Workin piepa · In-house review | v1.0 | 2026-11-20

How much example ina Jamaican Patwa mek di gap between two language model get smaller?

A second look pon nine way fi train di model dem fi di JamPatoisNLI reasoning test

Adrian Dunkley | info@caribbeanaiassociation.com

Wa di study find

Wi check if one model weh train inna English keep im lead over one weh work wid plenty language, when both a dem get labelled Jamaican Patwa example fi learn from. Wi work out di difference between di percentage a answer weh each model get right across nine test condition. English RoBERTa did lead XLM-RoBERTa by 11.50 percentage point wid no local pair fi training, an by 1.33 point wid 249 pair. Di gap grow likkle bit at some a di size dem in between. Di test measure how di model dem reason bout written sentence. It cyaan tell wi how well dem recognise speech or carry on a conversation. [1]

Weh di risorch a aks?

When di two model get di same Jamaican Patwa sentence dem fi test, how di gap between English RoBERTa an multilingual XLM-RoBERTa change as dem get more labelled local pair fi training?

Data an metad

Armstrong, Hewitt an Manning put sentence together inna pair fi a test wid three choice. Di model read one statement, den a second one. It haffi decide if di first statement mean di second one haffi true, if dem clash, or if di first one nuh give enough fi decide. Di researcher dem get di first statement dem from published writing, mostly pon social media. One fluent Patwa speaker write di second statement dem. Di dataset explain di three class dem an how di team split up di example dem. [1, 2]

Wi tek di two RoBERTa column inna Table 9 weh all di model setting dem coulda change during training. Both model did already train fi reason wid English text. Wi keep every training size di table give fi dem two model. Each unit have one example fi each a di three class dem, so wi multiply di number by three. Each score a di average from three run pon di same 200 sentence pair weh di researcher dem put aside fi testing. [1] Di last row say 83 group a three, weh come to 249 example by dis rule. Table 5 list 250 training pair available (84, 83 an 83 inna di three class), an di paper also call di full condition 250. Wi total count di balanced group a three. Di table nuh tell wi if di extra pair did go inna any one run. [1]

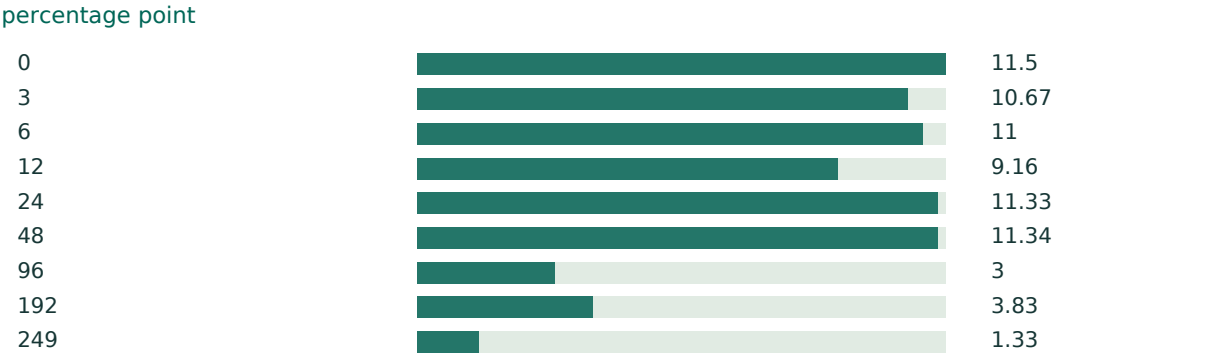
Fi each training size, wi tek weh di multilingual model percentage from di English model percentage. Wi find di first size dem test weh each model average reach 70% an den 75%. Dem two mark deh help describe weh di score dem do. Dem nuh set a standard fi every app.

Rizolt dem

Wid no local pair, English RoBERTa get 67.50% right an XLM-RoBERTa get 56.00%. Wid 249 pair, dem get 76.50% an 75.17%. Di English model gain 9.00 point an di multilingual one gain 19.17. Di difference between dem gain explain di 10.17 point weh come offa di original lead. [1]

Di gap reach 11.34 point wid 48 pair, drop to 3.00 wid 96, rise to 3.83 wid 192, den drop to 1.33 wid 249. Di first time di English model reach 70% a wid 48 pair, an fi 75% a wid 192. Di multilingual model reach dem same mark deh wid 96 an 249 pair. A size weh fall between di one dem test mighta reach di mark too. [1]

How far di English model lead at each local training size



Di number dem show how much Jamaican Patwa pair go inna training. Each bar a di English RoBERTa percentage right minus di XLM-RoBERTa percentage right. Both model do di same test. Source: Armstrong an colleague dem, Table 9 [1].

Wa di rizolt mean

If a team pick offa di result wid no local pair alone, di English model woulda look well ahead. If dem can label 249 example, di difference dem woulda see much smaller. Di result give dem reason fi try more than one amount a labelled data before dem choose a model fi dis work. It nuh tell dem which one fi use fi pick up di word dem when somebody call.

Di biggest drop between one training size an di next come between 48 an 96 pair. Somebody weh a repeat di study could try more size inna dat range. Drawing different group a example independently woulda help sort out weh come from di number a label an weh come from di particular sentence dem pick fi each run.

Limit a di study

Di model dem answer di same test item, an di published average join up repeated run. Fi work out a confidence interval weh keep each pair a prediction together, wi woulda need di individual prediction an di ID fi each run. Di 1.33-point lead a di gap wi observe at di end.

Di test measure reasoning pon one small collection a writing. Di sample never set out fi represent Jamaican speaker by parish, age or how dem use di language. Newer chat model, natural speech, different spelling an switching language inna one conversation need dem own test. [1, 2]

Run di analysis agen

Go a Table 9 inna [1]. Tek di last two model column fi every row. Keep dem inna di same order.

Multiply di first column by three. Tek weh di XLM-RoBERTa percentage right from di English RoBERTa percentage right. Round it to two place after di decimal point.

Read down each percentage column. Find di first row weh reach 70% or more, den di first weh reach 75% or more. Use di total local pair fi each comparison.

Data tiebl

Example fi each class	Local pair altogether	English RoBERTa answer right (%)	XLM-RoBERTa answer right (%)	Gap (percentage point)
0	0	67.5	56.0	11.5
1	3	68.17	57.5	10.67
2	6	69.17	58.17	11.0
4	12	66.83	57.67	9.16
8	24	68.83	57.5	11.33
16	48	70.17	58.83	11.34
32	96	73.0	70.0	3.0
64	192	76.33	72.5	3.83
83	249	76.5	75.17	1.33

Suos dem

[1] Armstrong, R.-A., Hewitt, J. and Manning, C. (2022). JamPatoisNLI: A Jamaican Patois Natural Language Inference Dataset. Findings of EMNLP, Tables 5 and 9.

[2] Armstrong, R.-A. JamPatoisNLI dataset and documentation.

Weh di data come fram

Armstrong, Hewitt & Manning (2022), Table 9. CAIA calculations. CC BY 4.0