Improving Neural Machine Translation for Low Resource Algerian Dialect by Transductive Transfer Learning Strategy

Amel Slim · Ahlem Melouah · Usef Faghihi · Khouloud Sahib

Abstract
This study is the first work on a transductive transfer learning approach for low-resource neural machine translation applied to the Algerian Arabic dialect. The transductive approach is based on a fine-tuning transfer learning strategy that transfers knowledge from the parent model to the child model. This strategy helps to solve the learning problem using limited parallel corpora. We tested the approach on a sequence-to-sequence model with and without the Attention mechanism. We first trained the models on a parallel multi-dialects Arabic corpus and then switch them to a low-resource of the Algerian dialect. Transductive transfer learning raises the BLEU score for the Seq2Seq model from 0.3 to more than 34, and for the Attentional-Seq2Seq model from less than 17 to more than 35. The obtained results prove the validity of this approach.

Keywords Neural machine translation · Transductive transfer learning · Algerian Arabic dialect · Low resource

1 Introduction
A strong demand for international communication in many sectors and forms fuels the field of machine translation research. Any application that matches a source sentence $X_i$ with a target sentence $Y_i$ is considered a machine translation system [1]. This matching is a complex process that deals with semantics, syntax, morphology and other grammatical complexities [2,3]. Recently, Neural Machine Translation (NMT) has emerged as a common technology as it has achieved maximum performance on several parallel corpora of multiple languages. The sequence-to-sequence network proposed in [4] is a popular neural machine translation model. It is composed of two sequences: the encoder and the decoder.

While the encoder encodes the input words into a vector of numbers, the decoder decodes the vector into another set of words. Both the encoder and the decoder are based on the well-known LSTM layer. This layer was successfully applied to solve several types of problems [5–9], but from our point of view, LSTM is more advantageous for translation problems, especially with the introduction of the attention mechanism [10]. That is why the application of LSTM has extended to the dialects translation.

A dialect is the written form of a spoken language based on a deviation from one or more standards languages according to a specific geographic area [11–13]. The dialect has great flexibility that allows anyone to write his message as he wishes. That explains its rapid spread in many communities. The Arab community is not an exception; like many other languages, Arabic is in the state of diglossia with about 22 dialects. The Arabic dialect has gained popularity that its automatic translation is now a necessity.

Most researches on Arabic dialect translation are based on rule-based machine translation [14–19] and statistical machine translation [3,20–24]. Probably due to the severe lack of data, there is little work on Arabic dialects neural machine translation, from our state of the art research, we found only four works. Two works applied a learning process via an RNN model, and two works focused on transfer learning to deal with data lack problem.
Guellil et al. [25] start by transliterating all Arabizi writing forms in Arabic forms. Then, they apply the RNN encoder-decoder model to translate the transliterated texts into modern standard Arabic (MSA). Slim et al. [26] translate Algerian dialect comments collected from the social media Facebook using an RNN encoder-decoder. Their work is a part of a study on the propagation of COVID-19 misinformation and its impact on Algerian society. These two works obtain a small BLEU Score value due to the low size of the used database. To overcome this problem, [2,27] introduce the transfer learning paradigm. Baniata et al. [2] propose a multitask transfer learning (MTL) model. The proposed model focuses on sharing one decoder for all tasks and using separate encoders for each source language. The proposed MTL model can leverage the useful information contained in multiple related tasks. In addition, the proposed MTL model can learn to generate the sentence in the correct order of the target language and make the translation clearer and smoother. Farhan et al. [27] adopt an unsupervised transfer learning approach based on common words in both standard Arabic language and Arabic dialect. They project words from both corpora into the same hyper-dimensional space, and they formulate an objective function to find a translation matrix to maps from one hyper-dimensional space to another. They apply this unsupervised approach to two models: Google NMT (GNMT) model and the Dialectal to Standard Language Translation (D2SLT) model. An inductive transfer learning strategy is adopted in [2], an unsupervised transfer learning strategy is adopted in [27], and in this work, we propose a transductive transfer learning strategy to overcome the data lack problem in Arabic dialect translation. The transductive transfer learning strategy has yielded very good results in the translation of formal languages. This work is the first one to test this paradigm on Arabic dialectal languages.

The rest of this paper is organized as follows: Section 2 details the proposed method by presenting the available Arabic dialect corpora, the used NMT model, the adopted transductive transfer learning strategy and the followed translation process. Section 3 presents and discusses the results obtained. Section 4 concludes the paper.

2 Methodology

Arabic dialects share many difficulties with modern standard Arabic (MSA). They are part of the Semitic languages whose derived morphology is complicated. In addition, Arabic dialects do not have a standard form of writing and are not restricted by any rules of grammar and morphology, making them difficult to translate. Another problem with Arabic dialect translation is its diversity even within the same country. In Algeria, for example, each region has its own dialect. This diversity is not supported by resources and corpora that are very rare and limited [28]. For that reason, the big problem encountered in the Arabic dialects machine translation is the availability of learning data, especially in the case of NMT where the quality of translation deteriorates if the data volume is limited. Therefore, to deal with each dialect, it is necessary to generate its corpus from zero. This is a costly task and may introduce redundancies, especially for the Arabic dialect. The ideal is to give an NMT model trained on a large corpus that covers all Arabic dialects the ability to translate a very specific dialect using a limited corpus of data. This is the goal of this work.

2.1 Dataset

The Algerian dialect is available only in MADAR and PADIC corpora. MADAR (Multi Arabic Dialect Applications and Resources) [29] contains 25 Arabic dialects (Beirut, Cairo, Doha, Rabat, Tunis, Aleppo, Alexandria, Amman, Aswan, Baghdad, Basra, Bengazi, Damascus, Fes, Jeddah, Jerusalem, Khartoum, Muscat, Riyadh, Salt, Sanaa, Sfax, Tripoli) with 2K sentences in each dialect. Therefore, the total MADAR size is 50K sentences. PADIC (A Parallel Arabic Dialect Corpus) [3] contains five dialects (two from Algeria, one from Tunisia, one from Syria, and one from Palestine) with 6400 parallel sentences by dialect.

2.2 NMT Model

We use the transductive transfer learning strategy with two NMT models:

- **Seq2Seq**: is a sequence-to-sequence learning model proposed in [4]. An embedding layer is the input of the model. Word embedding provides a dense representation of words and their relative meanings for the perfect presentation of data. The encoder and decoder are LSTM architecture. SoftMax is the last and output layer of the network. Figure 1 presents the LSTM architecture where:
  - $x_t$ is the input.
  - $h_t$ and $c_t$, respectively, hidden state and cell memory at time $(t-1)$.
  - $h_t$ and $c_t$, respectively, hidden state and cell memory at time $(t)$.
- **Attentional-Seq2Seq**: is an attentional sequence-to-sequence learning model proposed by [10]. The difference between the two models is the introduction of the attentional mechanism (Fig. 1). The attentional mechanism enhances neural machine translation (NMT) by concentrating selectively on parts of the source sentence during translation. The mathematical formulation of the attention unit based Luong score function [10] is:
where \( t^s \) is the transpose of the decoder output state \( s \) at time \((t-1)\).

**2.3 Transductive Transfer Learning Strategy**

The most common problem with Arabic dialects NMT is the severe lack of data. We propose a transductive transfer learning solution to this problem. First, we define the notion of domain and task as presented in [30]:

- **Domain** \( D = \{X, P(X)\} \) where the first component is the feature space \( X = \{x_1, x_2, \ldots, x_n\} \) and the second component is the marginal distribution of probability \( P(X) \).
- **Task** \( T = \{Y, f(.)\} \) where the first component is the label space \( Y \) and the second component is an objective predictive function \( f(.) \approx P(y/x) \) with \( x \in X \) and \( y \in Y \).

For a source \( S = (D_{Source}, T_{Source}) \) and a target \( T = (D_{Target}, T_{Target}) \), the transfer learning objective is to improve the learning of the target predictive function \( f_{Target}(.) \) using the knowledge in \( S \), where \( D_{Source} \neq D_{Target} \), or \( T_{Source} \neq T_{Target} \).

In the transductive transfer learning, \( T_{Source} = T_{Target} \), but \( D_{Source} \neq D_{Target} \) according to two different situations:

1. The feature spaces between the source and target domains are different, \( X_{Source} \neq X_{Target} \).
2. The marginal probability distributions between the source and target domains are different \( P(X_{Source}) \neq P(X_{Target}) \).

In our problem, the source is MADAR (25-Arabic dialect-MSA) and the target is PADIC (the Algerian dialect-MSA). In ADD-to-MSA translation, we have contain few available labeled data in the target domain PADIC and many available labeled data are contained in the source domain MADAR.

In the transductive transfer learning setting, to solve this problem we need to learn an optimal model for the target domain by minimizing the expected risk:

\[
\theta = \arg\min_\theta \sum_{i=1}^n P_T S (X) P(X_S) l(x_{Si}, x_{Si}, \theta)
\]  

with:

\[
P_T S (X) = \frac{P(X_T)}{P(X_S)}
\]

Where:

- \( n \) is the size of the training data,
- \( l(x_{Si}, x_{Si}) \) is a loss function that depends on the parameter \( \theta \).
We can solve the transductive transfer learning problems by estimate $P_T^S(X)$ [31] proposed an algorithm known as Kullback–Leibler Importance Estimation Procedure (KLIEP) to estimate $P_T^S(X)$ based on the minimization of the Kullback-Leibler divergence. KLIEP can be integrated with cross-validation to perform model selection automatically in two steps:

1. Estimating the weights of the source domain data.
2. Training models on the reweighted data.

2.4 Translation

The machine translation quality achieved by the NMT models is excellent for high-resource language pairs such as English–German and French–English. This is why recent research studies on machine translation have focused on improving the system’s efficiency for low-resource pairs as occurs with the Algerian dialect.

Because of the Algerian Arabic dialect data poverty, we applied transductive transfer learning to improve its neural machine translation. As shown in Fig. 2, our proposed method consists of two main steps:

1. Learning step: the Seq2Seq model and the Attentional-Seq2Seq model are trained using 25 Arabic dialects of the MADAR dataset.
2. Transfer learning step: using transfer learning the Seq2Seq model and the Attentional-Seq2Seq model are retrained using Algerian Arabic dialects of the PADIC dataset.

Using the large parallel corpus MADAR, we construct the parent model. Next, we generate a child model by retraining the parent model on the smaller parallel corpus of a low-dialect Algerian, PADIC. The child model inherits the knowledge from the parent model by reusing its parameters.

3 Results and Discussion

3.1 Experiments

Seq2Seq model and the Attentional-Seq2Seq model are implemented and trained using TensorFlow and Keras. In order accuracy metric, sparse categorical cross-entropy loss, and rmsprop optimizer are used in each experiment. We train the two models with 50 epochs in learning and transfer learning steps and with a selected batch of size 512 sequences for each iteration. All sequences within a lot must be padded to the longest sequence in that batch during training. From PADIC, we randomly take more than 10K sentences for training and 2729 sentences for testing. From MADAR, we use 40K sentences for training and 10 K for testing.

3.2 Results

The results of the Algerian translation into MSA are shown in Table 1. We note that:

- Seq2Seq completely failed, and it did not succeed in any translation. Even worse, it gave a very different translation compared to the content of the source sentence. The introduction of transfer learning remarkably improved the performance of the model. The Seq2seq child model successfully translates seven sentences among ten. This shows a significant boost in the efficiency of this model.
- Attentional-Seq2Seq successfully translated four sentences (1, 4, 8, and 9). It completely failed for two sentences (6, 10). For the remaining sentences (2, 3, 5, and 7), Attentional-Seq2Seq translation performance varies from good to worse. The introduction of transfer learning allowed a significant improvement in the attentional-Seq2Seq performance. From ten sentences, only three have an incorrect translation.

A second evaluation using the BLEU score [32] confirms the effectiveness of the transductive transfer learning approach. The BLEU score is a metric for evaluating a generated sentence by the translation model to an original sentence. If $c$ is the length of the candidate translation, $r$ is the effective reference corpus length, and $p_i$ is the i-gram precision on a multi-sentence test set, the BLEU score is defined by:

\[
\text{BLUE} = \text{BP} * e^{\sum_i w_i \log p_i} \\
\text{BP} = \begin{cases} 
1 & \text{if } c > r \\
\frac{e^{1-r/c}}{c} & \text{if } c \leq r 
\end{cases} \\
p_i = \frac{\sum_{C \in \text{candidates}} \sum_{i\text{-gram} \in C} \text{count}_{\text{clip}}(i - \text{gram})}{\sum_{C' \in \text{candidates}} \sum_{i\text{-gram} \in C'} \text{count}_{\text{clip}}(i - \text{gram}')} 
\]

where $w_i$ are the weights assigned to the $i$-gram precision, and $\text{count}_{\text{clip}}(i - \text{gram})$ is the number of occurrences of the $i$-gram in the clip.

Fig. 2  Learning and transfer learning steps
Table 1 Examples of Algerian dialect translation

| Sentences in the text | English translation |
|----------------------|--------------------|
| "I want a day off."  | "Je veux une journée de congés." |
| "What time is it?"    | "Quelle heure est-il?" |
| "How are you?"        | "Comment allez-vous?" |

Table 2 The BLEU score of the Seq2Seq and attentional-Seq2Seq models before and after transfer learning

|                      | Seq2Seq | Attentional seq2seq |
|----------------------|---------|----------------------|
| Before transfer learning | 10.24   | 10.56                |
| After transfer learning | 12.34   | 12.67                |

The transductive transfer learning strategy proposed in this paper consists of transferring knowledge from a parent model previously trained on a large campus to a child model by retraining it on a target dataset of a limited size. In other words, an inheritance of knowledge is possible through the use of a large dataset. To overcome this problem, several solutions are possible; one of these is to apply transfer learning.

3.3 Discussion

Translation using NMT methods involves the use of a large dataset. This is not always possible. In some situations, we have to deal with a small corpus as for the Algerian dialect. To overcome this problem, several solutions are possible; one of these is to apply transfer learning.

With or without transfer learning: The first intuitive question is: “Does transfer learning increase the performance of a translation system?” Tables 1 and 2 give a clear answer to this question. In Table 1, we notice that the number of sentences correctly translated by the seq2seq and attention-seq2seq models with transductive transfer learning is significantly more compared to the number of sentences translated by the same models without transfer learning. On the other hand, the BLEU score values in Table 2 of the seq2seq and attention-seq2seq models with transductive transfer learning are higher than the BLEU score values for both models without transfer learning. Therefore, transductive transfer learning improves the efficiency of the NMT model with or without the attention.
Table 3  Comparative of the BLEU score result of our system and other works

| Model                                      | Source dataset Name | Size | Target dataset Composed of                      | Size   | BLEU score |
|--------------------------------------------|---------------------|------|-------------------------------------------------|--------|------------|
| Multitask learning (MTL) model [2]         | MSA-ENG             | 10.185 | Maroccan, Algerian, and Tunisian                 | 17.736 | 30.00      |
| Unsupervised Attentional seq2seq [27]      | Saudi-MSA           | 143.518 | Jordanian                                       | 142.305 | 32.14      |
| Transductive Attentional-seq2seq (proposed) | MADAR-MSA           | 52.000 | Algerian                                        |        | 12.800     | 35.87      |

- Transductive transfer learning and the attention mechanism: the second question that comes to mind is: “What effect has transductive transfer learning on a model with or without attention?”. If we consider Table 2, we can see a gap between the BLEU Score values of a model before and after transductive transfer learning. However, this gap is wider for the Seq2Seq model compared to the Attention-seq2seq model. For the Seq2Seq model, the gap is equivalent to 34, while for a-seq2seq it does not exceed 20. More interestingly, the BLEU score of the model without attention is close to the BLEU score of the model with attention in the case of transductive transfer learning.

- Transductive transfer learning and sentence length: Another important question is: “Does transductive transfer learning effect long sentence translation performance?”

Table 1 and Fig. 3 do not reject the assumption of a positive impact of the transductive transfer learning on the
performance of a translation model on long sentences. With transductive transfer learning, Seq2Seq models can translate sentences longer than 25 words. In Fig. 2, the graph of the child seq2seq model, and the graphs of the parent and the child attentional seq2seq models, gradually rise to a peak and then fall down showing a degradation in the performance of the translation models. The peak differs from model to model. For the parent seq2seq model, the peak is 20 words (Fig. 2a), and for the parent and child attentional seq2seq models, the peak is 25 words (Fig. 2b).

- Transductive transfer learning and state-of-the-art transfer learning strategies: The last question is: “transductive transfer learning is it the best transfer learning strategy?” Table 3 gives a comparison of the obtained results with some state-of-the-art works. Because the three approaches use different databases, an explicit comparison of their BLEU scores is impossible. However, since the size of the used databases has a direct impact on the NMT approach performance, an implicit comparison of the BLEU score based on the size of the used database is conceivable. Table 3 shows the small difference between the BLEU scores of the three approaches which is equal to 30 points for the MTL approach, 32.14 points for the unsupervised approach and 35.85 points for the transductive approach. The performance of the proposed transductive approach is compared to the respective performances of the MTL and unsupervised approaches based on the source base size and the target base size. The transductive approach provides the best BLEU score for a target base of no more than 13,000 sentences and a source base of 52,000 sentences. The MTL approach uses a target base of 17,736 sentences, which is slightly larger than the one used by the transductive approach. However, the MTL approach requires a source base that is five times smaller than that used by the transductive approach to achieve a five-point lower BLEU score. The unsupervised approach is ranked third because it requires a target base of 142,305 sentences, which is eleven times the size of the target base of the transductive approach, and a source base twice as large as the one used by the transductive approach.

4 Conclusion

The Arabic dialect is commonly used for communication. Unlike the formal language, it is not bound by orthography or grammatical rules, which explains its popularity. Arabic dialect translation is more difficult than the formal Arabic language. Deep Learning is a relatively new machine translation method. In contrast to conventional machine translation, neural machine translation is a better alternative for more accurate translation and better efficiency.

For the Algerian dialect, PADIC is one of the largest datasets. However, it is insufficient for neural machine translation learning. For this reason, we applied transductive transfer learning to overcome this problem. The transductive transfer learning approach shifts knowledge from a parent model to a child model by fetching the values of its parameters. The transductive transfer learning strategy improves the translation performance of the NMT model.

Extending the application of the transductive approach to other dialects to consolidate the validity of this approach is a short-term perspective. Focusing on a hierarchical transfer learning strategy is a medium-term perspective of this work.

Acknowledgements We are grateful to the Direction Générale de la Recherche Scientifique et du Développement Technologique (DGRSDT) which kindly supported this research, as well as to the Laboratoire de Recherche Informatique (LRI) where this study was conducted.

References

1. Wu, L.; Tian, F.; Qin, T.; Lai, J.; Liu, T.Y.: A study of reinforcement learning for neural machine translation. In: Proceeding of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018, pp. 3612–3621 (2018). https://doi.org/10.18653/v1/d18-1397
2. Baniata, L.H.; Park, S.; Park, S.-B.: A neural machine translation model for Arabic dialects that utilizes multitask learning (MTL). Comput. Intell. Neurosci. 2018, 7534712 (2018). https://doi.org/10.1155/2018/7534712
3. Meftehou, K.; Harrat, S.; Jamoussi, S.; Abbas, M.; Smaili, K.: Machine translation experiments on PADIC: a parallel Arabic Dialect corpus. In: Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation, pp. 26–34 (2015)
4. Sutskever, I.; Vinyals, O.; Le, Q.V.: Sequence to sequence learning with neural networks. Adv. Neural. Inf. Process. Syst. 20(1), 3104–3112 (2014)
5. Malki, Z.; Atlam, E.; Dagniew, G.; Alzighaibi, A.R.; Ghada, E.; Gad, I.: Bidirectional residual LSTM-based human activity recognition. Comput. Inf. Sci. 13(3), 1–40 (2020)
6. Xu, H.; Liu, Q.; van Genabith, J.; Xiong, D.; Zhang, M.: Multi-head highly parallelized LSTM decoder for neural machine translation. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 273–282) (2021)
7. Nguyen, L.H.; Pham, V.H.; Dinh, D.: Improving neural machine translation with AMR semantic graphs. Math. Probl. Eng. (2021)
8. Shi, X.; Huang, H.; Jian, P.; Tang, Y.K.: Improving neural machine translation with sentence alignment learning. Neurocomputing 420, 15–26 (2021)
9. Sahib, K.; Melouah, A.; Touré, F.; Slim, A.: W-net and inception residual network for skin lesion segmentation and classification. Appl. Intell. (2021). https://doi.org/10.1007/s10489-021-02652-4
10. Luong, M.-T.; Pham, H.; Manning, C.D.: Effective approaches to attention-based neural machine translation. CoRR, vol. abs/1508.0 (2015)
11. Benmamoun, E.: The Feature Structure of Functional Categories: A Comparative Study of Arabic Dialects. Oxford University Press, Oxford (2000)

12. Hughes, A.; Trudgill, P.; Watt, D.: English Accents and Dialects: An Introduction to Social and Regional Varieties of English in the British Isles. Routledge, London (2013)

13. Wolfram, W.; Schilling, N.: American English: Dialects and Variation. Wiley, New York (2015)

14. Al-Gaphari, G.H.; Al-Yadoumi, M.: A method to convert Sana’ani accent to Modern Standard Arabic. Int. J. Inf. Sci. Manag. 8(1), 39–49 (2012)

15. Hamdi, A.; Boujelbane, R.; Habash, N.; Nasr, A.: The Effects of Factorizing Root and Pattern Mapping in Bidirectional Tunisian - Standard Arabic Machine Translation. In: MT Summit 2013, hal-00908761 (2013).

16. Mohamed, E.; Mohit, B.; Oflazer, K.: Transforming Standard Arabic to Colloquial Arabic. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 176–180 (2012)

17. Salloum, W.; Habash, N.: Elissa: a dialectal to standard arabic machine translation system. In: Proceedings of COLING 2012: Demonstration Papers, pp. 385–392 (2012)

18. Tachicart, R.; Bouzoubaa, K.: A hybrid approach to translate Moroccan Arabic dialect. In: 2014 9th International Conference on Intelligent Systems: Theories and Applications (SITA-14), pp. 1–5, (2014). https://doi.org/10.1109/SITA.2014.6847293

19. Hamada, S.; Marzouk, R.M.: Developing a transfer-based system for Arabic Dialects translation. In: Shaalan, K., Hassanien, A.E., Tolba, F. (eds.) Intelligent Natural Language Processing: Trends and Applications, pp. 121–138. Springer, Cham (2018)

20. Jeblee, S.; Feely, W.; Bouamor, H.; Lavie, A.; Habash, N.; Oflazer, K.: Domain and Dialect adaptation for machine translation into Egyptian Arabic. In: Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP), pp. 196–206 (2014). https://doi.org/10.3115/v1/W14-3627

21. Sajjad, H.; Darwish, K.; Belinkov, Y.: Translating Dialectal Arabic to English. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 1–6 (2013)

22. Zhib, R., et al.: Machine translation of Arabic dialects. In: Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 49–59 (2012)

23. Bakr, H.A.; Shaalan, K.; Ziedan, I.: A hybrid approach for converting written Egyptian colloquial dialect into diacritized Arabic. In: The 6th International Conference on Informatics and Systems, Infos2008. Cairo University (2008)

24. Sawaf, H.: Arabic dialect handling in hybrid machine translation. In: Proceedings of the Conference of the Association for Machine Translation in the Americas (AMTA), Denver, Colorado (2010)

25. Guellil, I.; Azouaouf, F.; Abbas, M.: Neural Vs statistical translation of Algerian Arabic Dialect written with Arabizi and Arabic letter Neural Vs statistical translation of Algerian Arabic Dialect written with Arabizi and Arabic letter. In: 31st Pacific Asia Conference on Language, Information and Computer (PACLIC) (2017)

26. Slim, A.; Melouah, A.; Faghihi, Y., et al.: Algerian Dialect translation applied on COVID-19 social media comments. In: International Conference in Artificial Intelligence in Renewable Energetic Systems, pp. 716–726. Springer, Cham (2020)

27. Farhan, W.; et al.: Unsupervised dialectal neural machine translation. Inf. Process. Manag. 57(3), 102181 (2020). https://doi.org/10.1016/j.ipm.2019.102181

28. Habash, N.Y.: Introduction to Arabic natural language processing. Synth. Lect. Hum. Lang. Technol. 3(1), 1–185 (2010). https://doi.org/10.2200/S00277ED1V01Y201008HLT010

29. Bouamor, H., et al.: The Madar Arabic dialect corpus and lexicon. In: Lr. 2018—International Conference on Language Resources and Evaluation, pp. 3387–3396 (2019)

30. Pan, S.J.; Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22(10), 1345–1359 (2009)

31. Sugiyama, M.; Nakajima, S.; Kashima, H.; Von Buenau, P.; Kawababe, M.: Direct importance estimation with model selection and its application to covariate shift adaptation. In: NIPS, vol. 7, pp. 1433–1440 (2007)

32. Papineni, K.; Roukos, S.; Ward, T.; Zhu, W.-J.: BLEU: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pp. 311–318(2002)