Overview of the Track on Author Profiling and Deception Detection in Arabic*

Francisco Rangel[1], Paolo Rosso[1], Anis Charfi[2], Wajdi Zaghouani[3], Bilal Ghanem[1], and Javier Sánchez-Junquera[1]

[1] PRHLT Research Center, Universitat Politècnica de València, Spain
[2] Carnegie Mellon University, Qatar
[3] Hamad Bin Khalifa University, Qatar

Abstract. This overview presents the Author Profiling and Deception Detection in Arabic (APDA) shared task at PAN@FIRE 2019. Two have been the main aims of this years task: i) to profile the age, gender and native language of a Twitter user; ii) to determine whether an Arabic text is deceptive or not in two different genres: Twitter and news headlines. For this purpose we have created three corpora in Arabic. Altogether, the approaches of 13 participants are evaluated.

Keywords: author profiling · deception detection · Arabic · Twitter · FIRE.

1 Introduction

PAN[4] lab is a series of scientific events and shared tasks on digital text forensics. This year at FIRE[5] we have organised the Author Profiling and Deception Detection in Arabic (APDA)[6] shared task. In this paper, we describe the resources that we have created and made available to the research community[7], illustrating the obtained results and highlighting the main achievements. The Author Profiling and Deception Detection in Arabic consists of two tasks. In the next section we will describe each of them.

1.1 Task 1. Author Profiling in Arabic Tweets

Author profiling distinguishes between classes of authors studying how language is shared by people. This helps in identifying profiling aspects such as age, gender, and language variety, among others. The focus of this task is to identify the age, gender, and language variety of Arabic Twitter users.

[4] http://pan.webis.de/
[5] http://fire.irsi.res.in/fire/2019
[6] https://www.autoritas.net/APDA/
[7] Following a methodology that accomplishes with the EU General Data Protection Regulation [24].

1.2 Task 2. Deception Detection in Arabic Texts

We can consider that a message is deceptive when it is intentionally written trying to sound authentic. The focus of the task is on deception detection in Arabic on two different genres: Twitter and news headlines.

The reminder of this paper is organised as follows. Section 2 covers the state of the art, Section 3 describes the corpora and the evaluation measures, and Section 4 presents the approaches submitted by the participants. Section 5 and 6 discuss results and draw conclusions respectively.

2 Related Work

In this section we briefly review the related work on author profiling (age, gender and language variety identification) and deception detection in Arabic.

2.1 Author Profiling

The investigation in age and gender identification in Arabic is scarce. The authors of [14] collect 8,028 emails from 1,030 native speakers of Egyptian Arabic. They propose 518 features and test several machine learning algorithms, and report accuracies between 72.10% and 81.15% respectively for gender and age identification. The authors of [4] approach the gender identification in well-known Arabic newsletters articles written in Modern Standard Arabic. With a combination of bag-of-words, sentiments and emotions, they report an accuracy of 86.4%. Subsequently, the authors of [3] extend their work by experimenting with different machine learning algorithms, data-subsets and feature selection methods, reporting accuracies up to 94%. The authors of [6] manually annotate tweets from Jordanian dialects with gender information. They show how the name of the author of the tweet can significantly improve the performance. They also experiment with other stylistic features such as the number of words per tweet or the average word length, achieving a best result of 99.50%.

The increasing interest in Arabic varieties identification is supported by the eighteen and six teams participating respectively in the Arabic subtask of the third [18] DSL track, the Arabic Dialect Identification (ADI) shared task [41], as well as the twenty teams participating in the Arabic subtask of the Author Profling shared task [27] at PAN 2017. However, as the authors of [29] highlight, there is still a lack of resources and investigations in that language. Some of the few works are the following ones. The authors of [38] use a smoothed word unigram model and report respectively 87.2%, 83.3% and 87.9% of accuracies for Levantine, Gulf and Egyptian varieties. The authors of [32] achieve 98% of accuracy discriminating among Egyptian, Iraqi, Gulf, Maghreb, Levantine, and Sudan with n-grams. The authors of [12] combine content and style-based features to obtain 85.5% of accuracy discriminating between Egyptian and Modern Standard Arabic.

2.2 Deception Detection

Despite the fact that deception detection research in Arabic is still very limited [29], there are some new initiatives focusing on this language. For example, in the context of fact check shared task[8] at CLEF[9] on automatic identification and verification of claims in political debates [21]. Nevertheless, the aforementioned shared task translate the contents from English to Arabic. Since the claims correspond to US politics, they are not representative of the idiosyncrasy of Arabs. In this sense, the CheckThat! shared task[10] on Automatic Identification and Verification of Claims [16] organised at CLEF 2019 includes a subtask only in Arabic. The authors of [2] collect a corpus in Arabic from 600 tweets and 179 news articles. They automatically annotate the credibility by measuring the cosine similarity between the tweets and the news articles. The authors of [7] complain about the automatic generation of the annotation and they collect and manually annotate two corpora from Twitter and Blogs. Regarding Twitter, they retrieve over 36 million tweets about four topics: *i)* The forces of the Syrian government; *ii)* Syrian revolution; *iii)* Syrian problems and concerns related to the Syrian revolution; and *iv)* The election of the Lebanese president. The annotation process is carried out by five annotators. According to the authors of [37] the obtained inter-annotator agreement (Fleiss' kappa 0.43) is moderate. The authors also propose a method to approach the credibility analysis of Twitter contents. The Credibility Analysis of Arabic Content on Twitter (CAT) [11] relies mainly on features obtained from the user who tweeted the content to be analysed. For example, the authors retrieve the user's timeline and extract features such as the number of retweets, the user's activity, or the user's expertise in the topic being discussed. They compare their approach with several baselines and show a significant improvement. In the framework of the project Arabic Author Profiling for Cyber-Security (ARAP)[11], we outperform with LDSE [26] (0.797 F-measure) the result obtained by the CAT method (0.701 F-measure) on the Credibility corpus [25].

3 Evaluation Framework

The purpose of this section is to introduce the technical background. We outline the construction of the corpora, as well as we introduce the performance measures.

3.1 Corpora

We have created the following corpora: the ARAP-Tweet corpus for author profiling, and the Qatar Twitter and Qatar News corpora for deception detection. We briefly describe them below.

[8] http://alt.qcri.org/clef2018-factcheck
[9] http://clef2018.clef-initiative.eu/
[10] https://sites.google.com/view/clef2019-checkthat/home?authuser=0
[11] http://arap.qatar.cmu.edu

ARAP-Tweet. This corpus was developed at the Carnegie Mellon University Qatar [39] with the aim at providing with a fine-grained annotated corpus in Arabic. It contains 15 dialectical varieties corresponding to 22 countries of the Arab League. For each variety, a total of 198 authors (150 for training, 48 for test) were annotated with age and gender, maintaining balance for both variables. The following groups were considered for the age annotation: Under 25, Between 25 and 34, and Above 35. For each author, more than 2,000 tweets were retrieved from her/his timeline. The included varieties are: Algeria, Egypt, Iraq, Kuwait, Lebanon Syria, Libya, Morocco, Oman, Palestine Jordan, Qatar, Saudi Arabia, Sudan, Tunisia, United Arab Emirates and Yemen. More information about this corpus is available in [40].

The Qatar Twitter corpus. In the context of the ARAP project, we created the Qatar Twitter corpus by retrieving during 2017 and annotating[12] tweets referring to the Qatar Blockade and the Qatar World Cup. Statistics about this corpus are shown in Table 1. The number of tweets for the blockade topic is completely balanced between credible and non-credible classes. For the World Cup topic the corpus is almost balanced, with a slightly smaller amount of credible tweets (48% / 52%).

The Qatar News corpus. We also created the Qatar News corpus by retrieving and annotating short contents such as headlines and/or excerpts from well-known Arabic newsletters. Statistics on this second corpus can be seen in Table 1. The number of documents is almost balanced, with a slightly smaller amount of credible news (47% / 53%).

Table 1. Distribution of credible and non-credible tweets per topic in the Qatar Twitter and Qatar News corpora.

Corpus	Topic	Credible	Non Credible	Total
Qatar Twitter	Blockade	115	115	230
	World Cup	262	281	543
	Total	377	396	773
Qatar News		889	999	1,888

3.2 Performance Measures

In this section we describe the performance measures used for evaluating the systems in the different tasks.

[12] For both the Qatar Twitter and Qatar News corpora, the annotators were 20 students at the Hamad Bin Khalifa University, representing various Arab countries. The inter-annotator agreement was about 80%.

Author Profiling Since the data is completely balanced, the performance is evaluated by accuracy, following what has been done in the author profiling tasks at PAN@CLEF. For each subtask (age, gender, language variety), we calculate individual accuracies. Systems rank by the joint accuracy (when age, gender and language variety are properly identified together).

Deception Detection As in this case the data is slightly imbalanced, we measure the performance with the macro-averaged F-measure.

4 Overview of the Submitted Approaches

Nineteen teams participated in the shared task and fifteen of them submitted the notebook paper[13]. We analyse their approaches from three perspectives: preprocessing, features to represent the authors texts, and classification approaches.

4.1 Preprocessing

The authors of [17, 9, 13, 30, 23, 20, 15] removed stop words commonly defined for Arabic, and one of the teams (Blat) also removed its own list containing the most frequent words in the vocabulary. Some teams removed punctuation signs [22, 15], special characters [13, 23, 36], numbers [13, 33, 20], or Twitter related items such as emojis, user mentions, urls or hashtags [36, 33, 23]. Tokenisation was applied by the authors of [5]. The authors of [20] lower cased the texts, the authors of [22, 20] treated character flooding, and the authors of [33, 23] removed non-Arabic words. Finally, the authors of [42] applied data augmentation.

4.2 Features

Most of the systems [9, 10, 5, 13, 22, 36, 33] relied on n-grams, some of them in its simplest representation: bag-of-words [17, 30, 20, 15]. The team MagdalenaYVino combined word n-grams with emoticons n-grams, and the authors of [17] combined bag-of-words with lists of the most discriminant words per class.

Some teams approached the task with stylistic features such as the occurrence of emoticons/emojis [17, 15], hashtags [15], tweets length [17], the number of mentions [17, 15], or the use of function words [33]. The authors of [5] combined content-based features (word and character n-grams, stems, lemmas, Parts-of-Speech) with style-based features (urls, hashtags, mentions, character flooding, the average tweet length, the use of punctuation marks). Finally, the authors of [23] used word embeddings, as well as the authors of [10] trained them with FastText.

[13] Although some of them were rejected due to their low quality.

4.3 Classification Approaches

The most used classifier has been Support Vector Machines [17, 9, 10, 5, 13, 30, 22, 23], followed by Multinomial Naive Bayes [33, 20, 15]. The authors of [36] used Logistic Regression, while the team MagdalenaYVino addressed the task with Random Forest. Finally, only two teams approached the task with deep learning: the authors of [42] used BERT pre-trained on Wikipedia and the authors of [35] used LSTM.

5 Evaluation and Discussion of the Results

Although we recommended to participate in both tasks, author profiling and deception detection, some participants approached only one problem. Following, we present the results separately.

5.1 Author Profiling

Thirteen teams have participated in the Author Profiling task, submitting a total of 28 runs. Participants have used different kinds of features: from classical approaches based on n-grams and Support Vector Machines, to novel representations such as BERT. The best overall result (45.56% joint accuracy) has been achieved by DBMS-KU [33] with combinations of word n-grams, character n-grams, and function words to train Support Vector Machines. The best result for gender identification (81.94%) has been obtained by MagdalenaYVino, with a combination of words and emoticons 2-grams and 3-grams. In case of age identification, the best result has been achieved by Yutong [36] (62.50%) with a Logistic Regression classifier trained with a combination of word unigrams with character 2 to 5-grams. Finally, in regards of language variety identification, the best result (97.78%) has been achieved also by DBMS-KU.

Table 2. Author Profiling: Statistics on the accuracy per task.

Measure	Gender	Age	Variety	Joint
Min	0.5111	0.2222	0.2444	0.0597
Q1	0.6496	0.5368	0.8858	0.3104
Median	0.7667	0.5486	0.9354	0.3756
Mean	0.7181	0.5282	0.8705	0.3425
SDev	0.1034	0.0917	0.1843	0.1153
Q3	0.7843	0.5771	0.9694	0.4174
Max	0.8194	0.6250	0.9778	0.4556
Skewness	-0.9919	-2.2173	-2.4694	-1.4303
Kurtosis	2.4632	7.2901	8.0241	3.9081
Normality (p-value)	3.01e-06	1.75e-08	1.681e-11	1.156e-05

It can be observed in Table 2 and Figures 1 and 2 that the highest results have been obtained in case of language variety identification, with most of the results very close to 100%, although with three outliers: two runs sent by Allaith (0.2444 and 0.3458), who did not send any description of their system, and the LSTM-based approach by Suman [35] (0.3458).

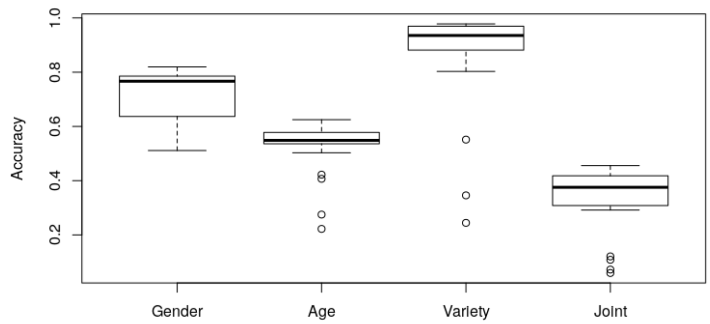

Fig. 1. Distribution of results for the author profiling task.

In this figures we can also observe that the lowest sparsity occurs with age identification, where most of the systems obtained very similar results. In this case, there are also four outliers: the two systems of Suman (0.2222 and 0.2750) based on LSTM, and the two systems of Allaith (0.4069 and 0.4222). In case of gender identification, results are more sparse, but there are no ourliers.

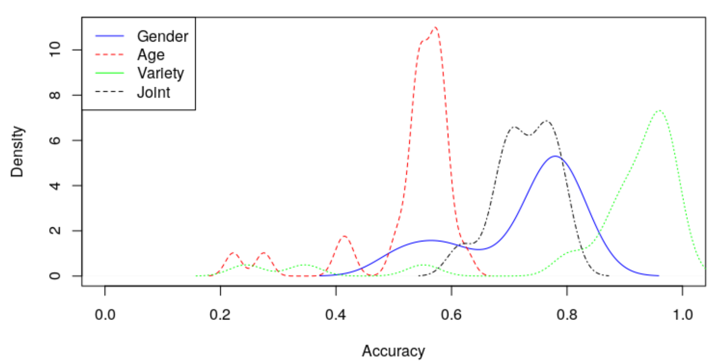

Fig. 2. Density of the results for the author profiling tasks.

Table 3. Author profiling: Overall ranking in terms of accuracy.

Ranking	Team	Gender	Age	Variety	Joint
1	DBMS-KU.2	0.7944	0.5861	0.9722	0.4556
2	Nayel.1	0.8153	0.5708	0.9750	0.4486
3	Nayel.3	0.8014	0.5792	0.9708	0.4486
4	DBMS-KU.3	0.7833	0.5819	0.9778	0.4444
5	DBMS-KU.1	0.7778	0.5792	0.9736	0.4347
6	KCE_DAlab.sub1	0.7667	0.5722	0.9583	0.4222
7	Nayel.2	0.7667	0.5764	0.9597	0.4194
8	MagdalenaYVino.1	0.8194	0.5653	0.9069	0.4167
9	KCE_DAlab.sub2	0.7458	0.5708	0.9694	0.4125
10	Chiyuzhang.maj2	0.8167	0.5472	0.9375	0.4097
11	Chiyuzhang.4	0.8167	0.5472	0.9264	0.4097
12	Blat.1	0.7875	0.5653	0.8722	0.3986
13	Chiyuzhang.2	0.7708	0.5472	0.9333	0.3875
14	Karabasz.1	0.7833	0.5403	0.9083	0.3819
15	KCE_DAlab.sub3	0.7444	0.5028	0.9583	0.3694
16	Alrifai.1	0.7708	0.5375	0.8903	0.3639
17	Alrifai.2	0.7681	0.5347	0.8917	0.3611
18	Kosmajac.1	0.7000	0.5417	0.9542	0.3583
19	Alrifai.3	0.7667	0.5139	0.8681	0.3431
20	SSN_NLP.1	0.7653	0.5500	0.8083	0.3403
21	Yutong.2	0.5111	0.6250	0.9694	0.3125
22	Karabasz.2	0.6111	0.5403	0.9083	0.3042
23	Yutong.3	0.5111	0.6000	0.9694	0.2944
24	Yutong.1	0.5111	0.5875	0.9694	0.2917
25	Allaith.1	0.5806	0.4069	0.3458	0.1208
26	Suman.LSTM_Features	0.6625	0.2222	0.8028	0.1083
27	Suman.LSTM	0.5764	0.2750	0.5514	0.0722
28	Allaith.2	0.5806	0.4222	0.2444	0.0597

5.2 Deception Detection

Thirteen teams have participated in the Deception Detection task, submitting a total of 25 runs. Participants have used different kinds of features such as classical approaches based on n-grams and Support Vector Machines. No novel approaches based on deep learning have been used, apart from some word embedding-based representations. The best overall result (0.8003 Macro F-measure) has been achieved by Nayel [22] with n-grams weighted with TF/IDF and Support Vector Machines. The best result on the Qatar News corpus (0.7542 Macro F-measure) has been also obtained by Nayel, while the best result on the Qatar Twitter corpus (0.8541 Macro F-measure) has been obtained by KCE_Dalab [10], who approached the task with a combination of word and character n-grams and Fast text embeddings to train a Support Vector Machine.

In Table 5 and Figures 3 and 4 we can observe that the highest results have been obtained on the Twitter corpus, with similar sparsity on both genres. Perhaps, it should be highlighted that the distribution of results on the News corpus is more skewed to the right, with the median higher than the mean, and most systems close to the best performing ones.

Table 4. Deception detection: Overall ranking in terms of macro F-measure.

Ranking	Team/Run	News	Twitter	Average
1	nayel.3	0.7542	0.8464	0.8003
2	nayel.1	0.7417	0.8463	0.7940
3	KCE_Dalab.sub1	0.7232	0.8541	0.7887
4	KCE_Dalab.sub2	0.7331	0.8293	0.7812
5	DBMS-KU.2	0.7352	0.8125	0.7739
6	nayel.2	0.7133	0.8337	0.7735
7	Allaith.2	0.7106	0.8289	0.7698
8	Allaith.1	0.7274	0.7950	0.7612
9	SSN_NLP.1	0.7108	0.8087	0.7598
10	DBMS-KU.1	0.7188	0.7877	0.7533
11	DBMS-KU.3	0.7188	0.7877	0.7533
12	Actimel.tfidf_svm	0.7235	0.7717	0.7476
13	RickyTonCar.1	0.6754	0.7748	0.7251
14	Cabrejas.2	0.6651	0.7699	0.7175
15	Actimel.tree_SVC	0.7043	0.7288	0.7166
16	Eros.1	0.6277	0.7924	0.7101
17	Blat.1	0.6675	0.7355	0.7015
18	Cabrejas.1	0.6566	0.7443	0.7005
19	Actimel.trigram_arab_dict_SVM	0.6572	0.7383	0.6978
20	RickyTonCar.2	0.6912	0.7008	0.6960
21	Sinuhe.SVM	0.6261	0.7627	0.6944
22	Eros.2	0.6277	0.7339	0.6808
23	KCE_Dalab.sub3	0.6613	0.6791	0.6702
24	Sinuhe.kNN	0.5640	0.6716	0.6178
25	Bravo.1	0.5827	0.6477	0.6152

Table 5. Deception detection: Statistics on the F-measure per task.

Measure	News	Twitter	Average
Min	0.5640	0.6477	0.6152
Q1	0.6572	0.7355	0.6978
Median	0.7043	0.7748	0.7251
Mean	0.6847	0.7713	0.7280
SDev	0.0502	0.0572	0.0505
Q3	0.7232	0.8125	0.7698
Max	0.7542	0.8541	0.8003
Skewness	-0.7928	-0.4649	-0.5946
Kurtosis	2.8250	2.4024	2.7434
Normality (p-value)	0.0501	0.5339	0.2214

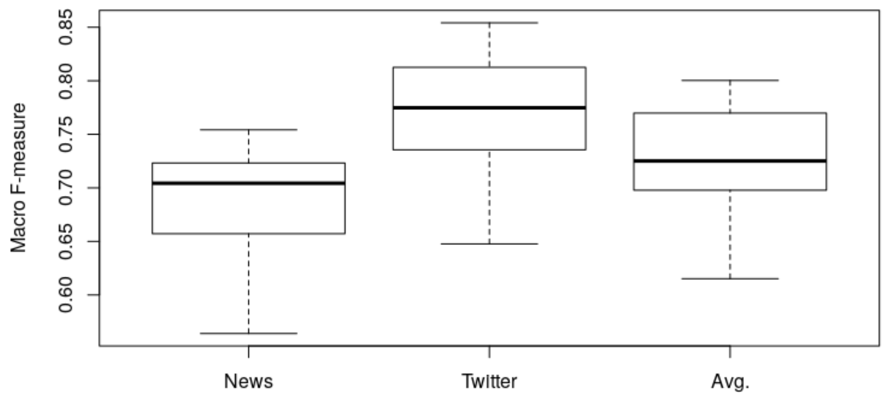

Fig. 3. Distribution of results for the deception detection task.

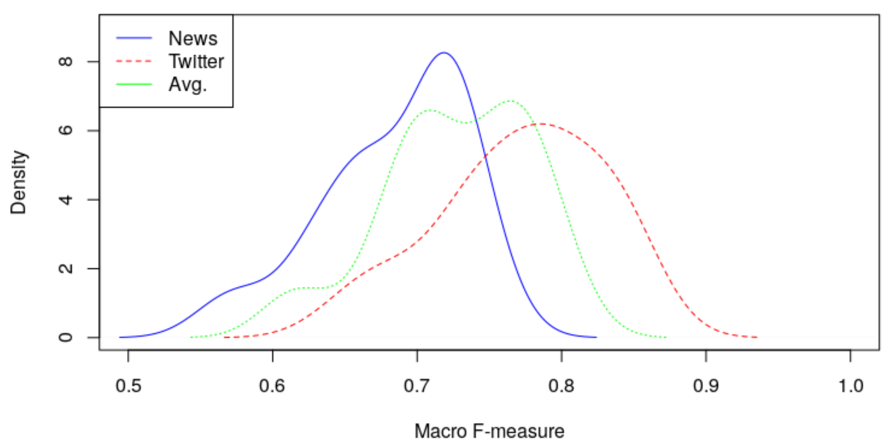

Fig. 4. Density of the results for the deception detection tasks on the different corpora.

6 Conclusions

In this paper we have presented the results of the Author Profiling and Deception Detection in Arabic (APDA) shared task hosted at FIRE 2019. Two have been the main aims: *i)* to profile the age, gender and native language of a Twitter user; *ii)* to determine whether an Arabic text is deceptive or not, in two different genres: Twitter and news headlines.

The participants have used different features to address the task, mainly: *i)* n-grams; *ii)* stylistic features; and *iii)* embeddings. With respect to machine learning algorithms, the most used one was Support Vector Machines. Nevertheless, a couple of participants approached the author profiling task with deep learning techniques. In such cases, they used BERT and LSTM respectively. According to the results, traditional approaches obtained better performances than deep learning ones. The best performing team in the author profiling task [33] used combinations of word and character n-grams with function words to train Support Vector Machines, while the best performing team in the deception detectin task [22] used n-grams weighted with TF/IDF and Support Vector Machines.

Acknowledgments

This publication was made possible by NPRP 9-175-1-033 from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors. The work of Paolo Rosso was also partially funded by Generalitat Valenciana under grant PROMETEO/2019/121.

References

1. A. Akbar-Maulana-Siagian, M. Aritsugi. DBMS-KU Approach for Author Profiling and Deception Detection in Arabic. In: Mehta P., Rosso P., Majumder P., Mitra M. (Eds.) Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019). CEUR Workshop Proceedings. CEUR-WS.org, Kolkata, India, December 12-15, 2019
2. R.M.B. Al-Eidan, H.S. Al-Khalifa, A.S. Al-Salman. Measuring the Credibility of Arabic Text Content in Twitter. In 2010 Fifth International Conference on Digital Information Management (ICDIM), 2010
3. K. Alsmearat, M. Al-Ayyoub, R. Al-Shalabi. An Extensive Study of the Bag-of-words Approach for Gender Identification of Arabic Articles. In: 11th International Conference on Computer Systems and Applications (AICCSA), pages 601608. IEEE/ACS, 2014
4. K. Alsmearat, M. Shehab, M. Al-Ayyoub, R. Al-Shalabi, G. Kanaan. Emotion Analysis of Arabic Articles and its Impact on Identifying the Authors Gender. In: 12th International Conference on Computer Systems and Applications (AICCSA), IEEE/ACS, 2015
5. K. Alrifai, G. Rebdawi, N. Ghneim. Arabic Tweeps Traits Prediction AT2P. In: Mehta P., Rosso P., Majumder P., Mitra M. (Eds.) Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019). CEUR Workshop Proceedings. CEUR-WS.org, Kolkata, India, December 12-15, 2019

6. E. Al Sukhni, Q. Alequr. Investigating the Use of Machine Learning Algorithms in Detecting Gender of the Arabic Tweet Author. In: International Journal of Advanced Computer Science & Applications, 1(7):319328, 2016
7. A. Al Zaatari, R. El Ballouli, S. Elbassuoni, W. El-Hajj, H.M. Hajj, K.B. Shaban, N. Habash, E. Yahya. Arabic Corpora for Credibility Analysis. In: Language Resources and Evaluation Conference (LREC), 2016
8. L. Cagnina, P. Rosso. Detecting Deceptive Opinions: Intra and Cross-Domain Classification Using an Efficient Representation. In: International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 25, Suppl. 2, pp. 151–174, World Scientific, 2017
9. J. Cabrejas, J.V. Mart, A. Pajares, V. Sanchis. Deception Detection in Arabic Texts Using N-grams Text Mining. In: Mehta P., Rosso P., Majumder P., Mitra M. (Eds.) Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019). CEUR Workshop Proceedings. CEUR-WS.org, Kolkata, India, December 12-15, 2019
10. S. Devi, S. Kannimuthu, G. Ravikumar, A. Kumar. KCE DALab-APDAFIRE2019: Author Profiling and Deception Detection in Arabic using Weighted Embedding. In: Mehta P., Rosso P., Majumder P., Mitra M. (Eds.) Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019). CEUR Workshop Proceedings. CEUR-WS.org, Kolkata, India, December 12-15, 2019
11. R. El Ballouli, W. El-Hajj, A. Ghandour, S. Elbassuoni, H. Hajj, K. Shaban. CAT: Credibility Analysis of Arabic Content on Twitter. In Proc. of the Third Arabic Natural Language Processing Workshop, 2017
12. H. Elfardy, M.T. Diab. Sentence Level Dialect Identification in Arabic. In: Association for Computational Linguistics (ACL), pp. 456–461 (2013)
13. F. Eros-Blázquez-del-Rio, M. Conde-Rodríguez, J.M. Escalante. Detection of Deceptions in Twitter and News Headlines Written in Arabic. In: Mehta P., Rosso P., Majumder P., Mitra M. (Eds.) Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019). CEUR Workshop Proceedings. CEUR-WS.org, Kolkata, India, December 12-15, 2019
14. D. Estival, T. and Gaustad, B. Hutchinson, S. Bao-Pham, W. Radford. Author Profiling for English and Arabic eMails, 2008
15. F.J. Fernández-Bravo Peñuela. Deception Detection in Arabic Tweets and News. In: Mehta P., Rosso P., Majumder P., Mitra M. (Eds.) Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019). CEUR Workshop Proceedings. CEUR-WS.org, Kolkata, India, December 12-15, 2019
16. M. Hasanain, R. Suwaileh, T. Elsayed, A. Barrn-Cedeno, P. Nakov. Overview of the CLEF-2019 CheckThat! Lab on Automatic Identification and Verification of Claims. Task 2: Evidence and Factuality. In CEUR Workshop Proceedings, CEURWS. org, 2019
17. I. Karabasz, P. Cellini, G. Galiana. Predicting Author Characteristics of Arabic Tweets through Author Profiling. In: Mehta P., Rosso P., Majumder P., Mitra M. (Eds.) Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019). CEUR Workshop Proceedings. CEUR-WS.org, Kolkata, India, December 12-15, 2019
18. S. Malmasi, M. Zampieri, N. Ljubešić, P. Nakov, A. Ali, J. Tiedemann. Discriminating between Similar Languages and Arabic Dialect Identification: A Report on the Third DSL Shared Task. In: Proceedings of the Third Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial3), pp. 1–14 (2016)

19. T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, Jeff. Distributed Representations of Words and Phrases and their Compositionality. In: Advances in Neural Information Processing Systems, 2013
20. A. Moreno, R. Navarro, C. Ruiz. UPV at Author Profiling and Deception Detection in Arabic: Task 2. Deception Detection in Arabic Texts. In: Mehta P., Rosso P., Majumder P., Mitra M. (Eds.) Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019). CEUR Workshop Proceedings. CEUR-WS.org, Kolkata, India, December 12-15, 2019
21. P. Nakov, A. Barrón-Cedeño, T. Elsayed, R. Suwaileh, L. Màrquez, W. Zaghouani, P. Atanasova, S. Kyuchukov, and G. Da San Martino. Overview of the CLEF-2018 CheckThat! Lab on Automatic Identification and Verification of Political Claims. In: International Conference of the Cross-Language Evaluation Forum for European Languages, 2018.
22. H.A. Nayel. NAYEL@APDA: Machine Learning Approach for Author Profiling and Deception Detection in Arabic Texts. In: Mehta P., Rosso P., Majumder P., Mitra M. (Eds.) Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019). CEUR Workshop Proceedings. CEUR-WS.org, Kolkata, India, December 12-15, 2019
23. A. Ranganathan, H. Ananthakrishnan, D. Thenmozhi, C. Aravindan. Arabic Author Profiling and Deception Detection using Traditional Learning Methodologies with Word Embedding. In: Mehta P., Rosso P., Majumder P., Mitra M. (Eds.) Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019). CEUR Workshop Proceedings. CEUR-WS.org, Kolkata, India, December 12-15, 2019
24. F. Rangel, P. Rosso. On the Implications of the General Data Protection Regulation on the Organisation of Evaluation Tasks. In: Language and Law= Linguagem e Direito, vol. 5 (2), pp. 95–117, 2019
25. F. Rangel, P. Rosso, A. Charfi, W. Zaghouani. Detecting Deceptive Tweets in Arabic for Cyber-Security. In: Proc. of the 17th IEEE Int. Conf. on Intelligence and Security Informatics (ISI), 2019
26. F. Rangel, P. Rosso, M. Franco. A Low Dimensionality Representation for Language Variety Identification. In: Proceedings of the 17th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing16), Springer-Verlag, LNCS(9624), pp. 156-169, 2018
27. F. Rangel, P. Rosso, M. Potthast, B. Stein. Overview of the 5th Author Profiling Task at PAN 2017: Gender and Language Variety Identification in Twitter. Working Notes Papers of the CLEF 2017 Evaluation Labs, Editors: Linda Cappellato and Nicola Ferro and Lorraine Goeuriot and Thomas Mandl, pp. 1613–0073, CLEF and CEUR-WS.org (2017)
28. H. Rheingold. Smart Mobs: the Next Social Revolution. Basic books, 2007
29. P. Rosso, F. Rangel, I. Hernández-Farías, L. Cagnina, W. Zaghouani, A. Charfi. A Survey on Author Profiling, Deception, and Irony Detection for the Arabic Language. Language and Linguistics Compass, vol. 12 (4), pp. e12275, Wiley Online Library (2018a)
30. F.I. Ruedas-Diaz, S. Martínez-Rodríguez, S. Muñoz-Lorenzo, V. Cristiny-Sá-de-Aráujo, C. Muñoz-Carrasco. Deception Detection in Arabic Texts. In: Mehta P., Rosso P., Majumder P., Mitra M. (Eds.) Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019). CEUR Workshop Proceedings. CEUR-WS.org, Kolkata, India, December 12-15, 2019
31. C. Russell, B. Miller. Profile of a Terrorist. Studies in Conflict & Terrorism, vol. 1 (1), pp. 17–34, Taylor & Francis, 1977

32. F. Sadat, F. Kazemi, A. Farzindar. Automatic Identification of Arabic Language Varieties and Dialects in Social Media. In: Proceedings of SocialNLP, pp. 22 (2014)
33. M. Siagian, A. H. Akbar, M. Aritsugi. DBMS-KU Approach for Author Profiling and Deception Detection in Arabic. In: Mehta P., Rosso P., Majumder P., Mitra M. (Eds.) Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019). CEUR Workshop Proceedings. CEUR-WS.org, Kolkata, India, December 12-15, 2019
34. A.B. Soliman, K. Eisa, S.R. El-Beltagy, AraVec: A set of Arabic Word Embedding Models for use in Arabic NLP. In: 3rd International Conference on Arabic Computational Linguistics (ACLing), 2017.
35. C. Suman, P. Kumar, S. Saha, P. Bhattacharyya. Gender, Age and Dialect Recognition Using Tweets in a Deep Learning Framework - Notebook for FIRE 2019. In: Mehta P., Rosso P., Majumder P., Mitra M. (Eds.) Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019). CEUR Workshop Proceedings. CEUR-WS.org, Kolkata, India, December 12-15, 2019
36. Y. Sun, H. Ning, K. Chen, L. Kong, Y. Yang, J. Wang, H. Qi. Author Profiling in Arabic Tweets: An Approach based on Multi-Classification with Word and Character Features. In: Mehta P., Rosso P., Majumder P., Mitra M. (Eds.) Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019). CEUR Workshop Proceedings. CEUR-WS.org, Kolkata, India, December 12-15, 2019
37. A.J. Viera, J.M. Garrett. Understanding Interobserver Agreement: the Kappa Statistic. Fam med journal, 37(5), pp.360-363. 2005
38. O.F. Zaidan, C. Callison-Burch. Arabic Dialect Identification. In: Computational Linguistics, vol. 40 (1), pp. 171–202, MIT Press (2014)
39. W. Zaghouani, A. Charfi. ArapTweet: A Large MultiDialect Twitter Corpus for Gender, Age and Language Variety Identification. In: Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC), Miyazaki, Japan, 2018
40. W. Zaghouani, A. Charfi. Guidelines and Annotation Framework for Arabic Author Profiling. In: Proceedings of the 3rd Workshop on Open-Source Arabic Corpora and Processing Tools, 11th International Conference on Language Resources and Evaluation (LREC), Miyazaki, Japan, 2018
41. M. Zampieri, S. Malmasi, N. Ljubešić, P. Nakov, A. Ali, J. Tiedemann, Y. Scherrer, N. Aepli. Findings of the Vardial Evaluation Campaign 2017. In: Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects, pp. 1–15 (2017)
42. C. Zhang, M. Abdul-Mageed. BERT-Based Arabic Social Media Author Profiling. In: Mehta P., Rosso P., Majumder P., Mitra M. (Eds.) Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019). CEUR Workshop Proceedings. CEUR-WS.org, Kolkata, India, December 12-15, 2019