An Annotated Corpus for Realis Event Detection in Short Stories Written in English and Low Resource Assamese Language

Chaitanya Kirti[1], **Pankaj Choudhury**[1], **Ashish Anand**[2,1], **Prithwijit Guha**[3,1]
[1]Centre for Linguistic Science and Technology
[2]Department of Computer Science and Engineering
[3]Department of Electronics and Electrical Engineering
Indian Institute of Technology Guwahati, Assam, India
{ckirti,pankajchoudhury,anand.ashish,pguha}@iitg.ac.in

Abstract

This paper presents an annotated corpora of Assamese and English short stories for event trigger detection. This marks a pioneering endeavor in short stories, contributing to developing resources for this genre, especially in the low-resource Assamese language. In the process, 200 short stories were manually annotated in both Assamese and English. The dataset was evaluated and several models were compared for predicting events that are actually happening, i.e., realis events. However, it is expensive to develop manually annotated language resources, especially when the text requires specialist knowledge to interpret. In this regard, TagIT, an automated event annotation tool, is introduced. TagIT is designed to facilitate our objective of expanding the dataset from 200 to 1,000. The best-performing model was employed in TagIT to automate the event annotation process. Extensive experiments were conducted to evaluate the quality of the expanded dataset. This study further illustrates how the combination of an automatic annotation tool and human-in-the-loop participation significantly reduces the time needed to generate a high-quality dataset.

1 Introduction

Event detection aims to find event instances, commonly called event triggers. An event is a particular instance of the occurrence of something at a specific time and place, thus indicating state change (Peinelt et al., 2020). Event detection helps extract relevant information from vast amounts of text. It is considered a crucial part of Information Extraction due to its importance in several downstream tasks like Question-Answering (Li et al., 2020), Knowledge-base construction and completion (Hürriyetoğlu et al., 2021), etc. Significant efforts have been put into event detection in texts from domains like newswire (Dasigi and Hovy, 2014) and biomedical (Wang et al., 2016). However, event detection from short stories remains under-studied, specifically in low-resource languages (Sims et al., 2019). The overarching aim of this work is to explore event detection in Assamese and English short stories written in the Indian context. Recognizing events within literary texts, such as short stories, poses substantial challenges. Challenges arise due to several factors. Literary narratives tend to feature more complex and intricate storytelling structures (Sims et al., 2019). Additionally, the language used in news articles or biomedical texts often describes real-life occurrences and establishes clear cause-and-effect connections between events (Sprugnoli and Tonelli, 2017). In contrast, literature is fundamentally a creative pursuit, and within most literary narratives, events are not necessarily grounded in factual reality. The coexistence of both actual or realistic events and non-realistic events within these texts further complicates the task of event detection. Exploring events in short stories would involve examining how events are conceptualized, encoded, and expressed.

Towards the primary objective, this work

1. Assamese:
তেওঁ গা **ধুইছিল** আৰু ৰাতিপুৱাৰ আহাৰ **খাইছিল**।
Gloss : "teo gaa dhuichil aaru ratipuwar ahar khaisil"
2. English:
Rancho **walked** up to him and **said**, "You are indeed stronger than all of us."

Figure 1: Examples illustrating the events present in Assamese and English short story dataset. Bold words are event triggers.

presents an annotated corpus for Assamese and English short stories. Two hundred short stories in both Assamese and English languages are manually annotated. Only actually happening events or realis events are considered for annotation. Examples of sentences with events in Assamese and English are shown in Figure 1. Several baseline models are trained on a subset of manually annotated datasets and evaluated for the event detection task. However, creating a large gold standard dataset takes time and effort. A silver standard corpus is an alternative to a gold standard annotated corpus (Rebholz-Schuhmann et al., 2010). The idea of employing automatic annotation systems and a method to synchronize the resulting annotations to produce the silver standard was first put forth (Rebholz-Schuhmann et al., 2011).

This work also opts for a silver standard dataset as it is still being used widely (Sousa et al., 2019). An automatic annotation tool, "*TagIT*", is developed to create this dataset. Using *TagIT*, the event detection dataset is expanded from 200 to 1,000 stories in each language. The quality of the silver standard augmented data is significantly enhanced when the process is integrated with human-in-the-loop participation. For an input document (short story), *TagIT* recommends the possible events. This aids in further verification (with greater ease) by the human annotator.

The major contributions of this work are as follows:

- This work introduces a manually annotated dataset containing 200 short stories in both Assamese and English, annotated for actually happening events, i.e., realis events. The dataset is evaluated through extensive experimentation.

- The paper presents "*TagIT*", an easy-to-use automatic annotation tool especially designed for automatic event annotation.

- The proposed dataset with 200 manually annotated short stories is further expanded to 1,000 stories in both Assamese and English using the automatic annotation tool, "*TagIT*". The quality of the resulting 1,000 annotated data is further enhanced by the human-in-the-loop technique.

2 Related Work

Event trigger detection in a text document is a necessary first step for identifying events (Liao and Grishman, 2010). Over time, there has been a steadily growing demand for corpora with extensive annotations (Sun et al., 2017). To study linguistic phenomena at different levels, a corpus in a machine-readable form is crucial. This corpus can be utilized in several domains. Also, supervised machine learning techniques depend on annotated corpora to create and assess NLP techniques. ACE 2005 marked the pioneering news corpus dedicated to addressing real events (Consortium et al., 2005).

The absence of biological corpus for event detection led to the development of the GENIA corpus (Abdullah et al., 2022). Annotators from Amazon Mechanical Turk developed a set of sentences extracted from news articles, covering reports on various topics such as events, science journalism, and finance, classifying them as either general or specific (Louis and Nenkova, 2012). The Gun Violence Database (GVDB) was unveiled as a fresh dataset containing articles on gun violence from local newspapers and television station reports, all of which have been annotated (Pavlick et al., 2016). Xiang et al. (Xiang and Wang, 2019) thoroughly explored numerous datasets for the purpose of event detection.

Most prior research focused on handling factual information. In contrast, very few have analyzed events within literary texts. The first attempt to identify events in literary content by generating a marked collection of real events from novels was proposed in (Sims et al., 2019). The scarcity of annotated data poses a significant challenge in event detection tasks, especially when dealing with Indian languages with limited linguistic resources (Roy et al., 2023). Several efforts have been made to detect events in Indian language documents written in Hindi (Sahoo et al., 2020), Tamil (Kuila et al., 2018), Bengali (Mishra, 2020), and Marathi(Dave et al., 2020). However, these works have primarily focused on news and social media text. Only a single work focused on the detection of natural calamities by processing Assamese language posts in social media (Kalita et al., 2021).

This work attempts to identify events within children's short stories. Literary works like short stories are longer than fact-based articles and exhibit unique event structures. Detecting real events within these imaginative narratives poses a significant challenge. The narrative structure caters to the psychology of children, featuring situations where inanimate objects or animals converse with each other. These distinctive scenarios set these stories apart from both factual and other literary texts.

3 Dataset

This section describes the dataset collection method (subsection 3.1), annotation guidelines (subsection 3.2), and the annotation process (subsection 3.3). Finally, the dataset statistics are summarized in subsection 3.4.

3.1 Collection

This work focuses on two languages: resource-abundant English language and resource-scarce Assamese language. The datasets are constructed by crawling short stories from various blogs available in the public domain. Accordingly, 1,000 short stories are collected for the two languages. These stories revolve around children and are set within the Indian cultural context. The stories are taken from the following sources: *Panchtantra*, *Tenali-Rama* and two famous Indian epics *Ramayana* and *Mahabharata*. *Panchtantra* consists of interrelated animal fables, while *Tenali-Rama* contains stories on the intelligence and problem-solving ability of a court jester. The diversity of story genres ensures a varying nature of stories that is essential for delving into different events, writing techniques, and narrative approaches.

3.2 Annotation Guidelines

The annotation guideline is adapted from the ones proposed in ACE (Consortium et al., 2005), HISTO Corpus (Sprugnoli and Tonelli, 2019), and Litbank (Sims et al., 2019) for event annotation. Every event is denoted by a word that elicits that specific event. This word is referred to as an event trigger. Usually event triggers are single verbs, but can also be nouns or adjectives. Recognizing events in short stories presents a distinctive challenge for following primary reasons. These stories are tailored for children as their target audience and they have specific narrative structures. In these narratives, it is typical to encounter animals or even inanimate objects participating in dialogues. The occurrence of events in such stories greatly depends on the storytelling style, genre, and overall approach to narration. This work adheres to the Light ERE (Aguilar et al., 2014) approach for annotating the realis events. Following are the aspects to capture realis events, adopted from Light ERE:

- `POLARITY`: Events should have positive polarity; events with negative polarity do not signify their occurrence.
- `SPECIFICITY`: Only specific events must be marked. Generic events that represent a persistent state should be ignored.
- `TENSE`: Past and present tense events are surely occurring. The surety of future events is uncertain.
- `MODALITY`: Events expressing beliefs, hypotheses, desires, promises, commands, requests, etc., are not necessarily bound to occur, and therefore, they are not tagged.

Along with these aspects, this work introduces the new aspect of `CONDITIONALITY`. This means that if an event is dependent upon a condition, and that condition is found to be invalid, then the event will not be tagged. For instance, in the sentence, "*If he comes, I will go to the market.*", the event trigger "*comes*" is not certain, at least with the limited context. So, neither "*comes*" nor "*go*" will be tagged as an event trigger.

In broad terms, the objective is to detect realis events. Realis events belong to the real world and are portrayed as existing within the imaginative realm of the literary work, occurring at a particular location and a specific moment. In literature, imaginary causalities exist, which are difficult to extract in contrast

Table 1: Basic statistics pertaining to words, sentences, and events in Assamese and English short story dataset.

Statistics	Assamese	English
Total words in the dataset	139,214	157,287
Total unique words in the dataset	14,410	10,558
Total sentences in the dataset	12,956	13,861
Average words per story	696	786
Average sentences per story	65	69
Average words per sentence	10	11
Total event triggers in the dataset	10,882	13,171
Average events per story	54	66

to the hard-coded causalities in fact-based reports like in news and bio-medical text (Saurí and Pustejovsky, 2009).

3.3 Annotation Process

Out of 1,000 stories in the dataset, only 200 stories have been manually annotated for realis events in each language. Two annotators carried out annotations using Brat rapid annotation tool (Stenetorp et al., 2012). We chose the two annotators who have expertise in the field of linguistics to make sure they can grasp the varied composition of short stories and annotate events correctly. To evaluate the inter-annotator agreement between annotators and assess the annotation guidelines, 20 short stories were annotated by the two annotators in each language. The inter-annotator agreement was found to be 84.1% and 86.5% for Assamese and English, respectively. Finally, the remaining short stories were thoroughly marked by these two annotators in each language, establishing them as the gold-standard annotation. Later, these 200 annotated stories are used to train and test the event detection models.

3.4 Dataset Description

Table 1 presents the basic statistics for both the datasets. The table clearly indicates that short stories are lengthier than news articles. These short stories tend to have an average length of 60–70 sentences. The average number of words per sentence is identical in Assamese and English. The average number of event triggers in English is more than Assamese. The number of events in the dataset is lesser than the number of sentences. Several sentences either do not contain events or they have non-realis events. The datasets also have

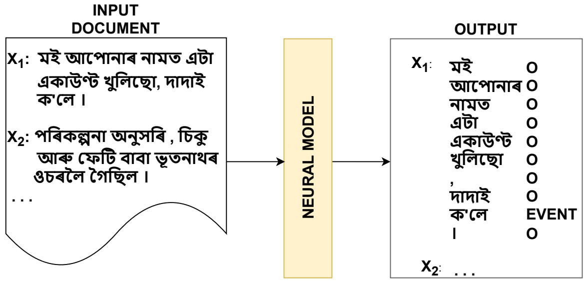

Figure 2: Block diagram illustrating the event detection framework.

sentences with two or more number of event triggers.

4 Experiment Settings

4.1 Baseline Models

The task of event trigger detection is framed as a sequence labeling task. Overall framework is shown in Figure 2. Neural model in this framework consists of the following three modules – (a) *Embedding Generation*, (b) *Sequence to Vector* and (c) *Dense Layer*. For the embedding generation module, this study has experimented with both static and contextual embeddings. Pre-trained Fasttext embeddings[1] (Grave et al., 2018) are used as static embedding, and BERT embeddings [2] (Kenton and Toutanova, 2019) are used as contextual embedding.

As baseline models, a set of different variants of LSTM and BiLSTM-based neural models with different choices of pre-trained embeddings are compared. The details of these neural models are provided in this section.

[1] https://fasttext.cc/docs/en/crawl-vectors.html
[2] https://huggingface.co/bert-base-multilingual-cased

Apart from LSTM and BiLSTM, the following three variations of BiLSTM are used – (a) *BiLSTM with Document Context*, (b) *BiLSTM with Sentence CNN*, and (c) *BiLSTM with Subword CNN*. All the variations of BiLSTM are trained separately using fasttext pretrained word vectors and BERT embeddings. Finally, the dense layer learns the function to predict the event status for each token of the input sentence. Specific descriptions regarding each baseline models are discussed below.

LSTM: A 100-dimensional, single-direction LSTM model is used as a baseline. LSTMs are good at capturing long-term dependencies, which makes them ideal for sequence prediction tasks.

BiLSTM: The accurate detection of a token as a label depends on the backward and forward context information. BiLSTM serves as a great baseline to model both backward and forward contexts. Further modification of the BiLSTM is done as described below to capture more information.

BiLSTM with Subword CNN: Subword character CNN is used as an intermediate layer between the embedding layer and BiLSTM. Subword character CNN captures meaningful representations of out-of-vocabulary words for learned embeddings. Each word embedding vector is represented as the result of a CNN having 100 filters, with max pooling resulting in a 100-dimensional character representation of a token.

The embedding of the token at that place is subsequently updated to include character representation, and the updated representation goes as an input to the subsequent LSTM layer.

BiLSTM with Sentence CNN: Several works have used a CNN on the sentence level (Nguyen and Grishman, 2015) for event detection. In a sentence $w = \{w_1, ..., w_n\}$ of n tokens, while determining the status of an event of a word at position i, the CNN (1-dimensional CNN) convolves over the sequence w with positional encodings $p = \{p_1, ..., p_n\}$. The positional encoding encodes the distance between the target token i and each token position $j \in [1, n]$. In this work, the work of Nguyen et al. (Nguyen and Grishman, 2015) was adopted, where the output of the CNN is then sent to a max-pooling layer to build a representation c_i for the target location i that is appended to the output of the BiLSTM o_i at the time step before generating the prediction. The CNN consists of 200 filters (Each word's bigrams and trigrams were scoped 100 times). Using signed bucketing $(\pm 1, \pm 2, \ldots, \pm 20, > 20)$, the positional arguments are encoded between the target token at position i and the token at position j. With its 5-dimensional embedding, each bucket refers to a unique choice of position.

Table 2: Dataset split for experiments.

Dataset Split	Train	Validation	Test
No. of stories	120	20	60

BiLSTM with Document Context: The main problem with BiLSTM is that the only information available is from the given sentence, which may not always be enough to say precisely whether a given token represents an event. Therefore, to boost the performance of BiLSTM, the document context was integrated with the sentence context already captured by the model. Based on previous research involving the global context, conclusions are drawn that the exact prediction of composite realis events requires a lot of document context across pages or documents (Liao and Grishman, 2010). Therefore, the entire document has been considered as a single sentence in this experiment.

4.2 Dataset Split

The dataset split for the experiments was the same in both Assamese and English. The manually annotated dataset of 200 short stories has been divided into three sets as shown in Table 2. Two sets of experiments were performed on both the datasets. In the first experiment, fasttext embedding was used. In the next experiment, BERT embedding with BiLSTM models was used.

5 Results

Comparative performance analysis of different baseline neural models are shown in Table 3 and Table 4 for the task of realis event trigger detection in Assamese and English datasets respectively. For both datasets, it was observed that the model BiLSTM+Document Context

Table 3: The performance of the experiments involving various models for event detection in Assamese, utilizing two distinct embeddings.

METHOD	STATIC EMBEDDINGS			BERT EMBEDDING		
	PRECISION	RECALL	F1	PRECISION	RECALL	F1
LSTM	76.3	53.0	62.5	-	-	-
BiLSTM	81.5	76.5	78.9	81.2	80.8	81.0
+Sentence CNN	81.1	71.8	76.2	82.3	78.5	80.4
+Subword CNN	83.3	75.7	79.3	82.9	77.6	80.2
+Document Context	80.0	77.4	78.7	82.0	81.8	**81.9**

Table 4: The performance of the experiments involving various models for event detection in English, utilizing two distinct embeddings.

METHOD	STATIC EMBEDDING			BERT EMBEDDING		
	PRECISION	RECALL	F1	PRECISION	RECALL	F1
LSTM	88.8	83.6	86.1	-	-	-
BiLSTM	87.9	85.9	86.9	86.8	94.9	90.7
+Sentence CNN	89.9	81.3	85.4	87.0	94.5	90.6
+Subword CNN	87.6	88.7	88.2	87.8	95.1	91.3
+Document Context	87.4	86.6	87.0	89.3	95.3	**92.2**

with BERT embedding performs best achieving 81.9% and 92.2% in terms of macro-F1 score in Assamese and English, respectively. The result is as per expectation. This is due to the model's ability to capture of forward and backward context and availability of richer context information by treating the whole document as a sentence. Moreover, the rich embedding of BERT further elevates the performance of the model.

5.1 Qualitative Analysis

This section provides an overview of the errors encountered when predicting realis events using the top-performing BiLSTM+Document context model. Most errors were rooted in misclassifying non-real events as real ones. In the sentence, এজন প্ৰকৃত মহাত্মাই এনেদৰে ক'ব: "আপুনি সদায় সঠিক পথ বাছি লব।", the word "ক'ব" is predicted as a realis event by the model. However, "ক'ব" is a future event and according to our annotation guideline, it should not be tagged. Similarly, in the sentence, "He wished to go to the field and play cricket.", "*play*" is misclassified as a realis event. Here, "*play*" is someone's wish and is not sure to happen. One of the possible reasons for this misclassification may be due to the lack of samples of this particular type in the dataset. Apart from the error, it is noteworthy that the model

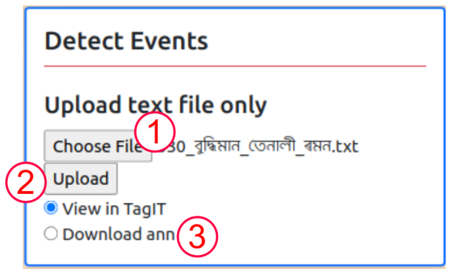

Figure 3: TagIT User Interface. A user can upload a text file and choose either to download the annotation or to modify it using the visualizer.

correctly predicted the words that the annotators missed during annotation. For instance, in the sentence, "She then went to Kittu spider asking for some sweet treats.", the annotators missed to label the real event "*asking*". Whereas the model correctly predicted "*asking*" as realis event.

6 Dataset Expansion

This section illustrates the process of expanding the dataset from 200 to 1000 using an automatic process. For this purpose, "*TagIT*", an easy-to-use automatic annotation tool, was developed and employed. Later, in this section, the process of data augmentation is also elaborated.

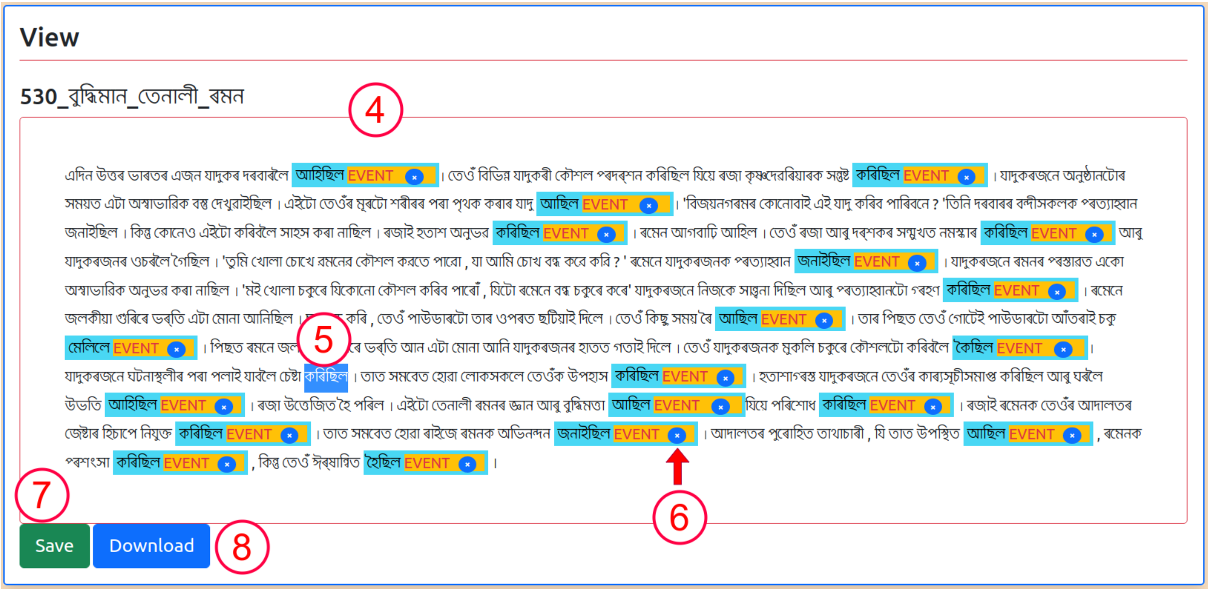

Figure 4: **TagIT Visualizer**: (4) Visualizer display (5) User selected word (6) TagIT auto recommended event word (7) Button to save annotation (8) Button to download annotation.

6.1 TagIT

Manually annotating linguistic resources for training and testing is expensive, especially for specialized text fields. A tool named "*TagIT*" was developed that helps to create a large corpus for event detection. The user-friendly interface of TagIT enables users with little to no prior experience with annotation tools to complete their tasks. It is customizable and extendable. It supports UTF-8 encoding and outputs BRAT standoff format. Such standardized output facilitates data transmission and allows users to easily transform annotated texts into alternate forms. Furthermore, since UTF-8 can encode all unicode characters, documents in multiple languages, including special characters are easily presented. Moreover, TagIT has a custom event detection model integration feature. Models trained in any language can be integrated with TagIT for the purpose of automatic annotation.

The user interface of TagIT is shown in Figure 3. The user can upload (1) a text file and submit it (2) to the system to generate annotations. The user has the choice to download it or modify it (3) in the TagIT visualizer. The TagIT visualizer highlights the detected events in the uploaded text. In the visualizer, the text is displayed in the central part (4), as shown in Figure 4. It also supports new annotation as well as manual corrections of existing annotations. The user can mark (5) a word as an event if it is missed by the event detection model. Also, the user can remove (6) the wrongly labeled event, if any. Finally, the user can save (7) the modified annotations and download (8) it.

6.2 Data Augmentation

For automatic annotation in TagIT, the previous best-performing model, i.e., BiLSTM+Document context with BERT embeddings, was used. This model was trained and tested using the 200 manually annotated short stories. This model, integrated with TagIT, was used to generate labels for the remaining 800 unannotated short stories in both datasets. To check the annotation quality of TagIT, a subset of 120 stories from both datasets was randomly picked from 800 stories labeled automatically using TagIT.

The previous best-performing model, BiLSTM+Document context with BERT embedding, was trained again using the above-mentioned subset. Subsequently, the trained models were tested on the test set of manually annotated 200 short stories. The results are presented in Table 5. From the table, it is evident that the F1-score for automatic label generation (silver standard dataset) is slightly low compared to the manually annotated dataset (gold standard dataset). How-

Table 5: Comparison of results of manual, automatic, and human-in-the-loop annotation in terms of precision, recall, and F1-scores.

Language	Training Data	Precision	Recall	F1 (macro)
Assamese	Manual	82.0	81.8	81.9
	Automatic	79.4	77.8	78.6
	Human-in-the-loop	81.9	80.9	81.4
English	Manual	89.3	95.3	92.2
	Automatic	90.1	89.7	89.9
	Human-in-the-loop	91.8	92.2	92.0

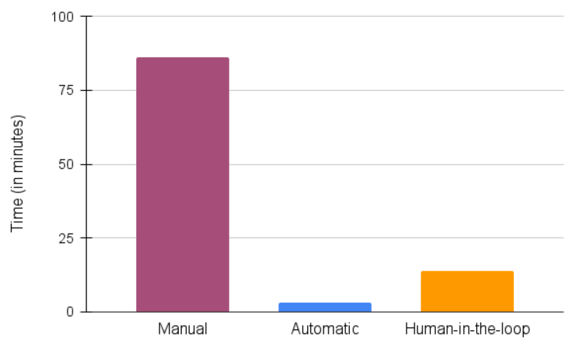

Figure 5: Comparison of time required to annotated short stories in manual, automatic, and human-in-the-loop mode.

ever, when human experts corrected the automatically generated labels and trained the BiLSTM+Document context with BERT embedding model, the F1 score almost resembled the F1 score of the gold standard dataset. This implies that the automatic mode combined with the human-in-the-loop mode of annotation can result in a gold standard-like dataset.

An experiment was also conducted to show the time spent on annotation. A total of ten Assamese short stories were annotated by an author in manual, automatic, and human-in-the-loop mode. It was observed that the annotation duration reduced from 86 minutes to 3 minutes from manual to automatic annotation mode. However, when the automatic annotations were further verified by a human expert (human-in-the-loop mode), the total duration of annotation was 14 minutes. A comparison of time required to annotate short stories by different modes is shown in Figure 5.

7 Conclusion

Datasets for event detection have predominantly emphasized on news and biomedical domains. This work focuses on developing resources (for event detection) in the domain of short stories. Accordingly, this paper introduces a novel corpora containing short stories in both low-resource Assamese and English language. Initially, a gold standard corpus was created, which contains 200 manually annotated short stories in the two languages. Further, this dataset is expanded to 1000 short stories using an automatic annotation tool, resulting in a silver standard dataset. For that purpose, this work proposes an automatic annotation tool called TagIT. Event detection models were separately trained using the gold and silver standard datasets. The experimental observations indicate that the model learned from the silver standard dataset yields comparable results. The silver standard dataset is further verified using human-in-the-loop process, improving the performance while reducing the annotation time significantly. In future, we aim to classify realis events short stories written in other Indian languages.

References

Mohd Hafizul Afifi Abdullah, Norshakirah Aziz, Said Jadid Abdulkadir, Emelia Akashah Patah Akhir, and Noureen Talpur. 2022. Event detection and information extraction strategies from text: A preliminary study using genia corpus. In *International Conference on Emerging Technologies and Intelligent Systems*, pages 118–127. Springer.

Jacqueline Aguilar, Charley Beller, Paul McNamee, Benjamin Van Durme, Stephanie Strassel, Zhiyi Song, and Joe Ellis. 2014. A comparison of the events and relations across ace, ere, tac-kbp, and framenet annotation standards. In *Proceedings of the second workshop on EVENTS: Definition, detection, coreference, and representation*, pages 45–53.

Linguistic Data Consortium et al. 2005. Ace (automatic content extraction) english annotation guidelines for events. version 5.4.3. *ACE*.

Pradeep Dasigi and Eduard Hovy. 2014. Modeling newswire events using neural networks for anomaly detection. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1414–1422.

Bhargav Dave, Surupendu Gangopadhyay, Prasenjit Majumder, Pushpak Bhattacharya, Sudeshna Sarkar, and Sobha Lalitha Devi. 2020. Fire 2020 ednil track: Event detection from news in indian languages. In *Proceedings of the 12th Annual Meeting of the Forum for Information Retrieval Evaluation*, pages 25–28.

Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. 2018. Learning word vectors for 157 languages. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*.

Ali Hürriyetoğlu, Erdem Yörük, Osman Mutlu, Fırat Duruşan, Çağrı Yoltar, Deniz Yüret, and Burak Gürel. 2021. Cross-context news corpus for protest event-related knowledge base construction. *Data Intelligence*, 3(2):308–335.

S Kalita, RR Deka, MP Bhuyan, K Kashyap, and SK Sarma. 2021. Detection of natural calamities from assamese posts in social media. In *IOP Conference Series: Materials Science and Engineering*, volume 1070, page 012055. IOP Publishing.

Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacL-HLT*, volume 1, page 2.

Alapan Kuila, Sarath chandra Bussa, and Sudeshna Sarkar. 2018. A neural network based event extraction system for indian languages. In *FIRE (Working Notes)*, pages 291–301.

Fayuan Li, Weihua Peng, Yuguang Chen, Quan Wang, Lu Pan, Yajuan Lyu, and Yong Zhu. 2020. Event extraction as multi-turn question answering. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 829–838.

Shasha Liao and Ralph Grishman. 2010. Using document level cross-event inference to improve event extraction. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 789–797.

Annie Louis and Ani Nenkova. 2012. A corpus of general and specific sentences from news. In *LREC*, volume 1818, page 10. Citeseer.

Shubhanshu Mishra. 2020. Non-neural structured prediction for event detection from news in indian languages. In *FIRE (Working Notes)*, pages 817–822.

Thien Huu Nguyen and Ralph Grishman. 2015. Event detection and domain adaptation with convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 365–371.

Ellie Pavlick, Heng Ji, Xiaoman Pan, and Chris Callison-Burch. 2016. The gun violence database: A new task and data set for nlp. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1018–1024.

Nicole Peinelt, Dong Nguyen, and Maria Liakata. 2020. tbert: Topic models and bert joining forces for semantic similarity detection. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7047–7055.

D Rebholz-Schuhmann, A Jimeno Yepes, C Li, S Kafkas, I Lewin, Ning Kang, P Corbett, D Milward, E Buyko, E Beisswanger, et al. 2011. Assessemnt of ner solutions against the first and second calbc silver standard corpus. *Journal of Biomedical Semantics*, 2(supplement).

Dietrich Rebholz-Schuhmann, Antonio José Jimeno Yepes, Erik M. van Mulligen, Ning Kang, Jan Kors, David Milward, Peter Corbett, Ekaterina Buyko, Katrin Tomanek, Elena Beisswanger, and Udo Hahn. 2010. The CALBC silver standard corpus for biomedical named entities — a study in harmonizing the contributions from four independent named entity taggers. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Aniruddha Roy, Isha Sharma, Sudeshna Sarkar, and Pawan Goyal. 2023. Meta-ed: Cross-lingual event detection using meta-learning for indian languages. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(2):1–22.

Sovan Kumar Sahoo, Saumajit Saha, Asif Ekbal, and Pushpak Bhattacharyya. 2020. A platform for event extraction in Hindi. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 2241–2250, Marseille, France. European Language Resources Association.

Roser Saurí and James Pustejovsky. 2009. Factbank: a corpus annotated with event factuality. *Language resources and evaluation*, 43(3):227–268.

Matthew Sims, Jong Ho Park, and David Bamman. 2019. Literary event detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3623–3634, Florence, Italy. Association for Computational Linguistics.

Diana Sousa, Andre Lamurias, and Francisco M. Couto. 2019. A silver standard corpus of human phenotype-gene relations. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1487–1492, Minneapolis, Minnesota. Association for Computational Linguistics.

Rachele Sprugnoli and Sara Tonelli. 2017. One, no one and one hundred thousand events: Defining and processing events in an inter-disciplinary perspective. *Natural language engineering*, 23(4):485–506.

Rachele Sprugnoli and Sara Tonelli. 2019. Novel event detection and classification for historical texts. *Computational Linguistics*, 45(2):229–265.

Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. Brat: a web-based tool for nlp-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107.

Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. 2017. Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE international conference on computer vision*, pages 843–852.

Jian Wang, Honglei Li, Yuan An, Hongfei Lin, and Zhihao Yang. 2016. Biomedical event trigger detection based on convolutional neural network. *International Journal of Data Mining and Bioinformatics*, 15(3):195–213.

Wei Xiang and Bang Wang. 2019. A survey of event extraction from text. *IEEE Access*, 7:173111–173137.