Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 142 (2018) 262–269

Procedia Computer Science

www.elsevier.com/locate/procedia

The 4th International Conference on Arabic Computational Linguistics (ACLing 2018), November 17-19 2018, Dubai, United Arab Emirates

Automatic Arabic Dialect Classification Using Deep Learning Models

Leena Lulu[a] and Ashraf Elnagar[b*]

[a]College of IT, Ajman University, Ajman, UAE
[b]Department of Computer Science, University of Sharjah, Sharjah, UAE

Abstract

Recently, the vast use of social media and the high availability of internet access have produced a considerably different textual data from the formal and standard data on the Web. This includes various Arabic dialectal languages, which are the native spoken languages of Arabic speakers. The presence of textual Arabic dialectal languages on the Web has brought many new opportunities as well as challenges for machine learning and Arabic language processing. The identification of this type of informal data has its crucial effect on several applications such as sentiment analysis and machine translation. However, the standard NLP tools developed for traditional data fall short due to nature of dialectal textual data. Deep learning tools have proven to be very effective in processing social Media dialectal text. In this paper, we consider a variety of deep learning models for the automatic classification of Arabic dialectal text. We use a free large manually-annotated dataset known as Arabic Online Commentary (AOC), which includes several Dialectal Arabic (DA) along with the Modern Standard Arabic (MSA), [3]. We consider the most frequent dialects in the dataset. Namely, the Egyptian (EGP), Levantine (LEV), and Gulf –including Iraqi - (GLF). Four different deep neural network models have been implemented to examine the Arabic dialectal classification problem for each pair of the 3 dialects (binary classification experiments) as well as one ternary-classification experiment including all dialects together. The results show a varying but promising performance of the models for each pair of dialects. Furthermore, a closer examination on the manually-annotated AOC dataset has been carried out and hence, we conclude that there is a serious demand for a thorough refinement and review of the AOC annotated sentences as it is an important benchmark dataset in the field.

* Corresponding author. Tel.: +971 6 5050402; fax: +971 6 5050102.
E-mail address: ashraf@sharjah.ac.ae

1877-0509
10.1016/j.procs.2018.10.489

Keywords: deep learning models, classification, Arabic dialects, AOC dataset.

1. Introduction

The nature of the Arabic language being a macro, morphological language with many varieties makes it one of the challenging languages for natural language processing and language identification. It has been classified by ISO Registration Authority as a language with 30 varieties including its formal Modern Standard Arabic (MSA) [1]. The Arabic native spoken languages that are mainly classified based on the geographical locations and known as the Arabic dialects form most of the varieties of the Arabic language. For example, some of the known dialects are Egyptian, Levantine, Gulf, Iraqi, Yemeni, Moroccan, etc. It is worth mentioning that the Classical Arabic is one of the known varieties of the Arabic language as the historical Arabic language from which the Modern Standard Arabic has stemmed from. The MSA is the only standardized and structured form of the Arabic language that is used formally in written communications, whereas the Arabic dialects considerably deviate from the formal MSA in terms of phonology, morphology, and syntax and are used in daily spoken communications [2]. However, a dialectal Arabic text can be presented in written form by following the spelling rules similar to MSA since they are typically phonetic.

It is worth noting that the Arabic dialects have become increasingly apparent as the language of informal communication on the web in blogs, forums, Social Media networks, etc. The enormous use of dialectal language presents further challenges to the language identification problem as the performance of the traditional NLP techniques are seriously degrading since they are mainly developed for traditional MSA language. Using the same Arabic characters in forming dialectal Arabic text, along with the non-standard spelling, the poor quality, the common vocabulary among the different varieties poses more complexities to the identification problem of the language dialectal varieties. Needless to say that identifying dialectal Arabic varieties manually by human annotators is a confusing task as will be elaborated later in this paper. Nonetheless, dialect identification has gained importance as it becomes essential to many natural language processing applications such as machine translation, sentiment analysis, and information retrieval from online contents.

The rest of the paper is organized as follows: Section 2 presents the related work. The dialect Arabic classification is reviewed in Section 3. Experimental results are presented in Section 4. Finally, we conclude the work in Section 5.

2. Literature Review

Arabic dialects identification is increasingly emerging as a new research area in Machine Learning and Natural Language Processing. This has been triggered by the substantial use of dialectal Arabic text (DA) as the mean of informal communication on the web. This new situation has constituted a new form of textual data on the web that is inseparably mixed with the standard MSA. Research work mainly addressed building dialectal corpora [3-4], in addition to language classification between MSA and the varieties of DA [5-8].

The early research by Habash et al. [5] described general strategies for recognizing DA text in MSA text using only 1.6K sentences. They produced annotated text at the sentence and word levels. The percentage of DA in MSA has been measured by Diab et al. [9].

Further research on DA has been significantly enriched by the construction of the public Arabic Online Commentary dataset (AOC) developed by Zaidan and Callison-Burch [3]. The dataset is collected from online Arabic forums. The annotation of the AOC records has been carried out via crowdsourcing to classify each record as one of 6 labels distinguishing the following categories: MSA, Egyptian, Gulf, Iraq, Levantine, and Maghrebi.

For instance, the research in [6] has utilized a small subset of AOC to identify whether a given text is either MSA or DA (i.e., Egyptian dialect – EGP) using a supervised SVM classifier that achieved an accuracy of 85.5%.

Another research in [10] used the same subset of AOC with more features to boost the performance of the classifier while producing a binary output (i.e., MSA or EGP). Using lexical and morphological features when compared to n-grams, [8] produced better results of their binary classifier but on twitter data. The work in [11] also utilized AOC along with Facebook posts to address the problem of multi-classification of sentences into 4 main categories. Namely EGP, GLF, LEV, and MSA. The proposed n-gram based classifier obtained an accuracy of 87.8% on AOC test-dataset.

The developers of AOC proposed an automated 4-class classifier in [4] using the same dataset, which attained an accuracy of 81%. Different classifiers based on bigrams with Naïve Bayes have been proposed by Sadat et al. [12] to characterize 18 different Arabic dialects with an overall F1-measure of around 76%. A comparable- work is proposed in [13] to classify 4 dialects, which are Levantine, Egyptian, Saudi and Iraqi. The dataset consists of 2 parts. The first one is a small corpus of manually annotated instances and the second one is a large collection crawled from the of web using word-marks. The classifier achieved F1-measure of 91%. With the exception of AOC, none released their data.

Our investigation is carried out using AOC dataset only to categorize a given sentence as one of 3 main dialects: EGP, GLF, and LEV, while utilizing deep-learning based classifiers. Early research on these dialects in addition to MSA for sentiment analysis purposes is reported in [14-16,26].

3. Automatic Dialect Classification

3.1. Dataset

We use the annotated Arabic Online Commentary (AOC) dataset, which is made of almost 110K labeled sentences. The dialect labels were Egyptian, Gulf, Iraqi, Levantine, Maghrebi, MSA, and some other categories for multiple dialects or others. The labels were assigned through crowdsourcing efforts. For this work, we limit our work to the 3 well represented dialects, which are Egyptian (EGP), Gulf -including Iraqi- (GLF), and Levantine (LEV). In fact, we did pre-process the dataset to identify a dominant dialect in case of tie or missing class. We dropped the Maghrebi dialect instances as it is very small (less than 250) compared to the selected ones. Furthermore, we combined Gulf and Iraqi dialects together as it is relatively close and belong to almost the same geographical region. In order to work with a balanced dataset, we limited the training dataset to 30K reviews (10K per dialect) and the testing dataset to 3K records. We use 10% of the training dataset for cross-validation. We implemented 3 binary classifiers to investigate the categorization of each pair of the 3 dialects as well as a ternary classifier for all of them.

It is worth noting that the overall, human annotators have a classification accuracy of 90.3%, with dialect recall at 89.0%, and MSA recall at 91.5%. Such recall rates do vary across annotators causing some accuracy rates to drop as low as 80% or 75%, [4].

3.2. Challenges

Using the character set, automatic classification of some languages is straightforward. Systems that identify languages of almost the same character also exist with high accuracy based on character-histograms. See, [17] for example. Recent works on language detection include [18-19] and geographical location [20] as an example. Classification of DA is different and rather harder to treat. This is because of the high similarity of lexical features of some Arabic dialects. Therefore, the histogram method is not practical. In fact, identifying dialects by region rather than by country is meant to elevate this commonality among them.

3.3. Classifiers

In line with recent research on the use of deep neural networks (DNNs) for NLP problems on the English language and their promising results, we explore the use of such models for identifying DA. We chose deep learning because of their proven high performance without much need for features engineering. Existing DNN models stem from 2 main architectures, which are hierarchical (convolutional neural network (CNN), [21] and sequential (recurrent neural network (RNN), [22]. Long short-term memory (LSTM), [23], is an enhanced version of RNN. As there is no clear agreement on which architecture is more suited for text classification problem, we implement both models in addition to 2 variations of both architectures. Namely,

1. long-short term memory (LSTM)
2. convolutional neural networks (CNN)
3. bidirectional LSTM (BLSTM)
4. convolutional LSTM (CLSTM)

- Long-Short Term Memory (LSTM)

We adopt a word-based LSTM, which models the word sequence x as follows:

$$
\begin{aligned}
i_t &= \sigma(x_t U^i + h_{t-1} W^i + b_i) \\
f_t &= \sigma(x_t U^f + h_{t-1} W^f + b_f) \\
o_t &= \sigma(x_t U^o + h_{t-1} W^o + b_o) \\
q_t &= \tanh(x_t U^q + h_{t-1} W^q + b_q) \\
p_t &= f_t \cdot p_{t-1} + i_t \cdot q_t \\
h_t &= o_t \cdot \tanh(p_t)
\end{aligned}
$$

The input sequence x is represented as a feature map of dimensionality $d \times n$, where n is the number of tokens and d is a dense vector. A specific token at position t is expressed by x_t. h_t is a hidden state at t computed based on the history $x_1, \dots, x_t$. U and W are parameters of dimensionality $d \times h$ and $h \times h$, respectively. The 3 main gates of LSTM are input gate, i_t, forget gate, f_t, and output gate, o_t. All gates are computed by a sigmoid function of the input x_t, preceding hidden state h_{t-1}, bias d-dimensional b, and parameters U and W. An updated history p_t computation depends on the input and forget gates, previous history step as well as a temporary computation, q_t, which is a hyperbolic tangent function of input x_t and h_{t-1}. At last, the output gate o_t and the updated history p_t yield the final hidden state, h_t.

- Convolutional Neural Network (CNN)

We adopt a common architecture that is based on [24] which has several layers among which are the core input and convolution layers. The input layer is like the input of feature map explained above. The input is initialized randomly.

The convolution Layer is designed to slide windows of w-grams. Given a sequence $x_1, \dots, x_n$ and filter size w, we compute the vector c_i (of dimension $w \times d$) as the concatenated embeddings of w tokens: $x_{\mathrm{i-w+1}}, \dots, x_i$, where $0 < i < s + w$. Using the convolution weights W and bias b, we compute:

$$p_i = tanh(W \cdot c_i + b)$$

Maxpooling is applied over the sequence output of p_i (i = 1 · · · s + w − 1) to produce the representation of input sequence x:

$$x_j = max(p_{1,j}, p_{2,j}, \dots), (j = 1, \dots, d)$$

The fully connected dense layer uses a ReLU non-linear activation function.

- Bidirectional LSTM (BLSTM)

BLSTMs are an extension of traditional LSTMs to improve performance on sequence classification problems. This architecture trains two LSTM models on the input sequence. While the first one treats the input sequence, the second LSTM works on an inverted copy of the input. This setup is expected to lead to faster and complete learning process.

- Convolution LSTM (CLSTM)

The CLSTM is proposed [25] as end-to-end trainable model. It is a variant of the CNN model in which the dense layer is replaced by an LSTM layer. The convolution output is fed to the LSTM layer and, hence, capture the advantages of each model.

4. Experimental Results

4.1. Setup and Pre-processing

Our objective is to explore the success of using DNN models to classify Arabic dialects. As mentioned earlier, we chose the most frequent dialects in AOC, which are EGP, GLF, and LEV. Therefore, we conducted 4 experiments on a balanced subset of AOC that comprises of 33K sentences. The first 3 experiments involve binary classification for every pair of the 3 dialects. The fourth experiment is designed to perform ternary classification for all 3 dialects. The implementation is in-line with existing works performed on English language and therefore we use similar set of hyper-parameters. Optimizing such parameters is beyond the scope of this work for the time being. We split the dataset into 80% for training, 10% for cross-validation, and 10% for testing. We report the accuracy on testing dataset for each of the 4 implemented DNN models. It should be noted that embeddings are initialized at random for the input layer. We chose Tensorflow framework for the implementation.

Table 1 Accuracy results for 3 pairs of dialects.

DA pair model	LSTM		CNN		BLSTM		CLSTM	
	CV	Test	CV	Test	CV	Test	CV	Test
EGP-GLF	97.1%	80.4%	96.8%	81.0%	93.3%	**83.8%**	89.6%	79.3%
EGP-LEV	91.3%	**81.3%**	96.0%	78.1%	87.7%	79.6%	90.8%	79.0%
LEV-GLF	91.7%	**77.2%**	96.8%	75.1%	95.3%	74.5%	89.5%	73.0%

Text pre-processing is a key pre-requisite to clean the dataset and, hopefully, improve the final results. Therefore, we first filter out non-Arabic content. This is particularly important when dealing with data collected from social media or the web in general. Although Arabic character set is somehow unique, it is easy to eliminate non-Arabic characters. However, some languages share a good deal of the character set such as Persian and Urdu.

The challenge was to detect Arabic text only and filter out anything else. Python provides a library (i.e.,) to serve this objective. Regular expressions are used to filter other character sets.

We further eliminate all diacritics, elongation (i.e., “جميــــل” is reduced to “جميل”), punctuation marks, extra spaces, etc. Another widely adopted practice is to apply normalization on Arabic characters. This involves replacing the letters “إ”, “أ” and “آ” with “ا”, letter “ة” with “ه”, and letter “ي” is replaced with “ى”. We didn’t implement normalization as we believe it can affect the contextual meaning for some words such as “فأر” and “فار” or “كرة” and “كره”.

4.2. Evaluation and Discussion

We report the accuracy results of the first 3 experiments on binary classification in **Table 1** for each of the 4 DNN models. All implementations beat the baseline. The table shows that BLSTM reported the best classification result for the pair EGP-GLF whereas the pairs EGP-LEV and LEV-GLF, LSTM model outperformed the rest.

Table 2 summarizes the accuracies of the ternary-classifiers on all 3 dialects. As expected, the accuracy is less than what we obtained in **Table 1**. LSTM attained the highest accuracy with a score of 71.4% followed by CLSTM with 71.1% and BLSTM with 70.9%. We observed that the CNN model suffers from overfitting problem as shown by the difference between the cross-validation and the test results. **Fig. 1** shows the accuracy results as well.

Table 2 Accuracy results for all 3 dialects together.

Model	CV	Test
LSTM	84.5%	**71.4%**
CNN	96.0%	68.0%
BLSTM	84.4%	70.9%
CLSTM	90.4%	71.1%

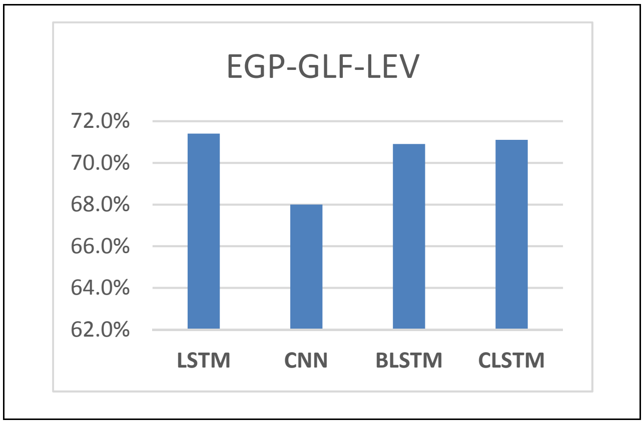

Fig. 1 Accuracy of ternary-classifiers on all 3 DA pairs.

Fig 2 depicts the confusion matrix of the true dialects and the predicted ones of the LSTM model. We observe that miss-classification rate of 25% are produced by predicting GLF dialect as LEV. Likewise, the miss-classification percentage of the LEV dialect as GLF is another source of errors with 15%. Besides, 17% classification error is reported for EGP dialect as LEV. This high confusion rates suggests that the dialect-pair share good amount of vocabulary, which is acceptable for EGP-LEV compared to LEV-GLF. This observation motivated us to re-examine the AOC dataset, which resulted in the following findings:

- A good majority of the annotators who speak LEV reside in the Gulf region. As a result, they could mix both dialects and therefore consider one for the other.
- A good percentage of annotations were wrongly classified or the sentences is acceptable in more than one dialect. We show a sample of such annotations in **Table 3**, which includes the true dialect class, the predicted dialect class, expert opinion, expert confidence, and the sentence. We selected 10 sentences for each pair. We asked 12 experts who are familiar with the 3 dialects to classify the sentences in the table. The experts' classification results were collected and then a decision is made on each sentence by simple voting. The confidence level reflects the number of votes for the selected class. Some experts voted for both dialects whenever it is felt that the sentence belongs to both dialects. We believe that a confidence level between 50 and 60 may be interpreted as neutral and therefore the sentence can be classified to either dialect. The 12 experts' decision agrees with the annotated classification of the 10 sentences for the dialects pair LEV-GLF with success rate of 60%. This means the remaining 40% indicate miss-classification in the existing annotations of the AOC dataset. As for the pairs EGP-LEV and EGP-GLF, the experts' decision came in accordance with the predicted dialects reported by the LSTM classifier with agreement rates of 90% and 80%, respectively (See **Table 3**). Of course, there are sentences that are clearly miss-classified by the LSTM classifier, which we did not include in the **Table 3**.

In conclusion, we strongly believe that the AOC dataset requires a thorough revision to minimize the problem mentioned above whenever possible. Although AOC is an important good benchmark dataset in the field of DA, inconsistent data would negatively affect the performance of classifiers.

Fig. 2 Heatmap of true versus predicted dialects.

Table 3 Samples of actual annotated dialects vs. predicted ones. Red-coloured words are vocabulary that belongs to the predicted dialect.

DA Annotation	DA Prediction.	Experts Decision	Conf. vote	Sentence	#
GLF	LEV	GLF	75.0%	هاي صارت موضة الهوش والتهجم على مراكز الأمن!	1
GLF	LEV	**LEV**	61.5%	رخصوا البنزين ولا ما رخصوا احنا مش سائلين	2
GLF	LEV	**LEV**	69.2%	اذا الشاب بيكون على علاقة ببنت وبيروح وبيطلبها والاهل بيرفضوا ...	3
GLF	LEV	**LEV**	80.0%	قال ايش ؟؟	4

GLF	LEV	GLF	81.8%	**ماتن** من البرد !!	5
GLF	LEV	GLF	61.1%	ان شاء الله نشوف ابداعاتكم	6
GLF	LEV	GLF	61.1%	الله يعين بس	7
GLF	LEV	**LEV**	52.4%	الله يهدي الجميع	8
GLF	LEV	GLF	52.9%	الله يرحمهم جميع ويتجاوز عنهم.	9
GLF	LEV	GLF	61.1%	عيني عليك باردة والى الامام	10
EGP	LEV	**LEV**	100.0%	**مهابيل** وين الروح الرياضية	1
EGP	LEV	**LEV**	76.9%	اذا **ما كبرت ما تصغر** واليهود **مش** جايينها لبر ومش فارقة معهم عرب ولا **...**	2
EGP	LEV	**LEV**	58.3%	سلم **سلامات** درع الاتحاد للوحدات	3
EGP	LEV	**LEV**	57.1%	والله **مش** فاهم هالناس اللي قاعدين **بتفلسفو** وبقولو **شو هالفتاوي ومش** عارفين**...**	4
EGP	LEV	**LEV**	69.2%	الحلو حلو و**المخيس مخيس** وخلاص	5
EGP	LEV	EGP	60.0%	يا ريت كل منتخباتنا زي منتخب السلة.	6
EGP	LEV	**LEV**	73.3%	أخي في الله الحبيب انا بتكلم عن السياسة وهذا	7
EGP	LEV	**LEV**	68.8%	صياد رحت أصطاد صادونى	8
EGP	LEV	**LEV**	50.0%	وسهره حلووووه..	9
EGP	LEV	**LEV**	62.5%	كله **لجلط** المستهلك بأسرع وقت.	10
EGP	GLF	**GLF**	100.0%	**تبي** فلوووس ماطلبتها الا في هالوووقت عشان عارف بيعطوووك قدامهم الازرق **...**	1
EGP	GLF	**GLF**	58.3%	قسم القصه **ذي** صارت معي اتصلت على واحد من معارفي في نفس المدينة لما **...**	2
EGP	GLF	**GLF**	90.9%	**تحس** ان الطريقة بدائية وين تقنية المعلومات والا **هذي** بس وقت المخالفات !!!	3
EGP	GLF	**GLF**	91.7%	كذب المنجمون ولو صدقوا وهذا كله بعلم الغيب **محد** يدري	4
EGP	GLF	**GLF**	100.0%	خلاااص **فكونا تراكم ذبحتونا!!!**	5
EGP	GLF	**GLF**	70.6%	راحو الطيبين	6
EGP	GLF	EGP	52.9%	خساره على العقول والناس اللي بتفهم	7
EGP	GLF	EGP	68.8%	يا ساتر استر	8
EGP	GLF	**GLF**	58.8%	قنوات المجد برسوم وهي تستاهل	9
EGP	GLF	**GLF**	57.9%	تستاهل التكريم	10

5. Conclusions

The work describes a thorough investigation on dialect identification of Arabic language using deep neural network models on the Arabic Online Commentary benchmark dataset. We considered classifying 3 most frequent Arabic dialects. Namely, Egyptian (EGP), Levantine (LEV), and Gulf –including Iraqi (GLF). The resulting subset consists of 33K sentences that are almost distributed equally among the 3 dialects. We dropped the modern standard Arabic from our comparison, which is the bulk of the AOC dataset. We implement four deep neural network models. Namely, CNN, LSTM, Bi-directions LSTM, and Convolution LSTM for 2 tasks. The first is to perform binary classification for each pair of the dialects and the second to conduct a classification on all 3 dialects together. While BLSTM outperformed all models for the EGP-GLF pair, LSTM produced the best accuracies for the remaining pairs as well as the ternary classification task.

We further performed another experiment in which we invited 12 experts to manually annotate a small sample of the miss-classified sentences. The outcome of this investigation supports most of the findings of the LSTM model, which conflicted with the AOC annotations. As a result, we strongly believe that a further thorough examining and refinement of the AOC annotated sentences is required as it is an important benchmark dataset in the field. In the near future, we will embark on this study.

References

[1] ISO 639-3 Registration Authority, https://iso639-3.sil.org/, 2017.

[2] Biadsy F., Hirschberg J., and Habash N. 2009. Spoken arabic dialect identification using phonotactic modeling. In Proceedings of the Workshop on Computational Approaches to Semitic Languages at the meeting of the European Association for Computational Linguistics (EACL), Athens, Greece.

[3] Omar F Zaidan and Chris Callison-Burch. 2011. The arabic online commentary dataset: an annotated dataset of informal arabic with high dialectal content. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2, pages 37–41. Association for Computational Linguistics.

[4] Omar F Zaidan and Chris Callison-Burch. 2014. Arabic dialect identification. Computational Linguistics, 40(1):171–202.

[5] Habash N., Rambow O., Diab M., Kanjawi-Faraj R. (2008), Guidelines for annotation of Arabic dialectness, Proceedings of the LREC Workshop on HLT & NLP within the Arabic World, Marrakech, p.p. 49–53.

[6] Elfardy H., Diab M. (2013), Sentence level dialect identification in Arabic, Proceedings of the 51st Annual Meeting of the ACL, Sofia, Bulgaria, pp. 456–461.

[7] Ryan Cotterell and Chris Callison-Burch. 2014. A multi-dialect, multi-genre corpus of informal written arabic. In LREC, pages 241–245.

[8] Kareem Darwish, Hassan Sajjad, and Hamdy Mubarak. 2014. Verifiably effective arabic dialect identification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1465–1468.

[9] Diab M., Habash N., Rambow O., Altantawy M., Benajiba Y. (2010), COLABA: Arabic dialect annotation and processing, In LREC Workshop on Semitic Language Processing, Valleta, Malta, pp. 66–74.

[10] Christoph Tillmann, Saab Mansour, and Yaser Al-Onaizan. 2014. Improved sentence-level arabic dialect clas- sification. In Proceedings of the First Workshop on Applying NLP Tools to Similar Languages, Varieties and Dialects, pages 110–119.

[11] Fei Huang. 2015. Improved arabic dialect classification with social media data. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 2118–2126.

[12] 9. Sadat F., Kazemi F., Farzindar A. (2014), Automatic identification of Arabic language varieties and dialect in social media, Proceedings of the Second Workshop on Natural Language Processing for Social Media (SocialNLP) ,Dublin, Ireland, pp. 22–27.

[13] Durandin O. V. , Strebkov D. Y., Automatic Arabic Dialect Classification, Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2016", Moscow, June 1–4, 2016

[14] Elnagar, A. (2016). Investigation on Sentiment Analysis of Arabic Book Reviews. . 13th ACS/IEEE International Conference on Computer Systems and Applications AICCSA 2016.

[15] Elnagar, A. ; and Einea, O. (2016). BRAD: Book Reviews in Arabic Dataset. International Workshop for Data Science and Computing (DSC'2016); 13th ACS/IEEE International Conference on Computer Systems and Applications AICCSA 2016. Agadir: IEEE.

[16] Elnagar, A. ; Khalifa, Y; and Einea, A. (2017). Hotel Arabic-Reviews Dataset Construction for Sentiment Analysis Applications. In Intelligent Natural Language Processing: Trends and Applications, Springer International Publishing, pages:35—52.

[17] Souter, Clive & Churcher, G & Hayes, J & Hughes, J & Johnson, S. (1994). Natural language identification using corpus-based models. Hermes J. Linguist.. 13.

[18] David Jurgens, Yulia Tsvetkov, and Dan Jurafsky. 2017. Incorporating dialectal variability for socially equi- table language identification. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), volume 2, pages 51–57.

[19] Tommi Jauhiainen, Marco Lui, Marcos Zampieri, Timothy Baldwin, and Krister Linde´n. 2018. Automatic language identification in texts: A survey. arXiv preprint arXiv:1804.08186.

[20] Afshin Rahimi, Trevor Cohn, and Tim Baldwin. 2018. Semi-supervised user geolocation via graph convolutional networks. arXiv preprint arXiv:1804.08049.

[21] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, November 1998.

[22] Elman, J. L. Finding structure in time. Cognitive Science, 14(2), 179–211, 1990.

[23] Sepp Hochreiter and Ju¨rgen Schmidhuber. 1997. Long short-term memory. Neural computation, 9(8):1735–1780.

[24] Yoon Kim. 2014. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

[25] Chunting Zhou, Chonglin Sun, Zhiyuan Liu, and Francis Lau. 2015. A c-lstm neural network for text classifica- tion. arXiv preprint arXiv:1511.08630.

[26] Elnagar, A. ; Lulu, L; and Einea, O. (2018). An Annotated Huge Dataset for Standard and Colloquial Arabic Reviews for Subjective Sentiment Analysis. Accepted in Fourth International Conference On Arabic Computational Linguistics, Procedia Computer Science, 2018.