarizes the number of documents in each category of BBC-Arabic corpus.

6.2 | Testing strategy and other classifiers for comparison

To test analogical classifiers (as well as other compared classifiers), we apply a standard 10-fold cross-validation technique to build the training and testing datasets. In a k-cross validation procedure, a dataset is randomly divided into k groups with equal size. In each fold, one group is held out as a testing set and the remaining k-1 groups are used as a training dataset which yields one evaluation metric for each fold. Thus each result for each assessment metric shown below is obtained by averaging the k different values ($k = 10$ in our experiments) for each fold.

In order to evaluate the efficiency of analogical classifiers, we compare their results to some classic ML classifiers which are language-independent:

- **SVM**: a sequential minimal optimization (SMO) algorithm for training a support vector classifier. We use the Polynomial kernel and we tune its *degree d* ($d = 1, 2, ..., 10$) (also called $exponent$). We also tune the complexity parameter C as recommended by 88.
- **Naïve Bayes**: a probabilistic classifier with no KernelEstimator and no SupervisedDiscretization (as we treat only Boolean data).
- **J48**: generating a pruned or unpruned C4.5 decision tree. We tune the classifier with different confidence factors used for pruning $C = 0.1, 0.2, ..., 1$.
- **IBk**: a k-NN classifier, we use the Hamming distance most suitable for Boolean vectors and we tune the classifier on different values of the parameter $k = 1, 2, ..., 10$.

Results for ML classifiers are obtained using the freely available implementation of Weka[7] Toolkit. As in the case of ML classifiers, for analogical classifiers, we have to tune the parameter N ($N = 5, 10, 20, 30$) (the number of best triples to be considered for the final classification). We adopt

[1] https://antcorpus.github.io/
[2] http://www.alarabiya.net/ar/
[3] http://www.bbc.com/arabic/
[4] https://arabic.cnn.com/
[5] http://www.france24.com/ar/
[6] http://skynewsarabia.com/
[7] https://www.cs.waikato.ac.nz/ml/weka/

this testing strategy: we first tune the corresponding parameter for each classifier in a preliminary step using the training set. This allows us to choose the best value of the parameter, then we test the classifier using this selected value of this parameter. All classification results for different classifiers, displayed in the tables below, correspond to the best value of each tuned parameter.

To assess the performance of the proposed classifiers with regard to previous works, we also compare our obtained results to those of recent Arabic text classifiers in Section 6.4.3.

6.3 | Assessment metrics

The most commonly used metrics to assess text classifiers are precision, recall, F_1 and accuracy.

The precision indicates the percentage of correctly classified documents from those predicted as positive. The precision formula is defined as:

$$Precision = \frac{Documents\ correctly\ classified\ to\ the\ class\ C}{Total\ documents\ classified\ to\ class\ C} \tag{10}$$

The recall estimates the percentage of correctly classified documents from the real positive documents. The recall formula is as follows:

$$Recall = \frac{Documents\ correctly\ classified\ to\ the\ class\ C}{Total\ documents\ in\ class\ C} \tag{11}$$

F_1 represents the harmonic mean of precision and recall. The formula of F_1 is as follows:

$$F_1 = 2 * \frac{precision * recall}{precision + recall} \tag{12}$$

Accuracy is the percentage of correctly classified texts from the total number of texts defined as:

$$Accuracy = \frac{Total\ correctly\ classified\ documents}{Total\ number\ of\ documents} \tag{13}$$

6.4 | Results and comparative study

In this section, we conduct three comparative studies to assess the performance of Analogical Arabic Text Classifiers. In the first (resp. second) experiment, given in subsection 6.4.1 (resp. subsection 6.4.2), we provide the experimental results obtained with the two proposed classifiers AATC1 and AATC2 as well as with other ML classifiers tested on the large ANT corpus v2.1 43 (resp. small BBC-Arabic corpus 84). For a fair comparative study between AATC and other ML classifiers, we have applied the same feature selection technique (Term Frequency-based approach) and the same text representation method (document-term matrix model) as a pre-processing step for all the classifiers equally.

In addition to the previous comparative studies with classic ML classifiers (using both (ANT v2.1) and (BBC-Arabic) collections), we have also compared the proposed classifiers to some related works. Sub-section 6.4.3 presents these additional results. It is important to note that most of the existing Arabic text classifiers in the literature have been assessed using various evaluation metrics and different text collections. Nevertheless, performing a fair comparative study with these related works is hard because they lack a unified experimental protocol. In Sub-section 6.4.3, the compared Arabic text classifiers have either applied ML classifiers (most of the case SVM and DT) or DL techniques and have been tested on different data collections such as ANT corpus v1.1 (including 6114 documents and 8 categories), AlKhaleej-2004 Arabic text collection (including 5670 documents and 4 classes) and CNN-Arabic text collection (enclosing 5070 documents and 6 classes). Moreover, these compared approaches do not apply the same feature selection techniques that we used.

In all the following subsections, the best result(s) are highlighted in bold in each corresponding table.

6.4.1 | Results for the ANT corpus

Table 10 includes classification results sorted by sources for AATC1, AATC2 as well as for other ML classifiers.

From these results we can draw the following conclusions:

– Comparing first the accuracy rates for all sources (see Figure 2 for this comparison) we note that :

- AATC1 has the highest accuracy rate for AlArabiya source (95.31%), if compared to all other classifiers. In particular, AATC1, outperforms the robust SVM classifier which is the second-best classifier for this source (94.85%).
- AATC2 has the best accuracy for France24 source (62.52%) among all other classifiers.
- For CNN source, in terms of accuracy, AATC2 is the second ranked classifier (85.88%) following SVM (87.68%) and followed by AATC1 (83.77%).

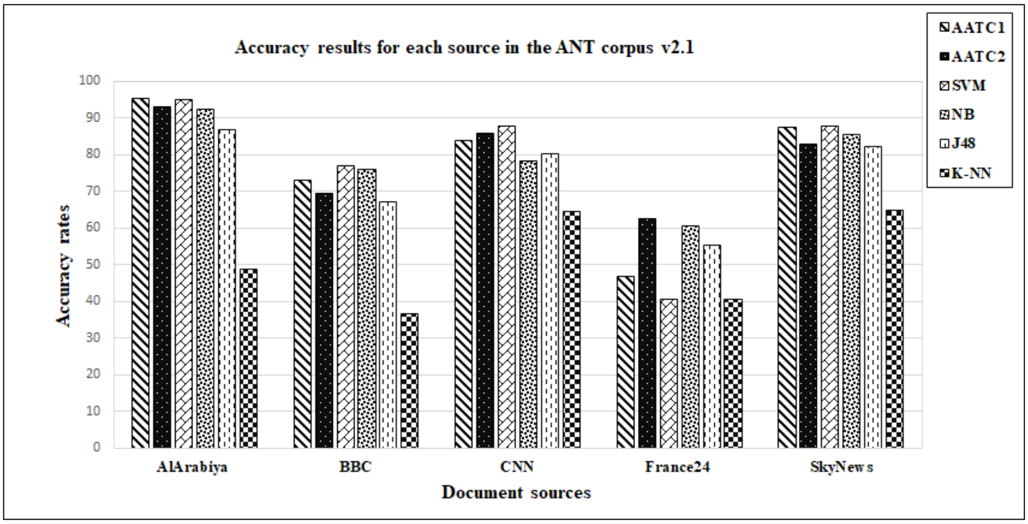

Figure 2 Accuracy results for each source in the ANT corpus v2.1

- AATC1 performs similarly as SVM for SkyNews source (accuracy for AATC1 and SVM are respectively 87.51% and 87.67%).
- For the BBC source, our proposed algorithm AATC1 has the third best accuracy (73.10%) following SVM (76.99%) and NB (75.81%).
- In terms of average accuracy over all sources, AATC2 is ranked first with a total accuracy of 78.78% while AATC1 has close overall results to SVM (AATC1: 77.31% and SVM: 77.55%).
- In terms of sources, AlArabiya shows the highest accuracy rate.

– Regarding the precision metric:

- AATC1 has the best precision value for AlArabiya, BBC, CNN and SkyNews with the highest value for AlArabiya (0.95).
- While NB has the best achievement for France24 (0.49).
- Overall, AATC2 has the best average precision (0.77) if compared to all other classifiers.

– Finally, comparing the recall and F_1 values, it is clear that:

- AATC1 has the best recall for AlArabiya source (0.93) and the best F_1 for AlArabiya and SkyNews sources (0.94 and 0.87 respectively) with the highest value for AlArabiya.
- SVM has the best recall for both CNN and SkyNews sources (0.85 and 0.88 respectively) and the best F_1 for BBC, CNN and SkyNews sources (0.75, 0.86 and 0.87 respectively).
- NB has the best recall and F_1 for both BBC and France24 sources (the recall are respectively 0.77 and 0.71, while the F_1 are respectively 0.75 and 0.55).

– If compared to other classifiers, k-NN seems the least efficient and provides the lowest accuracy rates for all sources with 36.54% as the lowest accuracy measure for the BBC source. It also shows the lowest precision, recall and F_1 for all the sources except France24.

– If we compare the results for different sources, AlArabiya has the highest precision, recall, F_1 and accuracy rates for all the classifiers.

– France24 has the lowest precision, recall and F_1 values for all the tested algorithms. In fact, this source has clearly unbalanced classes where most of the collected documents belong to the dominant category: InternationalNews (2793 docs. among a total of 4241) while some other categories have a very limited number of documents (for example, the Economy category has only 78 documents). This may cause the misclassification of many documents in most categories (except InternationalNews) which explains the low values obtained for France24 source. One way to deal with this problem is to enrich data sources, having unbalanced classes, with more documents to build more balanced data collections. This is out of the scope of this paper.

It is also relevant to examine, in more detail, the sensitivity of the obtained results to the ANT corpus v2.1 categories. For this purpose, we also present in Table 11 the average values of precision, recall and F_1 obtained for each category over all sources. Comparing results for different categories and different classifiers, we can see that results are sensitive to the tested categories and it is hard to derive a general conclusion over all categories and/or classifiers. We mention below some remarkable results:

- The Sport category shows the best results overall categories for most classifiers. NB has the highest recall and F_1 values (0.89 and 0.89 respectively) over all other classifiers for this category. While AATC2 and NB achieve the best precision (0.89) and ranked first before SVM (0.81) and AATC1 (0.80).
- For the Technology category, AATC1 outperforms all other classifiers for the precision, NB outperforms all other classifiers for the recall and SVM outperforms all other classifiers for the F_1. Similar results are obtained for the Economy category (AATC1 is the best classifier for the precision and NB for the recall and F_1).
- Still AATC1 has the best precision value and AATC2 has the best recall value in the case of the InternationalNews category.
- As noted before, k-NN algorithm shows the lowest performance.
- Comparing all categories, for all computed metrics the lowest results are obtained with the Culture category while the best results are obtained in general with sport and/or technology categories. It has to be noted that, in the case of sport and technology categories, the extracted keywords are in general very discriminative which helps to avoid overlapping between classes. While in the case of the culture category, the semantic richness of this domain and the limited number of documents in this category are the main cause of misclassification. For example, the term "Player" in Sport or "Computer" in Technology are so discriminative however the term "Manifestation" may be linked to the culture (cultural event) or to the news (political manifestation) depending on the context.
- Comparing all classifiers, the average measures show that AATC2 has the best precision (0.77) over all classifiers while NB has the best recall (0.81) and the best F_1 (0.76).

Finally, in order to statistically compare the proposed AATC1 and AATC2 classifiers to other ML classifiers, we apply the Wilcoxon Matched-Pairs Signed-Ranks Test as proposed by 89. It is a non-parametric alternative to the paired t-test that enables us to compare each pair of classifiers such as AATC1 versus SVM, AATC1 versus NB, etc. for a significance level of 0.05. Tables 12 and 13 display the computed p-values for AATC1 (respectively AATC2) when compared to other classifiers. For each comparison of C1 vs. C2, the (+) means that the first classifier C1 is statistically better than the second classifier C2 while (-) means the contrary.

The computed p-values, in Table 12, show that AATC1 algorithm is significantly better than J48 and k-NN in terms of precision measure. It is also significantly better than k-NN for other metrics: recall and F_1. SVM is only significantly better than AATC1 for F_1. While we cannot confirm any noticeable difference between other pairs of compared classifiers since the p-values > 0.05 (AATC1 vs. SVM according to both precision and recall metrics as an example).

Comparing AATC2 to other classifiers in Table 13, shows that AATC2 is significantly better than J48 and k-NN for all the computed metrics and no other significant difference can be noticed between other pairs of compared classifiers as the p-values > 0.05.

6.4.2 | Results for the BBC-Arabic corpus

Results for the BBC-Arabic corpus are saved in Table 14. We also compare the average metrics for the different classifiers in Figure 3.

From these results, we notice that:

- The average accuracy shows that SVM has the best results (84.40%) followed by our proposed classifier AATC1 (80.52%).
- If we look at the average scores, AATC1 has the best average precision (0.88) while SVM has the best average recall and F_1 scores (0.82 and 0.84 respectively).
- Regarding the precision, AATC1 has the best value for the category Sport and the second best value for three categories (Divers, Economy and InternationalNews) among seven categories, while SVM has the best value for one category (MiddleEast) and the second best value for the Journal category. k-NN has the best precision value for three categories (Economy, InternationalNews and Journal) and the second best value for one category (MiddleEast).
- If we consider the recall values, SVM has either the best or the second best value for almost all categories except for technology where it has the third best value. AATC1 has the best value for one category (MiddleEast) and the second best value for one category (InternationalNews).

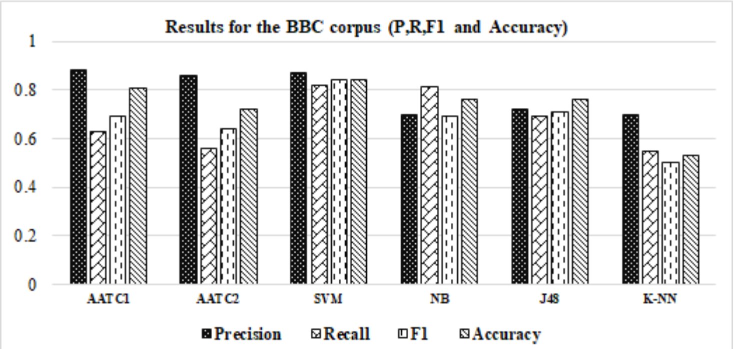

Figure 3 Results for the BBC corpus (P,R,F_1 and Accuracy)

- If we consider the F_1 values, SVM is the best for all categories followed by AATC1 and NB which have the second best F_1 for three categories each.
- J48 shows the lowest precision value for three categories (Divers, Economy and InternationalNews) and k-NN shows the lowest recall and F_1 values for five categories.

As for the ANT corpus, we also statistically compared each pair of classifiers using Wilcoxon Matched-Pairs Signed-Ranks Test. The computed p-values comparing AATC1 (respectively AATC2) to other classifiers are saved respectively in Tables 15 and 16.

These results show that AATC1 is significantly better than J48 in the case of the precision measure and is significantly better than k-NN in the case of F_1 metric. For AATC2, we cannot confirm any significant difference between this classifier and other algorithms in terms of precision. SVM and J48 are significantly better than AATC2 in terms of recall and F_1 metrics.

6.4.3 | Comparison with existing Arabic text classifiers

Table 17 shows further experimental results obtained with our analogical Arabic text classifiers and compared with different related works on Arabic text classification. In the first part of this table, we compare our results to existing Arabic text classifiers applying classic ML algorithms, in particular, these recent works 48, 69 and 19; that have been tested using AlKhaleej-2004 or CNN-Arabic collections. The second part of the table compares our results to Arabic classifiers applying DL, in particular, 22 and 24 that have been tested using CNN-Arabic and ANT v1.1 collections.

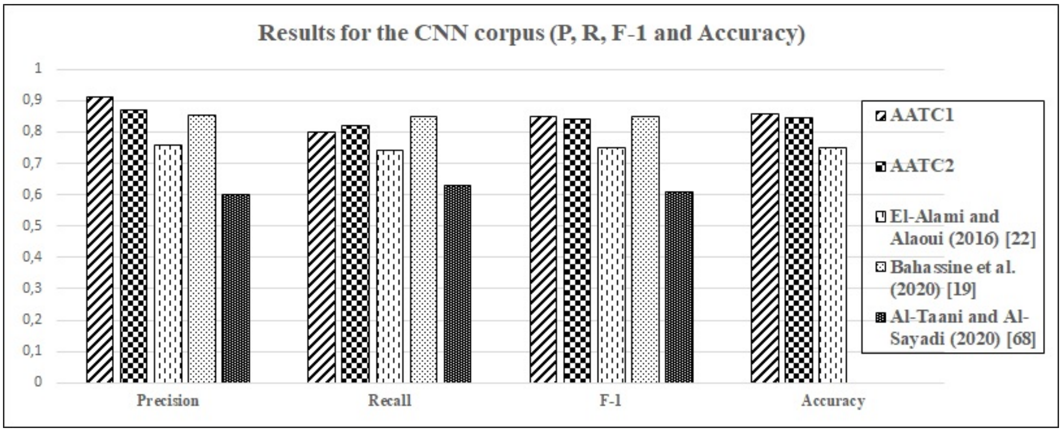

Figure 4 Results for the CNN corpus (P,R,F_1 and Accuracy)

Looking at the results in Table 17, we can derive the following comments:

- AATC1 and AATC2 outperforms the classifier in 48 in terms of Precision and F_1 while the later is better in terms of Recall for AlKhaleej-2004 Arabic text collection. Note that for all compared classifiers no accuracy rate is displayed by the authors.

- When applied to the CNN-Arabic text collection (see Figure 4 for a comparative study between various classifiers using this corpus), AATC1 and AATC2 perform largely better than the classic ML classifier 69 as well as the DL classifier 22 for all assessment metrics. If compared to 19, analogical classifiers perform better than this classifier in terms of Precision while 19 is better in terms of recall. They behave similarly for F_1.

- Almuzaini and Azmi (2020) 24 suggested 7 different DL classifiers with optimized parameters and applied 10 different stemming algorithms from the literature works in addition to one without stemming. We report here the results of the *best* and the *worst* DL classifiers using stemming or without stemming that have been tested on the ANT corpus v1.1. In Table 17, the labels $DL_*(NoStem)$, $DL_*(Stem)$, $DL^*(NoStem)$, $DL^*(Stem)$ refer respectively to the worst DL classifier CNN-GRU with no stemming, worst DL classifier CNN-GRU with stemming, best DL classifier BiGRU with no stemming and best DL classifier BiGRU with stemming.

  Note that all achieved results in 24 apply the weighted average metrics (P, R and F_1). For this reason and to keep a fair comparison with these DL classifiers, $only$ for this data collection we compute the weighted average measures, in the same way as defined by the authors: we compute each metric independently for each category then we take the average weighted by the number of true examples in this category. These results show that AATC1 i) outperforms the *worst* DL classifier: CNN-GRU either with or without stemming in terms of F_1, ii) outperforms the $best$ DL classifier: BiGRU without stemming in terms of F_1 and iii) performs better than this latter with stemming in terms of precision while BiGRU is better for the recall and F_1 metrics.

6.5 | Synthesis and discussion

In the previous subsections, we assess the efficiency of the proposed analogical Arabic text classifiers using a variety of Arabic news datasets. The experimental results highlight the efficiency of analogical proportions for text classification if compared either to language-independent classifiers (based on ML algorithms) or some other state-of-the-art Arabic text classifiers. To understand better the behaviour of analogical Arabic classifiers, we deeply investigate the characteristics of each tested corpus in the following.

It has been checked in both ANT corpus v2.1 and BBC-Arabic corpora that the two categories “InternationalNews" and “MiddleEast" are overlapping because documents in these categories contain a large set of common keywords. Besides, most of the “InternationalNews" documents are highly related to “MiddleEast" documents in terms of topics and contents. As noted above, AATC1 is the most efficient classifier using the sub-collection “AlArabiya" of ANT corpus v2.1 which excludes the “MiddleEast" category (see Table 8). This has contributed to avoid misclassification of test documents as “InternationalNews" or “MiddleEast" in AATC1 as well as AATC2 for this sub-collection.

We first note that AATC1 seems efficient in classifying documents from large sub-collections such as “AlArabiya" (6519 docs.) and “SkyNews" (13716 docs.) (see Table 8). Similar behaviour is noticed when using large categories such as “InternationalNews" (10352 docs.) of ANT corpus v2.1 (see Table 8) for which AATC1 achieves the best precision (0.8) and the best F_1 (0.77) (see Table 11) or categories “InternationalNews" (1489 docs.) and “MiddleEast" (2356 docs.) from the BBC-Arabic corpus (see Table 9). This success can be justified by the fact that, for these sub-collections, the classifier can easily find several triples of documents in the training set that build a valid analogical proportion with the test document.

In the contrary, although analogical classifiers maintain good precision, they show reduced recall values for categories with a small number of documents. We especially mention for example the categories “Culture" and “Technology" of ANT corpus v2.1 and “Divers", “Economy" and “Technology" of BBC-Arabic corpus. This is due to the limited number of $valid$ analogical proportions that can be built which causes a low score for this category and thus favoring other competitive categories. These results may indicate the sensitivity of AATC to unbalanced data collections. Applying suitable tools for preprocessing data to ensure balanced categories may help to improve results for categories with a reduced number of documents. This is the topic of a work in progress.

On the other hand, AATC1 seems more efficient in terms of recall than AATC2, in particular for categories “InternationalNews" with large documents and “Journal" categories with a very limited number of documents from BBC-Arabic corpus.

In fact, the “MiddleEast" category sub-collection, sharing a large number of keywords with the “InternationalNews" sub-collection (as focusing on common political issues), and having a larger number of documents seems to be a dominant class (the “MiddleEast" category has 2356 docs. while “InternationalNews" has only 1489 docs.). This may justify the very good recall (as well as precision) for the “MiddleEast" category for both AATC1 and AATC2 classifiers as displayed in Table 14. Since AATC2 extracts keywords from the whole class, those extracted keywords may not be discriminative enough for a particular category in case this class overlaps with other categories as in the case of “InternationalNews" and “MiddleEast" for example. This will affect “less" the results of AATC1 which extracts the keywords from the testing document rather than from the whole class. This also may justify why AATC1 seems more robust to the corpus size as well as to the number of common keywords in different categories. As shown in Table 14, AATC1 achieves good recall and precision values for those two categories “InternationalNews" and “MiddleEast" while AATC2 performs clearly better for the dominant class.

This justification can be confirmed by the good results (Recall, Precision and F_1) obtained with AATC2 in the case of the “Sport" category in both ANT and BBC-Arabic corpora. In fact, the keywords of this category are very specific and rarely used in documents of other categories, we mention for example the keywords Football (كرة القدم), League (الدوري), World Cup (كأس العالم), etc.

Thus, AATC seems less efficient in the case of overlapping categories such as in the case of “InternationalNews" and “MiddleEast" mentioned before. We believe that a suitable extension of the proposed model, considering analogies between sentences 90 rather than words, may help to improve the efficiency of the classifier by reducing the ambiguity between close keywords. This is still an open question that requires future investigation.

Finally, as can be seen in Figure 2, AATC1 appears clearly better than AATC2 for “AlArabiya" and “Sky News" sources of ANT corpus v2.1 having a large number of documents while AATC2 is better for “CNN" and “France24" sources with a reduced number of documents (see Table 8). As said before, since AATC2 extracts keywords from the whole class, it may have better control over each class label and thus seems especially efficient when only a small set of documents is available for each class (see for example results for France 24 in Figure 2).

7 | CONCLUSION

A great research effort has been devoted to investigate English text classification issues and only a limited number of works have been dedicated to deeply examine and enhance Arabic text classification, which is still limited by the challenges of Arabic language processing.

Most of the existing Arabic text classifiers are based on classic ML algorithms and are only tested on small data collections because of the lack of freely available large datasets. Few Arabic text classifiers have also applied DL techniques to improve the accuracy of the used classifiers. Although DL classifiers perform well when classifying large text collections, they show limited ability to treat small text collections, especially with high dimensionality. Hence the necessity to explore new classification techniques, able to treat as good as small and large Arabic text collections, which is the main goal of the current research.

In this work, we study the effectiveness of analogical proportions as a promising tool for text classification and we propose two analogical Arabic text classifiers for this purpose. These latter are based on the following principle: we first look at the analogical relationship between a new document to be classified in the testing set and each triple of pairs (document, class) existing in the training set. If such triple build a valid analogical proportion with the new document, this states the basis for predicting the class of this new document being classified. In this paper, we propose two kinds of procedures to quantify the relevance/irrelevance of a given keyword used to guess the document class. In the first algorithm, denoted AATC1, the classification process is based on a set of keywords extracted from the document being classified. Whereas in the second algorithm, denoted AATC2, we focus on the whole class set to extract the set of keywords. In this latter, we have based our intuition on the assumption that relevant keywords for the whole class set are generally useful for classifying any document belonging to this class.

To assess the efficiency of the proposed analogical Arabic text classifiers, we test them using a variety of Arabic news datasets. The experimental results highlight the efficiency of analogical proportions for text classification if compared to the previous Arabic text classifiers. In particular, the proposed classifiers perform *as good as* with large (ANT corpus v2.1) and small (BBC-Arabic corpus) Arabic news data collections which may confirm the objectives of this work. The best-achieved results of the first classifier AATC1 are P = 0.95, R = 0.93, F_1 = 0.94 and accuracy = 95.31% using “AlArabiya" news documents, a subset of the ANT corpus v2.1. It also achieves the best precision using the sub-collections BBC (0.81), CNN (0.88) and “SkyNews" (0.91).

Considering the BBC-Arabic corpus, AATC1 performs the best average precision (0.88) over all other classifiers and the best recall (0.97) for the sub-collection “MiddleEast" of this corpus. Besides, AATC1 (resp. AATC2) achieves the best precision (0.99) for the “Sport" sub-collection (resp. 1 for the sub-collection “Technology"). In terms of accuracy, AATC1 is ranked second just after the SVM for this corpus.

If we consider the overall ANT corpus v2.1, AATC2 outperforms either AATC1 as well as all other ML classifiers in terms of average precision (0.77) and average accuracy rate (78.78%), while NB classifier has the best recall (0.81) and F_1 (0.76).

In the following, let us summarize the main achievements of the proposed classifiers and make the link to the characteristics of each tested corpus. We also mention some limitations of such classifiers.

We first note that AATC1 performs very well when classifying documents from large sub-collections such as “AlArabiya" and “SkyNews" or large categories such as “InternationalNews" of ANT corpus v2.1. For such collections, the classifier can easily find several triples of documents in the example set that create a valid analogical proportion with the document being classified. In the contrary, AATC seems less efficient in classifying documents for categories with a small number of documents such as “Culture" and “Technology" of ANT corpus v2.1. This may indicate the sensitivity of AATC to unbalanced data collections and thus require further preprocessing tools to prepare balanced datasets. This is a subject of further investigation.

Comparing AATC1 to AATC2 (and also to other previous classifiers), the former seems more robust to the corpus size as well as to overlapping keywords in different categories than the latter. For example, AATC1 achieves good recall and precision values for the two overlapping categories

"InternationalNews" and "MiddleEast" from the BBC-Arabic corpus while AATC2 performs clearly better for the dominant class "MiddleEast". Moreover, AATC2 performs clearly better for non-overlapping categories such as "Sport" category in the two tested corpus.

It has to be mentioned that AATC classifiers are language-independent. It may be beneficial to investigate the effectiveness of analogical text classification using large benchmark datasets and collections in other languages. Given the encouraging results achieved by analogical proportions in the field of Arabic text classification and summarization 32 as well as in the field of domain-specific information retrieval 31, we wonder if such tools can be applied to other related fields such as Arabic NLP or Arabic mono- and cross-language information retrieval (IR/CLIR) 91. Analogical learning and reasoning have been already applied, for languages other than Arabic, in the domain of NLP (e.g. 92 93 94) and IR/CLIR (e.g. 95 96 97). We think that these works can state the basis for developing new analogical Arabic NLP, IR and CLIR tools which is still an open problem. Finally, we believe that analogical classifiers may be helpful in decision-making. Contrary to some other classic ML classifiers that may be seen as a black box, analogical classifiers have these strengths: not only they are competing with the previous works in terms of results but also show their ability to explain their results by demonstrating, for example, the most candidate features leading to such decision. Interested readers may refer to 40, 41 for a complete presentation on the explanation ability of analogical proportions.

8 | ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their invaluable comments and suggestions. Hussain would like to acknowledge the support of the UK Engineering and Physical Sciences Research Council (EPSRC) - Grants Ref. EP/M026981/1, EP/T021063/1, EP/T024917/1.

References

1. Deng Z, Sun C, Zhong G, Mao Y. Text Classification with Attention Gated Graph Neural Network. *Cogn. Comput.* 2022; 14(4): 1464–1473.

2. Alwaneen TH, Azmi AM, Aboalsamh HA, Cambria E, Hussain A. Arabic question answering system: a survey. *Artificial Intelligence Review* 2022; 55(1): 207–253.

3. Saeed RMK, Rady S, Gharib TF. Optimizing Sentiment Classification for Arabic Opinion Texts. *Cognitive Computation* 2021; 13(1): 164–178.

4. Fernández-Isabel A, Cabezas J, Moctezuma D, Diego dIM. Improving Sentiment Classification Performance through Coaching Architectures. *Cogn. Comput.* 2023; 15(3): 1065–1081.

5. Li Y, Nie X, Huang R. Web spam classification method based on deep belief networks. *Expert System with Applications* 2018; 96: 261–270.

6. Onan A. An ensemble scheme based on language function analysis and feature engineering for text genre classification. *Journal of Information Science* 2018; 44(1): 28–47.

7. Lulu L, Elnagar A. Automatic Arabic Dialect Classification Using Deep Learning Models. In: Proceedings of the Fourth International Conference On Arabic Computational Linguistics (ACLING). ; 2018: 262–269.

8. Elnagar A, Einea O. BRAD 1.0: Book reviews in Arabic dataset. In: Proceedings of the 13th IEEE/ACS International Conference of Computer Systems and Applications (AICCSA). ; 2016: 1–8.

9. Elnagar A, Khalifa YS, Einea A. Hotel Arabic-Reviews Dataset Construction for Sentiment Analysis Applications. In: Intelligent Natural Language Processing: Trends and Applications. Springer International Publishing. 2018 (pp. 35–52).

10. Elnagar A, Lulu L, Einea O. An Annotated Huge Dataset for Standard and Colloquial Arabic Reviews for Subjective Sentiment Analysis. In: Proceedings of the Fourth International Conference On Arabic Computational Linguistics (ACLING). ; 2018: 182–189.

11. El-Affendi MA, Alrajhi K, Hussain A. A Novel Deep Learning-Based Multilevel Parallel Attention Neural (MPAN) Model for Multidomain Arabic Sentiment Analysis. *IEEE Access* 2021; 9: 7508–7518.

12. Yan C, Liu J, Liu W, Liu X. Sentiment Analysis and Topic Mining Using a Novel Deep Attention-Based Parallel Dual-Channel Model for Online Course Reviews. *Cogn. Comput.* 2023; 15(1): 304–322.

13. Eldos T. Arabic text data mining: A root-based hierarchical indexing model. *International Journal of Modelling and Simulation* 2003; 23(3): 158–166.

14. Elnagar A, Debsi RA, Einea O. Arabic text classification using deep learning models. *Information Processing and Management* 2020; 57(1).

15. Elayeb B. Arabic Text Classification: A Literature Review. In: 18th IEEE/ACS International Conference on Computer Systems and Applications, AICCSA 2021, Tangier, Morocco. ; 2021: 1–8.

16. Hamdan MA. Arabic text classification: A review. *Modern Applied Science* 2019; 13(5): 88–104.

17. Abdullah AM, Hasan RA, Ali AH, Mohammed MA. The classification of the modern Arabic poetry using machine learning. *Telkominika* 2019; 17(5): 2667–2674.

18. Elnahas A, Nour M, El-Fishawy NA, Tolba M. Machine learning and feature selection approaches for categorizing Arabic text: Analysis, comparison and proposal. *Egyptian Journal of Language Engineering* 2020; 7(2): 1–19.

19. Bahassine S, Madani A, Al-Sarem M, Kissi M. Feature selection using an improved Chi-square for Arabic text classification. *Journal of King Saud University Computer and Information Sciences* 2020; 32(2): 225–231.

20. Chowdhury SA, Abdelali A, Darwish K, Jung gS, Salminen J, Jansen BJ. Improving Arabic Text Categorization Using Transformer Training Diversification. In: Proceedings of the fifth Arabic Natural Language Processing Workshop. ; 2020: 226-236.

21. Hadi W, Al-Radaideh QA, Alhawari S. Integrating associative rule-based classification with Naïve Bayes for text classification. *Applied Soft Computing* 2018; 69: 344–356.

22. El-Alami FZ, Alaoui SOE. An efficient method based on deep learning approach for Arabic text categorization. In: Proceedings of the International Arab Conference on Information Technology (IACIT). ; 2016: 1–7.

23. Alwehaibi A, Roy K. Comparison of Pre-Trained Word Vectors for Arabic Text Classification Using Deep Learning Approach. In: 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018, Orlando, FL, USA. ; 2018: 1471–1474.

24. Almuzaini HA, Azmi AM. Impact of stemming and word embedding on deep learning-based arabic text categorization. *IEEE Access* 2020; 8: 127913–127928.

25. Bounhas M, Prade H. Analogy-based classifiers: An improved algorithm exploiting competent data pairs. *International Journal of Approximate Reasoning* 2023; 158: 108923.

26. Bounhas M, Prade H, Richard G. Analogy-based classifiers for nominal or numerical data. *International Journal of Approximate Reasoning* 2017; 91: 36–55.

27. Essid M, Bounhas M, Prade H. Continuous Analogical Proportions-Based Classifier. In: Proceedings of the 18th International Conference of Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU). Springer, CCIS 1237; 2020: 541–555.

28. Miclet L, Prade H. Handling Analogical Proportions in Classical Logic and Fuzzy Logics Settings. In: Proceedings of the 10th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU). Springer, LNCS 5590; 2009: 638–650.

29. Prade H, Richard G. From Analogical Proportion to Logical Proportions. *Logica Universalis* 2013; 7(4): 441–505.

30. Stroppa N, Yvon F. An Analogical Learner for Morphological Analysis. In: Proceedings of the Ninth Conference on Computational Natural Language Learning. ; 2005: 120–127.

31. Bounhas M, Elayeb B. Analogy-based Matching Model for Domain-specific Information Retrieval. In: Proceedings of the 11th International Conference on Agents and Artificial Intelligence (ICAART), Volume 2. ; 2019: 496–505.

32. Elayeb B, Chouigui A, Bounhas M, Khiroun OB. Automatic Arabic Text Summarization Using Analogical Proportions. *Cognitive Computation* 2020; 12(5): 1043–1069.

33. Bayoudh S, Miclet L, Delhay A. Learning by Analogy: A Classification Rule for Binary and Nominal Data. In: Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI). ; 2007: 678–683.

34. Miclet L, Bayoudh S, Delhay A. Analogical Dissimilarity: Definition, Algorithms and Two Experiments in Machine Learning. *Journal of Artificial Intelligence Research* 2008; 32: 793–824.

35. Prade H, Richard G, Yao B. Enforcing regularity by means of analogy-related proportions-A new approach to classification. *International Journal of Computer Information Systems and Industrial Management Applications* 2012; 4: 648–658.

36. Bounhas M, Prade H. An analogical interpolation method for enlarging a training dataset. In: Proceedings of the 13th International Conference on Scalable Uncertainty Management (SUM). Springer, LNCS 11940; 2019: 136–152.

37. Bounhas M, Pirlot M, Prade H, Sobrie O. Comparison of analogy-based methods for predicting preferences. In: Proceedings of the 13th International Conference on Scalable Uncertainty Management (SUM). Springer, LNCS 11940; 2019: 339–354.

38. Bounhas M, Prade H. Logical Proportions-Related Classification Methods Beyond Analogy. In: . 13562 of *Proceedings of the 15th International Conference of Scalable Uncertainty Management (SUM) 2022,*. Springer; 2022: 219–234.

39. Sander E. *L'analogie, du Naïf au Créatif*. Editions l'Harmattan . 2000.

40. Hüllermeier E. Towards Analogy-Based Explanations in Machine Learning. *CoRR* 2020; abs/2005.12800.

41. Lim S, Prade H, Richard G. Using Analogical Proportions for Explanations. In: . 13562 of *Proceedings of the 15th International Conference of Scalable Uncertainty Management (SUM), 2022,*. Springer; 2022: 309–325.

42. Hesse M. On defining analogy. *Proceedings of the Aristotelian Society* 1959; 60: 79–100.

43. Chouigui A, BenKhiroun O, Elayeb B. An Arabic multi-source news corpus: Experimenting on single-document extractive summarization. *Arabian Journal for Science and Engineering* 2021; 46(4): 3925–3938.

44. Hmeidi I, Al-Ayyoub M, Abdulla NA, Almodawar AA, Abooraig R, Mahyoub NA. Automatic Arabic text categorization: A comprehensive comparative study. *Journal of Information Science* 2015; 41(1): 114–124.

45. Elhassan R, Ahmed M. Arabic text classification review. *International Journal of Computer Science and Software Engineering* 2015; 4(1): 1–5.

46. Alabbas W, Al-Khateeb HM, Mansour A. Arabic text classification methods: Systematic literature review of primary studies. In: Proceedings of the 4th IEEE International Colloquium on Information Science and Technology (CiSt). ; 2016: 361–367.

47. Sbou AMFA. A survey of Arabic text classification models. *International Journal of Electrical and Computer Engineering* 2018; 8(6): 4352–4355.

48. Hassanein AMDE, Nour M. A proposed model of selecting features for classifying Arabic text. *Jordanian Journal of Computers and Information Technology* 2019; 5(3): 275–290.

49. Elghannam F. Text representation and classification based on bi-gram alphabet. *Journal of King Saud University Computer and Information Sciences* 2021; 33(2): 235–242.

50. Basabain S, Cambria E, Alomar K, Hussain A. Enhancing Arabic-text Feature Extraction Utilizing Label-semantic Augmentation in Few/Zero-shot Learning (In press). *Expert Systems*. doi: 10.1111/exsy.13329

51. Tan CC, Eswaran C. Performance Comparison of Three Types of Autoencoder Neural Networks. In: Second Asia International Conference on Modelling and Simulation, AMS 2008, Kuala Lumpur, Malaysia. ; 2008: 213–218.

52. AbuZeina D, Al-Anzi FS. Employing fisher discriminant analysis for Arabic text classification. *Computers and Electrical Engineering* 2018; 66: 474–486.

53. Kanan T, Hawashin B, Alzubi S, et al. Improving Arabic Text Classification Using P-Stemmer. https://doi.org/10.2174/2666255813999200904114023. *Recent Advances in Computer Science and Communications* 2020; 13(1).

54. Pasupa K, Ayutthaya TSN. Hybrid Deep Learning Models for Thai Sentiment Analysis. *Cognitive Computation* 2022; 14(1): 167–193.

55. Diwali A, Dashtipour K, Saeedi K, Gogate M, Cambria E, Hussain A. Arabic sentiment analysis using dependency-based rules and deep neural networks. *Appl. Soft Comput.* 2022; 127: 109377.

56. Abdullah M, Shaikh S. TeamUNCC at SemEval-2018 Task 1: Emotion Detection in English and Arabic Tweets using Deep Learning. In: Proceedings of The 12th International Workshop on Semantic Evaluation (SemEval@NAACL-HLT). ; 2018: 350–357.

57. Samy AE, El-Beltagy SR, Hassanien E. A Context Integrated Model for Multi-label Emotion Detection. In: Proceedings of the Fourth International Conference On Arabic Computational Linguistics (ACLING). ; 2018: 61–71.

58. Jabreel M, Moreno A. A Deep Learning-Based Approach for Multi-Label Emotion Classification in Tweets. *Applied Sciences* 2019; 9(6): 1123.

59. Al-Ayyoub M, Khamaiseh AA, Jararweh Y, Al-Kabi MN. A comprehensive survey of Arabic sentiment analysis. *Information Processing and Management* 2019; 56(2): 320–342.

60. Oueslati O, Cambria E, HajHmida MB, Ounelli H. A review of sentiment analysis research in Arabic language. *Future Generation Computer Systems* 2020; 112: 408–430.

61. Nassif AB, Elnagar A, Shahin I, Henno S. Deep learning for Arabic subjective sentiment analysis: Challenges and research opportunities. *Applied Soft Computing* 2020; 98: 106836.

62. Al-Ayyoub M, Nuseir A, Alsmearat K, Jararweh Y, Gupta B. Deep learning for Arabic NLP: A survey. *Journal of Computational Science* 2018; 26: 522–531.

63. Galal M, Madbouly MM, El-Zoghby A. Classifying Arabic text using deep learning. *Journal of Theoretical and Applied Information Technology* 2019; 97(23): 3412–3422.

64. Alhawarat M, Aseeri AO. A Superior Arabic Text Categorization Deep Model (SATCDM). *IEEE Access* 2020; 8: 24653–24661.

65. Chaturvedi I, Cambria E, Cavallari S, Welsch RE. Genetic Programming for Domain Adaptation in Product Reviews. In: ; 2020: 1-8

66. Al-Smadi M, Hammad MM, Al-Zboon SA, AL-Tawalbeh S, Cambria E. Gated recurrent unit with multilingual universal sentence encoder for Arabic aspect-based sentiment analysis. *Knowl. Based Syst.* 2023; 261: 107540.

67. Al-Radaideh QA, Al-Abrat MA. An Arabic text categorization approach using term weighting and multiple reducts. *Soft Computing* 2019; 23(14): 5849–5863.

68. Hawalah A. Semantic ontology-based approach to enhance Arabic text classification. *Big Data and Cognitive Computing* 2019; 3(53): 1–14.

69. Al-Taani AT, Al-Sayadi SH. Classification of Arabic Text Using Singular Value Decomposition and Fuzzy C-Means Algorithms. In: Johri P, Verma JK, Paul S., eds. *Applications of Machine Learning*Springer Singapore. 2020 (pp. 111–123).

70. Vulli A, Srinivasu P, Sashank M, Shafi J, Choi J, Ijaz M. Fine-Tuned DenseNet-169 for Breast Cancer Metastasis Prediction Using FastAI and 1-Cycle Policy. *Sensors* 2022; 22(8).

71. Kumar M, Kavita , Verma S, Kumar A, Ijaz M, Rawat D. ANAF-IoMT: A Novel Architectural Framework for IoMT-Enabled Smart Healthcare System by Enhancing Security Based on RECC-VC. *IEEE Transactions on Industrial Informatics* 2022; 18(12): 8936-8943.

72. Mandal M, Singh P, Ijaz M, Shafi J, Sarkar R. A Tri-Stage Wrapper-Filter Feature Selection Framework for Disease Classification. *Sensors* 2021; 21(16).

73. Juraev F, El-Sappagh S, Abdukhamidov E, Ali F, Abuhmed T. Multilayer dynamic ensemble model for intensive care unit mortality prediction of neonate patients. *Journal of Biomedical Informatics* 2022; 135: 104216.

74. El-Sappagh S, Ali F, Abuhmed T, Singh J, Alonso J. Automatic detection of Alzheimer's disease progression: An efficient information fusion approach with heterogeneous ensemble classifiers. *Neurocomputing* 2022; 512: 203-224.

75. Ali S, Abusabha O, Ali F, Imran M, Abuhmed T. Effective Multitask Deep Learning for IoT Malware Detection and Identification Using Behavioral Traffic Analysis. *IEEE Transactions on Network and Service Management* 2022: 1-1. doi: 10.1109/TNSM.2022.3200741

76. Bhattacharya D, Sharma D, Kim W, Ijaz M, Singh P. Ensem-HAR: An Ensemble Deep Learning Model for Smartphone Sensor-Based Human Activity Recognition for Measurement of Elderly Health Monitoring. *Biosensors* 2022; 12.

77. Bayoudh S, Miclet L, Delhay A, Mouchère H. De l'utilisation de la proportion analogique en apprentissage artificiel. In: Actes des Journées Inteligence Artificielle Fondamentale (IAF'07). ; 2007.

78. Prade H, Richard G. Homogeneous Logical Proportions: Their Uniqueness and Their Role in Similarity-Based Prediction. In: Proceedings of the 13th International Conference on Principles of Knowledge Representation and Reasoning. ; 2012: 402–412.

79. Prade H, Richard G. Reasoning with Logical Proportions. In: Principles of Knowledge Representation and Reasoning: Proceedings of the 12th International Conference. ; 2010: 545–555.

80. Bounhas M, Prade H, Richard G. Analogical classification: A new way to deal with examples. In: Proceedings of the 21st European Conference on Artificial Intelligence (ECAI). ; 2014: 135–140.

81. Bounhas M, Prade H, Richard G. Analogical Classification: Handling Numerical Data. In: Proceedings of the 8th International Conference on Scalable Uncertainty Management (SUM). Springer, LNCS 8720; 2014: 66–79.

82. Larkey LS, Ballesteros L, Connell ME. Light Stemming for Arabic Information Retrieval. In: Soudi A, Bosch AVD, Neumann G., eds. *Arabic Computational Morphology: Knowledge-based and Empirical Methods*Springer Netherlands. 2007 (pp. 221–243).

83. Fahandar MA, Hüllermeier E. Learning to rank based on analogical reasoning. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18). ; 2018: 2951–2958.

84. Saad M, Ashour W. OSAC: Open-Source Arabic Corpora. In: Proceedings of the 6th International Conference on Electrical and Computer Systems (EECS). ; 2010: 1–6.

85. Chouigui A, Khiroun OB, Elayeb B. ANT Corpus: An Arabic News Text Collection for Textual Classification. In: Proceedings of the 14th IEEE/ACS International Conference on Computer Systems and Applications (AICCSA). ; 2017: 135–142.

86. Chouigui A, Khiroun OB, Elayeb B. Related Terms Extraction from Arabic News Corpus Using Word Embedding. In: On the Move to Meaningful Internet Systems: OTM Workshops - Confederated International Workshops: EI2N, FBM, ICSP, and Meta4eS, Revised Selected Papers. Springer, LNCS 11231; 2018: 230–240.

87. Chouigui A, Khiroun OB, Elayeb B. A TF-IDF and Co-occurrence Based Approach for Events Extraction from Arabic News Corpus. In: Proceedings of the 23rd International Conference on Applications of Natural Language to Information Systems (NLDB). Springer, LNCS 10859; 2018: 272–280.

88. Hsu CW, Chang CC, Lin CJ. A practical guide to support vector classification. 2010.

89. Demsar J. Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research* 2006; 7: 1-30.

90. Afantenos SD, Kunze T, Lim S, Prade H, Richard G. Analogies Between Sentences: Theoretical Aspects - Preliminary Experiments. In: . 12897 of *Proceedings of the 16th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU 2021*,. Springer; 2021: 3–18.

91. Elayeb B, Bounhas M. Mono- and Cross-Language Information Retrieval based on Analogical Proportions: A Review. In: . 216 of *Proceedings of Sixth International Congress on Information and Communication Technology - ICICT 2021, London, Volume 3*. Springer; 2021: 629–653

92. Lavallée J, Langlais P. Moranapho: a multilingual system for morphological analysis based on formal analogy. *Traitement Automatique des Langues* 2011; 52(2): 17–44.

93. Langlais P, Yvon F. *Issues in Analogical Inference Over Sequences of Symbols: A Case Study on Proper Name Transliteration*. 548 of *Studies in Computational Intelligence*. : 59-82; Springer Berlin Heidelberg . 2014.

94. Lim S, Prade H, Richard G. Solving Word Analogies: A Machine Learning Perspective. In: Proceedings of the 15th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU). Springer, LNCS 11726; 2019: 238–250.

95. Moreau F, Claveau V, Sébillot P. Automatic Morphological Query Expansion Using Analogy-Based Machine Learning. In: Proceedings of the 29th European Conference on IR Research Advances in Information Retrieval (ECIR). Springer, LNCS 4425; 2007: 222–233.

96. Langlais P, Patry A. Translating Unknown Words by Analogical Learning. In: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL). ; 2007: 877–886.

97. Denoual E. Analogical translation of unknown words in a statistical machine translation framework. In: Proceedings of the Machine Translation Summit XI. ; 2007: 135–141.

How to cite this article: Bounhas M., Elayeb B., Chouigui A., Hussain A., and Cambria E. (2023), Arabic Text Classification based on Analogical Proportions

Table 1 Analogical proportions truth table

a	b	c	d	$a : b :: c : d$
0	0	0	0	1
1	1	1	1	1
0	0	1	1	1
1	1	0	0	1
0	1	0	1	1
1	0	1	0	1

Table 2 Solving analogical equation on classes

$class(\vec{a})$	$class(\vec{b})$	$class(\vec{c})$	x
t	t	t	$class(\vec{d}) = t$
t	t	v	$class(\vec{d}) = v$
t	v	t	$class(\vec{d}) = v$

Table 3 Example of Analogical inference for text classification

Doc.	t_1="*Goal*"	t_2="*Inflation*"	t_3="*Competition*"	t_4="*Winner*"	class(doc.)
Doc_1	1	1	0	0	Economy
Doc_2	1	0	0	1	Sport
Doc_3	1	1	1	0	Economy
Doc_x	1	0	1	1	?

Table 4 Valid analogical proportions used in case of AATC1 and AATC2

	d_1^j	d_2^j	d_3^j	d_x^j	$d_1^j : d_2^j :: d_3^j : d_x^j$	$Rel.for AATC1$	$Rel.for AATC2$
$P1$	1	1	1	1	1	$true$	$true$
$P2$	0	1	0	1	1	$true$	$true$
$P3$	0	0	1	1	1	$true$	$true$
$P4$	1	0	1	0	1	$false$	$true$

Table 5 Score computation for each triple t = (d_1, d_2, d_3) in AATC2

Keywords	d_1^j	d_2^j	d_3^j	d_x^j	$d_1^j : d_2^j :: d_3^j : d_x^j$	$d_1^j : d_2^j :: d_3^j : d_x^j \wedge d_x^j$	$d_1^j : d_2^j :: d_3^j : d_x^j \wedge \neg d_x^j$
w_1	1	1	1	1	1	1	0
w_2	1	0	1	0	1	0	1
w_3	0	0	1	1	1	1	0
w_4	0	1	0	1	1	1	0
w_5	1	1	1	1	1	1	0
w_6	0	0	1	1	1	1	0
w_7	1	1	0	0	1	0	1
w_8	1	0	0	1	0	0	0
w_9	0	0	0	1	0	0	0
w_{10}	0	1	1	0	0	0	0
p_t^+	-	-	-	-	-	**=5/10 = 0.5**	-
p_t^-	-	-	-	-	-	-	**=2/10=0.2**
pt	-	-	-	-	-	**=0.5-0.2=0.3**	

Table 6 Score computation for each triple t = (d_1, d_2, d_3) in AATC1

Keywords	d_1^j	d_2^j	d_3^j	d_x^j	$d_1^j : d_2^j :: d_3^j : d_x^j$
w_1	1	1	1	1	1
w_3	0	0	1	1	1
w_4	0	1	0	1	1
w_5	1	1	1	1	1
w_6	0	0	1	1	1
w_8	1	0	0	1	0
w_9	0	0	0	1	0
w_{11}	0	1	0	1	1
w_{12}	0	1	1	1	0
w_{13}	1	0	0	1	0
p_t	-	-	-	-	**=6/10 = 0.6**

Table 7 Computed scores for each class of the input document 1 using AATC1 and AATC2

Class	score(c_k) using AATC1	score(c_k) using AATC2
Economy	0.0019	0.04
InternationalNews	0.0017	0.04
MiddleEast	0.002	0.05
Sport	**0.0028**	**0.09**
Technology	0.0013	0.03

Table 8 Description of ANT corpus v2.1

Category	**AlArabiya**	**BBC**	**CNN**	**France24**	**SkyNews**	**Total**
Culture	606	338	0	126	0	**1070**
Economy	1071	281	463	78	1479	**3372**
International News	2030	1220	734	2793	3575	**10352**
MiddleEast	0	1131	1640	947	3972	**7690**
Sport	2443	385	483	297	2975	**6583**
Technology	369	460	187	0	1715	**2731**
Total	**6519**	**3815**	**3507**	**4241**	**13716**	**31798**

Table 9 Description of BBC-Arabic corpus

Category	**Number of documents**
Divers	122
Economy	296
International News	1489
Journal	49
MiddleEast	2356
Sport	219
Technology	232
Total	**4763**

Table 10 Results for each source and each classifier using ANT corpus v2.1

Classifier		AlArabiya	BBC	CNN	France24	SkyNews	Average
AATC1	Precision	**0.95**	**0.81**	**0.88**	0.12	**0.91**	0.73
	Recall	**0.93**	0.65	0.72	0.14	0.85	0.66
	F_1	**0.94**	0.69	0.75	0.13	**0.87**	0.68
	Accuracy	**95.31%**	73.10%	83.77%	46.85%	87.51%	77.31%
AATC2	Precision	0.93	0.79	0.86	0.39	0.86	**0.77**
	Recall	0.89	0.61	0.78	0.37	0.81	0.69
	F_1	0.91	0.66	0.80	0.37	0.83	0.71
	Accuracy	93.08%	69.47%	85.88%	**62.52**%	82.93%	**78.78**%
SVM	Precision	0.93	0.76	0.87	0.21	0.87	0.73
	Recall	0.92	0.75	**0.85**	0.20	**0.88**	0.72
	F_1	0.93	**0.75**	**0.86**	0.2	**0.87**	0.72
	Accuracy	94.85%	**76.99%**	**87.68%**	40.54%	**87.67%**	77.55%
NB	Precision	0.88	0.75	0.74	**0.49**	0.87	0.75
	Recall	0.92	**0.77**	0.78	**0.71**	0.86	**0.81**
	F_1	0.9	**0.75**	0.74	**0.55**	0.86	**0.76**
	Accuracy	92.19%	75.81%	78.33%	60.44%	85.48%	78.45%
J48	Precision	0.83	0.64	0.77	0.26	0.81	0.66
	Recall	0.81	0.63	0.73	0.25	0.81	0.65
	F_1	0.82	0.63	0.75	0.25	0.81	0.65
	Accuracy	86.67%	67.18%	80.24%	55.18%	82.05%	74.26
k-NN	Precision	0.57	0.44	0.68	0.18	0.81	0.54
	Recall	0.45	0.29	0.67	0.18	0.59	0.44
	F_1	0.43	0.26	0.63	0.18	0.62	0.42
	Accuracy	48.89%	36.54%	64.44%	40.39%	64.74%	51%

Table 11 Results sorted by category from all news sources of ANT corpus v2.1

Category		AATC1	AATC2	SVM	NB	J48	k-NN
Culture	Precision	0.56	**0.67**	0.56	0.59	0.46	0.13
	Recall	0.47	0.51	0.55	**0.83**	0.43	0.35
	F_1	0.51	0.57	0.55	**0.69**	0.44	0.18
Economy	Precision	**0.7**	0.67	0.65	0.68	0.54	0.63
	Recall	0.57	0.57	0.65	**0.77**	0.53	0.33
	F_1	0.6	0.59	0.65	**0.69**	0.54	0.41
International News	Precision	**0.8**	0.73	0.76	0.78	0.72	0.50
	Recall	0.75	**0.82**	0.79	0.74	0.76	0.67
	F_1	**0.77**	**0.77**	**0.77**	0.76	0.74	0.56
MiddleEast	Precision	0.57	0.68	0.68	**0.76**	0.68	0.64
	Recall	0.72	0.69	0.67	**0.76**	0.66	0.41
	F_1	0.64	0.67	0.67	**0.75**	0.67	0.49
Sport	Precision	0.8	**0.89**	0.81	**0.89**	0.79	0.66
	Recall	0.76	0.87	0.79	**0.89**	0.79	0.48
	F_1	0.78	0.88	0.80	**0.89**	0.79	0.53
Technology	Precision	**0.94**	0.93	0.86	0.70	0.69	0.46
	Recall	0.61	0.59	0.81	**0.85**	0.62	0.27
	F_1	0.71	0.71	**0.83**	0.75	0.65	0.20

Table 12 Results for the Wilcoxon matched-pairs signed-ranks test for AATC1 using ANT corpus v2.1

Metrics	AATC1 vs. SVM	AATC1 vs. NB	AATC1 vs. J48	AATC1 vs. k-NN
Precision	0.53	0.9	$0.01^{(+)}$	$0.02^{(+)}$
Recall	0.05	0.15	0.36	$0.0007^{(+)}$
F_1	$0.01^{(-)}$	0.39	0.08	$0.00001^{(+)}$

Table 13 Results for the Wilcoxon matched-pairs signed-ranks test for AATC2 using ANT corpus v2.1

Metrics	AATC2 vs. AATC1	AATC2 vs. SVM	AATC2 vs. NB	AATC2 vs. J48	AATC2 vs. k-NN
Precision	0.58	0.32	0.72	0.0004$^{(+)}$	0.001$^{(+)}$
Recall	0.9	0.12	0.08	0.03$^{(+)}$	0.0001$^{(+)}$
F_1	0.77	0.06	0.06	0.0002$^{(+)}$	0.00001$^{(+)}$

Table 14 Results sorted by category for the BBC-Arabic corpus

Category	Metrics	AATC1	AATC2	SVM	NB	J48	k-NN
Divers	Precision	0.95	0.71	0.85	**1**	0.70	0.91
	Recall	0.49	0.44	**0.76**	0.52	0.64	0.60
	F_1	0.65	0.54	**0.80**	0.68	0.67	0.73
Economy	Precision	0.93	0.91	0.83	0.73	0.61	**0.95**
	Recall	0.46	0.44	**0.87**	0.86	0.64	0.27
	F_1	0.61	0.59	**0.85**	0.79	0.63	0.42
International News	Precision	0.85	0.83	0.78	0.77	0.72	**0.89**
	Recall	0.71	0.47	**0.80**	0.67	0.69	0.41
	F_1	0.77	0.60	**0.79**	0.72	0.71	0.56
Journal	Precision	0.67	0.90	0.93	0.20	0.65	**1**
	Recall	0.77	0.44	0.78	**0.92**	0.61	0.35
	F_1	0.72	0.59	**0.84**	0.34	0.63	0.51
MiddleEast	Precision	0.76	0.67	**0.88**	0.85	0.81	0.85
	Recall	**0.97**	**0.97**	0.88	0.79	0.84	0.60
	F_1	0.85	0.79	**0.88**	0.82	0.82	0.70
Sport	Precision	**0.99**	0.98	0.97	0.98	0.90	0.37
	Recall	0.85	0.81	**0.92**	0.87	0.88	0.65
	F_1	0.91	0.89	**0.94**	0.92	0.89	0.47
Technology	Precision	0.99	**1**	0.77	0.47	0.65	0.10
	Recall	0.23	0.26	0.72	**0.90**	0.56	0.83
	F_1	0.37	0.41	**0.75**	0.62	0.60	0.18
Average	Precision	**0.88**	0.86	0.87	0.70	0.72	0.70
	Recall	0.63	0.56	**0.82**	0.81	0.69	0.55
	F_1	0.69	0.64	**0.84**	0.69	0.71	0.50
	Accuracy	80.52%	71.91%	**84.40%**	76.21%	76.29%	53.13%

Table 15 Results for the Wilcoxon matched-pairs signed-ranks test for AATC1 using BBC-Arabic corpus

Metrics	AATC1 vs. SVM	AATC1 vs. NB	AATC1 vs. J48	AATC1 vs. k-NN
Precision	0.73	0.08	0.04 $^{(+)}$	0.37
Recall	0.5	0.87	0.4	0.18
F_1	0.24	0.94	0.08	0.03 $^{(+)}$

Table 16 Results for the Wilcoxon matched-pairs signed-ranks test for AATC2 using BBC-Arabic corpus

Metrics	AATC2 vs. AATC1	AATC2 vs. SVM	AATC2 vs. NB	AATC2 vs. J48	AATC2 vs. k-NN
Precision	0.28	0.87	0.4	0.09	0.87
Recall	0.075	0.03 $^{(-)}$	0.06	0.04 $^{(-)}$	0.55
F_1	0.06	0.02 $^{(-)}$	0.24	0.03 $^{(-)}$	0.13

Table 17 Comparison of AATC with existing Arabic text classifiers

Arabic Text Classifier			ANT corpus v2.1	ANT corpus v1.1	AlKhaleej-2004	CNN-Arabic
1. Classic ML Arabic text classifiers						
Hassanein and Nour (2019) 48		Precision	-	-	0.83	-
		Recall	-	-	0.82	-
		F-1	-	-	0.82	-
		Accuracy	-	-	-	-
Al-Taani and Al-Sayadi (2020) 69		Precision	-	-	-	0.60
		Recall	-	-	-	0.63
		F-1	-	-	-	0.61
		Accuracy	-	-	-	-
Bahassine et al. (2020) 19		Precision	-	-	-	0.852
		Recall	-	-	-	0.851
		F-1	-	-	-	0.849
		Accuracy	-	-	-	-
2.DL Arabic text classifiers						
El-Alami and Alaoui (2016) 22		Precision	-	-	-	0.76
		Recall	-	-	-	0.74
		F-1	-	-	-	0.75
		Accuracy	-	-	-	75.10%
Al Muzaini et al. (2020) 24	$DL_*(NoStem)$	Weighted F-1	-	0.786	-	-
	$DL_*(Stem)$	Weighted F-1	-	0.801	-	-
	$DL^*(NoStem)$	Weighted F-1	-	0.815	-	-
	$DL^*(Stem)$	Precision	-	0.836	-	-
		Recall	-	0.838	-	-
		Weighted F-1	-	0.836	-	-
		Accuracy	-	-	-	-
3. Analogical proportions-based Arabic text classifiers						
AATC1		Precision	0.73	0.84	0.92	0.91
		Recall	0.66	0.81	0.80	0.80
		F-1	0.68	0.82	0.86	0.85
		Accuracy	77.31%	80.00%	85.06%	85.64%
AATC2		Precision	0.77	0.767	0.91	0.87
		Recall	0.69	0.764	0.78	0.82
		F-1	0.71	0.758	0.84	0.84
		Accuracy	78.78%	79.00%	79.81%	84.46%