

Deep Learning Approach for Semantic Indexing of Animal Experiments Summaries in German Language

S. Kayalvizhi, D. Thenmozhi and Chandrabose Aravindan

Department of CSE, SSN College of Engineering, Chennai
kayalvizhi1704@cse.ssn.edu.in{theni_d,aravindanc}@ssn.edu.in

Abstract. Semantic indexing of animal experiment summaries is the process of annotating the summaries with its medical codes. Semantic indexing is helpful in reducing time and performance in knowing the context and finding relevant summaries. Indexing the Non-Technical Summaries (NTP)s using codes from the German version of the International Classification of Diseases (ICD-10) is a challenging task. ICD-10 codes, which is a comprehensive way of storing the health conditions are useful for the identification of many disorders, diseases and other health related problems. Thus, annotating the NTPs with codes will make the way of storing, organising, retrieval and comparing the health information more easier. In our paper, we have approached the problem using deep neural network. This work is evaluated on the dataset given by eHealth@CLEF2019. The test set given by the task is used to evaluate our methodology which attains precision, recall and f1 score of 0.19, 0.27 and 0.23 for Run 1 , 0.19, 0.27 and 0.22 for Run 2 and 0.13, 0.34 and 0.19 for Run 3 respectively. The performance of our method can further be increased by considering other recurrent units.

Keywords: Semantic indexing · Deep neural network · Text Mining · Deep Learning · Non-technical summaries.

1 Introduction

In the current generation, storing the health reports in digital way helps us in great way to diagnose and reduce medical complications at easier way. eHealth@CLEF2019¹ [7] is a shared task done on those electronically available medical records which consists of three tasks namely Task 1 - Multilingual Information Extraction, Task 2 - Technologically Assisted Reviews in Empirical Medicine and Task 3 - Consumer Health Search, in which this paper focuses on

Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CLEF 2019, 9-12 September 2019, Lugano, Switzerland.

¹ <http://clef-ehealth.org/>

Task 1 - Multilingual Information Extraction.[11]

In this task, Non Technical Summaries(NTPs) are given which are to be indexed with their International Classification of Diseases (ICD-10) in German version. ICD-10 codes are helpful in many ways to diagnose various diseases and identify its drugs [4, 12, 6, 3, 1, 14, 5]. NTPs are short summaries which are currently publicly available in the AnimalTestInfo database², as part of the approval procedure for animal experiments in Germany. The database currently contains more than 10,000 NTPs, many of which have been manually indexed by experts. We have built a deep neural network with LSTM to generate the codes for the summaries.Task 1 of eHealth@CLEF2019 focuses on automatic indexing of NTPs with ICD-10 codes.

2 Dataset Description

The dataset is given for the task 1 of eHealth-CLEF@2019[8]. The training set contains 7544 document ids in which 5854 are along with its annotations, development set has 842 ids in which 654 are along with its annotations and the test set contains 407 ids whose annotations are to be found out. Each document of animal summaries in the dataset has six lines of text which has the following information:

1. title of the document;
2. uses (goals) of the experiment;
3. possible harms caused to the animals;
4. comments about replacement (in the scope of the 3R principles);
5. comments about reduction (in the scope of the 3R principles);
6. comments about refinement (in the scope of the 3R principles).

The example for ICD-10 codes:

C50-C50|C00-C97|C00-C75|II

where '|' separates one code from another one.

3 Proposed Methodology

A Deep neural network model [10, 9] is used in our work for generating the ICD-10 codes for the NTPs given in German language. The data is prepared for giving as input to Seq2Seq deep learning algorithm. The input documents are split up into sentences (six for each document) and the corresponding ICD-10 codes are generated for the documents. The vocabulary for all the input documents and output labels are all formed.

A deep neural network model was built using a multi-layer RNN (Recurrent Neural Network) in which LSTM (Long Short Term Memory) as its recurrent

² <http://www.animaltestinfo.de>

unit. Layers namely embedding layer, encoder-decoder layer, projection layer and loss layer are used to build the deep neural network. The input lines in the document and its corresponding code labels in the embedding layer are used to learn the weight vectors based on their vocabulary. Two hidden layers are used for encoding and decoding. The attention mechanism such as Normed Bahdanau(NB) and Scaled Luong(SL) models [2,9] are used. Softmax is used as activation function in the projection layer to obtain the ICD-10 codes for the summaries. Loss computed in the loss layer is reduced by back propagation while building the model. Thus, ICD-10 codes is obtained by using the built Seq2Seq model 1. The TensorFlow code based on tutorial code released by

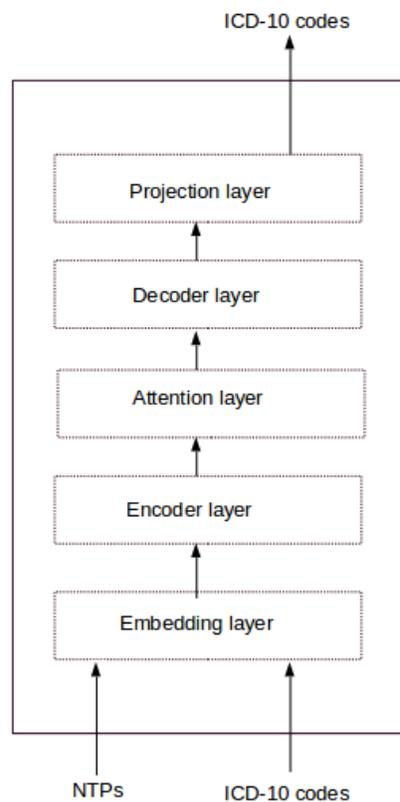


Fig. 1. System Architecture for Semantic Indexing of Animal Experiment Summaries.

Neural Machine Translation³ [9] that was developed based on Seq2Seq models [13,2,10] is used to implement our deep learning approach for Semantic

³ <https://github.com/tensorflow/nmt>

indexing. We have implemented two broad variations of the Seq2Seq model by varying the attention models with a batch size of one hundred and twenty eight, two encoder-decoder layers, dropout of 0.2 and bi-directional. Further variations are done by different considerations in post processing and type of input(attention model) given. The different variations are explained in the Table 1.

Table 1. Different variations

Model No.	Attention model	Consideration of occurrences
1	NB	all
2	SL	all
3	NB	minimum of two
4	SL	minimum of two
5	NB	minimum two and whole if nothing is generated
6	SL	minimum two and whole if nothing is generated
7	NB	minimum two and whole if only one code is generated
8	SL	minimum two and whole if only one code is generated
9	NB	minimum two and whole if only one or nothing is generated
10	SL	minimum two and whole if only one or nothing is generated

The performance obtained for these variations are evaluated using the evaluation script provided by the CLEF-ehealth@2019 in shown Table 2.

Table 2. Performance of development set

Models	Precision	Recall	F1-score
1	0.24	0.81	0.37
2	0.24	0.81	0.37
3	0.61	0.67	0.64
4	0.61	0.67	0.64
5	0.56	0.68	0.61
6	0.55	0.68	0.61
7	0.52	0.68	0.59
8	0.52	0.69	0.59
9	0.48	0.70	0.57
10	0.48	0.70	0.57

From the above models, model 3, 4 and 6 are submitted as Run 1, Run 2 and Run 3 respectively for the shared task.

The evaluation script provided by the CLEF-eHealth@2019⁴ is used to evaluate our models with respect to both development set and test set. The devel-

⁴ <https://github.com/mariananeves/clef19ehealth-task1>

Table 3. Result of evaluation script

Runs	Test set			Development set		
	True positive	False positive	False negative	True positive	False positive	False negative
Run 1	213	889	570	1127	700	555
Run 2	210	871	573	1128	694	554
Run 3	265	1788	518	1148	884	534

opment set has more accuracy than compared to the test set. Testing accuracy can also be improved by building the model using test set as its development set. From the Table 3, Run 3 has more true positive score than the other runs for both development set and test set.

4 Results

The final evaluation is done on the dataset provided by CLEF-eHealth@2019. The test set contains 407 summaries which should be annotated.

Table 4. Final evaluation for Test Data

Teams and Runs	Precision	Recall	F-score
SSN_NLP Run 1	0.19	0.27	0.22
SSN_NLP Run 2	0.19	0.27	0.23
SSN_NLP Run 3	0.13	0.34	0.19
DEMIR Run 1	0.46	0.50	0.48
DEMIR Run 2	0.49	0.44	0.46
DEMIR Run 3	0.46	0.49	0.48
IMS_UNIPD Run 1	0	0	0
IMS_UNIPD Run 2	0.009	0.50	0.017
IMS_UNIPD Run 3	0.10	0.05	0.07
MLT-DFKI	0.64	0.86	0.73
TALP_UPC	0.37	0.35	0.36
WBI Run1	0.83	0.77	0.80
WBI Run2	0.84	0.74	0.79
WBI Run3	0.80	0.78	0.79

From the Table, it seems that our approach (SSN_NLP) does not outperform the other approaches. The performance may be improved by using other recurrent units.

5 Conclusions

Semantic indexing is the indexing of animal experiment summaries with ICD-10 codes in German version. We have made use of deep neural network with two different attention models for the indexing the summaries with their medical codes such as C50-C50|C00-C75|II. We have splitted the documents into sentences and generated the codes with respect to the documents and combined them according to minimum of two occurrences with NB attention for Run 1 which has 0.19, 0.27 and 0.23 and the same with SL attention for Run 2 which has 0.19, 0.27 and 0.22 and considering minimum of two occurrences and considering the whole codes generated if nothing is generated as its code as Run which has 0.13, 0.34 and 0.19 as its precision, recall and F1 score which is evaluated on the test set of eHealth@CLEF2019. Further improvements can be done by considering Google Neural Machine Translation (GNMT), Gated Recurrent Unit (GRU) as recurrent units instead of LSTM.

References

1. Alotaibi, G.S., Wu, C., Senthilselvan, A., McMurtry, M.S.: The validity of icd codes coupled with imaging procedure codes for identifying acute venous thromboembolism using administrative data. *Vascular medicine* **20**(4), 364–368 (2015)
2. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)
3. Bergen, D.C., Beghi, E., Medina, M.T.: Revising the icd-10 codes for epilepsy and seizures. *Epilepsia* **53**, 3–5 (2012)
4. Germaine-Smith, C.S., Metcalfe, A., Pringsheim, T., Roberts, J.I., Beck, C.A., Hemmelgarn, B.R., McChesney, J., Quan, H., Jette, N.: Recommendations for optimal icd codes to study neurologic conditions: a systematic review. *Neurology* **79**(10), 1049–1055 (2012)
5. Hohl, C.M., Karpov, A., Reddekopp, L., Stausberg, J.: Icd-10 codes used to identify adverse drug events in administrative data: a systematic review. *Journal of the American Medical Informatics Association* **21**(3), 547–557 (2013)
6. Jette, N., Beghi, E., Hesdorffer, D., Moshé, S.L., Zuberi, S.M., Medina, M.T., Bergen, D.: Icd coding for epilepsy: past, present, and future—a report by the international league against epilepsy task force on icd codes in epilepsy. *Epilepsia* **56**(3), 348–355 (2015)
7. Kelly, L., Goeuriot, L., Suominen, H., Neves, M., Kanoulas, E., Spijker, R., Azzopardi, L., Li, D., Palotti, J., Zuccon, G., et al.: Clef ehealth 2019 evaluation lab. In: *European Conference on Information Retrieval*. pp. 267–274. Springer (2019)
8. Kelly, L., Suominen, H., Goeuriot, L., Neves, M., Kanoulas, E., Li, D., Azzopardi, L., Spijker, R., Zuccon, G., Scells, H., ao Palotti, J.: Overview of the CLEF eHealth evaluation lab 2019. In: *Crestani, F., Braschler, M., Savoy, J., Rauber, A., et al. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Tenth International Conference of the CLEF Association (CLEF 2019)*. Lecture Notes in Computer Science. Springer, Berlin Heidelberg, Germany (2019)
9. Luong, M., Brevdo, E., Zhao, R.: Neural machine translation (seq2seq) tutorial. <https://github.com/tensorflow/nmt> (2017)

10. Luong, M.T., Pham, H., Manning, C.D.: Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025 (2015)
11. Neves, M., Butzke, D., Dörendahl, A., Leich, N., Hummel, B., Schönfelder, G., Grune, B.: Overview of the CLEF eHealth 2019 Multilingual Information Extraction. In: Crestani, F., Braschler, M., Savoy, J., Rauber, A., et al. (eds.) *Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Tenth International Conference of the CLEF Association (CLEF 2019)*. Lecture Notes in Computer Science. Springer, Berlin Heidelberg, Germany (2019)
12. Skull, S.A., Andrews, R.M., Byrnes, G.B., Campbell, D.A., Nolan, T.M., Brown, G.V., Kelly, H.A.: Icd-10 codes are a valid tool for identification of pneumonia in hospitalized patients aged 65 years. *Epidemiology & Infection* **136**(2), 232–240 (2008)
13. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: *Advances in neural information processing systems*. pp. 3104–3112 (2014)
14. Thygesen, S.K., Christiansen, C.F., Christensen, S., Lash, T.L., Sørensen, H.T.: The predictive value of icd-10 diagnostic coding used to assess charlson comorbidity index conditions in the population-based danish national registry of patients. *BMC medical research methodology* **11**(1), 83 (2011)