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

Chwa tèks pou tès la chanje nòt yon tradiksyon soti nan anglè pou rive nan kreyòl ayisyen

Yon konparezon nòt ki pibliye pou fraz òdinè, tèks sou relijyon ak istwa pou lekòl

Adrian Dunkley | info@caribbeanaiassociation.com

Rezime

Menm sistèm tradiksyon an ka jwenn nòt diferan selon tèks li resevwa pou tès la. Nou konpare nòt OPUS pou tradiksyon soti nan anglè pou rive nan kreyòl sou Tatoeba, JW300 ak MIT-Haiti. Apre sa nou konpare de modèl sou MIT-Haiti sèlman. Lè istwa pou lekòl yo ranplase Tatoeba, nòt OPUS la bese 30.5 pwen BLEU ak 23.4 pwen chrF. Lè nou chanje modèl sou menm istwa yo, diferans yo se 7.3 ak 8.1 pwen. Chif sa yo montre yon risk nan fason yon lekòl ka prepare tès tradiksyon li. Yo pa di nou ki tradiksyon elèv yo konprann. [1, 2]

Kesyon rechèch la

Pou tradiksyon anglè pou kreyòl, ki gwosè diferans chwa tèks tès la bay lè nou konpare li ak diferans ant de modèl sou menm tèks pou lekòl yo?

Done ak metòd

Nou kenbe yon sèl direksyon pou tout tradiksyon yo. Fich OPUS la bay nòt anglè pou kreyòl sou Tatoeba ak JW300. Lent ak kolaboratè li yo bay rezilta OPUS ak CreoleM2M sou menm 1,559 liy anglè pou kreyòl MIT-Haiti yo. Yo prepare materyèl edikatif nan plizyè lang apati 60 istwa ki sou Platfòm MIT-Ayiti. Nou sèvi ak nòt total yo. Fraz yo ak tradiksyon yo rete bò kote sous ki te bay yo. [1, 2, 3]

BLEU gade ki gwoup mo ki parèt ni nan tradiksyon modèl la ni nan referans lan. Li retire pwen pou tradiksyon ki twò kout. chrF konpare gwoup karaktè yo epi li konsidere ki pati nan tradiksyon an ki koresponn ak referans lan ak ki pati nan referans lan tradiksyon an jwenn. Fich OPUS la bay chrF kòm yon fraksyon. Nou miltipliye valè sa yo pa 100 pou sèvi ak menm echèl ak CreoleVal. Okenn nan de nòt yo pa bay pousantaj fraz ki byen tradui. [1, 2, 4, 5]

Nou kalkile twa diferans pou chak mezi. Nou wete nòt OPUS sou tèks lekòl yo nan nòt li sou Tatoeba. Nou refè kalkil sa a ak JW300. Finalman nou wete nòt OPUS sou tèks lekòl yo nan nòt CreoleM2M sou menm tèks yo. Nou konpare chak mezi apa sou pwòp echèl li.

Rezilta

Nan BLEU, OPUS jwenn 45.2 sou Tatoeba, 38.3 sou JW300 epi 14.7 sou MIT-Haiti. Nan chrF, li jwenn 59.2, 54.5 ak 35.8. Toude nòt yo desann lè tèks lekòl yo ranplase youn nan de lòt seri tèks yo. Diferans ant Tatoeba ak tèks lekòl yo se 30.5 pwen BLEU ak 23.4 pwen chrF. Lè nou sèvi ak JW300, diferans yo pi piti: 23.6 ak 18.7 pwen. [1, 2]

Sou MIT-Haiti, CreoleM2M jwenn 22.0 pwen BLEU ak 43.9 pwen chrF. Li devan OPUS sou menm liy yo ak 7.3 pwen BLEU ak 8.1 pwen chrF. Pou chak mezi, toude diferans ant seri tèks

yo pi gwo pase diferans ant modèl yo. Rezilta sa a kenbe lè JW300 pran plas Tatoeba nan konparezon an. [1, 2]

Nòt BLEU pou anglè pou kreyòl selon seri tèks la ak modèl la

pwen BLEU



Yon nòt ki pi wo vle di tradiksyon an gen plis gwoup mo ki koresponn ak referans lan. De dènye ba yo sèvi ak menm materyèl pou lekòl la. Tablo a bay tou nòt chrF ki konpare karaktè. Sous: CreoleVal, tablo 5 [1], ak fich OPUS la [2].

Diskisyon

Yon lekòl k ap chwazi yon lojisyèl bezwen konnen sa l ap bay sou materyèl lekòl la pral tradui. Nòt sou yon seri fraz òdinè ka bay yon move lide sou fason menm modèl la pral tradui istwa pou elèv. Menm lè nou sèvi ak JW300 pou konparezon an, diferans ant seri tèks yo pi gwo pase diferans ant de sistèm yo sou istwa yo.

Nan yon lòt evalyasyon, chèchè yo ta ka bay pwofesè ayisyen pasaj antye, epi poze de kesyon apa. Èske tradiksyon an kenbe sans lide y ap anseye a? Èske yon elèv ka suiv sa yo mande li fè a? Lè yon pwofesè li tout pasaj la, li ka wè si yon mo nan yon fraz ap fè referans ak move bagay nan yon lòt fraz. Yon tès fraz pa fraz ka rate sa.

Limit etid la

Seri tèks yo diferan nan sijè yo, nan fason yo chwazi fraz yo ak nan fason tradiktè referans yo ekri. Diferans yo pa izole yon efè ki ta soti nan nati edikatif tèks yo sèlman. Sous yo pa etabli si yo te sèvi ak menm lojisyèl ak menm paramèt pou bay nòt nan twa tès yo.

Sous yo pibliye nòt total. San tradiksyon ki mache pa pè, idantifyan istwa yo ak referans ki aliyen nan analiz sa a, nou pa kalkile entèval ensètitid ni konte move tradiksyon ki ta ka fè moun mal. Ansyen sistèm sa yo ak istwa sa yo pa etabli kalite zouti ki la kounye a, liv lekòl oswa itilizasyon yo nan klas.

Refè analiz la

Pran liy Tatoeba ak JW300 sou fich OPUS anglè pou kreyòl la [2]. Miltipliye chrF pa 100.

Nan tablo 5 CreoleVal la [1], pran de liy ki gen anglè kòm lang depa ak kreyòl ayisyen kòm lang sib. Kenbe toude nòt yo.

Pou chak mezi, wete nòt OPUS sou MIT-Haiti a nan nòt chak nan twa lòt liy yo. Bay diferans lan ak yon chif apre pwen an epi bay non mezi a.

Tablo done

Seri tèks pou tès la	Modèl	Pwen BLEU	Pwen chrF (0–100)
Tatoeba	OPUS	45.2	59.2
JW300	OPUS	38.3	54.5
MIT-Haiti	OPUS	14.7	35.8
MIT-Haiti	CreoleM2M	22.0	43.9

Referans

[1] Lent, H. et al. (2024). CreoleVal: Multilingual Multitask Benchmarks for Creoles. Transactions of the Association for Computational Linguistics, Section 4.2 and Table 5.

[2] Helsinki-NLP. OPUS-MT English-to-Haitian model card, benchmark scores for Tatoeba and JW300.

[3] Platfòm MIT-Ayiti. Educational resources in Haitian Creole.

[4] Papineni, K. et al. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation.

[5] Popović, M. (2015). chrF: Character n-gram F-score for automatic MT evaluation.

Sous done yo

Lent et al. (2024), Table 5. CAIA comparisons. CC BY 4.0