



**HAL**  
open science

# Malayalam Language Textual Inference Identification Through Attention over Embeddings

Sara Renjit, Sumam Mary Idicula

► **To cite this version:**

Sara Renjit, Sumam Mary Idicula. Malayalam Language Textual Inference Identification Through Attention over Embeddings. 6th International Conference on Computer, Communication, and Signal Processing (ICCCSP), Feb 2022, Chennai, India. pp.3-11, 10.1007/978-3-031-11633-9\_1. hal-04388162

**HAL Id: hal-04388162**

**<https://inria.hal.science/hal-04388162v1>**

Submitted on 11 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

# Malayalam Language Textual Inference Identification through Attention over Embeddings

Sara Renjit <sup>\*\*1</sup> and Sumam Mary Idicula<sup>2</sup>

<sup>1</sup> Department of Computer Science, Cochin University of Science and Technology  
sararenjit.g@gmail.com

<sup>2</sup> Department of Computer Science, Muthoot Institute of Technology and Science

**Abstract.** This work focus on natural language inference between two text fragments, which is a uni-directional relationship. Natural language inference is attempted for a language called Malayalam. The Malayalam language is a South Indian, low-resource language. NLI is a subtask in every language because of its importance in summarization, information retrieval, and many other applications. There are only a few attempts in NLI for the Malayalam language. In this work, the application of the attention mechanism helped to enhance the classification performance for binary and multiclass systems. A densenet with additive attention is implemented for softmax and sigmoid classification, increasing accuracy, recall, and F1-score.

**Keywords:** Natural language inference · Malayalam · Attention

## 1 Introduction

A directional relationship between sentence pairs is known as textual entailment. Natural language inference is another term used commonly now for recognizing textual entailment. Recognizing textual entailment is one of the two types of textual entailment, where the idea is to identify the entailments in text. Another type of textual entailment is generating textual entailment, where the entailed sentence is generated from a text.

Text(T) and hypothesis(H) are sentences or text fragments, where text T entails hypothesis H if the meaning of hypothesis can be inferred from the meaning of the text. Text contradicts with hypothesis if the meaning of text and hypothesis is opposite. The hypothesis is neutral with text if the meaning of the hypothesis cannot be inferred/ remains neutral with the meaning of the text.

Textual entailment has its definition in different senses. Most classically, it is defined as the text entails hypothesis, if the hypothesis is valid in all possible circumstances in which text is true. In a more applied sense, text entails a hypothesis if the hypothesis is true when a human reads it. Mathematically,

---

<sup>\*\*</sup> Corresponding author: sararenjit.g@gmail.com

entailment is defined as the probability of the hypothesis being true with respect to text is greater than the likelihood of hypothesis as true [7].

NLI is a binary classification task identifying entailment and contradiction classes. It can also be a multiclass classification with entailment, neutral, and contradiction classes. Identifying these distinct classes for sentence pairs is vital for many language processing applications. In multi-document summarization, redundant sentences can be removed. In information retrieval, it helps in the faster retrieval of entailed sentences. In question answering systems, it helps in the efficient retrieval of answers for query input. NLI is also considered as an evaluation task for various sentence models and transformer models.

The contribution of this work is the application of the attention mechanism over a dense net with LASER sentence representation. We have obtained a performance improvement on the best so far results with additive attention applied. The subsequent sections include related works in Section 2. Dataset information is in Section 3. Section 4 details the proposed system, and Section 5 details the experimental evaluations. Section 6 discusses the results and their observations, and Section 7 concludes the work.

## 2 Background

Textual entailment started as a challenge in 2005 for the English language to develop systems that can identify similar meanings between text fragments. This challenge continued in the following years by the name RTE-1, RTE-2, RTE-3, RTE-4, RTE-5 with different sized datasets in the order of 1k sentence fragments. The approaches used with the RTE dataset classification are mainly cross pair similarity, word and phrase alignments, dependency tree-based transformation, logical inference, ontology-based learning, and machine learning algorithms like Support Vector Machines (SVM) [5].

**Table 1.** RTE challenges [6]

RTE	T-H pairs	Features
1	287	Mostly lexical systems.
2	800	Question Answering domain.
3	800	Longer sentences.
4	1000	3-way tasks.
5	600	Unedited real world texts.
6	15955	221 hypothesis.
7	21,420	Paragraph long texts.

Since the inception of the RTE challenge in 2005, there have been numerous attempts for RTE in English and other languages like French, German, and Italian. The evolution of RTE challenges is described in Table1. Recognizing textual entailment is now known by the term natural language inference. Later

in 2015, the Stanford Natural Language Inference dataset [3] was introduced, which is a large dataset of size 570k sentence pairs collected through Amazon mechanical trunk. With this dataset, deep learning techniques are also used for RTE tasks. Sum of words, RNN, and LSTM based embeddings are used to obtain the sentence representations for classification.

MNLI (Multi-genre natural language inference dataset) [13] and XNLI (cross-lingual natural language)[4] datasets are newer versions of NLI dataset. MNLI has sentences derived from multiple genres like face-to-face, letters, telephone, government, slate. XNLI is a cross-lingual dataset derived from MNLI by automatic translations to 15 languages.

With the XNLI dataset, identifying entailments in the text is attempted in other languages like French, German, Japanese, Italian, Spanish, Hindi, and low resource languages like Swahili and Urdu. Different approaches using LSTM (Long Short Term Memory networks), BiLSTMs, GRU, XLM-R, and transformers are used. [9] gives an overview of the deep learning approaches used for textual entailment with sentence encoding models, match encoding models, tree-based CNN models, and other hybrid models.

Some of the existing works for textual entailment approaches include generating inferences from texts using residual LSTM [8], recognizing entailments using word by word neural attention [11], using sentence embeddings and decomposable attention model for Japanese language [12]. Textual entailment in languages other than English is also attempted by automatic translation into English and using the best approaches for English language [10].

For the Malayalam language in the context of this work, there are very few works for entailment. The MaNLI (Malayalam Natural Language Inference) dataset is introduced, which are translated pairs from the SNLI dataset. In their work, language agnostic sentence representation (LASER) based embedding encodes the sentence pairs, which are then classified through a dense net with Softmax or sigmoid classifier. They have obtained an accuracy of 64% for multiclass and 77% for binary classification.

### 3 Dataset

In this work, the MaNLI (Malayalam Natural Language Inference) dataset is used, created through the human and google translations of text hypothesis pairs from the SNLI (Stanford Natural Language Inference) dataset. The linguistic corrections and semantic consistency are manually verified by linguists from the Thunchath Ezhuthachan Malayalam University, Kerala.

The dataset consists of 12k pairs of text hypothesis sentence pairs labeled as entailment, contradictory and neutral. The same dataset as binary classes has 7989 pairs of sentences with classes entailment and contradiction. The statistics of the dataset are given in Table2.

A sample from the Malayalam language dataset and its English translations is in Figure1 and Table3 respectively.

Table 2. MaNLI dataset

Classes	No of Sentence pairs
Entailment	4026
Contradiction	3963
Neutral	4011

Text	Hypothesis	Label
രണ്ട് ആളുകൾ സൈക്കിളിൽ മത്സര ഓട്ടം നടത്തുന്നു.	ആളുകൾ സൈക്കിളിൽ സഞ്ചരിക്കുന്നു.	entailment
രണ്ട് ആളുകൾ സൈക്കിളിൽ മത്സര ഓട്ടം നടത്തുന്നു.	ആളുകൾ തെരുവിൽ സൈക്കിളിൽ സഞ്ചരിക്കുന്നു.	neutral
രണ്ട് ആളുകൾ സൈക്കിളിൽ മത്സര ഓട്ടം നടത്തുന്നു.	കുറച്ച് ആളുകൾ മീൻ പിടിക്കുന്നു.	contradiction

Fig. 1. Sample from the MaNLI dataset

## 4 Proposed System

Natural language inference is regarded as a two-way (binary) and three-way (multiclass) classification task. The design of the system implemented is shown in Figure2. The input sentences, namely text, and hypothesis are embedded using LASER. LASER embedding based representations are fed to the dense net with relu activations, and then attention is applied and fed to the final dense layer with sigmoid/ softmax activations to obtain probability distributions to each class.

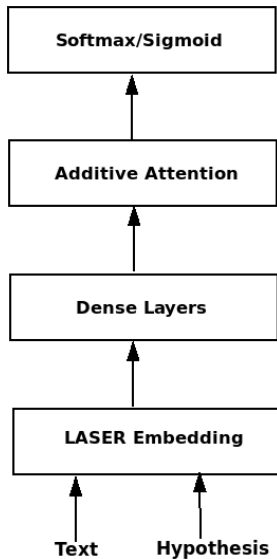
The proposed system for this classification consist of the following modules: **LASER embeddings for sentence representation:** Language agnostic sentence representations [1] are used for representing text and hypothesis pairs in a meaningful numerical format. It is an encoder-decoder architecture where the input words are encoded using byte pair encoding and trained on parallel corpora. Sentence embeddings are obtained by max-pooling over the BiLSTM layer. In the previous work, LASER representation showed higher performance as compared to other representations namely, Doc2Vec, mBERT and fastText. This work improves the LASER representation based classifier with the application of attention.

**Dense layers:** The attention outputs are then passed through feed forward network consisting of multiple dense layers. Dense layers are deeply connected layers that apply the activation function on the summation of weighted input (WX) with bias (B), (WX+B), where W is the weight matrix, and X is the input.

**Additive attention:** Additive attention [2] computes attention score using the tanh activation on a feed forward layer for attention alignment (Eq. 1). Softmax operation is applied to the attention score and attention weights are

**Table 3.** Sample Dataset in English (SNLI [3])

Text	Hypothesis	Label
Two men on bicycles competing in a race.	People are riding bikes. Men are riding bicycles on the street. A few people are catching fish.	entailment neutral contradiction



**Fig. 2.** System Design

obtained (Eq. 2). These weights are probabilities that indicate the importance of words in a sequence. Finally context vector is derived as weighted sum of previous hidden representation (Eq. 3).

$$e_{t,i} = \tanh(W[h_i; s_{t-1}]), \tag{1}$$

where  $e_{t,i}$  is the attention score,  $h_i$  is the hidden representation and  $s_{t-1}$  is the previous state.

$$\alpha_{t,i} = \text{softmax}(e_{t,i}) \tag{2}$$

where  $\alpha_{t,i}$  is the attention weight.

$$c_t = \sum_{i=1}^T \alpha_{t,i} h_i \tag{3}$$

where  $c_t$  is the context vector.

**Softmax/Sigmoid:** Softmax activation is used in the final dense layer to predict probability distribution for classification into three classes, namely entailment, contradiction, and neutral. Sigmoid activation is used to classify the input into two classes, entailment, and contradiction for binary classification.

## 5 Experiments and Evaluation

*Experiments* : The implementations are conducted in Google Colab and using Python3 with Tensorflow and Keras. The Spyder environment is also used for training. The sentences are embedded using the pretrained LASER model. Text and hypothesis representations are concatenated with their difference and dot product before it is input to the densenet. The dense layers have neuron configurations of 4096X512X384. The activations used in hidden layers is rectified leaky unit (ReLU). Attention mechanism is implemented using the dense layer with tanh activation followed by softmax of the output. This softmax output is then multiplied with the previous hidden state and fed to the final dense layer.

*Evaluation* : The evaluation metrics used in this classification are accuracy, precision, recall, and F1-score.

- Accuracy is defined as the metric that defines the performance of the model across all classes.  $\text{Accuracy} = \frac{\text{No of correct predictions}}{\text{Total no of predictions}}$ .
- Precision (P) is defined as the number of correct predictions for a class divided by the total predictions for that class.
- Recall (R) is the number of correct predictions for a class divided by all relevant samples for that class.
- F1-score (F1) measures the precision of the classifier. It is the harmonic mean of precision and recall.
- Support (S) is the number of the actual occurrence of each class in the dataset.

Table 4 shows the results for binary classification. Table 5 shows the results obtained for multiclass classification.

**Table 4.** Classification report for binary classification

Class	Precision	Recall	F1-score	Support
Contradiction	0.84	0.72	0.78	798
Entailment	0.76	0.86	0.81	800
Accuracy			0.79	1598
Macro average	0.80	0.79	0.79	1598
Weighted average	0.80	0.79	0.79	1598

## 6 Results and Observations

Table 6 shows the comparison of weighted average precision, recall and F1-scores. An improvement in recall F1-measure is obtained with attention. Table 7 gives accuracy based comparison with the previous LASER based system. Attentive



**Table 5.** Classification report for multiclass classification

Class	Precision	Recall	F1-score	Support
Contradiction	0.65	0.63	0.64	1651
Entailment	0.71	0.68	0.70	1682
Neutral	0.59	0.63	0.61	1667
Accuracy			0.65	5000
Macro average	0.65	0.65	0.65	5000
Weighted average	0.65	0.65	0.65	5000

**Table 6.** Weighted average precision, recall and F1-score comparison of LASER embedding based classification with and without attention

LASER Model	Class	Precision	Recall	F1-score
Densenet	Binary	0.79	0.77	0.77
	Multiclass	0.65	0.64	0.64
Densenet + Attention	Binary	0.80	0.79	0.79
	Multiclass	0.65	0.65	0.65

**Table 7.** Accuracy of models with and without attention for binary and multiclass classification

Method	Binary	Multiclass
LASER+Densenet	0.77	0.64
LASER + Densenet + <b>Attention</b>	<b>0.79</b>	<b>0.65</b>

context vector in the final representation also helped in increasing accuracy measures.

The application of attention mechanism better classifies the text hypothesis pairs with improved scores. This highlights that sentence representation is improved through attentive context vector representation.

The semantic content differs between sentence pairs, and generalizations are difficult in this type of dataset. Also, the dataset has a dynamic nature representing different factual information and inherit properties of the language such as inflectional nature, agglutination, word compounding, multiple words having similar meanings. These distinct properties make the dataset and its classification a challenging task, and hence these are few factors that affect the performance.

## 7 Conclusion

This work has been an attempt to improve the existing performance of textual entailment for Malayalam language using attention mechanism. Attentive sentence representations through multiple layers in the existing LASER based dense-net system gained an improvement in accuracy and other classification metrics. This is also an application and understanding of the performance of

attention in NLI context for a Dravidian language like Malayalam. The semantic context variability and neutral category data are challenging elements in the dataset, on which attention has improved the performance.

## References

1. Artetxe, M., Schwenk, H.: Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics* **7**, 597–610 (2019)
2. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014)
3. Bowman, S., Angeli, G., Potts, C., Manning, C.D.: A large annotated corpus for learning natural language inference. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. pp. 632–642 (2015)
4. Conneau, A., Rinott, R., Lample, G., Williams, A., Bowman, S., Schwenk, H., Stoyanov, V.: Xnli: Evaluating cross-lingual sentence representations. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. pp. 2475–2485 (2018)
5. Dagan, I., Dolan, B., Magnini, B., Roth, D.: Recognizing textual entailment: Rational, evaluation and approaches—erratum. *Natural Language Engineering* **16**(1), 105–105 (2010)
6. Ghuge, S., Bhattacharya, A.: Survey in textual entailment. *Center for Indian Language Technology*, retrieved on April (2014)
7. Glickman, O., Dagan, I., Koppel, M.: A probabilistic lexical approach to textual entailment. In: *IJCAI*. vol. 5, pp. 1682–1683 (2005)
8. Guo, M., Zhang, Y., Zhao, D., Liu, T.: Generating textual entailment using residual lstms. In: *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*, pp. 263–272. Springer (2017)
9. Mishra, A., Bhattacharyya, P.: Deep learning techniques in textual entailment. *Survey Paper, Center For Indian Language Technolgy* (2018)
10. Pakray, P., Neogi, S., Bandyopadhyay, S., Gelbukh, A.: Recognizing textual entailment in non-english text via automatic translation into english. In: *Mexican International Conference on Artificial Intelligence*. pp. 26–35. Springer (2012)
11. Rocktäschel, T., Grefenstette, E., Hermann, K.M., Kočiský, T., Blunsom, P.: Reasoning about entailment with neural attention. *arXiv preprint arXiv:1509.06664* (2015)
12. Son, N.T., Phan, V.A., Nguyen, L.M.: Recognizing entailments in legal texts using sentence encoding-based and decomposable attention models. In: *COLIEE@ ICAIL*. pp. 31–42 (2017)
13. Williams, A., Nangia, N., Bowman, S.: A broad-coverage challenge corpus for sentence understanding through inference. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. pp. 1112–1122 (2018)