تحديد اللهجات المغاربية المنطوقة باستخدام طريقة التعلم المنقول

Identification of Spoken Maghrebi Dialects using a Transfer Learning Approach

مراد عباس[1]، محمد ليشوري[2]، خالد لوناس[2]

[1] المجلس الأعلى للغة العربية (الجزائر)، abbas.hcla@gmail.com

[2] جامعة هواري بومدين للعلوم والتكنولوجيا (الجزائر)، mlichouri@usthb.dz

[3] جامعة هواري بومدين للعلوم والتكنولوجيا (الجزائر)، Klounnas@usthb.dz

تاريخ الإرسال: 2024/05/21 تاريخ النشر: جوان/2024

ملخص

نقدم في هذا المقال طريقة التعلم المنقول التي انتهجناها لحل مسألة تحديد بعض اللهجات المنطوقة ضعيفة الموارد في منطقة المغرب العربي، تتمثل في العربية والبربرية في كل من الجزائر والمغرب.

استخدمنا نموذجين للتعلم المنقول، الشبكة العصبية المتبقية (Resnet101)، والمجموعة الهندسية المرئية (VGG16) باستخدام مدونة قمنا بإنجازها لكل لهجة. وتتكون المدونة من عشرة أرقام منطوقة باللهجات المذكورة آنفا من قبل ستة متحدثين أصليين مكررة عشر مرات.

حصلنا على نتائج مختلفة وفقًا لأسباب متنوعة منها عدد الدورات، والخلايا العصبية، وحجم الحزمة، وكذلك مجموعات البيانات المستخدمة في مراحل التدريب والاختبار. كانت النتيجة التي حققناها بشكل عام باستخدام النموذج VGG16 بمعدل تحديد متوسط قدره 62.7%.

الكلمات المفتاحية: تحديد اللهجة، ف ج ج 16، ريسنيت 101

Abstract:

This paper investigates a transfer learning approach to solve the spoken dialects identification problem for some under-resourced dialects of the Maghrebi region, including Arabic and Berber Dialects in Algeria and Morocco.

In our experiments, we used two Transfer learning models, namely: Residual Neural Network (Resnet101) and Visual Geometric Group (VGG16) using an in-house corpus that we built for each dialect. The corpus is composed of ten digits recorded for each of the aforementioned dialects, repeated ten times by six native speakers.

The results vary according to different reasons: the number of epochs, neurons, batch size, and the datasets combinations used in training and test phases. Overall, the results show the robustness of our system based on the VGG16 model with an average identification rate of 62.7%.

Keywords : Dialect Identification, Resnet101, VGG16

المؤلف المرسل: مراد عباس ، الإيميل المهني: abbas.hcla@gmail.com

مقدمة

يتم تحديد اللغة أو اللهجة المنطوقة باستخدام التعلم الآلي الذي ينسب الكلام المنطوق إلى لهجته المناسبة من بين مجموعة من اللهجات. إن التعامل مع اللهجات أكثر صعوبة منه بالنسبة للغات، خاصة عندما تكون اللهجات قريبة جدًا من بعضها البعض ومتشابهة في سياقات مختلفة، بما في ذلك المستويات الصوتية والصرفية والمعجمية والنحوية. لهذا الغرض، انصب اهتمامنا على بناء مدونة للهجات المغاربية، لاستخدامها بغرض إجراء مجموعة من التجارب القائمة على طرق التعلم المنقول من أجل الوصول إلى تكوين الشبكة العصبية الأكثر جودة.

لقد أجريت أبحاث عديدة لتحديد اللهجات المنطوقة، استخدم في بعضها النموذج الكلاسيكي القائم على التصنيف الإحصائي، وفي البعض الآخر تم استعمال تقنيات التعلم العميق.

باتباع النموذج الأول وقصد العثور على متجه الخصائص المناسب والتصنيف الهرمي، قام باحثون بدمج المعلومات الصوتية والنغمية للتمييز بشكل أفضل بين العربية والقبائلية [1-3]. وفي نفس السياق، أدى استخدام الإشارات النغمية إلى أداء أفضل لتحديد أربع لهجات عربية رئيسية، هي: الخليجية، والعراقية، والشامية، والمصرية. إذ ثبت أن استخدام هذا النوع من الواصفات لتدريب نموذج الخليط الجاوسي (GMM) مع نموذج الخلفية الشامل (UBM) بإمكانه تحديد هذه اللهجات بطريقة أكثر فعالية، وذلك اعتمادا على كلام منطوق لا تتجاوز مدته دقيقتان [4]. كما أجريت أبحاث لتحديد اللهجة واللكنة العربية باستخدام المتجهات الفائقة للتقسيم الصوتي، والتي تقوم أساسا على بناء دالة النواة التي تحسب التشابهات الصوتية، بغرض تدريب مصنف آلات المتجهات الداعمة، وقد أدت هذه الطريقة إلى معدل خطأ متساو (EER) قدره 12.9% [5]. وقد تم العمل على تحديد اللهجات العربية المنطوقة وهي: المصرية، والخليجية، والشامية، والمغاربية بالإضافة إلى العربية الفصحى الحديثة بدقة تساوي 51.36% [6]، باستخدام آلات المتجهات الداعمة الخطية التي تم تدريبها بالاعتماد على متجه للخصائص يحتوي على معلومات وتفاصيل نصية.

وفيما يتعلق بأنظمة التعلم العميق لتحديد اللهجات العرصؤبية المنطوقة واللهجات الفرعية، اقترح الباحثون معاملات نغمية لنمذجة بعض اللهجات الجزائرية شحيحة الموارد [7-8]، حيث أجروا دراسة مقارنة بين الطرق الإحصائية وطرق الشبكة العصبية الاصطناعية، واستنتجوا أن الشبكة العصبية الترشيحية CNN أكثر جودة من الطرق الإحصائية.

نظام تحديد اللهجات

يعتمد نظامنا على نماذج التعلم المنقول، بما في ذلك الشبكة العصبية المتبقية (ResNet) والمجموعة الهندسية المرئية (VGG). قمنا بإجراء تجاربنا باستخدام معاملات متنوعة ومجموعات مختلفة من مجموعة البيانات. بادئ ذي بدء، استخرجنا الصور الطيفية لاستخدامها كمدخل لنظام تحديد اللهجات.

ثم شكلنا عشر مجموعات تدريبية واختبارية مختلفة باستخدام مجموعات متعددة من المتحدثين. تتكون كل مجموعة من أربعة متحدثين للتدريب (70%) ومتحدثين اثنين للاختبار (30%)، كما هو موضح في الجدول 1، حيث تمثل S_i المتحدث رقم i.

الجدول 1. مجموعة بيانات متعددة

المجموعة	مجموعة التدريب	مجموعة الاختبار
01	S_1، S_2، S_3 ، S_4	S_5 ،S_6
02	S_5،S_6 ، S_1، S_2	S_3 ،S_4
03	S_3،S_4 ، S_6، S_5	S_1 ،S_2
04	S_3،S_4 ، S_6، S_2	S_1 ،S_5
05	S_2،S_4 ، S_1، S_5	S_3 ،S_6
06	S_2،S_1 ، S_3، S_5	S_4 ،S_6
07	S_3،S_1 ، S_4، S_6	S_2 ،S_5
08	S_4،S_2 ، S_5، S_6	S_1 ،S_3
09	S_3،S_2 ، S_5، S_6	S_1 ،S_4
10	S_1،S_3 ، S_5، S_6	S_2 ،S_4

1.2تحضير المدونة

المدونة (أو مجموعة البيانات) هي حجر الأساس الذي نعتمد عليه لتنفيذ مجموعة من التجارب بغرض إنجاز نظام تحديد اللهجات المغاربية المنطوقة، لذلك قمنا ببناء مدونة تحتوي على أربع لهجات من منطقة المغرب العربي وهي: اللهجة العربية الجزائرية (AAD)، اللهجة البربرية الجزائرية (ABD)، اللهجة العربية المغربية (MAD) واللهجة البربرية المغربية (MBD). حيث تم تسجيل الأرقام من 0 إلى 9 عشر مرات من طرف متحدثين جزائريين. أما بيانات اللهجة المغربية فهي مأخوذة من المجموعة المستعملة في [9-10]. يبين الجدول 2 خصائص المدونة.

الجدول 2. خصائص المدونة

معدل أخذ العينات	16 كيلوهرتز
عدد البتّات	16 بت
عدد القنوات	1، mono
الملف الصوتي	wav
# الناطقين	24
#الناطقين لكل لهجة	6
# اللهجات	4
#التسجيلات لكل ناطق	100
#الناطقين حسب الجنس	12ذكور و 12 إناث
مجموع التسجيلات	400
عدد الأرقام المسجلة	10أرقام (اللهجة العربية الجزائرية) 10 أرقام (اللهجة البربرية الجزائرية) 10 أرقام (اللهجة العربية المغربية) 10 أرقام (اللهجة البربرية المغربية)
عدد التكرارات لكل كلمة	10
نافذة هامينج	25.6ميلي ثانية
التداخل	10ميلي ثانية

2.2الشبكة العصبية المتبقية

أجرينا سلسلة من التجارب لاختبار معاملات الضبط المختلفة لشبكة الطبقتين الأخيرتين (NN2). تُستخدم هذه الشبكة لإعادة تدريب نماذج الشبكة العصبية المتبقية ResNet101 المدربة مسبقًا. في إطار هذه التجارب، قمنا بتغيير عدد الخلايا العصبية وعدد الدورات وحجم الحزمة، كما هو موضح في الجدول 3.

الجدول 3: معاملات التكوينات (عدد الدورات والخلايا العصبية وحجم الحزمة في الطبقة الأولى من شبكة NN2)

التكوين	# الدورات	حجم الحزمة	#الخلايا العصبية
Cf1	2	20	128
Cf2	2	20	256
Cf3	10	10	128
Cf4	10	10	256

قمنا بتدريب شبكة ResNet بعدد كبير من الطبقات (101 طبقة). أفضل النتائج التي تم تحقيقها باستخدام تكوينات cf1 وcf2 هي 49.56% و63.59% باستخدام المجموعة الثالثة والأخيرة، على التوالي. بالنسبة لتكوينات cf3 وcf4، لاحظنا تحسنًا كبيرًا في كلتا الحالتين في المجموعة الثانية مقارنة بتكوينات cf2، مع تحسين بنسبة 15% و20% على التوالي. يمكن ملاحظة النتائج في الجدول 4 التي تم الحصول عليها في التكوينات الأربعة (cf1 وcf2 وcf3 وcf4)، إذ أن أداء تحديد اللهجة المستند إلى ResNet يرتبط بمعاملات الضبط للشبكة المذكورة أعلاه بالإضافة إلى عدد الخلايا العصبية في الطبقة الأولى من شبكة NN2. في الواقع، إن زيادة عدد الخلايا العصبية (من 128 إلى 256) له تأثير إيجابي على أداء النظام. كما أن زيادة عدد الطبقات له نفس التأثير عند استخدام تكوين cf2.

الجدول 4. أداء الشبكة العصبية المتبقية (مقياس إف1) ResNet101

المجموعة	Cf1	Cf2	Cf3	Cf4
01	33.59	35.32	54.99	56.58
02	33.95	51.40	77.54	83.83
03	49.56	48.05	62.42	52.37
04	33.30	50.94	61.04	60.15
05	33.66	33.99	61.73	58.45
06	23.90	33.27	56.27	59.38
07	41.11	32.97	60.84	34.35
08	39.76	37.50	39.59	62.03
09	25.05	58.26	60.87	56.98
10	39.73	63.59	38.50	58.52
متوسط	35,36	44,53	57,38	58,26

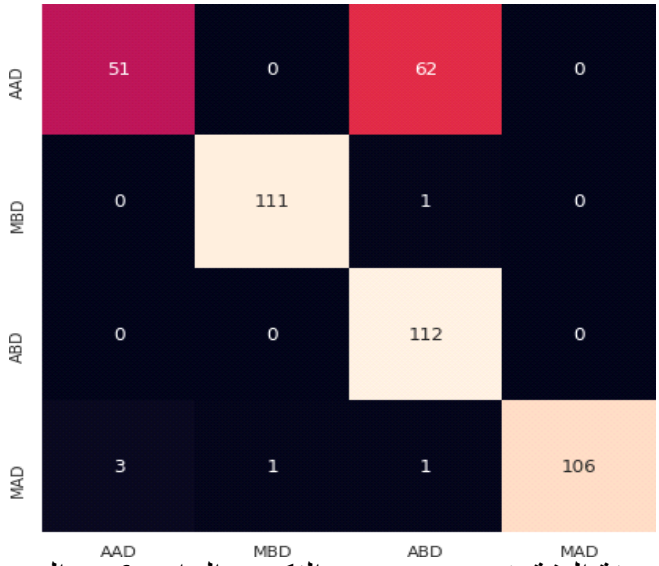
3.2تحديد اللهجة باستخدام المجموعة الهندسية المرئية (VGG)

المجموعة الهندسية المرئية عبارة عن شبكة عصبية ترشيحية ذات بنية محددة تم اقتراحها في [11] تتكون من أجزاء، وكل جزء يتكون من طبقات ترشيح ثنائية الأبعاد وطبقات تجميع قصوى. ويعكس الرقمان 16 و19 عدد الطبقات في النموذجين VGG16 وVGG19. نعرض في هذا الجزء التجارب التي قمنا بإجرائها باستخدام نماذج VGG16 المدربة مسبقا والتي اعتمدنا من أجلها نفس التكوينات الواردة في الجدول 3.

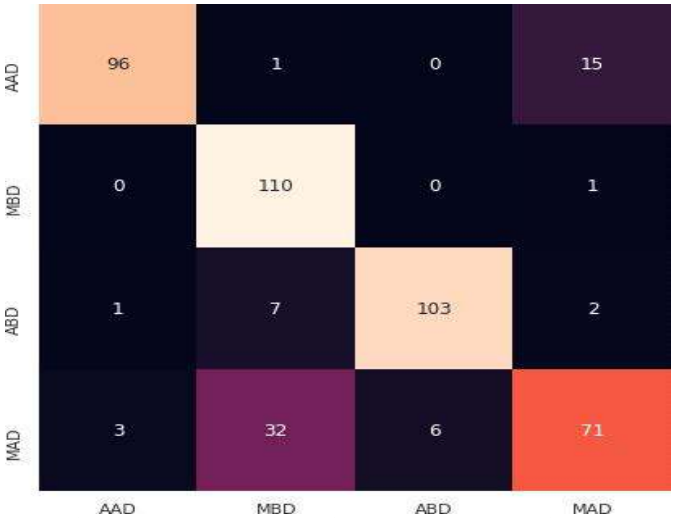
بالنسبة لنموذج VGG16+NN2، تم الحصول على أفضل نتيجتين مع التكوين الرابع والثالث باستخدام المجموعة الثانية، بنتيجة 84.95% و79.43% على التوالي. إذ أن استخدام المجموعة التاسعة أدى إلى نتيجة 54.37% باعتماد التكوين الأول. علاوة على ذلك، سمح التكوين الثاني بتحسين حوالي 7% باستخدام مجموعة مختلفة، وهي المجموعة الرابعة. بناء على هذا الأداء الذي حققناه باستخدام التكوينات الأربعة المذكورة أعلاه، نلاحظ بوضوح أنه كلما كان عدد الخلايا العصبية منخفضا، كلما كان النظام أقل أداء. كما أن أفضل نتيجة تم الحصول عليها باستخدام التكوين الرابع (الحد الأقصى: 84.95%، المتوسط: 55.47%) بينما تم الحصول على أفضل متوسط أداء مع التكوين الثالث (الحد الأقصى: 79.43%، المتوسط: 62.70%). ومع ذلك، يبدو أن VGG16+NN2 بالتكوين الثالث أكثر ملاءمة وأكثر استقرارًا مقارنة بالتكوينات الثلاثة الأخرى. يمثل الجدول 5 قيم إف1 التي تم الحصول عليها. كما يوضح الشكلان 1 و2 مصفوفتي الدقة لكل من الطريقتين ResNet101 و VGG16.

الجدول 5. أداء VGG16 باستخدام مقياس إف1

المجموعة	Cf1	Cf2	Cf3	Cf4
01	48,41	44,97	57,45	49,39
02	45,07	27,52	**79,43**	**84,95**
03	20,50	44,42	60,64	37,61
04	49,35	**61,48**	53,31	49,93
05	18,43	34,08	58.06	55.40
06	37,65	26,02	59.53	51.10
07	18,83	35,93	57.49	47.27
08	25,39	23,39	49.71	62.97
09	**54,37**	48,30	57.47	59.60
10	34,50	35,61	40.79	33.08
متوسط	35,25	38,17	**62,70**	55,47



الشكل :1 مصفوفة الدقة (ResNet101 التكوين الرابعcf4 ، المجموعة الثانية)



الشكل :2 مصفوفة الدقة) VGG16 التكوين الرابعcf4 ، المجموعة الثانية(

أظهرت النتائج أن أداء نموذج VGG16+NN2 الذي يستخدم عددًا أكبر من" معاملات التعلم " مقارنة بنموذج ResNet101+NN2 ، أفضل من أداء هذا الأخير. يمكن الاطلاع على الفرق في عدد المعاملات في الجدول.6

الجدول 6. عدد المعاملات لكل نموذج

النموذج	عدد المعاملات
ResNet101+NN2	M44.5
VGG16+NN2	M138

الخلاصة

قمنا في هذا العمل بتطبيق منهج التعلم المنقول لتحديد اللهجات المنطوقة في المغرب العربي، وتحديدا اللهجات الجزائرية والمغربية، حيث قمنا باستخدام خصائص المخطط الطيفي لإعادة تدريب بعض النماذج المدربة مسبقًا. تم إجراء سلسلة من التجارب باستخدام النموذجين (ResNet 101 وVGG 16، على التوالي) وتغيير القيم المتعلقة بعدد الدورات وحجم الحزمة. كما قمنا أيضًا بتنويع عدد الخلايا العصبية للشبكة ذات الطبقتين (NN2) والتي تم استخدامها لإعادة تدريب النموذجين المدربين مسبقًا. وتظهر النتائج أنه على الرغم من صغر حجم مجموعة البيانات المستخدمة، فإن VGG تفوق على النماذج الأخرى. ومع ذلك، فإن الأداء الذي تم تحقيقه باستخدام VGG يعتمد على تكلفة وقت الحساب ومساحة الذاكرة. عكس نماذج ResNet المدربة مسبقًا.

References:

1-Lounnas ,K ,.Demri ,L ,.Falek ,L & ,.Teffahi ,H. (2018, October .(automatic lan -guage identification for berber and arabic languages using prosodic features .In 2018 International Conference on Electrical Sciences and Technologies in Maghreb) (CISTEM) pp .(4-1 .IEEE.

2- Lounnas, K., Abbas, M., Teffahi, H., & Lichouri, M. (2019, March). A Language Identification System Based on Voxforge Speech Corpus. In International Confer- ence on Advanced Machine Learning Technologies and Applications (pp. 529-534). Springer, Cham.

3- Lounnas, K., Satori, H., Teffahi, H., Abbas, M., & Lichouri, M. (2020, April). CLIASR: A Combined Automatic Speech Recognition and Language Identification System. In 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET) (pp. 1-5). IEEE.

4- Biadsy, F., & Hirschberg, J. (2009). Using prosody and phonotactics in arabic di- alect identification. In Tenth Annual Conference of the International Speech Com- munication Association.

5- Biadsy, F., Hirschberg, J., & Ellis, D. P. (2011). Dialect and accent recognition using phonetic-segmentation supervectors. In Twelfth Annual Conference of the International Speech Communication Association.

6- Eldesouki, M., Dalvi, F., Sajjad, H., & Darwish, K. (2016, December). Qcri@ dsl 2016: Spoken arabic dialect identification using textual features. In Proceedings of the Third Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial3) (pp. 221-226).

7- Bougrine, H. C. S., & Abdelali, A. (2018, April). Spoken arabic algerian dialect identification. In 2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP) (pp. 1-6). IEEE.

8- Lounnas, K., Abbas, M., & Lichouri, M. (2019). Building a Speech Corpus based on Arabic Podcasts for Language and Dialect Identification. In Proceedings of the 3rd International Conference on Natural Language and Speech Processing (pp. 54-58).

9- Satori, H., Harti, M., & Chenfour, N. (2007). Syst`eme de Reconnaissance Automa- tique de l'arabe bas´e sur CMUSphinx. Corpus, 22, 25.

10- Satori, H., & Elhaoussi, F. (2014). Investigation Amazigh speech recognition using CMU tools. International Journal of Speech Technology, 17(3), 235-243.

11- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large- scale image recognition. arXiv preprint arXiv:1409.1556.