Papye travay · Verifikasyon entèn | v1.0 | 2026-11-20

Konbyen egzanp nan patwa jamayiken ki fè diferans ant de modèl lang diminye?

Yon dezyèm analiz sou nèf kondisyon fòmasyon nan tès rezònman JamPatoisNLI a

Adrian Dunkley | info@caribbeanaiassociation.com

Rezime

Nou egzamine si yon modèl ki te aprann an anglè konsève avantaj li sou yon modèl ki sèvi ak plizyè lang lè yo toude resevwa egzanp patwa jamayiken ki gen etikèt. Nou rekalkile diferans ant pousantaj bon repons yo nan nèf kondisyon tès JamPatoisNLI a. RoBERTa anglè te devan XLM-RoBERTa ak 11.50 pwen pousantaj lè pa t gen pè fraz lokal pou fòmasyon. Avèk 249 pè, avantaj la te 1.33 pwen. Diferans lan te monte yon ti kras nan kèk etap. Tès la verifye sa yon modèl ka dedui nan fraz ekri. Li pa mezire jan modèl la rekonèt pawòl oswa jan li fè konvèsasyon. [1]

Kesyon rechèch la

Sou menm fraz tès an patwa jamayiken yo, ki jan diferans ant RoBERTa anglè ak XLM-RoBERTa ki sèvi ak plizyè lang te chanje lè kantite pè fraz lokal ki gen etikèt yo te ogmante?

Done ak metòd

Armstrong, Hewitt ak Manning prepare pè fraz pou yon tès ki gen twa chwa. Modèl la li yon premye deklarasyon ak yon dezyèm. Li deside si premye deklarasyon an pèmèt li dedui dezyèm nan, si de deklarasyon yo kontredi youn lòt, oswa si premye a pa sifi pou li deside. Premye deklarasyon yo soti nan tèks ki te pibliye, sitou sou rezo sosyal. Yon moun ki pale patwa jamayiken byen ekri dezyèm deklarasyon yo. Dokiman ki akonpaye done yo esplike kategori yo ak fason ekip la separe egzanp yo. [1, 2]

Nan tablo 9 la, nou pran de kolòn RoBERTa kote tout paramèt modèl yo te ka ajiste. Toude modèl yo te deja aprann fè dediksyon ak tèks anglè. Nou pran tout gwosè echantiyon fòmasyon yo bay pou de modèl sa yo. Chak inite gen yon egzanp pou chak nan twa kategori yo. Nou miltipliye kantite inite yo pa twa. Chak pousantaj se mwayèn twa esè ki te sèvi ak menm 200 pè fraz yo te mete apa pou tès la. [1] Dènye liy lan bay 83 gwoup twa, sa vle di 249 egzanp dapre règ sa a. Tablo 5 konte 250 pè ki disponib pou fòmasyon (84, 83 ak 83 pou chak kategori), epi otè yo pale tou de 250 egzanp pou kondisyon konplè a. Total nou yo konte gwoup twa ki balanse yo. Tablo a pa di si pè anplis la te antre nan yon esè. [1]

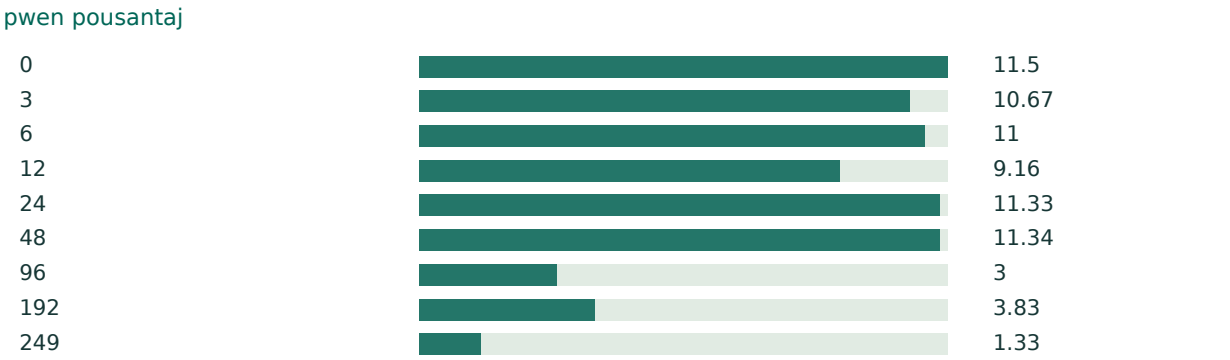
Pou chak gwosè, nou wete pousantaj bon repons modèl ki sèvi ak plizyè lang lan nan pousantaj modèl anglè a. Nou chèche tou premye gwosè yo teste kote mwayèn nan rive 70% epi 75% pou chak modèl. Pousantaj sa yo ede nou dekri rezilta yo. Yo pa fikse yon objektif pou yon aplikasyon.

Rezilta

San pè lokal, RoBERTa anglè bay 67.50% bon repons, epi XLM-RoBERTa bay 56.00%. Avèk 249 pè yo rive 76.50% ak 75.17%. Premye modèl la pran 9.00 pwen epi dezyèm nan pran 19.17. Diferans ant de ogmantasyon sa yo esplike poukisa avantaj la diminye 10.17 pwen. [1]

Diferans lan rive 11.34 pwen ak 48 pè, desann a 3.00 ak 96, monte a 3.83 ak 192 epi desann a 1.33 ak 249. Premye fwa modèl anglè a rive 70%, li te resevwa 48 pè. Pou 75%, li te resevwa 192. Modèl ki sèvi ak plizyè lang lan rive nan menm nivo yo ak 96 epi 249 pè. Yon kantite ant de gwosè yo pa t teste te ka bay menm nivo a tou. [1]

Avantaj modèl anglè a pou chak kantite pè lokal



Chif ki anba ba yo bay kantite pè patwa jamayiken modèl yo te sèvi pou fòmasyon. Chak ba bay pousantaj bon repons RoBERTa anglè mwens pa XLM-RoBERTa. Toude pase menm tès la. Sous: Armstrong ak kolaboratè li yo, tablo 9 [1].

Diskisyon

Yon ekip ki chwazi dapre rezilta san egzanp lokal la sèlman ta wè yon gwo avantaj pou modèl anglè a. Yon ekip ki kapab mete etikèt sou 249 egzanp ta wè yon pi piti diferans. Li ta bon pou ekip la teste plizyè kantite egzanp anvan li chwazi modèl pou travay sa a. Rezilta yo pa montre ki modèl ki ta rekonèt pawòl yon moun nan telefòn pi byen.

Pi gwo bès ant de etap ki suiv youn lòt rive ant 48 ak 96 pè. Moun k ap refè eksperyans lan ka teste plis gwosè nan entèval sa a. Si yo chwazi plizyè seri egzanp apa pou fòmasyon an, yo ka pi byen separe efè kantite egzanp yo ak efè fraz an patikilye ki te antre nan chak esè.

Limit etid la

Toude modèl yo reponn menm kesyon yo. Mwayèn ki pibliye yo melanje plizyè esè. Pou kalkile yon entèval konfyans sou diferans lan ki pran pè repons yo an kont, nou ta bezwen chak prediksyon ak nimewo esè li soti ladan an. Avantaj final 1.33 pwen an se yon diferans nou obsève.

Tès la mezire dediksyon nan yon ti ansanm tèks yo chwazi. Yo pa t fè echantiyonaj pou reprezante moun ki pale nan chak pawas Jamayik, chak gwoup laj oswa chak fason moun sèvi ak lang lan. Modèl konvèsasyon ki pi resan, pawòl espontane, varyasyon nan òtograf ak chanjman lang nan menm pawòl la mande lòt tès. [1, 2]

Refè analiz la

Nan tablo 9 sous [1], pran de dènye kolòn modèl yo pou chak liy. Kenbe menm lòd la.

Miltipliye premye kolòn nan pa twa. Wete pousantaj XLM-RoBERTa a nan pousantaj RoBERTa anglè a. Awondi jiska de chif apre pwen an.

Nan chak kolòn pousantaj, chèche premye liy ki rive omwen 70% epi premye liy ki rive omwen 75%. Sèvi ak kantite total pè lokal yo nan chak konparezon.

Tablo done

Egzanp pou chak kategori	Kantite pè lokal an total	Bon repons RoBERTa anglè (%)	Bon repons XLM-RoBERTa (%)	Diferans (pwen pousantaj)
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

Referans

[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.

Sous done yo

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