An Arabic Multi-Domain Spoken Language Understanding System

Mohamed Lichouri, Mourad Abbas
Computational Linguistics Dept.,CRSTDLA
Algeria
m.lichouri@crstdla.dz
m.abbas@crstdla.dz

Rachida Djeradi, Amar Djeradi
USTHB University
Algeria
rdjeradi@usthb.dz
adjeradi@usthb.dz

Abstract

In this paper, we suggest the generalization of an Arabic Spoken Language Understanding (SLU) system in a multi-domain human-machine dialog. We are interested particularly in domain portability of SLU system related to both structured (DBMS) and unstructured data (Information Extraction), related to four domains. In this work, we used the thematic approach for four domains which are School Management, Medical Diagnostics, Consultation domain and Question-Answering domain (DAWQAS). We should note that two kinds of classifiers are used in our experiments: statistical and neural, namely: Gaussian Naive Bayes, Bernoulli Naive Bayes, Logistic Regression, SGD, Passive Aggressive Classifier, Perceptron, Linear Support Vector and Convolutional Neural Network.

1 Introduction

With the increasing spread of internet content, there is a mutually growing number of web applications pushing human being in a race against time to exploit and to master all of these applications. In such a situation, a human-machine dialogue system is needed to assist humans for acquiring information efficiently and accurately. However, the existing dialogue systems cannot cover all application domains. That is why, we tackle in this paper the multi-domain task. We should note that a little initial work with regard to the multi-domain problem has been presented in (Minker, 1998; Liu and Lane, 2016), which remains an open issue. We have witnessed recently a renewed interest in the extension of application domain, where some systems use Latent Semantic Mapping (LSM) for the identification of any abrupt change towards another application (Nakano et al., 2011). In other works, a Markovian decision-making process was considered for the selection of an application among several ones (Wang et al.) or the extension to a new application in the Web (Komatani et al., 2008). While in (Jung et al., 2009), a study related to comparable applications (within the same domain) has been conducted. In the case of more than two applications, we can mention task-based applications (where the dialogue is finalized and specific to a given domain) as presented in (Lee et al., 2009) or managing specific applications of the Web (Jiang et al., 2014). In (Jaech et al., 2016; Chelba and Acero, 2006; Daumé-III, 2007; Daumé-III and Marcu, 2006), the principle of adaptation from application to another has been applied, where the system is trained in the first application and tested in the second one (Daumé-III and Jagarlamudi, 2011; Kim and Sarikaya, 2015). The majority of researches done on multi-domain are dealing with domains structured within DBMS(Lefevre et al., 2012) such as (Information on the schedules of trains, planes, tourism, car navigation, weather information, Guide of TV program, chat, etc). We aim to provide a portable system, with minimal intervention from experts, across four domains. Three domains are based on information extraction, which are *Medical Diagnostic*, *Diverse Consultation* and *Question-Answering (DAWQAS)*[1] domains (Ismail and Homsi, 2018), in addition to the *University Schooling Management* domain which is based on database information retrieval. In this paper, we first present, in section 2, an SLU system based on thematic approach, followed by a description of the feature selection process as well as the dataset we prepared. In section 3, we present experiments and the corresponding results, and we conclude in section 4.

[1]A Dataset for Arabic Why Question Answeing System

DBMS Information Retrieval	University Schooling Management Domain	على كم تحصلت في مادة المجال الكهرومغناطيسي
		How much I got in the electromagnetic field module
Information Extraction	Medical Diagnostic Domain	لقد أغمي علي وانا أشعر بالتوتر لأن نبضات قلبي سريعة
		I'm fainting and I feel nervous because my heartbeat is fast
	Consultation Domain	هل يمر الإنسان بدورات نفسية متعاقبة ليس لها علاقة بالظروف ؟
		Does the person undergo successive psychological courses that have nothing to do with the circumstances?
	Question-Answering Domain (DAWQAS)	لماذا التحدث عن نقاط ضعفك خلال مقابلة العمل أمر رائع بالنسبة لك
		Why talking about your weaknesses during a job interview is great for you

Table 1: Samples of requests related to the four domains.

2 Spoken Language Understanding

The SLU system is based on some of the cognitive properties of humans which is tendency to understand an utterance in two different ways: Slot Filling and Intent Identification. Note that Slot Filling consists in identifying significant terms of this utterance followed by the identification of relationships between these terms, which leads him to understand the meaning of the utterance. While Intent Identification aims to identify the subject of the utterance without understanding the words one by one. In this work, we adopt Intent Identification to implement the SLU system, using text categorization (Lichouri et al., 2015, 2018b). The techniques used include statistical and neural methods: Multinomial Naive Bayes(MNB), Bernoulli Naive Bayes(BNB), Logistic Regression, Stochastic Gradient Descent(SGD), Passive Aggressive Classifier, Perceptron, Linear Support Vector Classification(LSVC) and Convolutional Neural Networks(CNN).

2.1 Feature Selection

We first processed the requests by removing all the punctuation. Then we conducted experiments, with and without stop words, in order to show the impact of Arabic stop words on intent identification which yields the request (sentence) intent. Second we used both word and character analyzers (Lichouri et al., 2018a) as an input to the vectorization process either by using TF-IDF for statistical classification or One hot encoder for CNN. We should note that we applied n-grams as features in the case of word analyzer.

2.2 Data acquisition and description

In this section, we will present a description of the corpus related to the four domains. For *University Schooling Management* which is a DBMS Information Retrieval Domain, We collected from around 300 students which formulated their requests to access their information from the education office. After discarding the repeated requests, we obtained a corpus made of 127 different requests expressed in French. The collected corpus, which was initially in French, was translated manually by experts to Arabic (**?**). Some examples of these queries are given in the table 1. These queries express what do students request from the office of education such as Marks, Certificates and Diplomas. The second domain which is *Medical Diagnostic*, We collected a corpus from a medical care forum known as Doctissimo (Alexandre, 2000). Some examples of these queries are also given in the table 1. These queries express the symptoms and feelings of ill people describing their health states to a doctor on the forum so that he could administer their treatment or the advice to give. We choose seven diseases, namely: Allergy, Anemia, Bronchitis, Diarrhea, Fatigue, Flu and Stress. For the *Consultation* domain, We collected the dataset from Islamtoday website (Today, 2000). It contains four main tasks which are: Educational, Psychological, Social and Religion Consulting. An example of this corpus is presented in table 1. We have shared the first two corpora (University Schooling Management and Medical Diagnostic) in a github repository[2]

[2]https://github.com/licvol/Arabic-Spoken-Language-Understanding

for research purpose, where as the third will be shared in our future works. The fourth corpus related to Question-Answering domain, we used the DAWQAS[3] corpus which contains a set of QA couples including 13 tasks, which are: Animal, Art and Celebrities, Community, Food, Health, Nature, Philosophy, Politics, Religion, Science and Technology, Space, Sports, and Women. More details of the datasets related to the four domains are summarized in table 2.

Corpus	School	Medical	Consultation	DAWQAS
#Sentence	126	152	3541	2525
#word	700	866	400.972	19.836
#class	3	7	4	13

Table 2: Description of the four used corpora.

3 Experiments and results

We conducted experiments on SLU portability between two kinds of domains: DBMS Information Retrieval and Information Extraction. The request is considered to be well understood if it is assigned a correct category. We achieved a comparison between statistical methods (Pedregosa et al., 2011) and neural method[4]. The training has been achieved on 70% of the shuffled datasets and the testing on the rest of dataset. For the CNN, we considered two tests with 10 and 100 iteration, respectively. We compared the performance of the classifiers by combining the different sets of features. Figures 1 and 2 represent the different values of F1-score obtained using the different classifiers, where SW_r, CA, WA_u, WA_b, WA_t and SW stand for, respectively, stop words removal, using character analyzer, word analyzer on unigram, bigram, trigram and using stop words. We should note that each combination of the aforementioned features is attributed a number (from 1 to 8) where: 1=SW+WA_u, 2=SW+WA_b, 3=SW+WA_t, 4=SW+CA, 5=SW_r+WA_u, 6=SW_r+WA_b, 7=SW_r+WA_t and 8=SW_r+CA.

We can see that the average of F1 measure is around 63%, 25%,39% and 32% for the School, Medical, Consultations and DAWQAS domains, respectively. Whereas the maximum values of F1 scored for the four domains are: 100%, 54%, 74% and 63%. In addition, it is noticeable through results shown in tables 3, 4, 5 and 6 that it is unclear which features combination yields the best performance. For instance, the absence of stop words gives the best performance for SGD while it doesn't for other classifiers.

[3]https://github.com/masun/DAWQAS

[4]https://github.com/tensorflow/workshops/blob/master/extras/keras-bag-of-words/keras-bow-model.ipynb

	Best results(%)			Feaures		
	Prec	Recall	F1	Stop Words	Analyzer	n-gram
MNB	86	84	84	Yes	Word	1
BNB	90	89	89	Yes/No	Char	-
LSVC	98	97	97	Yes/No	Word	1
LogReg	81	71	67	Yes	Word	1
SGD	98	97	97	Yes/No	Word	1
PassAgg	98	97	97	Yes/No	Word	1
Perceptron	100	100	100	Yes/No	Word	1
CNN	95	95	95	No	Word	1

Table 3: Best performance for the School domain

As shown in table 3, in the case of School application, the best performance was achieved by the Perceptron classifier, with a perfect result by using a word analyzer with or without Arabic stop words. Whereas in table 4, for the medical application, the best result was performed by the SGD classifier, with an F1-score of 54% by also using the word analyzer and without removing the Arabic stop words.

	Best results(%)			Features		
	Prec	Recall	F1	Stop Words	Analyzer	n-gram
MNB	64	46	42	Yes	Word	1
BNB	21	26	23	Yes/No	Char	-
LSVC	66	52	49	No	Word	1
LogReg	60	43	39	Yes	Word	1
SGD	66	57	54	No	Word	1
PassAgg	61	52	52	Yes	Word	1
Perceptron	53	46	46	No	Word	1
CNN	74	39	47	No	Word	1

Table 4: Best performance for the Medical domain

Table 5 shows results for the Consultations domain. Note that both SGD and Logistic Regression classifiers achieved the best F1-score of 74% by using word analyzer. The SGD has performed equally by using either a unigram or bigram as input for the word analyzer, where the Logistic Regression has performed better with the trigram as an input.

For the last application related to DAWQAS corpus, the best results have been achieved with both LSVC and Passive Aggressive classifiers with F1-score of 63%. The first one has achieved equally by either filtering or not the Arabic stop words in plus to applying the word analyzer with a unigram as input. For the latter classifier, the same analyzer was used but without filtering the

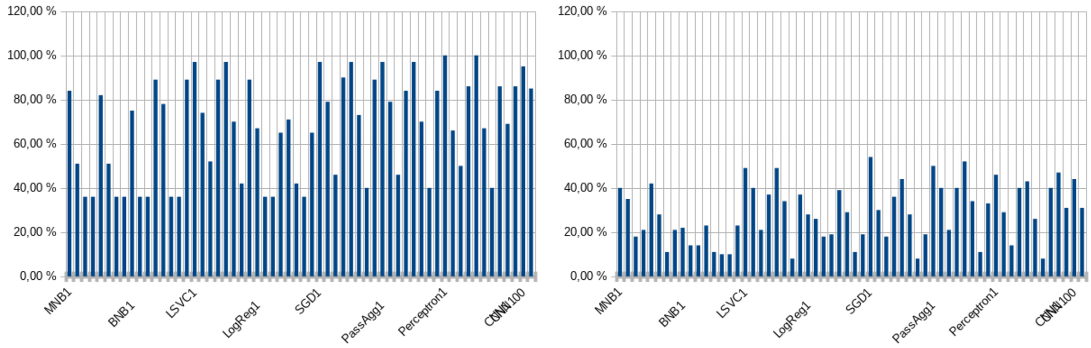

Figure 1: F1-score of the two domains: School (above), Medical (below).

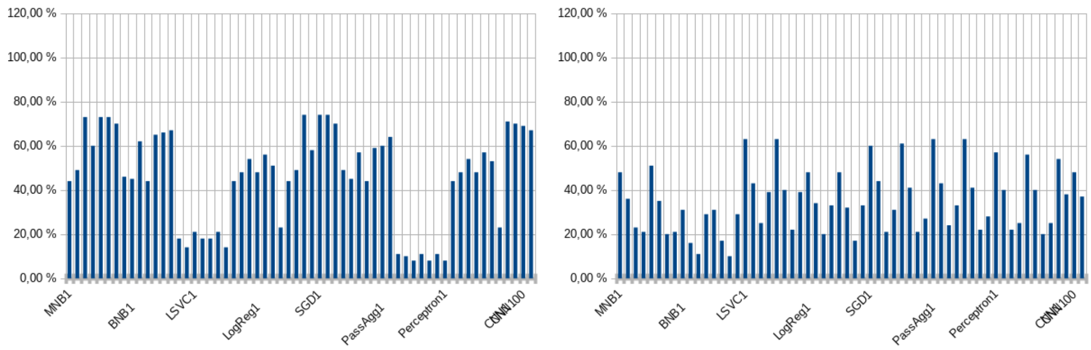

Figure 2: F1-score of the two domains: Consultations (above), DAWQAS (below).

	Best results(%)			Features		
	Prec	Recall	F1	Stop Words	Analyzer	n-gram
MNB	73	75	73	No	Word	2;3
BNB	69	67	67	Yes	Word	2
LSVC	55	62	54	Yes	Char	-
LogReg	74	75	74	Yes	Word	3
SGD	74	75	74	No	Word	1;2
PassAgg	65	65	64	No	Word	2
Perceptron	55	60	57	Yes	Word	2
CNN	73	69	71	No	Word	1

Table 5: Best performance for the Consult domain

Arabic Stop words.

	Best results(%)			Features		
	Prec	Recall	F1	Stop Words	Analyzer	n-gram
MNB	60	57	51	Yes	Word	1
BNB	40	42	31	No	Word	1
LSVC	64	64	63	Yes/No	Word	1
LogReg	57	54	48	No	Word	1
SGD	62	62	61	Yes	Word	1
PassAgg	64	64	63	No	Word	1
Perceptron	58	58	57	No	Word	1
CNN	57	53	54	No	Word	1

Table 6: Best performance for the DAWQAS domain

By comparing the performance of the different classifiers for the four domains, we can conclude that (i) the Arabic Stop words change the meaning or intent of utterance according the task and the domain. (ii) There is no perfect classifier to perform an acceptable SLU portability across domains, especially for the Arabic language, which is known for its richness at the lexical level.(iii) There is not a perfect size for a corpus to be considered when porting to a new domain. Indeed, performance for Consult domain is better than DAWQAS though the Consult corpus is smaller.

4 Conclusion and Perspective

This paper is a modest contribution to the ongoing research about the generalization of a Spoken Language Understanding System in a multi-domain Human-Machine Dialog. To our knowledge, this is the first study to investigate the possibility of a portable SLU system across domains, especially for the Arabic Language. The findings were quite interesting since the F1 scores obtained from experiments to adapt the Schooling Management domain to Medical, Consultations and DAWQAS were 54%, 74% and 63%, respectively.

References

Drs Claude Malhuret Laurent Alexandre. 2000. Sant et bien łtre avec doctissimo. http://www.doctissimo.fr/. [Online; accessed 07/08/2018].

C. Chelba and A. Acero. 2006. Adaptation of maximum entropy capitalizer: Little data can help a lot. *Computer Speech & Language*, 20(4):382–399.

H. Daumé-III. 2007. Frustratingly Easy Domain Adaptation. In *Proc. of the Annual Meeting of the Association for Computational Linguistics (ACL)*, Columbus, Ohio, USA.

Hal Daumé-III and Jagadeesh Jagarlamudi. 2011. Domain adaptation for machine translation by mining unseen words. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*, pages 407–412.

Hal Daumé-III and Daniel Marcu. 2006. Domain adaptation for statistical classifiers. *Journal of Artificial Intelligence Research*, 26.

Walaa Saber Ismail and Masun Nabhan Homsi. 2018. Dawqas: A dataset for arabic why question answering system. *Procedia computer science*, 142:123–131.

A Jaech, L Heck, and M Ostendorf. 2016. Domain adaptation of recurrent neural networks for natural language understanding. In *http://arxiv.org/abs/1604.00117*.

Ridong Jiang, Rafael E Banchs, Seokhwan Kim, Kheng Hui Yeo, Arthur Niswar, and Haizhou Li. 2014. Web-based multimodal multi-domain spoken dialogue system. In *Proceedings of 5th International Workshop on Spoken Dialog Systems*.

Sangkeun Jung, Cheongjae Lee, Kyungduk Kim, Minwoo Jeong, and Gary Geunbae Lee. 2009. Data-driven user simulation for automated evaluation of spoken dialog systems. *Computer Speech & Language*, 23(4):479–509.

Y Kim and R Sarikaya. 2015. New Transfer Learning Techniques For Disparate Label Sets. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, Beijing, China.

Kazunori Komatani, Satoshi Ikeda, Tetsuya Ogata, and Hiroshi G Okuno. 2008. Managing out-of-grammar utterances by topic estimation with domain extensibility in multi-domain spoken dialogue systems. *Speech Communication*, 50(10):863–870.

Cheongjae Lee, Sangkeun Jung, Seokhwan Kim, and Gary Geunbae Lee. 2009. Example-based dialog modeling for practical multi-domain dialog system. *Speech Communication*, 51(5):466–484.

Fabrice Lefevre, Djamel Mostefa, Laurent Besacier, Yannick Esteve, Matthieu Quignard, Nathalie Camelin, Benoit Favre, Bassam Jabaian, and Lina Maria Rojas Barahona. 2012. Leveraging study of robustness and portability of spoken language understanding systems across languages and domains: the portmedia corpora. In *The International Conference on Language Resources and Evaluation*.

Mohamed Lichouri, Mourad Abbas, Abed Alhakim Freihat, and Dhiya El Hak Megtouf. 2018a. Word-level vs sentence-level language identification: Application to algerian and arabic dialects. *Procedia Computer Science*, 142:246–253.

Mohamed Lichouri, Amar Djeradi, and Rachida Djeradi. 2015. A new automatic approach for understanding the spontaneous utterance in human-machine dialogue based on automatic text categorization. In *Proceedings of the International Conference on Intelligent Information Processing, Security and Advanced Communication*, page 50. ACM.

Mohamed Lichouri, Rachida Djeradi, and Amar Djeradi. 2018b. Combining topic-based model and text categorisation approach for utterance understanding in human-machine dialogue. *International Journal of Computational Science and Engineering*, 17(1):109–117.

Bing Liu and Ian Lane. 2016. Joint online spoken language understanding and language modeling with recurrent neural networks. *arXiv preprint arXiv:1609.01462*.

Wolfgang Minker. 1998. *Speech Understanding for Spoken Language Systems: Portability Across Domains and Languages*. Hänsel-Hohenhausen.

Mikio Nakano, Shun Sato, Kazunori Komatani, Kyoko Matsuyama, Kotaro Funakoshi, and Hiroshi G Okuno. 2011. A two-stage domain selection framework for extensible multi-domain spoken dialogue systems. In *Proceedings of the SIGDIAL 2011 Conference*, pages 18–29. Association for Computational Linguistics.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830.

Islam Today. 2000. Alistisharat. http://www.islamtoday.net/istesharat/index.htm. [Online; accessed 21/11/2018].

Zhuoran Wang, Hongliang Chen, Guanchun Wang, Hao Tian, Hua Wu, and Haifeng Wang. Policy learning for domain selection in an extensible multi-domain spoken dialogue system.