

XARXES NEURONALS I APRENENTATGE PROFUND

GRUP 04 - Machine Translation

Laura Marín (1636713), Mar Blazquez (1641111), Pau Clavera (1568315), Jaime Rueda (1566320)

INDEX

INTRODUCCIÓ	3
OBJECTIUS	4
DADES	5
MODEL	6
Mètriques	8
EXPERIMENTS	10
HIPERPARÀMETRES	10
1. Learning Rate	10
2. Hidden Size	11
3. Optimizers	12
DADES	13
1. Bidireccionalitat	13
2. Frases curtes vs llargues	15
3. eng-esp vs eng-cat	17
RESULTAT I MODEL FINAL	19
CONCLUSIONS	23
BIBLIOGRAFIA	24

INTRODUCCIÓ

En els últims anys, el camp de l'aprenentatge profund ha experimentat un creixement espectacular. En aquest projecte, no només hem explorat l'aplicació del model Seq2Seq a la tasca de traducció automàtica, sinó que també hem dut a terme una sèrie d'experiments exhaustius per millorar-ne el rendiment.

Hem investigat diverses configuracions del model, incloent-hi la mida de les unitats amagades (hidden size), el learning rate, els optimitzadors com Adam, SGD i RMSprop, la mida del batch, i altres factors com la bidireccionalitat del model i la longitud de les frases. A través d'aquests experiments, hem intentat comprendre millor com cada paràmetre afecta el rendiment del model i com podem optimitzar-lo per a la tasca específica de traducció automàtica.

El model Seq2Seq es basa en una arquitectura d'encodificació-decodificació que permet transformar una seqüència d'entrada en una seqüència de sortida de longitud variable. Aquest tipus d'arquitectura és especialment potent per a la tasca de traducció automàtica, ja que pot manejar frases de diferents mides i estructures complexes.

OBJECTIUS

Els objectius d'aquest projecte són diversos i amplis, abordant tant l'optimització del nostre model com la comprensió teòrica del model Seq2Seq en el context de la traducció automàtica.

En primer lloc, aspirem a aconseguir que el model Seq2Seq funcioni de manera òptima per a la tasca de traducció automàtica. Això implica assegurar que les traduccions generades siguin no només coherents i precises, sinó també fluïdes i naturals. Per a assolir aquest objectiu, ens enfoquem en la millora contínua del model, ajustant els seus paràmetres i provant noves tècniques per augmentar la qualitat de les traduccions.

En segon lloc, busquem entendre profundament el funcionament de l'arquitectura Seq2Seq. Això implica estudiar detalladament com el model processa les dades d'entrada i genera les traduccions de sortida. Ens interessa comprendre els mecanismes interns del model, com l'encodificació de les frases d'entrada, el procés de decodificació, la capa d'attention i com es maneja la informació.

A més a més, busquem trobar els millors hiperparàmetres per al nostre model fent diversos experiments. Els hiperparàmetres, com la mida de les unitats amagades, el learning rate, els optimitzadors, la mida del batch, i altres factors, tenen un impacte significatiu en el rendiment del model. Mitjançant experiments sistemàtics, provem diferents configuracions per identificar aquells paràmetres que maximitzen l'eficàcia del model.

Finalment, ens proposem dur a terme una sèrie d'experiments per provar el model amb dades de diverses llargades i en diferents idiomes. Aquesta investigació té l'objectiu de comprendre com el model Seq2Seq maneja la diversitat lingüística, així com les variacions en la longitud de les frases. Analitzarem com el model processa frases curtes i llargues en diferents llengües, i com aquestes diferències afecten la precisió i la coherència de les traduccions generades.

DADES

En aquest projecte hem utilitzat principalment el spa-eng.zip(141543) que conté un fitxer anomenat spa.txt. Aquest fitxer conté parelles d'idiomes a cada línia, primer la frase en anglès i després traduïda al castellà. Com que hi havia masses dades en aquest fitxer, hem agafat 100.000 parells de frases de manera aleatòria pels nostres experiments.

Per representar cada paraula en un idioma, es fa servir una tècnica anomenada vectorització "one-hot". Aquesta tècnica consisteix a crear un vector de zeros amb un únic un a la posició corresponent a cada paraula.

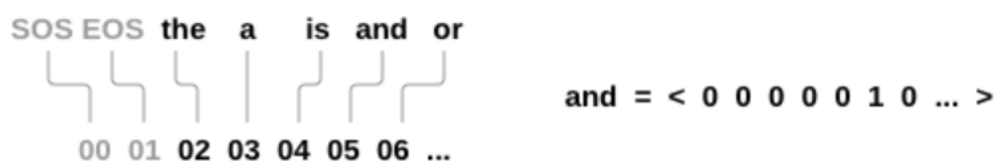


Fig 1 Vectorització One Hot

S'utilitza una classe auxiliar anomenada `Lang` per gestionar els índexs únics de cada paraula. Aquests índexs s'usaran com a entrades i objectius de les xarxes neuronals en etapes posteriors. La classe `Lang` conté diccionaris per a les conversions de paraula a índex i d'índex a paraula, així com un comptador de cada paraula.

El processament dels fitxers de text es realitza en diverses etapes:

Primer, tots els fitxers estan en format Unicode i es converteixen a ASCII per simplificar el processament. Després, el fitxer de dades es llegeix i es divideix en línies, i després cada línia es divideix en parells de traducció. Com que els fitxers contenen traduccions de l'anglès al castellà, s'afegeix l'opció d'invertir els parells si voleu traduir del castellà a l'anglès. Finalment, es normalitza el text, convertint-lo a minúscules i eliminant caràcters no desitjats. Després, es filtren els parells de traducció per incloure només frases relativament curtes amb el parametre MAX_LENGTH.

Per l'entrenament, es prepara dades transformant frases en parells d'idiomes diferents en índexs numèrics i organitzant-los després en tensors. Aquests tensors representen tant la frase d'entrada com la de sortida, amb un token especial (EOS) afegit al final de cadascuna per indicar la seva conclusió. Finalment, aquestes dades s'agrupen en lots i es passen a través d'un DataLoader, que en

facilita la iteració durant l'entrenament del model. Aquest procés assegura que les dades estiguin en el format correcte i estructurat per a l'entrenament efectiu del model de traducció.

MODEL

L'arquitectura del nostre model es basa en els algorismes Sequence to Sequence (Seq2Seq). El model Seq2Seq és una arquitectura de xarxa neuronal recurrent (RNN) especialitzada en tasques de seqüència a seqüència. Aquesta arquitectura està composta per dos components principals: l'encoder i el decodificador. L'objectiu principal d'aquest model, com ja s'ha esmentat, és traduir una seqüència d'entrada a una seqüència de sortida.

L'encoder rep com entrada la seqüència d'entrada i la processa passant-la a través d'un RNN. El seu objectiu és transformar la seqüència d'entrada en un vector fixe de dimensions determinades. Aquest vector codifica la "significació" de la seqüència d'entrada en un punt únic d'un espai multidimensional. La idea és que aquest vector conté tota la informació necessària per a la traducció sense tenir en compte l'ordre exacte o la longitud de la seqüència original.

El decoder, per la seva banda, rep aquest vector codificat de l'encoder i el converteix de nou en una seqüència. A diferència de l'encoder, el decodificador opera en ordre seqüencial, generant cada paraula de la seqüència de sortida un pas després de l'altre. Aquest procés continua fins que es genera una paraula d'atzar o es completa un límit predefinit.

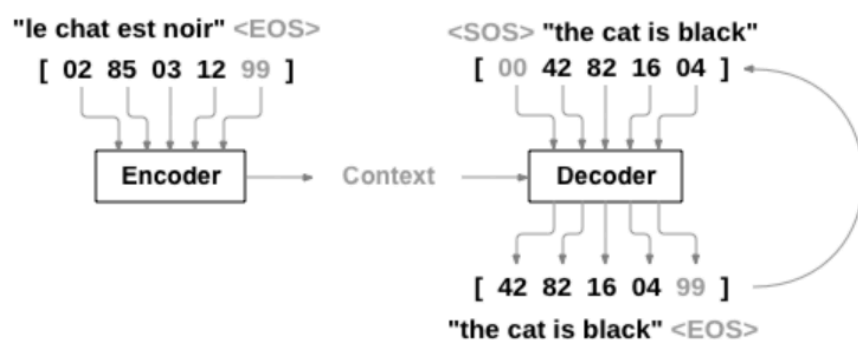


Fig 2 Encoder i decoder

La implementació del model seq2seq inclou tant el component d'encoder com el de decoder, així com una capa d'atenció Bahdanau per millorar la traducció al centrar-se en parts rellevants de la seqüència d'entrada durant la generació de la seqüència de sortida. L'EncoderRNN defineix un mida oculta, inicialitza una capa d'incrustació per convertir les entrades categòriques en vectors densos, i una unitat GRU per processar aquestes seqüències. També inclou una capa de dropout per regularitzar el model. En el seu procés, cada entrada es incrustada, s'aplica dropout i després es passa a través de la unitat GRU, que actualitza el seu estat ocult basat en la seqüència d'entrada. Finalment, retorna l'última sortida de la GRU i el seu estat ocult.

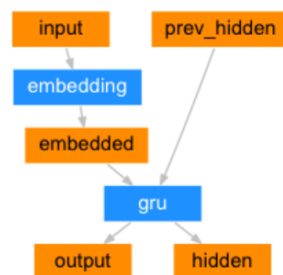


Fig 3 Implementació Encoder

La capa d'atenció Bahdanau implementa tres capes lineals que transformen les entrades per calcular puntuacions d'atenció. En el seu procés, calcula les puntuacions d'atenció entre una consulta (l'estat ocult del decoder) i totes les claus (les sortides del encoder). Aquestes puntuacions es fan servir per ponderar les claus, creant així un context que combina informació rellevant de la seqüència d'entrada. Retorna aquest context i els pes d'atenció. L'AttnDecoderRNN és similar a l'encoder, però també inclou una instància de BahdanauAttention. Utilitza una GRU per processar les entrades decodificades juntes amb el context d'atenció, i una capa lineal final per reduir la dimensionalitat de la sortida. En el seu procés, inicialitza el tensor d'entrada del decoder i el seu estat ocult amb el darrer estat ocult del encoder. Per cada pas de temps, calcula el context d'atenció basat en l'estat ocult actual i les sortides del encoder, després concatena aquest context amb la entrada incrustada abans de passar-la a través de la GRU. Si hi ha una seqüència de destí proporcionada, es fa servir per alimentar el següent pas; de cas contrari, es fa servir la pròpia predicció del decoder. Al final, concatena totes les sortides i els pes d'atenció, aplicant softmax a les sortides per obtenir probabilitats sobre les possibles sortides.

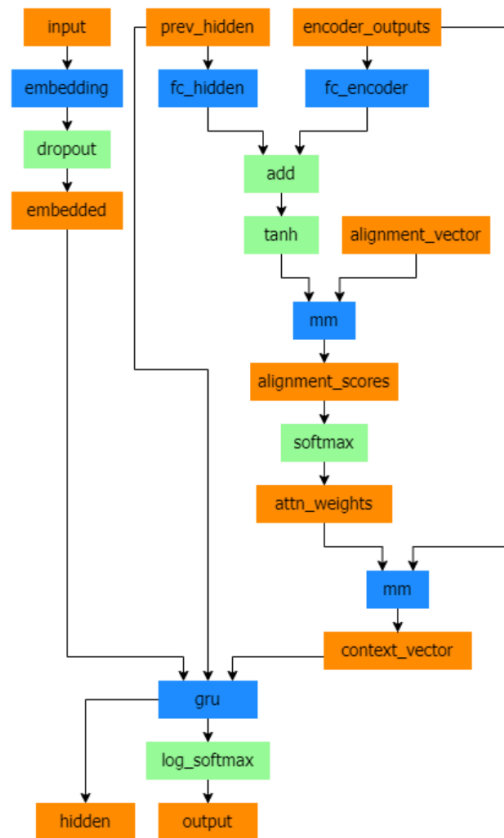


Fig 4 Implementació Decoder

Mètriques

En el desenvolupament de models de traducció automàtica, les mètriques serveixen per avaluar la qualitat de les prediccions generades pel model. Les tres mètriques principals que s'han utilitzat són el training loss, el validation loss i el score de Bleu.

Durant el procés d'entrenament, el training loss mesura la diferència entre les prediccions del model i les etiquetes reals. Aquesta mètrica és crucial per monitoritzar si el model està aprenent correctament. Un valor baix de training loss indica que el model està fent prediccions precises en comparació amb les etiquetes reals durant el entrenament.

Per a la validació, es calcula el validation loss de manera similar al training loss, però utilitzant les dades de validació en lloc de les dades d'entrenament. Aquesta mètrica serveix per estimar com el model es comportarà en nous dades no vistes. Un valor baix de validation loss indica que el model té una bona generalització i pot fer bones prediccions en dades no vistes durant el entrenament.

Una altra mètrica important és el score de Bleu (Bilingual Evaluation Understudy), que és popular per avaluar la qualitat de les traduccions automàtiques. A diferència del training loss i el validation loss,

que es basen en la diferència directa entre les prediccions i les etiquetes reals, el score BLEU compara les prediccions del model amb les traduccions de referencia.

El càlcul del score BLEU implica comparar les paraules generades pel model amb les paraules de referència. El score BLEU es calcula com la mitjana ponderada de les precisions de n-grams de diferents ordres. Una precisió de n-gram és la proporció de n-grams que apareixen en la traducció generada que també apareixen en la traducció humana. Aquests n-grams poden variar des de 1-gram (paraules individuals) fins a 4-grams (grups de quatre paraules).

El score BLEU varia entre 0 i 1, on 1 indica una perfecta correspondència amb la traducció humana i 0 indica que no hi ha cap correspondència.

EXPERIMENTS

En aquest apartat es descriuen els experiments realitzats per optimitzar el rendiment del model. L'objectiu és determinar els paràmetres òptims per aconseguir una millor generalització i evitar el sobreajustament (overfitting).

HIPERPARÀMETRES

1. Learning Rate

El learning rate és un paràmetre crucial en l'entrenament dels models. Controla la magnitud dels ajustaments que el model fa als seus pesos en resposta a l'error comès en cada iteració. En altres paraules, determina quant de ràpid o lentament un model aprèn. Si el learning rate és massa alt, el model pot aprendre de manera massa brusca, oscil·lant al voltant de la solució òptima sense arribar a estabilitzar-se. Si és massa baix, l'aprenentatge pot ser molt lent, trigant molt de temps a convergir o quedant-se atrapats en mínims locals. Per tal d'identificar el valor òptim del learning rate, s'han provat diverses configuracions: 0.001, 0.0001, i 0.01.



Fig 5 Gràfica del Training loss i Validation loss en funció del learning rate

En la gràfica anterior, s'observa que el learning rate de 0.0001 és el que proporciona els millors resultats. Aquest valor permet un equilibri adequat entre la velocitat d'aprenentatge i la capacitat de generalització, minimitzant el sobreajustament i assegurant un millor rendiment en dades no vistes. Així doncs, s'ha determinat que el learning rate òptim per aquest model és 0.0001.

2. Hidden Size

El paràmetre “hidden size” es refereix al nombre de neurones presents a cada capa oculta d’una xarxa neuronal. Aquest paràmetre és clau en la capacitat del model per capturar patrons complexos en les dades. Si el hidden size és massa petit, el model pot no tenir prou capacitat per aprendre relacions importants, provocant un infraajustament (underfitting). Pel contrari, si és massa gran, el model pot aprendre detalls específics de les dades d’entrenament, el que pot conduir a un sobreajustament (overfitting). Per tal d’identificar el valor òptim del hidden size, s’han provat diverses configuracions: 64, 128, 256, 512.



Fig 6 Gràfica del Training loss i Validation loss en funció del hidden size

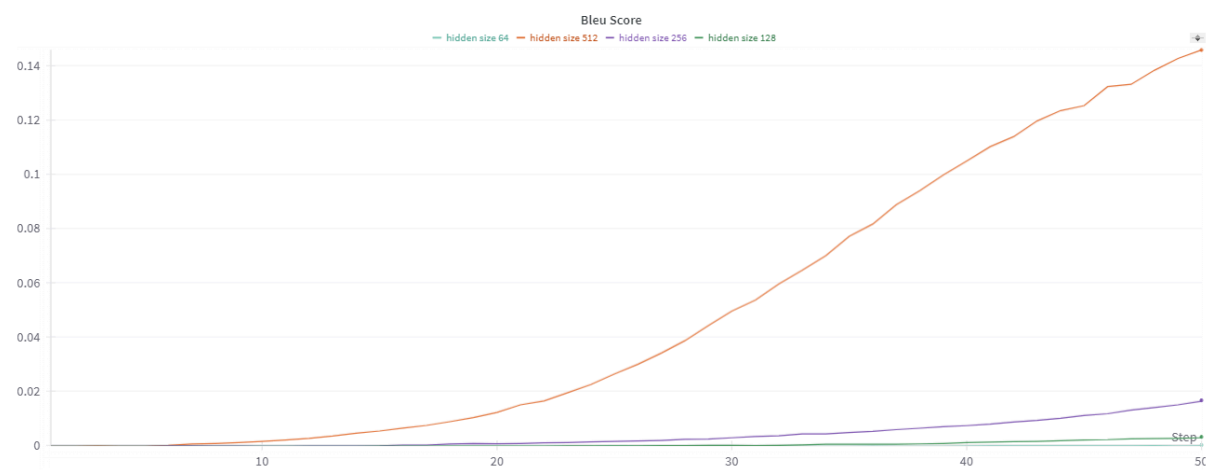


Fig 7 Gràfica del Bleu Score en funció del hidden size

Després de provar les diferents configuracions i d'observar les gràfiques anteriors, s'ha determinat que el valor de hidden size de 512 neurones per capa oculta és l'òptim per aquest model. Aquest ajust proporciona un bon equilibri entre la capacitat d'aprenentatge del model i la seva capacitat de generalització, contribuint a un millor rendiment del conjunt de dades de validació.

3. Optimizers

Un altre dels factors clau que s'han investigat són els optimitzadors utilitzats durant l'entrenament del model. Els optimitzadors són algorismes utilitzats per ajustar els pesos del model durant l'entrenament amb l'objectiu de minimitzar la funció de pèrdua. En aquest experiment, s'han provat tres optimitzadors diferents: Adam, SGD i RMSprop.

- **Adam:** Aquest optimitzador combina els avantatges dels optimitzadors adaptatius i els mètodes de moment estocàstic.
- **SGD (Stochastic Gradient Descent):** És un dels optimitzadors més bàsics utilitzats en l'aprenentatge profund. Actualitza els pesos del model mitjançant la derivada de la funció de pèrdua respecte als pesos, ajustant-los en direcció al mínim de la funció.
- **RMSprop (Root Mean Square Propagation):** Aquest optimitzador adapta la taxa d'aprenentatge de manera individual per a cada paràmetre. Això ajuda a superar els problemes d'ajustament excessiu del learning rate en les direccions de pèrdua que varien molt.



Fig 8 Gràfica del Training loss i Validation loss en funció dels optimizers

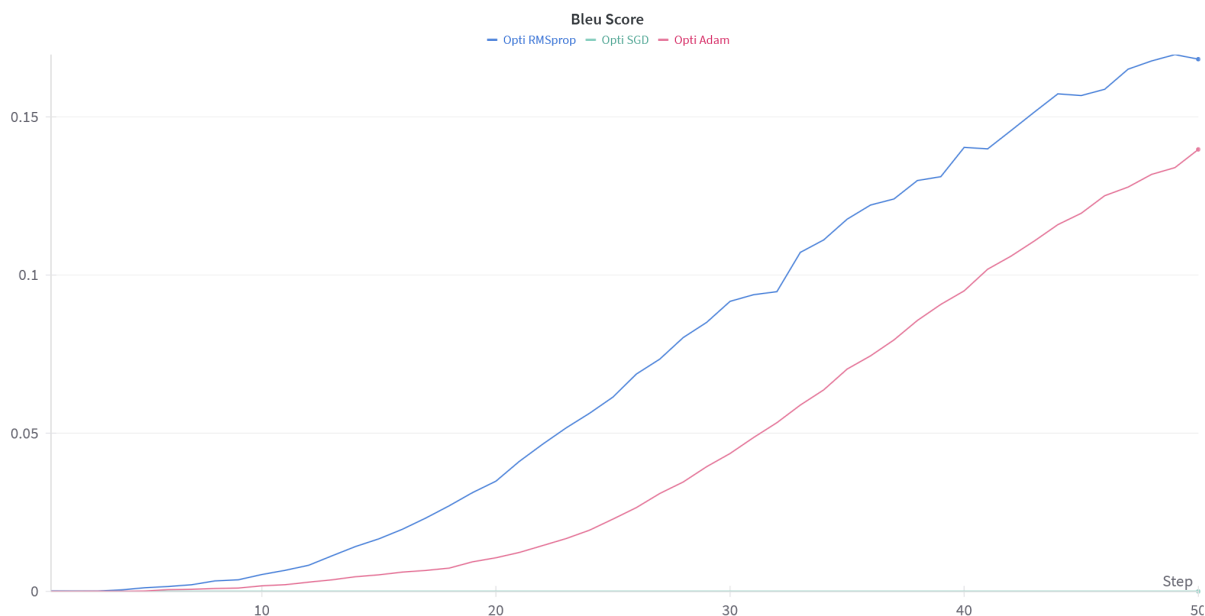


Fig 9 Gràfica del Bleu Score en funció dels optimizers

En les gràfiques anteriors, s'observa que tots els optimitzadors, menys el SGD, proporcionen resultats satisfactoris en termes de pèrdua d'entrenament i validació. No obstant això, en la gràfica de puntuació BLEU (Fig 9), el optimitzador RMSprop mostra el millor rendiment, amb la puntuació més alta. Així doncs, RMSprop destaca com el millor optimitzador per a aquesta tasca, proporcionant una millor puntuació BLEU, un indicador clau en tasques de traducció automàtica.

DADES

Un cop s'han trobat els hiperparàmtres adients pel nostre model, s'ha decidit seguir experimentant amb les dades. Mirant com varia l'eficàcia del model segons les diferents dades proporcionades en aspectes com: la bidireccionalitat, la longitud de les frases, i les traduccions entre anglès-espanyol i anglès-català.

1. Bidireccionalitat

A més dels hiperparametres, s'ha investigat la bidireccionalitat de les traduccions per millorar els resultats.

Inicialment, el model es va entrenar per traduir de l'anglès al castellà. Posteriorment, es va provar de canviar la direcció de la traducció, passant a traduir del castellà a l'anglès, per analitzar les diferències en termes de pèrdua d'entrenament, pèrdua de validació i puntuació BLEU.

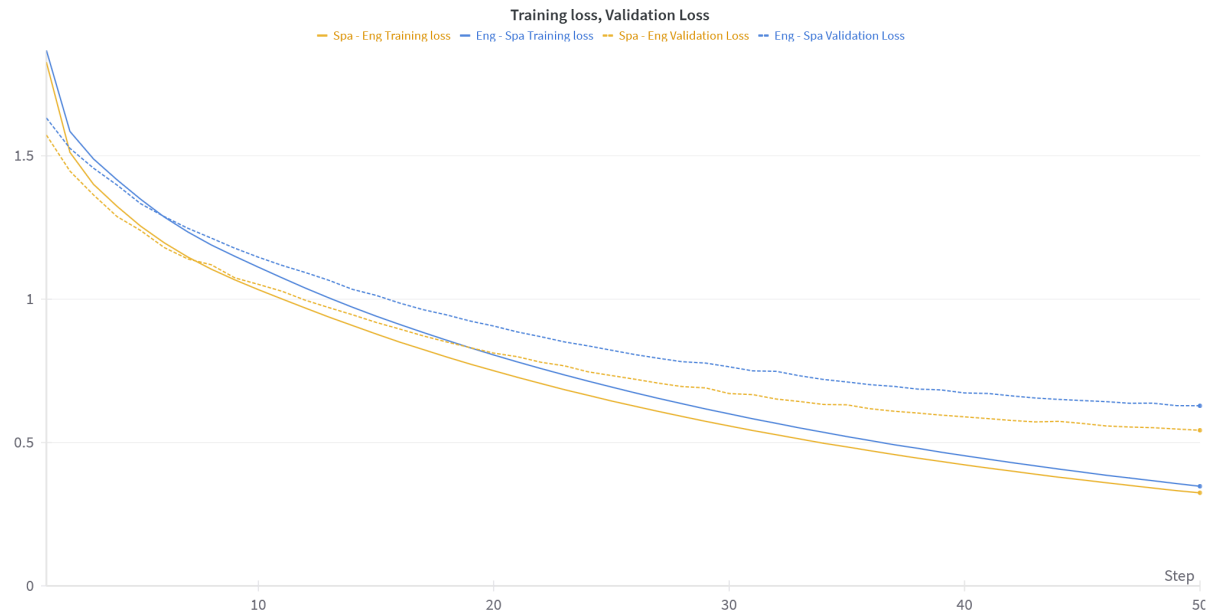


Fig 10 Gràfica del Training loss i Validation loss en funció de la bidireccionalitat

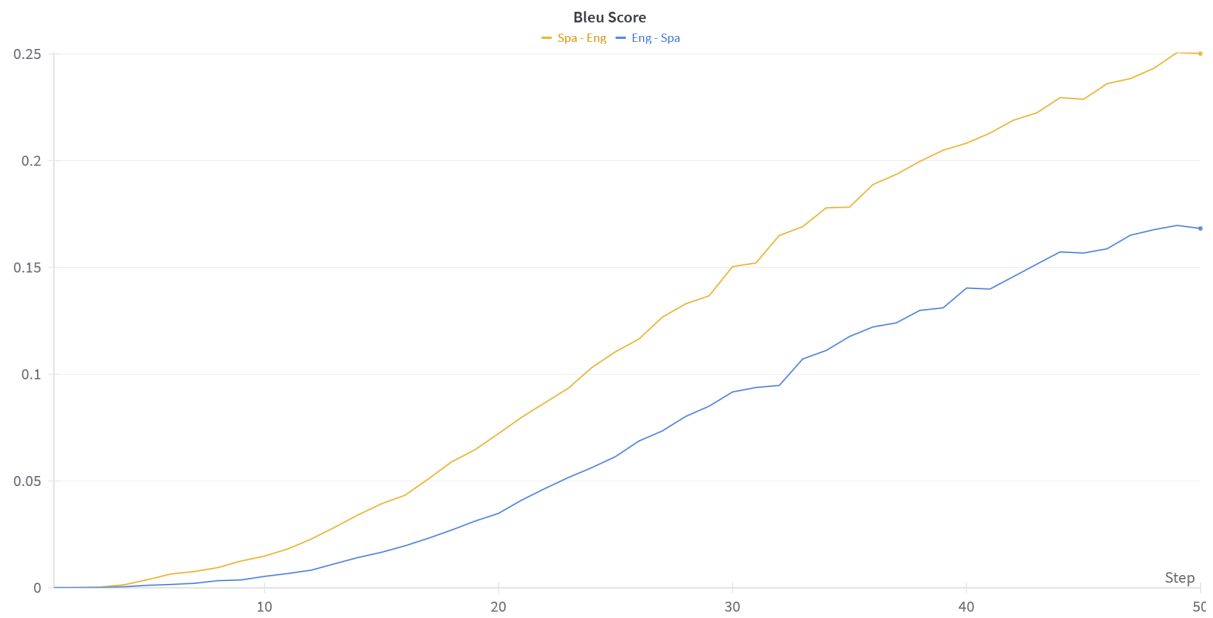


Fig 11 Gràfica del Bleu Score en funció de la bidireccionalitat

En les gràfiques anteriors, es pot observar que els resultats de traduir del castellà a l'anglès (spa-eng) són millors que els de traduir de l'anglès al castellà (eng-spa). Això es reflecteix en les mètriques següents:

La gràfica (Fig 10) mostra una pèrdua d'entrenament i de validació més baixa per a les traduccions del castellà a l'anglès, la qual cosa indica que el model s'ajusta millor i generalitza millor en aquesta direcció. A més a més, la gràfica (Fig. 11) mostra que el score BLEU de les traduccions del castellà a l'anglès és més alt. Això indica que les traduccions generades en aquesta direcció són més properes a les traduccions de referència.

Les diferències observades poden ser atribuïdes a diversos factors. En primer lloc, l'anglès té una estructura gramatical més simple i menys complexa que el castellà, cosa que pot facilitar que el model aprengui patrons de traducció consistents quan es tradueix del castellà a l'anglès. La simplicitat gramatical de l'anglès, amb conjugacions verbals menys variades i regles sintàctiques més flexibles, redueix la complexitat del procés d'aprenentatge per al model. Això contrasta amb el castellà, on la major variabilitat en l'ús dels temps verbals, les estructures gramaticals més rígides i les nombroses excepcions en les regles augmenten la dificultat d'entrenament i la probabilitat d'errors de traducció.

2. Frases curtes vs llargues

A més, a més s'ha analitzat com la longitud de les frases afecta l'eficàcia del model. Es vol determinar si el model funciona millor amb frases curtes o llargues, ja que la complexitat i el context necessari per comprendre correctament una frase poden variar significativament amb la seva longitud. Per analitzar això, s'han creat dos files addicionals.

- spa_sample_frases_cortas.txt: Aquest fitxer conté 70.000 parells de frases amb menys de 11 paraules escollides aleatòriament.
- spa_sample_frases_largas.txt: Aquest fitxer conté 70.000 parells de frases amb més o igual de 11 paraules escollides aleatòriament.



Fig 12 Gràfica del Training loss i Validation loss en funció de la llargària de la frase

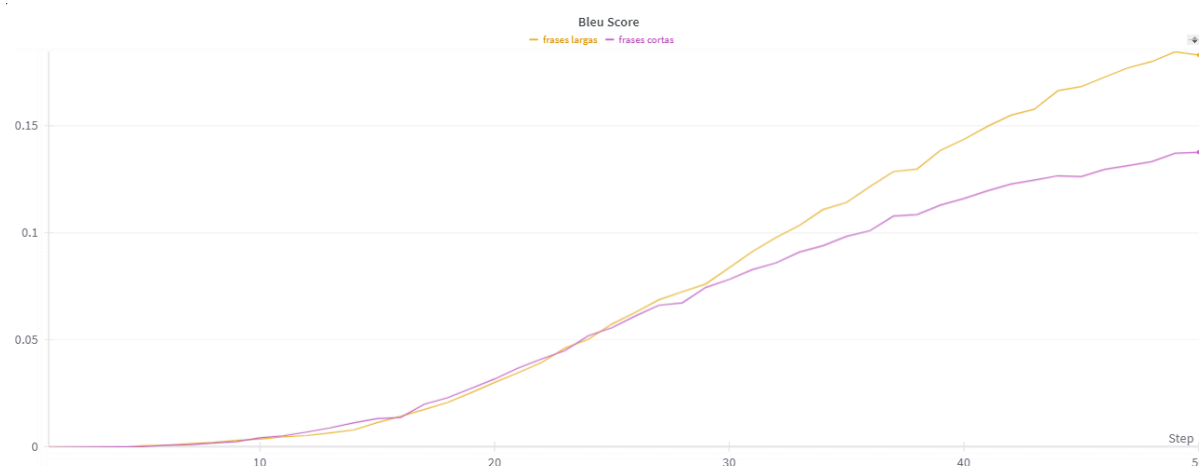


Fig 13 Gràfica del Bleu Score en funció de la llargària de la frase

Els resultats mostren que el model té un millor rendiment amb frases llargues comparat amb frases curtes, com es pot observar en el Bleu Score que és més alt a les frases llargues en comparació amb les frases curtes. Això suggereix que les traduccions de frases llargues són més properes a les traduccions de referència. En canvi, respecte la Training Loss i la Validation Loss, les pèrdues són més baixes per a les frases curtes en comparació amb les frases llargues. Això indica que, tot i que el model s'ajusta millor a les frases curtes durant l'entrenament, generalitza millor amb frases llargues durant la validació.

Aquestes diferències poden ser atribuïdes a la naturalesa de les frases llargues i curtes:

- **Frases Llargues:** Les frases llargues contenen més context i informació, cosa que permet al model captar millor el significat global i produir traduccions més precises. Això pot explicar el millor Bleu Score observat.

- **Frases Curtes:** Les frases curtes tenen menys context i poden ser més ambigües, la qual cosa pot dificultar la tasca del model a l'hora de generar una traducció precisa. Això es reflecteix en el menor Bleu Score tot i tenir una Training Loss i Validation Loss més baixes

En resum, tot i que el model sembla ajustar-se millor a les frases curtes durant l'entrenament, les frases llargues proporcionen traduccions més precises en termes de similitud amb les traduccions de referència.

3. eng-esp vs eng-cat

També s'ha volgut comparar el model entrenat amb el [spa-eng.zip](#) que conté 141543 parells de frases, amb un de moltes menys dades, un idioma no tan popular i que no tingués tantes traduccions. En aquest cas, s'ha agafat el [cat-eng.zip](#) que consta de tan sols 1375 parells de frases al arxiu cat.txt el qual també ha sigut desordenat per a què no influeixi la ordenació de les frases al experiment.



Fig 14 Gràfica del Training loss i Validation loss en funció de diferent quantitat de dades

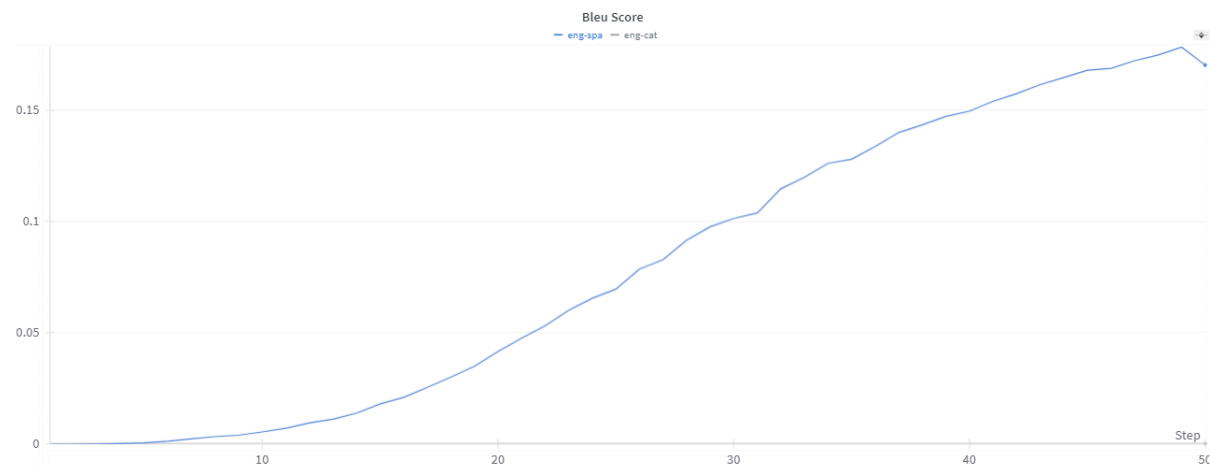


Fig 15 Gràfica del Bleu Score en funció de diferent quantitat de dades

En les gràfiques anteriors, es pot observar que els resultats de traduir de l'anglès al castellà (eng-spa) són molt millors que els de traduir de l'anglès al català (eng-cat). Això es reflecteix en les mètriques següents:

La gràfica (Fig. 14) mostra una pèrdua d'entrenament i de validació més baixa per a les traduccions de l'anglès al castellà, la qual cosa indica que el model s'ajusta millor i generalitza millor amb grans quantitats de dades. A més a més, la gràfica (Fig. 15) mostra que el score BLEU de les traduccions de l'anglès al castellà és bastant més alt.

Les diferències observades poden ser atribuïdes al volum de dades, i és que la disponibilitat de dades per entrenar el model és significativament més gran per a l'anglès-castellà (141543 parells de frases) en comparació amb l'anglès-català (1375 parells de frases). Aquesta disparitat en la quantitat de dades d'entrenament influeix directament en la capacitat del model per aprendre patrons de traducció consistents i precisos.

RESULTAT I MODEL FINAL

Després de realitzar diversos experiments amb l'objectiu d'optimitzar el rendiment del model de traducció automàtica, s'ha arribat a la conclusió que la configuració següent proporciona els millors resultats. A continuació, es detallen els paràmetres seleccionats i les raons que justifiquen la seva elecció:

- **Learning rate = 0.0001:** Aquest valor permet que el model aprengui de manera estable i efectiva, evitant els problemes d'oscil·lació associats a una taxa d'aprenentatge massa alta i la convergència lenta associada a una taxa massa baixa. Així, el model aconsegueix una bona generalització, minimitzant el sobreajustament.
- **Batch size = 500:** La mida del lot, o batch size, determina el nombre d'exemples que el model processa abans d'actualitzar els seus pesos. Una mida de lot de 500 s'ha demostrat ser ideal, ja que proporciona una bona estabilitat durant l'entrenament. A més, aquest valor optimitza l'ús de la memòria i els recursos computacionals disponibles, equilibrant la càrrega entre eficiència i rendiment.
- **Hidden size = 512:** La mida de la capa oculta és un altre paràmetre clau que afecta la capacitat del model per capturar i aprendre patrons complexos en les dades. Un valor de 512 per a la mida de la capa oculta ha demostrat ser suficient per capturar les complexitats inherents dels patrons de traducció sense sobrecarregar el model amb una complexitat innecessària.
- **Optimitzador = RMSprop:** L'optimitzador RMSprop ha estat seleccionat després de comparar diversos optimitzadors com Adam i SGD. RMSprop adapta la taxa d'aprenentatge per a cada paràmetre individualment, el que ajuda a superar els problemes d'ajustament excessiu i convergència lenta. En els nostres experiments, RMSprop ha mostrat un rendiment superior en termes de pèrdua d'entrenament i validació, així com en la puntuació BLEU, una mesura clau de la qualitat de les traduccions. Aquest optimitzador ha demostrat proporcionar traduccions més precises i consistents.
- **Nombre d'èpoques = 50:** El nombre d'èpoques determina quantes vegades el model processa l'ensenyament complet durant l'entrenament. Un valor de 50 èpoques s'ha demostrat ser adequat per permetre al model aprendre els patrons necessaris per a una traducció precisa.

A continuació, es presenten les quatre gràfiques corresponents a les mètriques obtingudes durant l'entrenament del model de traducció. Aquestes mètriques són essencials per avaluar el rendiment del model tant durant el procés d'entrenament com en la seva validació posterior

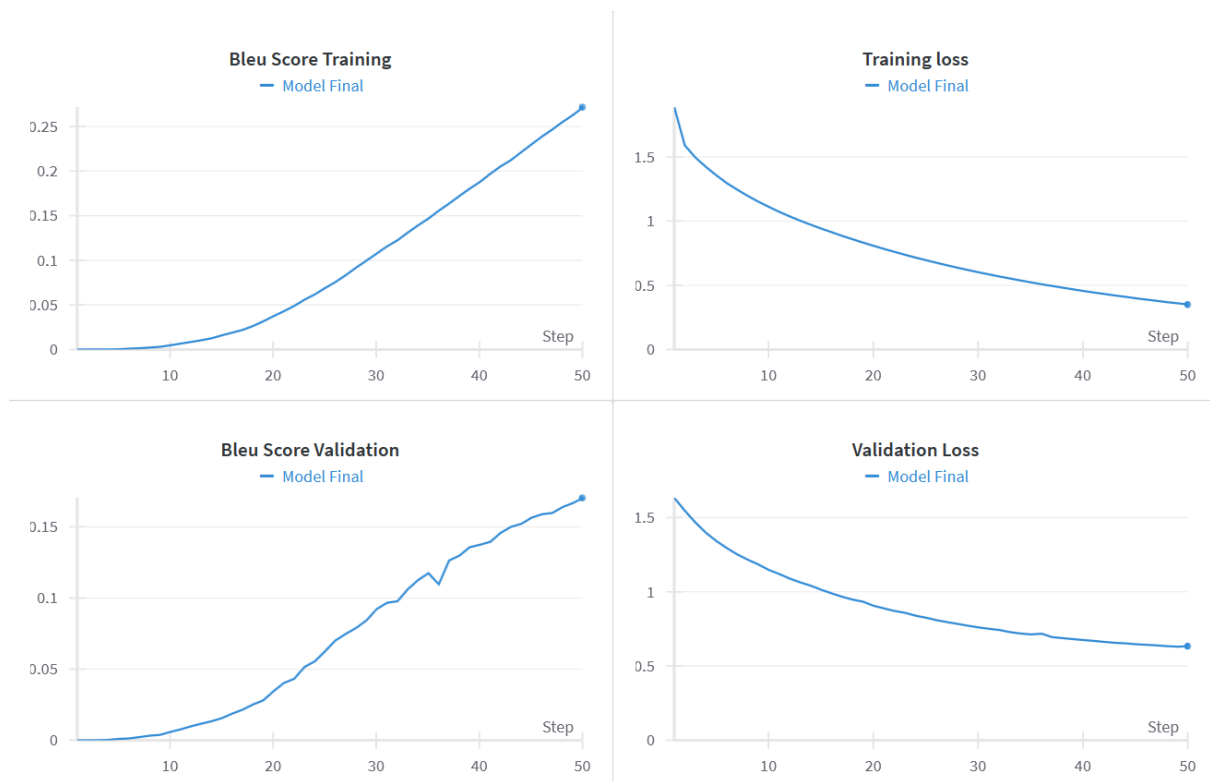


Fig 16 Gràfica de les quatre mètriques del model final

Per facilitar una millor comprensió dels resultats, hem combinat les dues gràfiques de BLEU Score i les dues gràfiques de Loss en dues gràfiques addicionals. Aquesta combinació permet visualitzar més clarament la relació entre les mètriques d'entrenament i validació, facilitant la identificació de possibles problemes.



Fig 17 Gràfica de les quatre mètriques del model final juntades

En les gràfiques, es pot veure que la pèrdua de entrenament (training loss) és més baixa que la pèrdua de validació (validation loss), el que és un comportament normal. També es pot observar que les puntuacions durant l'entrenament són més altes que en la validació, el que també és normal, ja que estem treballant amb les dades les quals s'ha entrenant el model.

Cada vegada que entrenem un model, ens guardem un arxiu JSON que conté algunes frases seleccionades de la validació de cada època. Això, ens permet observar com evolucionen les traduccions a mesura que l'entrenament progressa. Es pot veure clarament que, en les primeres èpoques, les traduccions no són encara bones. No obstant això, a mesura que avancen les èpoques, les qualitats de les traduccions milloren. Aquest són alguns exemples de les traduccions de l'última època.

"Input Sentence": "how long does it take to get from here to your house on foot ?",

"Target Sentence": "cuanto se tarda en ir desde aqui hasta tu casa andando ?",

"Predicted Sentence": "cuanto se tarda en llegar a aqui a que casa ? ?",

"Bleu Score": 0.24384183193426084

"Input Sentence": "tom has three adult sons",

"Target Sentence": "tomas tiene tres hijos grandes",

"Predicted Sentence": "tom tiene tres hijos",

"Bleu Score": 6.725854833444237e-78

"Input Sentence": "she has good handwriting",

"Target Sentence": "ella tiene buena letra",

"Predicted Sentence": "ella tiene buena letra",

"Bleu Score": 1.0

"Input Sentence": "the situation worries me very much",

"Target Sentence": "la situacion me preocupa mucho",

"Predicted Sentence": "la situacion me recuerda muy",

"Bleu Score": 6.86809206056511e-78

A partir de les dades proporcionades es poden extreure les següents conclusions: La precisió de les traduccions varia significativament segons la frase. Per exemple, la traducció de "she has good handwriting" a "ella tiene buena letra" és completament precisa, amb una puntuació BLEU de 1.0, indicant una traducció perfecta. No obstant això, algunes frases mostren errors importants en la traducció, com en el cas de "tom has three adult sons" que es tradueix a "tom tiene tres hijos" en lloc de "tomas tiene tres hijos grandes", reflectint-se en una puntuació BLEU molt baixa. Les frases més complexes o amb més context semblen ser més difícils de traduir amb precisió. Per exemple, la frase "how long does it take to get from here to your house on foot?" es tradueix incorrectament amb la puntuació BLEU resultant de 0.24384183193426084, indicant errors en la comprensió del context i la sintaxi. Encara que hi ha exemples de traduccions correctes, com la de "she has good handwriting", hi ha una manca de consistència en altres exemples, com "the situation worries me very much" que es tradueix incorrectament a "la situacion me recuerda muy", resultant en una puntuació BLEU negligible. En resum, es destaca que el model de traducció mostra variabilitat en la seva precisió, amb una clara diferència en la qualitat de les traduccions segons la complexitat de les frases. Tot i que les puntuacions són molt baixes, les traduccions són bones i coherents.

CONCLUSIONS

En aquest projecte, hem explorat la traducció automàtica utilitzant un model Seq2Seq, ajustant diversos hiperparàmetres per optimitzar-ne el rendiment. Descobrim que amb els hiperparàmetres adients es proporciona un bon equilibri entre la velocitat d'aprenentatge, la capacitat d'aprenentatge i la capacitat de generalització. Entre els optimitzadors avaluats, RMSprop va demostrar ser el millor, oferint la millor puntuació BLEU i, per tant, traduccions de més qualitat. Les traduccions del castellà a l'anglès van ser més precises que les inverses, possiblement per la simplicitat gramatical de l'anglès. Les frases llargues van proporcionar traduccions més precises en comparació de les curtes probablement perquè al haver més paraules augmenta la probabilitat d'encert, encara que el model es va ajustar millor a les frases curtes durant l'entrenament. La comparació entre els conjunts de dades eng-esp i eng-cat subratlla la importància de la quantitat i qualitat de les dades de formació.

Si haguéssim tingut més temps i recursos d'Azure, podríem haver provat altres configuracions per millorar encara més el rendiment del model. Algunes d'aquestes proves inclouen experimentar amb altres idiomes que continguin caràcters especials, cosa que hauria avaluat la capacitat del model per manejar diferents alfabetes i estructures lingüístiques. També hauria estat interessant treure la capa d'atenció per veure com això afecta la qualitat de les traduccions. Finalment, afegir més èpoques d'entrenament podria haver permès al model aprendre més profundament de les dades, millorant potencialment les puntuacions BLEU i la qualitat general de les traduccions.

En resum, la configuració adequada dels hiperparàmetres i una acurada selecció de dades són crucials per millorar el rendiment dels models de traducció automàtica. L'experimentació sistemàtica i l'adaptació a les especificitats de les dades utilitzades destaquen com a elements claus per obtenir resultats robusts. Amb més recursos i temps, podríem haver explorat aquestes vies addicionals per obtenir resultats encara més detallats i precisos.

BIBLIOGRAFIA

1. ManyThings.org. (n.d.). Anki: A friendly, intelligent flash card program. Retrieved from <http://www.manythings.org/anki/>
2. PyTorch. (n.d.). Sequence to sequence translation tutorial. Retrieved from https://pytorch.org/tutorials/intermediate/seq2seq_translation_tutorial.html
3. PyTorch. (n.d.). Optim: A package implementing various optimization algorithms. Retrieved from <https://pytorch.org/docs/stable/optim.html>
4. La Máquina Oráculo. (n.d.). Seq2Seq: De secuencia a secuencia. Retrieved from <https://lamaquinaoraculo.com/deep-learning/seq2seq-de-secuencia-a-secuencia/>
5. ResearchGate. (2019). The GRU Seq2Seq model structure. Retrieved from https://www.researchgate.net/figure/The-GRU-Seq2Seq-model-structure_fig2_335984301