International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 906-910 www.ijaem.net ISSN: 2395-5252

Turkish-English Neural Machine Translation Using an Encoder-Decoder Architecture with Attention Mechanism

[1]Aydin Yildiz, [2]Asst Prof.AdemOzyavas
[1,]Master ThesisStudent, ISTANBUL AYDIN UniversityDepartment of ComputerEngineering, Florya, Istanbul,Turkey
[,2]AssistantProf,ISTANBUL ATLASUniversityDepartment of SoftwareEngineering,Kağıthane, Istanbul, Turkey.
Corresponding Author: Aydın Yıldız

Submitted: 10-07-2022 Revised: 17-07-2022 Accepted: 21-07-2022

ABSTRACT:Neural machine translation (NMT) of natural languages has gained popularity in recent years. As a relatively new method of translation it addresses the problems faced with earlier methods but suffers from different types of problems. For instance, as the length of the source sentence increases, the quality of the translation drops. The most widely used architecture Encoder-Decoder is not an exception. To remedy the issue and increase the quality of translation, an attention mechanism is attached to this architecture which helps the translator remember the important keywords that will help during the translation. The proposed model is applied to a not-much-studied Turkish- English language pair. The results are encouraging. With more training data and modifications in the architecture to include more context should improve the quality of the machine translation. The BLEU scores of the machine-translated examples are shared in the tables in the proposed method section.

I. INTRODUCTION

Earlymachinetranslationsareperformedusingstatisticalmethods. A goodhistoricalreview can be found at [1]. IBMI 1 and 2 aresimplealgorithmsthatpavedthewayforearlywork in thefield of naturallanguagetranslation. This trend of usingstatisticalmethodsstillcontinuesbut neuralnetworksapplicationshaveemergedwithconsiderablesuccess in recentyears.

Artificial Neural Networks (ANN) have been recently applied as a method to natural language translations. The basic architecture of Neural Machine Translation (NMT) can bedescribed as two RecurrentNeural Networks (RNN) attachedbacktoback. Thefirstonethatencodesthesourcelanguagesentencein to a vectorthatextractsfeaturesfromitsinputsentence, aka a contextvectorwhich is an abstractrepresentation of thesourcelanguagesentence. Then, thisvector is fed intothesecond RNN which is todecode it intothetargetlanguagesentence.

Duringdecodingthetranslation is performedoneword at a time. The first RNN is aptlycalledthe Encoder andtheonethat is attached is calledtheDecoder.

Both Encoder anddecodergenerallyhavethesamestructure.

Forthoseinterested in thestructure of Encoder andDecoderneuralnetworks can referto [4]. As it can be seenthatdifferenttypesrecurrentneuralnetworks (RNN) could be usedwiththefundamentalonesbeingLong-Short-Term Memory (LSTM), GatedRecurrentUnit (GRU) withconsiderablevariations of those.

II. LITERATURE REVIEW

El-Kahloutand Kemal Oflazerproposed a statisticaltranslationmethod [12]. Theirproposedworkperformedsomemorphemeanalysis of suffixes of Turkishtext in the training data in ordertoincreasethesuccess of the English-Turkishmachinetranslationsystemwithsomeminorimprovements.

YenitreziandOflazer [13], in addition to the data obtained from superficial forms of the words, their morphological analysis on sentences produced new data on the word type, root, suffixesetcwhichtookthestatisticalmachinetranslationto a differentlevel. Theirmethodshowed an improvement of 38%.

Withtheapplication of ANN inmachinetranslation, theiruse has gainednew momentum in thisfield. Inadditiontoproposingnew ANN models, manystudieshavebeencarriedoutthatanalyzeexisting modelswith ANN whichhelp increasetheirspeedandquality. Junczys-Dowmunt et al. [2] performedanalysis on theexistingtranslationlanguagesdeveloping a newfasterdecoder. They alsocomparedthewellknownrule-basedand NMT systems.

Niehue et al. [3] separatedmodelingandsearchalgorithms in machinetranslationsystemsandanalyzedtheexistingtranslationsystemsmeasuringtheiradequacy in terms of translationquality. They utilized NTM in theirstudies. Errorsthatoccur in rule-basedtranslationanditsimprovement in NMT areexamined. The model of NMT isexplainedanditsdifferencesfromthestatisticalmachinetranslation is described. Improvements in the NMT model arebelievedto be promising. They created two modelsnamed TED and MSLT forcomparisonwithin NMT. Inthesemodels, a comparison has beenmadewithSingleandEnsemblesystems. Inthe TED model, theBigVocsystemwassuccessful in theSinglesystem. Theperformance of theSinglesystemwasalsofoundto be better in the MSLT model.

One of themostadvancedandstudiedmodels, calledthe “Transformer”, is proposedbyVaswani et al [3]. Thetransformer model allowsforfurtherparallelization. It is a simple network architecturebased on attentionmechanisms. Like LSTM, thetransformer is an architectureusedtoconvertonestringtoanotherwiththe help of an Encoder andDecoder. However, it differsfromexistingmodelsthatcan convertfromarraytoarraybecause it uses RNN.

There have been many studies on the analysis of the converter model, increasing its speed and improving the translation quality [5, 6, 7, 8, 9]. There have been quite research focused on Recurrent Machine Translation and Transformer Machine Translationcomparisons in terms of performance.

Koehn et al. [9] discussproblemsencountered in machinetranslationmodels, includingtheconverter model. They claimedthat increasingthe size of thebeamsearchdecreasesthetranslationqualityafter a certainbeamvalue. Six NMT challengesaredescribedanddiscussedbriefly in thestudywhichare domain mismatch, amount of training data, rarewords, longsentences, wordalignment, andbeamsearch. Thesechallengesarecompared in thefield of NMT and SMT. A positivecorrelationwasfoundbetweenincreasingthebeamlengthparameterandtheBilingual Evaluation Understudy (BLEU) scorewhich is one of themetricsusedtomeasurethetranslationquality.

Yang et al. [10] analyzedtheavailablebeamsearchvariationsanddeveloped a methodthatincreasesthe BLEU score at higherbeamsizes.

Huang et al. [11] workedandpresentedtheirresults on anotherwork on beamsearch . They wereabletoincreasethe BLEU scoreby 2 points in theChinese-English translation.

III. THE PROPOSED METHOD

Theproposedmethoduses a traditionalencoder-decoderarchitecturewithattentionmechanismtoimprovethetranslationquality. Interestedreaders can refertotheseminalpaper on theattentionmechanisim [4].

Thefollowing is thepseudo-code of thealgoriithmusedfortheTurkish-English translation. A moredetaileddescription of the main stepsfollowthepseudocode.

1. Read data file thatcontainstheparallelcorpus of textandpreprocessthe data
2. Createthe NMT model usingencoder, attentionanddecoder
3. Train the model
4. Makeinferences

The file thatcontainsthebilingualtext is readfirstandparsedwiththesourceandtargetsentencesalignedintotheparallellists. Somepreprocessing of the data followssuch as lowercasingthewordsandremovingnumerical data, removingpunctuationmarks, etc… Two dictionariesarecreated, thefirstoneforconvertingeverywordinto a uniqueintegerandthesecondfromintegerstotheircorrespondingwordsusedafterthedecodingprocess. As neuralnetworksrequiretheir data to be in number form andthesamelengthpadding is usedtoconvertallsentencestothesamelengthintegervalues. Then a more dense representation of sentencesareapplied as a distributionalrepresentation of them. This is thelast main step beforesupplyingthe data intotheencoder.

Encoder, decoderandtheattentionstructuresarecreatedfollowingthetraditionalapproach. Forthoseinterested in how thisarchitectureworks can referto [4]. The model is trainedwiththeconverted data obtainedfromthetext file. Because of the RAM limitations on the GPU, a batch size of 16 is utilized. Boththeencoderanddecoderstructuresused LSTM.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 906-910 www.ijaem.net ISSN: 2395-5252

Forthetranslation step, unseen data (exceptfor a couple of randomlychosenonesfromtheseen data duringtraining) areutilizedfortesting.

Some of thetranslationresultsareshown in thetablesbelow. Table 1 startswithshortsentences in thesourcelanguageTurkish. Source sentences, theircorresponding reference translations in English, theinferredtranslationsbythe model andthe BLEU scoreareallshown in theirrespectiverows. BLEU scoresrangebetween 0 and 1 inclusive; 1 denoting a perfectmatch (translation) and 0 nomatchbetweenthe reference translationandinferredone. Alsotoremedythefactthat BLEU scoreusedfromthe Natural Language Toolkit (NLTK) of Python usesuptoandincluding 4-grams andanymissing n-gram wouldresult in obtaining a scoreof 0, somesmoothingfunction is utilized, namely, method1 from NLTK library.

As it can be seen in Table 1, BLEU scoresare far from ideal. Eventhough “I have a bike” anditsmachinetranslation “I have a plane” has oneworddifferencewith a score of 0.4, “I’m reading a book” anditsmachinetranslation “ill’mreading a book” withthesameoneworddifference has a BLEU score of 0.17. Also, “Areyouinsane?” and “Areyoucrazy?” havethesamemeaning but because BLEU does not takesynonymsintoaccount, itsscoreresulted in a relativelylowscoreforthetranslationeventhoughthemachinetranslation is almostperfect. Similarpenaltyforusing a synonym can be seen in thethirdrowwhere “allow” is used in themachinetranslationinstead of “let”. AlsoStill BLEU is themostwidelyusedmetricformachinetranslationevaluationsothat it is used in thiswork.

Table2showssometranslationsperformedbytheproposedmethodemployinglongerandsomewhatmorecomplexsourcesentences. Thethree boldrows in thetableareselectedrandomlyfromthetraining data. Thereasonforthat is tosee how themachinetranslationwill be performedforthe data seenduringthetraining. As thetableshows, theresultsaremixed. The rest of thesourcesentencesare not usedfortraining. Therandomlyselectedsourcetext “Kızarmış pirinçli etin nasıl yapılacağını öğrenmek zorunda kalacağım” is machine-translated in such a waythat “need” is usedinstead of “willhaveto”. Thetranslationsemanticallyfoundthecorrecttranslationwhich is not reflected in the BLEU score. As expected, the BLEU scoresbecomelower as thesourcesentencelengthbecomeslonger.

Afterobservingthetraining data it can be seenthatmachinetranslationsare betterwithpatternsfrequentlyobserved. Forinstance, “wouldyouliketo” pattern in English seenoften in thetraining data is matchedwiththe “ister misiniz” pattern in Turkish. Therearequite a fewsentenceswith it in thetraining data sothatthelearning is accomplishedto a greatextent.

Source Sentence	Reference Translation	Inferred (Machine) Translation	BLEU Score
Ben zenginim.	I'm rich.	I’m bananas.	0.22
Bir motosikletim var.	I have a bike.	I have a plane.	0.40
Kazanmana izin verdim.	I letyouwin.	I allowguests.	0.08
Kahveyi severim	I likecoffee.	I likecoffee.	1.00
Bir kitap okuyorum.	I’m reading a book.	ill’mreading a book.	0.17
Kayak yapmaya gittim.	I wentskiing.	I wentskiing.	1.00
Ben bir hayalperestim.	I'm a dreamer.	I’m not a fool.	0.09
Deli misin?	Areyouinsane?	areyoucrazy?	0.32

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 906-910 www.ijaem.net ISSN: 2395-5252

Table 1. Short Source Sentence BLEU Scores

Source Sentence	Reference Translation	Inferred (Machine) Translation	BLEU Score
doktora gitmek ister misin	wouldyouliketosee a doctor	wouldyouliketogotothesee	0.36
eve gelir gelmez uyudum	islept as soon as i gothome	islept as soon as i gothome	1.00
bir kitaba ihtiyacim var	I need a book	ihave a book of reading a book	0.06
bize gelmek ister misin	wouldyouliketocomewith us	wouldyouliketocomewith us	1.00
tom iyi bir dosttur	tom is a goodfriend	tom is a goodgardener	0.67
ona bir kedi getirdim	ibrought her a cat	ibrought a bag of guests	0.05
Tom'un bana yapmamı söylediği her şeyi yapamadım	iwas not ableto do everythingtomtold me to do	iwas not ableto do everythingtomtold me to do	1.00
O şarkının ismini hatırlayamamıştım	I was not abletorememberthetitle of thatsong.	icouldhaveplayed her news on thatsong	0.04
Kızarmış pirinçli etin nasıl yapılacağını öğrenmek zorunda kalacağım	I willhavetolearn how tomakebeeffriedrice	ineedtolearn how tomakebeeffriedrice	0.68
arkadaşlarımahoşçakal demek istiyorum	I wantto say goodbyetomyfriends	imeanthatthesameway it smells	0.00
alışveriş yapmak için süpermarkete gidiyorum	I am goingtothe supermarket to do shopping	i am goingtothe supermarket to do a shoppingbag	0.60
Eve varmama 10 dakika kaldı	It is 10 minutesbefore I gethome	itwas ten minuteshomeyougethome	0.07
paranı geri vermek istiyorum	i'dliketogiveyourmoneyback	i'dliketogiveyourmoneyback	1.00
evimin önünde kazalar gördüm	I sawaccidents in front of myhouse	isawmymovingfrontdoor	0.04

Table2. Longe Source Sentence BLEU Scores

IV. CONCLUSION

Totheauthors' knowledge, machinetranslationfromTurkishto English is not studiedmuch. Especiallyresearchshowingtranslationresultswith BLEU scoresoranyothermetricdid not exist in theliteratur. Therefore, applying an encoder-decoderarchitecturewiththeattentionmechanismforTurkhs-English languagewasenough of motivation. Theproposed model is not a newone. On thecontrary it is a wellknown model. Thiswork is intendedtoseetheresults of theapplicationto a languagepairthat is not studiedmuch. Theresultsshowthat NMT withlimited data achievedsuccessto an acceptableextent. As mentionedabove, thepatternpairsthat morefrequentlyexist in thetraining data helpedtheproposed NMT model toperformbetter. TheNMT model is implementedusing Python andTensorflowwith a Nvidia GPU. Theparallelcorpustext is obtainedfromhttp://www.manythings.org/anki/whichcontainstextfiles of paralleltext of manylanguagepairs.

REFERENCES

[1]. Hutchins, J.(2007). Machine translation: A concise history. In Computer aided translation: theory and practice, C. S. Wai, Ed.

[2]. Junczys-Dowmunt, M., Dwojak, T., andHoang, H. (2016). Is Neural Machine Translation Ready for Deployment? A Case Study on 30 TranslationDirections. In: CoRRabs/1610.01108.

[3]. Niehues, J. et al. (2017). AnalyzingNeural MT Searchand Model Performance. In: First Workshop on Neural Machine Translation, pp. 11–17.

[4]. Vaswani, A. et al. (2017). Attention is AllyouNeed. In: Proceedings of theThirtyfirst Conference on Neural Information ProcessingSystems (NIPS), pp. 5998–6008.

[5]. Lakew, S. M., Cettolo, M., and Federico, M. (2018). A Comparison of TransformerandRecurrentNeural Networks on MultilingualNeural Machine Translation. In: Proceedings of the 27th International Conference on ComputationalLinguistics, pp. 641–652.

[6]. Kaiser, L. et al. (2018). FastDecoding in SequenceModels Using Discrete Latent Variables. In: Proceedings of theThirty- fth International Conference on Machine Learning (ICML), pp. 2395–2404.

[7]. Shazeer, N. and Stern, M. (2018). Adafactor: Adaptive Learning RateswithSublinear Memory Cost. In: Proceedings of theThirty-fth International Conference on Machine Learning (ICML), pp. 4603–4611.

[8]. Chen, M. X. et al. (2018). The Best of BothWorlds: CombiningRecent Advances in Neural Machine Translation. In: Proceedings of the 56th Annual Meeting of theAssociationforComputationalLinguistics, pp. 76–86.

[9]. Koehn, P. andKnowles, R. (2017). SixChallengesforNeural Machine Translation. In: First Workshop on Neural Machine Translation, pp. 28–39.

[10]. Yang, Y., Huang, L., and Ma, M. (2018). BreakingtheBeamSearchCurse: A Study of (Re-)ScoringMethodsandStoppingCriteriaforNeural Machine Translation. In: Proceedings of the 2018 Conference on EmpiricalMethods in Natural LanguageProcessing, pp. 3054–3059

[11]. Huang, L., Zhao, K., and Ma, M. (2017). WhentoFinish? Optimal BeamSearchforNeuralTextGeneration (modulobeam size). In: Proceedings of the 2017 Conference on EmpiricalMethods in Natural Language Processing, pp. 2134–2139.

[12]. Durgar El-Kahlout, İ., ve Oflazer K., (2006) "Türkçe-İngilizce için istatistiksel bilgisayarlı çeviri sistemi." FifteenthTurkishSymposium on ArtificialIntelligenceandNeural Networks, Akyaka, Muğla, Türkiye

[13]. Yeniterzi R., ve Oflazer K., (2010). "Syntax-to-morphologymapping in factoredphrase-basedstatisticalmachinetranslationfromEnglish toTurkish." 48. Annual Meeting of theAssociationforComputationalLinguistics. AssociationforComputationalLinguistics.