# Reinforced NMT for Sentiment and Content Preservation in Low-resource Scenario

DIVYA KUMARI and ASIF EKBAL, Indian Institute of Technology Patna, India
REJWANUL HAQUE, ADAPT Centre, School of Computing, National College of Ireland, Ireland
PUSHPAK BHATTACHARYYA, Indian Institute of Technology Patna, India
ANDY WAY, ADAPT Centre, Dublin City University, Ireland

The preservation of domain knowledge from source to the target is crucial in any translation workflows. Hence, translation service providers that use machine translation (MT) in production could reasonably expect that the translation process should transfer both the underlying pragmatics and the semantics of the source-side sentences into the target language. However, recent studies suggest that the MT systems often fail to preserve such crucial information (e.g., sentiment, emotion, gender traits) embedded in the source text in the target. In this context, the raw automatic translations are often directly fed to other natural language processing (NLP) applications (e.g., sentiment classifier) in a cross-lingual platform. Hence, the loss of such crucial information during the translation could negatively affect the performance of such downstream NLP tasks that heavily rely on the output of the MT systems.

In our current research, we carefully balance both the sides (i.e., sentiment and semantics) during translation, by controlling a global-attention-based neural MT (NMT), to generate translations that encode the underlying sentiment of a source sentence while preserving its non-opinionated semantic content. Toward this, we use a state-of-the-art reinforcement learning method, namely, *actor-critic*, that includes a novel *reward combination* module, to fine-tune the NMT system so that it learns to generate translations that are best suited for a downstream task, viz. sentiment classification while ensuring the source-side semantics is intact in the process. Experimental results for Hindi–English language pair show that our proposed method significantly improves the performance of the sentiment classifier and alongside results in an improved NMT system.

CCS Concepts: • **Computing methodologies → Machine translation**;

Additional Key Words and Phrases: Machine translation, neural machine translation, sentiment preservation, actor-critic, reinforcement learning, BERT

Authors duly acknowledge the support from Hindi to English Machine Aided Translation (HEMAT) System in Judicial domain, sponsored by the Ministry of Electronics and Information Technology (MeiTY), Government of India. Asif Ekbal acknowledges the Young Faculty Research Fellowship (YFRF), supported by Visvesvaraya Ph.D. scheme for Electronics and IT, Ministry of Electronics and Information Technology (MeitY), Government of India, being implemented by Digital India Corporation (formerly Media Lab Asia).

Authors' addresses: D. Kumari, A. Ekbal, and P. Bhattacharyya, Department of Computer Science and Engineering, Indian Institute of Technology Patna; emails: divya.24aug@gmail.com, asif@iitp.ac.in, pb@cse.iitp.ac.in; R. Haque, ADAPT Centre, School of Computing, National College of Ireland, Dublin, Ireland; email: rejwanul.haque@adaptcentre.ie; A. Way, ADAPT Centre, Dublin City University, Dublin, Ireland; email: andy.way@adaptcentre.ie.

2375-4699/2021/06-ART70 $15.00
https://doi.org/10.1145/3450970

**ACM Reference format:**
Divya Kumari, Asif Ekbal, Rejwanul Haque, Pushpak Bhattacharyya, and Andy Way. 2021. Reinforced NMT for Sentiment and Content Preservation in Low-resource Scenario. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 20, 4, Article 70 (June 2021), 27 pages.
https://doi.org/10.1145/3450970

## 1 INTRODUCTION

In an industrial setting, the preservation of domain-knowledge from source to the target is arguably the most crucial factor for the corporate customers. In other words, the pragmatics of text such as sentiment, emotion, politeness, gender traits, as well as the semantics of the source texts are required to be transferred in the target language. This is seen as the customers' prime expectation in the translation industries. As of today, a large amount of raw translations are directly being used in a variety of areas, as an input to the downstream **Natural Language Processing (NLP)** applications in the cross-lingual platforms. However, many studies [1, 30, 31, 37, 41–43] have found a significant loss of pragmatics in the translations produced by the state-of-the-art MT systems, which could indeed adversely affect the performance of these NLP applications that use unedited raw **machine translation (MT)** output. A suitable solution to this problem would certainly be a breakthrough in the MT research, and a blessing for the translation industry.

Tebbifakhr et al. [47] demonstrated that an MT system can be customised to produce controlled translations that encode those source-side properties (i.e., sentiment) that can essentially improve the performance of the downstream task (i.e., sentiment classification). However, they often fail to encode the semantics of the source sentence. Tebbifakhr et al. [47] sample translations with the highest polarity from a set of *k*-sampled candidate translations that are associated with the rewards from their sentiment classifier. Finally, the best candidate is chosen to update the model parameters. This contradicts the idea of one candidate sampling from the model's output distribution [50]. Nonetheless, such a stricter fine-tuning criterion can significantly change the model parameters, and eventually it focuses on producing strings that are only best-suited for the downstream task, and can even generate translations having nearly zero adequacy [47]. A number of recent studies have demonstrated how indispensable the non-opinionated semantic content can be for improving the quality of a sentiment classifier [53]. Accordingly, the transfer of such information from the source to the target can be pivotal too for the quality of sentiment classifier in the cross-lingual platforms.

In recent years, the deep language models trained on a large volume of unlabeled data have proven to be highly successful [13, 28, 29]. In particular, the **bidirectional encoder representations from transformers** (BERT) [13] model produces state-of-the-art performance for several NLP tasks. The BERT-based pre-trained model can be used with one additional output layer to fine-tune for the downstream NLP tasks, such as sentiment classification. However, such high-performance NLP tools are often available only for a few languages, e.g., English. In this context, the translation-based approach becomes an inevitable choice to harness the resource-richness of other languages.

In this article, unlike Tebbifakhr et al. [47], which focuses on the preservation of sentiment at the expense of translation quality, aiming at improving the performance of sentiment classifier, we apply a **reinforcement learning (RL)** strategy [3, 38] to fine-tune an MT model [4] so that it can learn to (i). generate translations that are best-suited for the sentiment classification task; and (ii). keep the source-side semantics intact in the translations. In particular, we optimize a multi-purpose reward function that includes (i) a function that performs element-wise dot product between the ground-truth sentiment distribution and the predicted sentiment distribution taken from the softmax layer of the classifier and (ii) a **sentence-BLEU (SBLEU)** [10] of the predicted translation given the reference translation. Eventually, this multi-purpose reward optimization resulted in a better NMT system and elevates the performance of sentiment classifier.

We summarize the key contributions/characteristics of our current research as follows:

(1) We create a sentiment (polarity) labeled Hindi–English parallel corpus named as **Sentiment Corpus (SC)** in the tourism domain. The dataset is much cleaner, structured and balanced to study the effect of automatic translation on the sentiment, unlike the noisy tweets used in References [6, 31, 37]. Tweets may have a lot of noises due to the presence of hashtags, urls, phonetic substitutions, e.g., "b4—before"; inappropriate use of white spaces, and so on. As the nature of these tweets significantly varies from the NMT training data, which easily explains why an NMT system easily falters when presented with these tweets. Comparatively, our dataset is more appropriate to study the other possible reasons for sentiment loss in the MT context. Second, none of these released parallel datasets are balanced with respect to the sentiment classes, viz. positive, negative, and neutral.
(2) We investigate the possibility of using a state-of-the-art RL technique (i.e., actor critic method) to improve the quality of NMT output in accordance with a machine-oriented criterion (i.e., improving sentiment classifier performance) without much affecting the human-oriented criterion (i.e., corpus level BLEU [40]).
(3) We define a novel reward function for this, which is the harmonic mean of the following two task-specific metrics, viz. (a) SBLEU as a reward for the content preservation and (b) classifier-assigned probability score as a reward for the sentiment preservation.
(4) Our ablation studies suggest that the harmonic reward is more robust to weigh the actions taken under a policy training than their individual counterparts.
(5) The empirical results show that our critic-based RL method with the novel harmonic reward is not only capable of preserving the sentiment in the translated text but also improves its overall translation quality. Through this work, we emphasise that while developing the MT system, the focus should be not only that the source text is better translated but also better paired, i.e., their sentiment relation is preserved.
(6) To evaluate our model performance, we use two widely used automatic evaluation metrics, viz. (i) one being the weighted $F_1$ score that measures whether the translated texts and source texts are consistent with respect to the sentiment; and the (ii) other is the BLEU metric, used for evaluating the content preservation or semantic transfer from the source to the target.

The remainder of the article is organised as follows. In Section 2, we discuss the related work. Section 3 presents the detailed steps of our in-domain dataset creation process and its statistics, along with the statistic of other datasets used in the experiments. In Section 4, we formulate the problem, present our proposed approach and the experimental setups. In Section 5, we demonstrate the evaluation results obtained, along with the findings, analysis, and our observations on the translations. Finally, Section 6 concludes our work with the avenues for future research.

## 2 RELATED WORK

The state-of-the-art MT approaches, be it **phrase-based statistical machine translation (PB-SMT)** or NMT, exploit the statistical dependencies learned usually from a large training corpus by maximizing the likelihood of translations that are statistically the most likely variation of the input source sentences but in the target language. The **maximum likelihood estimation (MLE)**-based objective used in training the current state-of-the-art neural MT [4, 49] is often detrimental to the crucial attributes of the source text, e.g., politeness, sentiment orientation of textual data. Banea et al. [8] in their empirical study demonstrated that the sentence-level subjectivity is preserved in translations generated by the human translators, which, however, is often not seen in the translations by the MT systems [2, 7, 15, 18, 33, 36, 37, 44, 46, 48].

Past studies suggest that efforts were made to preserve the crucial features, e.g., politeness, gender, sentiment in the automatic translation of the text. In one of the prior work on politeness, Mima et al. [35] proposed to incorporate the extra-linguistic information such as speaker's role, social rank, and gender information available from the context, situation, and environment. It improves the performance of a rule-based dialog MT system. In a similar work, Sennrich et al. [44] also tried to control politeness in the output text by providing it as an additional input feature to the NMT system in training. During testing, they assume that the desired level of politeness is specified by the user. Results suggest that incorporating politeness as a side-constraint via an extra source token is effective in controlling politeness in the translation. Unlike politeness, Wintner et al. [51] tried to preserve the gender traits in the automatic translation. For this, they relied on the data selection techniques from domain adaptation. In another such work, Niu et al. [39] focus on the preservation of formality of the text. As opposed to References [35, 39, 44, 51], where authors tried to explicitly model the speaker traits (e.g., politeness and gender), Michel and Neubig [34] modeled the speakers themselves through an additional bias vector in the output layer of the softmax. They learned this bias directly or through a factored model that treats each user as a mixture of a few prototypical bias vectors.

Given the context of this work, we also looked into research that studied the loss of sentiment during automatic translation. In particular, Mohammad et al. [37] pointed out that one of the causes for the loss of sentiment in automatic translation is that sentiment-bearing words are mistranslated or transferred to carry a neutral sentiment. They also found that the text in a resource-poor language (i.e., Arabic) when translated to a resource-rich language (i.e., English), and sentiment tagging using the state-of-the-art sentiment classifier in the resource-rich language produces competitive results when compared to the performance of the classifier in the resource-poor language. This finding can be viewed as an inspiring use case of the MT system as far as the cross-language sentiment analysis is concerned. Accordingly, we present here a survey of the related work that pursued this line of research. Kanayama et al. [21] employed the MT system to translate the sentiment units extracted from the source text toward building a more accurate sentiment analyzer in the target language. Similarly, Balahur and Turchi [5] also employed different MT systems and translated the training data in a language other than English to build the language-specific machine learning models. They found that for some language pairs, sentiment analysis systems built using the translated resources obtain a performance similar to the systems built-in English.

Chen and Zhu [11] investigated the use of a lexicon-based consistency approach to design features such as subjectivity, polarity, intensity, negation, and incorporated these into the PB-SMT *t*-table. They used these features in re-raking the candidates of the *t*-table and found, it improved the translation quality in a NIST Chinese-to-English task. Lohar et al. [30] prepared the positive, negative and neutral sentiment-specific MT systems to ensure cross-lingual sentiment consistency. In the German-to-English MT task, they found that their approach may help preserve the sentiment at the expense of translation quality. They also found that when the sentiment is altered by the MT, it often gets transformed to its nearest class (i.e., positive to neutral or vice versa; or negative to neutral or vice versa; but rarely from positive to negative). Being motivated by their findings, Lohar et al. [31] built the nearest neighbor-based sentiment translation systems (i.e., negative to neutral and positive to neutral MT models) by training them on the combined parallel data of two nearest sentiment classes. Using this approach they, to a certain extent, were able to balance the translation quality and the sentiment preservation.

Recently, Tebbifakhr et al. [47] posit a new perspective of a use case of the MT system, where the translations produced by the model are used to enhance the performance of a (binary) sentiment classifier. They pursued a *machine-oriented criterion* (i.e., the end-user of the MT output is a downstream NLP application) as opposed to a *human-oriented criterion* (i.e., the end-user of the MT output is a human) in the development of the NMT system. To achieve this, Tebbifakhr et al. [47] adapted REINFORCE of Williams [50] by incorporating an exploration-oriented $k$-candidates-based sampling strategy and exploits a weak feedback signal obtained from the downstream application. In other words, they sample $k$-candidate translations corresponding to a source sentence and update the policy (i.e., parameters of the NMT system) based on the highest rewarding candidates as per the feedback from the sentiment classifier.

Although they achieved a performance boost for the downstream task, they had to compromise with the translation quality to a greater extent. In contrast to Tebbifakhr et al. [47], our proposed framework aims at carefully maintaining the trade-off between the two sides, viz. the MT translation quality and performance of the downstream task. Further, we choose a *multi-class* sentiment classification task, which is difficult, as opposed to the *binary* task [47].

In the other closely related RL-based works, authors [26, 27, 38, 52] applied the RL techniques to adapt the NMT parameters toward the human-oriented criterion, which is to preserve the adequacy in the translation. Nguyen et al. [38] used the popular AC (actor-critic) method and focused on preserving the semantics of a text. Lam et al. [26, 27] investigated a different use case of MT, viz. the interactive-predictive NMT (INMT), where the output of an MT system is corrected by a human agent based on the observation of the partial translation in an interactive-predictive platform. Lam et al. [27] extended Lam et al. [26] work in the sense that, they applied *imitation learning* in the INMT framework via the demonstration of expert actions. In another interesting work, Wu et al. [52] used RL to boost the performance of an NMT system by the use of monolingual (both source and target) data in the RL training.

However, in this work, we focus on investigating how reinforcement learning-based strategy could be used for sentiment preservation in the MT task, which, in turn, improves the performance of the downstream sentiment classifier. The performance of the sentiment classifier heavily relies on the output of the NMT system. Hence, we use the output of the sentiment classifier as a weak feedback signal to guide the NMT system to produce an output that primarily results in improved classifier performance. The primary objective is to generate a translation, $\hat{t}_e = \{y_1, y_2, \ldots, y_m\}$ that is consistent with the gold sentiment besides ensuring the adequacy.

Empirical evaluation results show that our proposed model outperforms the following baselines (see Section 4.5) in the sentiment classification task : (i) a vanilla NMT system prepared for the traditional human-oriented criterion (i.e., maximising the adequacy of the text via the MLE-based training); (ii) Continued Training [17], and (iii) **Machine-oriented (MO)** Reinforce model [47].

## 3 DATA PREPARATION

In this section, we present the details of the dataset that we create for our experiments. To the best of our knowledge, there is no existing (freely available) sentiment annotated parallel data for Hindi–English. Therefore, we manually label the sentences of a domain-specific parallel corpus with three sentiment orientations, namely, positive, neutral, and negative. The sentences of our domain-specific parallel corpus were, in fact, taken from two different data sources: (i) an already available ILCI corpus[1] and (ii) a new parallel corpus created by crawling sentences from Tripadvisor.[2] For the latter, we first automatically crawl the English reviews from Tripadvisor, and then the sentences were manually translated into Hindi. From now on, we call this corpus **review corpus**

[1] ILCI Hindi–English tourism text corpus.

[2] www.tripadvisor.com.

**(RC)**. These two parallel corpora (ILCI and RC) are combined to form a bigger polarity-labeled corpus, which we call **sentiment corpus (SC)**.

The SC thus obtained has a total of 9,004 sentences, out of which nearly 3,164 (~35%) sentences are from newly created RC and 5,840 (~65%) sentences from the ILCI corpus. The subsequent sections explain the steps of our in-domain data creation process including data collection, cleaning, translation, and annotations.

### 3.1 Data Collection

Tripadvisor, a popular online travel company provides a web-based platform to their customers to post reviews on their experiences. The reviews of Tripadvisor are usually more structured compared to those found in the other tourism websites. Although there are no strict rules, the review texts usually follow a formal structure, i.e., they provide at least a few basic pieces of information such as the types of attractions, a five-point rating scale, the identity (e.g., name) of the reviewer, a title for the review and the text of the review. These are the reasons why we chose Tripadvisor for data collection. Since one of our objectives is also to perform analysis of sentiment bias in machine translation, we decide to create a balanced data set of positive, negative, and neutral classes to observe the NMT induced biases toward a sentiment class, if any, as claimed by Mohammad et al. [37] and Lohar et al. [31] for the neutral class.

Our investigation based on an *exploratory data analysis*[3] on one of the sub-corpus (i.e., ILCI) reveals a *class-imbalance problem*. In other words, we found that majority of the sentences of the ILCI corpus belong to the neutral class followed by the positive class and then the negative class (with the negative class containing a very small number of sentences). In this context, sentences of ILCI are about Indian tourist attractions, e.g., historical sights and landmarks, scenic beauty, nature, and parks. Hence, to obtain a balanced distribution for each sentiment class in the newly compiled dataset, we follow the following procedure:

**Step 1**: First, a list of famous tourist attractions in India is prepared manually. In addition to this, a supplementary list of tourist attractions was automatically extracted from the ILCI corpus. For this, we use the Stanford **Named Entity Recogniser (NER)** toolkit.[4] The toolkit tags the English sentences of the corpus at the word level with seven **named entity (NE)** classes. We consider only the words tagged as LOCATION and ORGANIZATION, and create a list of such NEs. In Table 1, we illustrate this simple process with two example sentences. We further manually verify each item from the resultant list of NEs, and remove the noisy items, if any.

Thus, we obtain a list of tourism-related valid attractions, which are then appended to the list of manually created NEs. The NEs of the final list serve as our set of **seed NEs (SNE)**. SNE is then used to filter the reviews of the "famous tourist attractions," crawled from Tripadvisor.

**Step 2**: Step 1 provides a list of reviews. In this step, we apply another filtering criterion, and remove those reviews that do not contain any word of the NRC Affect Intensity Lexicon[5]. The idea underpinning this is to keep positive and negative sentences as many as possible so that we can create a balanced polarity-labeled sentiment dataset.

**Step 3**: Finally, we merge the filtered sentences with those from the ILCI. As mentioned above, this resultant combined corpus is referred to as SC.

[3] Sentence clustering followed by word-cloud visualization of each cluster.

[4] https://nlp.stanford.edu/software/CRF-NER.html#Download.

[5] https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm.

Table 1. Extracting Place Names Using Stanford NER

| | |
|---|---|
| Sentence 1 | This is a unique museum, which is only in Vishakhapattanam |
| Sentence 2 | In 1955–’56 it was renamed as Jim Corbett National Park |
| Tagged Sentence 1 | [(“This,” “O”), (“is,” “O”), (“a,” “O”), (“unique,” “O”), (“museum,” “O”), (“which,” “O”), (“is,” “O”), (“only,” “O”), (“in,” “O”), (“Vishakhapattanam,” “LOCATION”)] |
| Tagged Sentence 2 | [(“In,” “O”), (“1955,” “DATE”), (“-,” “O”), (“56,” “DATE”), (“it,” “O”), (“was,” “O”), (“renamed,” “O”), (“as,” “O”), (“Jim,” “ORGANIZATION”), (“Corbett,” “ORGANIZATION”), (“National,” ”ORGANIZATION”), (“Park,” “ORGANIZATION”)] |
| Seed NE list (automatically extracted) | {“Vishakhapattanam,” “Jim Corbett National Park” } |

### 3.2 Annotation

A human translator with proficient English and Hindi language skills was hired to translate the English sentences of RC to Hindi. One more translator was asked to verify the translation. Then, three annotators who are bilingual and experts in both Hindi and English took part in the annotation task. The annotators have a post-graduate qualification in linguistics and a good knowledge of English and Hindi both. They have prior experience in judging the quality of machine translation and sentiment annotation.

For translation, the following were the instructions: (i) experts were asked to read the Hindi sentence carefully; (ii) source and target sentence should carry the same semantic and syntactic structure; (iii) they were instructed to carefully read the translated sentences and see whether the fluency (grammatical correctness) and adequacy are preserved; (iv) they made the correction in the sentences if required; (v) transliteration of a Hindi word can also be used, especially if this is a NE.

For sentiment labeling, annotators were asked to follow the guidelines as follows: (i) they were instructed to look into the sentiment class of the source sentence, locate its sentiment bearing tokens; (ii) they were asked to observe both of these properties in the translated sentences; (iii) they were asked to annotate the source sentences (Hindi) in SC into four classes, namely, positive, negative, neutral, and others; (iv) they were instructed to tag those sentences only with the first three labels for which there is a clear connotation for three orientations of polarity (i.e., negative, positive, and neutral), and to mark otherwise with the fourth label, i.e., other.

The further detailed instructions for the sentiment annotations are given as follows:

(1) Select the option that best captures the sentiment being conveyed in the sentences: positive-negative-neutral-others.
(2) Select positive if the sentence shows a clear positive attitude (possibly toward an object or event). Example: “I hope every year I get a chance to visit this mesmerizing beauty.”
(3) Select negative if the sentence shows a clear negative attitude (possibly toward an object or event). Example: “The priest’s behaviour was terrible. He didn’t let us enter the temple.”
(4) Select neutral if the sentence shows a neutral attitude (possibly toward an object or event) or is an objective sentence. Objective sentences are sentences that do not carry any opinion,

Table 2. Statistics of the SC

| | Language | RL trainset | RL testset | RL devset |
|---|---|---|---|---|
| Sentences | | 6,000 | 1,134 | 900 |
| Tokens | English | 97,870 | 19,215 | 14,527 |
| | Hindi | 115,740 | 24,143 | 17,289 |
| Vocabulary | English | 14,008 | 4,643 | 4,147 |
| | Hindi | 11,853 | 4,079 | 3,871 |

e.g., facts are objective expressions about entities, events and their properties. Example: (i) "Red fort was built in year 1565 A.D." (objective), (ii) "I visited this place last year and plan to visit it in future too" (neutral).

(5) Others: These denote the sentences that do not fall in above three categories, e.g., (i) the sentence is highly ungrammatical and hard to understand, (ii) the sentence expresses both positive and negative sentiment, i.e., mixed polarity. Example: "The overall journey was exhausting, also garbage all around on the sideways but reaching the temple of vaishno maa tooks all our exhaustion. Such a divine energy. Everyone must visit this peaceful place" (mixed polarity).

These annotation guidelines were decided after thorough discussions among ourselves. After we had drafted our initial guidelines, the annotators were asked to perform the verification of the translated sentences, and sentiment annotation for the 100 sentences. The disagreements cases were thereafter discussed and resolved through discussions among the experts and annotators. Finally, we came up with the set of instructions as discussed above to minimize the number of disagreement cases.

Inter-annotator agreement measured via Fleiss'Kappa [16] across three annotators is 0.78. All the English sentences are lower-cased and the Hindi sentences are normalized. Statistics of SC are presented in Table 2. To perform tokenization for English, we use the standard tokenization tool[6] of the Moses toolkit [24]. For tokenizing the Hindi words, we use the IndicNLP[7] library.

As pointed out by References [31, 37], one of the causes for the loss of sentiment in translation is that the sentiment bearing expressions are either translated to the neutral expressions or do not appear in the translations. To avoid such situation, we decided not to introduce any biases for a particular class in the RL training. Hence, to pursue this objective, we randomly sample sentences from the SC and create the training, development and test sets (refer to the parallel corpus statistics in Table 2) as follows:

- Training set: It includes 6,000 examples distributed equally over the positive, negative and neutral classes. We call this set, RL trainset.
- Development set: It includes 900 examples distributed equally over the positive, negative and neutral classes. We call this as RL devset.
- Test set: From the remaining 2,104, we include 1,134 examples distributed equally among the positive, negative and neutral classes in our testset. We call this RL testset. Please see Table 4 for the derived corpus class-wise statistic.

[6] https://github.com/moses-smt/mosesdecoder/blob/master/scripts/tokenizer/tokenizer.perl.

[7] https://github.com/anoopkunchukuttan/indic_nlp_library.

Table 3. The Generic Domain NMT(s) Training and Development Datasets

| Training Set | Datasets | No. of sentences |
|---|---|---|
| Hi – En | IITB parallel corpus | 1.6M |
| It – En | Europarl | 2M |
| | JRC | 0.8M |
| | Wikipedia | 1M |
| | ECB | 0.2 M |
| | TED | 0.2M |
| | KDE | 0.3M |
| | News11 | 0.04M |
| | News | 0.02M |
| | Total | 4.56M |
| Development set | Datasets | No of sentences |
| Hi – En | IITB dev | 520 |
| It – En | Newstest2009 | 2,525 |

*MLE training data:* We perform experiments for two language pairs, viz. Hindi–English, and Italian–English. For training and validation of our vanilla NMT(s) (i.e., It–En and Hi–En), we use the corresponding parallel datasets from Table 3. For Italian–English, all the source and target sentences are tokenized. As an extra pre-processing step, all the English and Italian sentences are lowercased. As mentioned previously (Section 3.2), in the Hindi–English pre-processing steps, we lowercase all English and normalize the Hindi sentences. Tokens in the training sets are segmented into sub-word units using the Byte-Pair Encoding (BPE) technique [45] with 15,500 merge operations [25] for Hindi–English and 32,000 for Italian–English [45]. All the sentences exceeding more than 90 tokens are removed from the training and development sets.

*RL dataset:* In the RL-based fine-tuning of Hindi–English and Italian–English models, we use datasets as mentioned in Table 4. For Italian–English language pair, annotated tweets released for the Italian sentiment analysis task at Evalita 2016 [9] have been used. All the objective, neutral and mixed polarity tweets are filtered out (refer to the final statistics in Table 4, row (ii)). Please note that this dataset is not a parallel dataset. Hence, we use it only in reference to the (binary) sentiment classification task.

## 4 METHODOLOGY

In this section, we define our problem mathematically, and then present the proposed methodology in details.

### 4.1 Problem Definition

Let us consider a source sentence $s_h = \{x_1, x_2, \ldots, x_n\}$ and its sentiment class $\{l : l \in [l_1, l_2, l_3]$ for multi-class $\vee\ l \in [l_1, l_2]$ for binary class$\}$, where $l$ being the gold source sentiment (e.g., *positive*, *negative*, *neutral*) represented as a one-hot vector. The objective is to generate a translation, $\hat{t}_e = \{y_1, y_2, \ldots, y_m\}$ that is consistent with the gold sentiment besides ensuring adequacy.

It is emphasised that the translation $\hat{t}_e$ is to be used for consumption by a downstream NLP component, i.e., sentiment classifier. The consistency of objective ($\hat{t}_e$ as per the sentiment of $s_h$) is achieved by using a state-of-the-art reinforcement learning (RL) technique, namely, the actor-critic (AC) method, with a novel reward function.

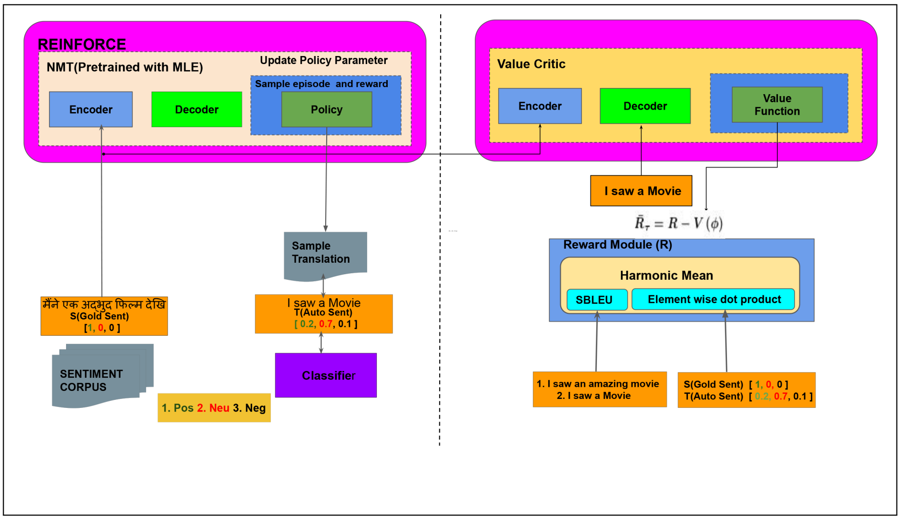

Fig. 1. Architecture of our proposed RL Model.

## 4.2 Model Overview

This section presents our proposed approach briefly. First, we perform pre-training (discussed in Section 4.3) of two NMT system(s) (i.e., Hi–En, It–En) using the generic domain datasets (cf. Table 3). The parameters of the NMT(s) act as policies. We refer to a pre-trained NMT policy as our *actor*. The parameters of the NMT system(s) are then fine-tuned using a popular *policy gradient* method, namely, actor-critic [3, 38]. The subsequent discussion is with reference to one of the actors—Hindi–English NMT unless specified otherwise. The RL training involves, fine-tuning the actor's policy (i.e., the parameters of the NMT system) via observing a discrete reward that constitutes a weighted harmonic mean of the rewards corresponding to two individual sub-tasks (i.e., sentiment classification and MT). The architecture of our proposed RL method is illustrated in Figure 1.

We see from the lower-left side of Figure 1 that input to the actor is a Hindi sentence, its output is an English translation and the expectation is that the English translation would preserve the sentiment of the source Hindi sentence. The *reward module* has two rewards: $R_1$ and $R_2$. $R_1$ is the SBLEU score of the translation given the reference. $R_2$ is the inner product of the ground-truth sentiment distribution of source and the sentiment distribution of the translation obtained from the softmax layer of the sentiment classifier. Finally, we take the harmonic mean of $R_1$ and $R_2$ to obtain an intermediate reward R($\hat{t}_e$). The *critic* network is used to produce an estimate of the expected (future) reward, which is then subtracted from the intermediate reward, R. In other words, it implicitly assigns a reward, $\bar{\text{R}}_\tau$ to the translation, which is finally used to weigh the actor's action at a given time step $\tau$. In Section 4.3, we discuss the setups for the different stages of training of the proposed model. Section 4.4 discusses the methodology for reward computation while Section 4.5 provides the details of our baseline models.

## 4.3 Experimental Setups

This section is divided into the following subsections. In Section 4.3.1, at first, we describe our experimental setups to pre-train (warm-up) the policy (constituting parameters of the actor) and the critic. Then Section 4.3.2 discusses the steps to prepare the classifier used to obtain the sentiment

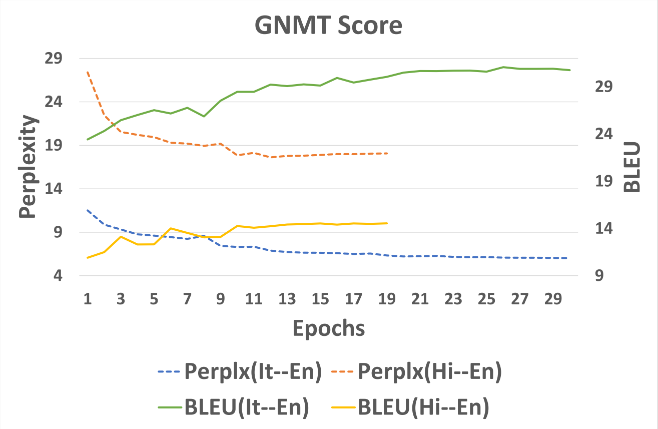

Fig. 2. Performance of the GeNMT systems in terms of BLEU and Perplexity on the NMT development sets at the MLE steps of the actors.

scores. Section 4.3.3 discusses the actor's fine-tuning strategy via critic. Finally, in Section 4.3.4, we summarise the complete workflow.

*4.3.1 The NMT System.* For policy implementation, we use a sequence-to-sequence global-attention-based architecture of Bahdanau et al. [4] with single-layer bidirectional long short-term memory units [20]. Similarly, our critic network is also an encoder-decoder model with the exact parameters configuration as of the actor network. Actor and critic parameters are uniformly initialized in $[-0.1, 0.1]$. The sizes of embedding and hidden layers are 256 and 512, respectively. The Adam optimizer [22] with $\beta_1 = 0.9, \beta_2 = 0.99$ is used and gradient vector clipped to magnitude 5. We set the dropout to 0.2. The NMT model is trained under the MLE objective with input feeding [32] with learning rate (lr) and batch size (bs) set to 0.001 and 64, respectively. For sampling a candidate translation in the RL training (cf. Section 4.3.3), we use *multinomial sampling*, with the sampling length of 50 tokens for Hindi–English and 30 tokens for Italian–English (memory constraint imposed due to large vocabulary size), and use the greedy search for decoding. The NMT model is trained and validated on the generic domain corpus (cf. Table 3). We monitor the model's perplexity on the development set and the lr is decayed by 0.5 after the eighth epoch if perplexity on the development set increases. For Hindi–English, the model converges after 12 MLE epochs, whereas for Italian–English, the training is stopped after 28 MLE epochs (cf. Figure 2) as no significant drop (more than 0.02) in the perplexity is observed on the development set. As mentioned previously, we refer to the NMT systems/actors (Hi–En, It–En) thus obtained on the generic domain corpus as generic NMT systems (GeNMT(s)) that act as trained policies in the RL training (refer to the upper left side of Figure 1). Given an actor, we keep the corresponding actor fixed and warm-up the critic for one epoch with the learning rate of 0.001 and batch size of 64 on the generic corpus from Table 3 with SBLEU as a reward.

*4.3.2 Classifier Training.* To simulate the BERT-based sentiment classifier as a service to be used to obtain sentiment scores for the translations during the RL training, we fine-tune the pre-trained *bert-base-uncased* model [14] using the task-specific dataset (cf. Table 4). For Hindi–English multi-class sentiment classification task, we fine-tune our English classifier (target) on the RL devset[8] (900 English sentences; row (i)). We fine-tune the classifier on the development set to

[8] one may use complete RL training set.

Table 4. Class-wise Distribution of Polarity-tagged Hindi–English (Parallel) and Italian (Mono-lingual) RL Training Dataset

| Tasks | RL trainset | | | RL devset | | | RL testset | | |
|---|---|---|---|---|---|---|---|---|---|
| | Pos | Neu | Neg | Pos | Neu | Neg | Pos | Neu | Neg |
| (i). Hi–En | 2,000 | 2,000 | 2,000 | 300 | 300 | 300 | 378 | 378 | 378 |
| (ii). It–En | 1,450 | | 2,289 | 161 | | 254 | 316 | | 734 |

simulate a real word setting, where the cross-lingual classifier already exists and hence might not have seen the vocabulary of the NMT training corpus. For Italian–English binary twitter sentiment classification task, we use a balanced set of 1.6M negative and positive tweets [19]. In addition to two *bert-base-uncased* target language classifiers, two more source language classifiers are fine-tuned, i.e., uncased *multilingual variations* of BERT on the source-side (Hindi and Italian) sentences of the RL trainset.

Following nomenclatures are used for our classifiers to give more meaningful names. The first letter shows the language (source or target) on which the classifier is fine-tuned and the second letter shows the case or uncased variation of BERT. For example,

(1) h-uBERT: This corresponds to our *H*indi sentiment classifier, a multilingual *uncased* model.
(2) i-uBERT: This corresponds to the *I*talian sentiment classifier, a multilingual *uncased* model.
(3) e-uBERT: Our *E*nglish sentiment classifier, a monolingual *uncased* model. This will also be used as a fixed sub-component to obtain reward, $R_2$ for sentiment preservation in our harmonic reward model.

We fine-tune the BERT models for 3 epochs with a learning rate of *2e-5* as suggested in Reference [14] and batch size of 4. For a fair comparison, the recommended best setup of Devlin et al. [14] is used for all our experiments.

*4.3.3 Reinforce Training.* This section explains the experimental setups for the RL training of Hindi–English, and Italian–English referenced NMT(s).
*Hindi–English:*

The actor model (pre-trained Hi–En NMT), which can be thought of as a pre-trained policy is fine-tuned using the AC method, where the critic learned *value function* is used to center an intermediate (here, it is the harmonic average-based reward $R$, discussed in Section 4.4) under the assumption that bitext is available.
*Italian–English:*

The actor model (pre-trained It–En NMT) is also fine-tuned using the AC method, but the critic learned *value function* is used to center an intermediate reward (here, it is a classifier-based reward, $R_2$ discussed in Section 4.4) as the task dataset is not parallel.

*4.3.4 Training Setups.* The complete workflow of the whole system can be described in the following steps:

(1) Before we start the RL training, we prepare a fixed actor (GeNMT; Section 4.3.1). The actor is pre-trained on the generic domain training corpus (cf. Table 3) using MLE loss.
(2) The actor is used to warm up the critic. We keep the actor fixed and warm-up the critic for one epoch with the learning rate of 0.001 and batch size of 64 using generic domain sentences (cf. Table 3) with SBLEU reward ( see Equation (1) below).

(3) In the RL epochs, the actor and the critic networks are trained jointly by setting a learning rate of 0.0001 and a batch size of 4 (Hi–En) and 3 (It–En) for the 10 RL epochs.
(4) As for the RL epochs, for a given source sentence $s_h$ and the corresponding sequence of translated text $\hat{t}_e$ sampled under the actor's policy, an intermediate reward $R(\hat{t}_e)$ is calculated by (a). taking harmonic mean of two rewards $R_1$ and $R_2$ in case of Hindi–English (using bitext) vs. (b). calculating reward $R_2$ only in case of Italian–English (using polarity tagged monolingual source text). It is discussed in the subsequent section.
(5) The final reward, $\bar{\mathrm{R}}_\tau$ used to weigh the actor's action is a centered reward using the critic future reward prediction. For the critic training, the same source, $s_h$ is given to the encoder and their corresponding sampled action, $\hat{t}_e$ to the decoder as shown in Figure 1. The critic loss is calculated using the ***mean-squared error*** **(MSE)** between the true (intermediate) reward and the critic predicted value.

### 4.4 Reward Computation

This section explains how we compute the rewards that weigh the NMT(s) actions.

*4.4.1 Hindi–English.* For fine-tuning of the Hindi–English model, we optimize the harmonic mean of two rewards $R_1$ and $R_2$ (see Equations (1) and (3)). Please note that $R_1$ computation requires access to the referenced translation. For this, we use the bitext from Table 2. Using $R_1$ and $R_2$, an intermediate harmonic reward R is calculated, which is centered via critic predicted value. The final reward is, thus, obtained, $\bar{\mathrm{R}}_\tau(\hat{y}, x)$ is used to weigh the actor's action taken at time step $\tau$ under policy $\mathrm{G}_\theta(\hat{t}_\tau | \hat{t}_{<\tau}, s_h)$, where G symbolizes generic, $\theta$ represents the parameters of the policy. We explain below why we propose this formulation for the reward combination.

(1) The reason to choose weighted harmonic mean over other averaging methods is that our objective was to give equal weight to each reward component (i.e., $R_1$ as SBLEU and $R_2$ as the element-wise dot product). For harmonic mean to be high, both $R_1$ and $R_2$ are required to be high. Using other averaging methods (e.g., arithmetic mean) might make the obtained reward biased toward the extreme value of $R_1$ or $R_2$. This may bias the final reward magnitude. The harmonic mean, however, ensures a more stable and small reward for gradient updates. We believe this reward should encourage the model to collectively perform well for the two tasks, i.e., sentiment preservation and MT. This was not the case with the ***machine-oriented reinforce*** **(MO-Reinforce)** [47], which is closely related to ours. As mentioned in Section 1, the technique proposed in Reference [47] observed a significant drop in the BLEU score.
(2) A pre-trained NMT model is already available and can be fine-tuned for a few epochs using the task-specific rewards without worrying about hyper-parameter tuning. This is much easier compared to the linear weighted combination that requires deciding good weights to be placed on the individual rewards, which is indeed a cumbersome task.

The detailed steps to calculate the harmonic reward are discussed below. As mentioned above, for a translation $\hat{t}_e$ corresponding to a given source $s_h$, rewards $R_1$ and $R_2$ are calculated as follows. The reward $R_1$ is measured by SBLEU for one full episode of the sampled action $\hat{t}_e$ and its reference translation $t_e$, as in Equation (1):

$$\mathrm{R}_1 = \mathrm{SBLEU}(\hat{t}_e, t_e). \tag{1}$$

In other words, to understand how SBLEU is different than corpus BLEU, we need to understand how BLEU is calculated. The corpus-level BLEU has two components: (a) **brevity penalty (BP)** and (b) **precision component (PC)**, the geometric mean of n-gram precisions $p_n$, $1 \leq n \geq N$, see

Equation (2), where BP is defined as 1 if $c > r$ and $\exp(1 - \frac{r}{c})$ otherwise:

$$BLEU = BP * \left(\prod_{n=1}^{N} p_n\right)^{\frac{1}{N}}. \tag{2}$$

Here, $c$ is the length of the candidate and $r$ is the effective reference corpus length. BLEU is defined at the corpus-level, and $p_n$, $r$ and $c$ are sums over all the corpus sentences. Similarly, an SBLEU for each sentence can be obtained by re-defining $p_n$, $r$, and $c$ to look at sentence $i$ only, i.e., $p_n = \frac{m_{in}}{h_{in}}$, $c = c_i$, and $r = r_i$, where $m_{in}$ is the number of $n$-gram matches between a translation and the references for sentence $i$, and $h_{in}$ is the number of $n$-grams in the hypothesis. However, such a formulation of SBLEU is problematic though, since it can easily get a zero score. Hence, SBLEU is used with several smoothing techniques [10]. We use the add-one smoothing.

Our second reward, $R_2$ is measured as the element-wise dot product of the ground-truth sentiment distribution of the source sentence $s_h$, denoted as $\mathrm{P}(s_h)_{gold}$ and the sentiment distribution given to the corresponding episode of sampled actions $\hat{t}_e$ by the classifier, denoted as $\mathrm{P}(\hat{t}_e)_{bert}$, as in Equation (3):

$$\mathrm{R}_2 = \mathrm{P}(s_h)_{gold} \bullet \mathrm{P}(\hat{t}_e)_{bert}. \tag{3}$$

In particular, $R_2$ is interpreted as a softmax probability given by the classifier to the gold sentiment class. For example, it is with respect to the positive class, which is the gold class in Figure 1, i.e., 0.2. The idea is to maximize the scores of the gold sentiment class. If the model shows low confidence in sampled actions with respect to the gold class, then the reward will be low, and the other way round. We want the model to sample actions that increase the classifier's confidence in the gold class as compared to other classes. For example, say, a gold distribution is $[1, 0, 0]$ and the predicted distribution of the sampled translation is $[0.2, 0.7, 0.1]$ (cf. Figure 1). We see that the classifier's confidence in the gold class is low (i.e., 0.2 in Figure 1). Hence, the model should be encouraged to sample the translation for which its confidence on the gold class increases, e.g., positive class in Figure 1. To pursue this objective, we compute the dot product [9] of the gold sentiment distribution and the predicted sentiment distribution obtained from the softmax layer.

The harmonic reward (as an intermediate reward) is calculated using Equations (1) and (3), as in Equation (4):

$$\mathrm{R}(\hat{t}_e, s_h) = (1 + \beta^2)\frac{(2 \cdot \mathrm{R}_1 \cdot \mathrm{R}_2)}{(\beta^2 \cdot \mathrm{R}_1) + \mathrm{R}_2}, \tag{4}$$

where $\beta$ is a harmonic weight, which is set to 0.5, and $\mathrm{R}(\hat{t}_e, s_h) \equiv \sum_{\tau=1}^{n} \mathrm{R}(\hat{y}_\tau, x_\tau)$ is interpreted as true terminal reward observed only at the end of the actor episode, i.e., time step $\tau = n$. During the RL training, we collect $(\hat{t}_e^i, s_h^i)$ pairs for batches of input and pass them to the critic model to train the value estimator $\mathrm{V}\,(\phi)$.

Finally, the critic estimated value function, $\mathrm{V}(\phi)$, where $\phi$ being the parameters of critic, is used to center the reward $\mathrm{R}(\hat{t}_e, s_h)$ using $\mathrm{V}(\hat{y}_{<\tau}, s_h)$ predicted by the critic for the time step $\tau$ through Equation (5):

$$\bar{\mathrm{R}}_\tau(\hat{y}, x) = \mathrm{R}(\hat{t}_e, s_h) - \mathrm{V}(\hat{y}_{<\tau}, s_h). \tag{5}$$

We use $\bar{\mathrm{R}}_\tau$ to train the actor in the **policy gradient (pg)** methods. For a fixed $s_h = \{x_1, x_2 \ldots x_n\}$ one will approximate the expectation of reward using a single sample estimate [50], which is usually seen as a common practice. Hence, the simplified pg loss estimate used to update the

[9] Please note, before using the resultant distribution as reward, we sum the vector to a scalar.

actor's policy, $\mathrm{G}_\theta$ is given, as in Equation (6)[10]:

$$\nabla_\theta \mathrm{L}^{\mathrm{pg}}_{\mathrm{actor}}(\theta) \approx \sum_{\tau=1}^{n} \bar{\mathrm{R}}_\tau(\hat{y}) \nabla_\theta \log \mathrm{G}_\theta(\hat{y}_\tau | \hat{y}_{<\tau}, s_h^l). \tag{6}$$

The critic optimizes the MSE loss between its estimates and the true intermediate rewards. For a fixed $s_h$ and $\hat{t}_e \sim \mathrm{G}_\theta(.|s_h)$, we use a gradient approximation to update the critic, as in Equation (7):

$$\nabla_\phi \mathrm{L}_{crt}(\phi) \approx \sum_{\tau=1}^{n} \left[\mathrm{V}_\phi(\hat{y}_{<\tau}) - \mathrm{R}(\hat{t}_e)\right] \nabla_\phi \mathrm{V}_\phi(\hat{y}_{<\tau}). \tag{7}$$

Algorithm 1 summarises the entire fine-tuning steps.

**ALGORITHM 1:** Proposed algorithm (fine-tuning process for Hindi–English with harmonic mean as an intermediate reward). To run this actor-critic algorithm, we follow the mini-batch update rule

```
Initialize the actor model ($\mathrm{G}_\theta$) with uniform weights $\theta \in [-0.1, 0.1]$.
Pre-train the actor with MLE objective.
Initialize the critic model ( $\mathrm{V}_\phi$) with uniform weights $\phi \in [-0.1, 0.1]$.
Pre-train the critic with MSE loss for one epoch using sample from the actor with SBLEU as a reward in Equation (7).
for each iteration $i = 1, 2, \ldots, N$ do
    Sample a translation sequence $\hat{t}_e^i$ given the source sequence $s_h^i$.
    Observe reward $R_1$ and $R_2$ using Equation (1) and Equation (3).
    Compute the intermediate reward (here, harmonic mean of $R_1$ and $R_2$) using Equation (4).
    Feed the source $s_h^i$ to the critic encoder and the sampled translation $\hat{t}_e^i$ to the critic decoder.
    Obtain the value estimation,$\mathrm{V}_\phi$, of the expected (future) reward using the critic.
    Update the critic's parameter using Equation (7).
    Obtain final reward $\bar{\mathrm{R}}$ using Equation (5).
    Update the actor's parameter using pg loss in Equation (6).
end for
```

*4.4.2 Italian–English.* In the absence of in-domain parallel data for Italian–English, we use the polarity-tagged source sentences from Table 4, row (ii) in the RL fine-tuning, where $R_2$ is used as the intermediate reward, $R$. This implies the critic predicted value is then used to center the reward, $R_2$.

### 4.5 Baseline Models

We validate our hypothesis that the translations generated by the proposed NMT framework can be used to further improve the performance of the target language sentiment classifier. Accordingly, we compare the performance of the e-uBERT(s) (binary for It–En or multi-class for Hi–En) sentiment classifier by feeding it with the translations of the corresponding RL testsets. These translations are generated by the GeNMT(s) (our first baseline) and the following fine-tuned models obtained using two different types (parallel vs. monolingual) of the training resources:

[10] Although shown like this, the generic model is not conditioned on the sentiment label $s_h^l$ directly here, it is used only for calculating rewards after the post-sampling stage.

*4.5.1 GeNMT Tuned using Parallel Sentences.* As mentioned previously, the in-domain parallel sentences are available for GeNMT (Hi–En). Accordingly, we compare the performance of e-uBERT (multi-class) by feeding it with the translations from,

(1) Our first baseline- the pre-trained MLE-based NMT- trained on the generic domain data, which we also referred as GeNMT.
(2) Continued Training- our second baseline- obtained via fine-tuning of the referenced GeNMT from (1) until convergence on the in-domain bitext.
(3) GeNMT- fine-tuned using our critic-based approach- with harmonic reward, which we referred to as AC-HAR, our proposed framework.

To tune and validate the given referenced GeNMT model,[11] we use the parallel sentences of RL trainset and RL devset (cf. Table 3, row (i)). The model is fine-tuned for a maximum of 10 epochs, and the average reward is maximised on the development set. For continued training, perplexity is used to select the best epoch.

*4.5.2 GeNMT Tuned using Corresponding Monolingual Polarity-tagged Sentences.* Alongside, for comparison with the recent approach, we also adapted the idea of Tebbifakhr et al. [47] to fine-tune our GeNMT model (Hi–En or It–En) to feed the corresponding target classifier (multi-class or binary, respectively). Please note that, Tebbifakhr et al. [47]'s MO Reinforce uses only source sentences with the task-specific labels during the fine-tuning. For this, they sampled $k$-candidates translations (with $k = 5$) associated with their downstream rewards corresponding to a given source sentence. Finally, to update the GeNMT parameters, they choose the highest rewarding candidate among the sampled $k$-candidates. To make a fair comparison with MO Reinforce, we also used our actor-critic framework under the assumption of the availability of similar resource, i.e., the GeNMT has access to only the sentiment-tagged source sentences in the fine-tuning steps. Accordingly, we use the corresponding target classifier to optimize the BERT-based reward, $R_2$. We call this variation of our proposed framework as AC-BERT.

Furthermore, to validate the claim of References [37, 47] that for a low-resource language, performing cross-lingual sentiment analysis via the translation-based approach to a resource-rich language and subsequently using the resource-rich language classifier is a better option than using the language-specific classifier. To this end, we additionally perform the following experiments,

(1) Recorded the performance of the source classifiers (h-uBERT, i-uBERT) on the source-side of RL testsets (cf. Table 6, row (v), (vi) for the results).
(2) Recorded the performance of the target sentiment classifier, e-uBERT (multi-class),[12] on the gold standard English test set, i.e., target-side of the RL testset (cf. Table 6, row (iv) for the result).

## 5 RESULTS AND ANALYSIS

In this section, we present the evaluation results along with necessary discussions and analysis. Section 5.1 first discusses the results for the language pair, Hindi–English (cf. Table 5), which compares the performance of the baseline models (GeNMT, Continued training and, MO Reinforce) obtained by fine-tuning the referenced GeNMT, with our proposed model (AC-HAR). To evaluate the NMT(s) performance, we use BLEU, a widely-used automatic evaluation metric for measuring

[11] Please note that for all the baselines implementation, our encoder-decoder architecture is RNN-based.

[12] Please note, we do not report e-uBERT (binary) classifier gold standard English test set performance, as the corresponding gold English translations for the Italian task is not available.

Table 5. Performance of the MT Systems (Hi–En) and the (Multi-class) Target Sentiment Classifiers in Terms of BLEU and $F_1$ Score

| | Training Resource | MT system | BLEU | BERT | $F_1$ Score |
|---|---|---|---|---|---|
| (i). | | GeNMT | 19.99 | e-uBERT | 88.51 |
| (ii). | Parallel data | Continued | 26.88 | e-uBERT | 89.68 |
| (iii). | | AC-HAR | 24.56 | e-uBERT | **91.99** |
| (iv). | Labelled data | MO Reinforce | 16.70 | e-uBERT | 91.19 |

Table 6. Performance of the Multi-class and Binary-class Sentiment Classifiers in terms of $F_1$ Score for Hi–En and It–En, respectively

| | Language | MT system | BERT | $F_1$ Score (Hi–En) | $F_1$ score (It-En) |
|---|---|---|---|---|---|
| (i). | Target | GeNMT | e-uBERT | 88.51 | 53.15 |
| (ii). | Target | MO Reinforce | e-uBERT | 91.19 | 57.92 |
| (iii). | Target | AC-BERT | e-uBERT | **92.17** | 55.42 |
| (iv). | Target | Gold set | e-uBERT | 94.03 | x |
| (v). | Source | | h-uBERT | 71.00 | x |
| (vi). | Source | | i-uBERT | x | 68.00 |

the translation quality. Additionally, we obtain the $F_1$ scores to evaluate the sentiment classifier (multi-class e-uBERT ) performance on the translated RL testset across all the models.

Further, to evaluate the NMT(s) monolingual setting performance, we present the classification $F_1$ scores (cf. Table 6) of Tebbifakhr et al. [47] MO Reinforce, and AC-BERT for two topologically dissimilar source languages, i.e., Hindi, Italian and a common target language, English. As mentioned earlier, for fine-tuning and validation of the referenced GeNMT(s) (Hi–En, It–En), we use the RL datasets (polarity-tagged source Hindi or Italian sentences from Table 4) with $R_2$ as the reward (see Equation (3)) in case of MO Reinforce and $R_2$ as the intermediate reward centered via critic in case of AC-BERT (see Equation (5)).

Moreover, with respect to the language pair, Hindi–English, we perform the following ablation studies and the human evaluation of the model's output. In particular, Section 5.2 discusses the contributions of different components of AC-HAR and in Section 5.3, we present the qualitative error analysis. This section also discusses why an increase in BLEU does not correlate with $F_1$ score.

### 5.1 Automatic Evaluation

Taking randomness of the deep learning algorithms into consideration, we repeat the experiments three times with different seed values, and record the $F_1$ and BLEU scores to evaluate the sentiment analysers and the MT systems, respectively, on the RL testset. We perform the statistical significance test using bootstrap resampling method [23] to validate the differences in BLEU scores. Additionally, student's t statistics was calculated to validate differences between the $F_1$ scores.

We first discuss our observation on the quality of sentiment transfer in Hindi–English. As can be seen from Table 5, when we apply e-uBERT to the translated text (obtained through AC-HAR), we obtain a $F_1$ score as 91.99 point, which is +3.48 points absolute gain over the GeNMT score (88.51) and is also the highest observed score. This difference is significant with 95% confidence level ($p < 0.05$).

Table 7. DISTRIBUTION of POSITIVE, NEGATIVE, and NEUTRAL CLASSES. HERE, at REFERS to AUTOMATIC TRANSLATION, (A) RL TESTSET (GOLD-STANDARD), (B) ENGLISH SIDE of RL TESTSET, PREDICTED by E-UBERT, (C) TRANSLATIONS of RL TESTSET (GENMT), PREDICTED by E-UBERT,(D) TRANSLATIONS of RL TESTSET (CONTINUED), PREDICTED by E-UBERT, (E) TRANSLATIONS of RL TESTSET (MO REINFORCE), PREDICTED by E-UBERT, and (F) TRANSLATIONS of RL TESTSET (AC-HAR), PREDICTED by E-UBERT

| | | | Positive (%) | Neutral (%) | Negative (%) |
|---|---|---|---|---|---|
| (a) | | Gold-standard | 33.33 | 33.33 | 33.33 |
| (b) | e-uBERT | RL testset (English) | 35.36 | 30.86 | 33.78 |
| (c) | e-uBERT | GeNMT, AT | 34.22 | 33.95 | 31.83 |
| (d) | e-uBERT | Continued, AT | 34.92 | 32.45 | 32.63 |
| (e) | e-uBERT | MO Reinforce, AT | 36.60 | 30.33 | 33.06 |
| (f) | e-uBERT | AC-HAR, AT | 35.63 | 31.39 | 32.98 |

Table 8. Percentage of Labels that Deviated from Its Gold Sentiment after Translation

| MT systems | Gold | changed (%) | changed (%) |
|---|---|---|---|
| (a) Generic | 0 | 1 (2.30) | 2 (0.88) |
| | 1 | 2 (1.50) | 0 (2.80) |
| | 2 | 1 (2.50) | 0 (1.30) |
| (b) MO Reinforce | 0 | 1 (1.00) | 2 (0.70) |
| | 1 | 2 (1.00) | 0 (4.00) |
| | 2 | 1 (1.00) | 0 (1.00) |
| (c) Continued | 0 | 1 (1.80) | 2 (0.97) |
| | 1 | 2 (1.40) | 0 (3.30) |
| | 2 | 1 (1.80) | 0 (1.20) |
| (d) AC-HAR | 0 | 1 (1.80) | 2 (0.97) |
| | 1 | 2 (1.40) | 0 (3.30) |
| | 2 | 1 (1.80) | 1 (1.20) |

Here, 0 is positive, 1 is neutral, and 2 is negative class.

On comparing the AC-HAR and the continued training, we see that for the translation of AC-HAR, we obtain +2.31 points absolute improvement in the $F_1$ score over continued training despite the highest observed BLEU score (26.88) for the latter. The second best BLEU improvement is observed from our harmonic model. It suggests that our proposed AC-HAR is best suited for the sentiment preservation task compared to the baselines (GeNMT, Continued and MO Reinforce). It further suggests that despite the high BLEU score obtained with the Continued training (inspired by a human-oriented criterion, i.e., maximizing likelihood of in-domain training data) may not necessarily implies an optimal sentiment projection. This is an interesting observation based on the empirical results only, and needs to be investigated thoroughly to understand the limitations of MLE training along with the limitations of automatic evaluation metric BLEU. We now refer the reader to Table 6 for discussion on the cross-lingual sentiment analyser performance for Hindi–English and Italian–English in the monolingual setting.

For Hindi–English, when we run e-uBERT on the target-side of the RL testset (gold), we obtain $\mathrm{F}_1$ score of 94.03 (see row (iv)). This may be viewed as the upper bound in this task. The last row of Table 6 represents the $\mathrm{F}_1$ score that we obtain by applying the Hindi sentiment classifier on the

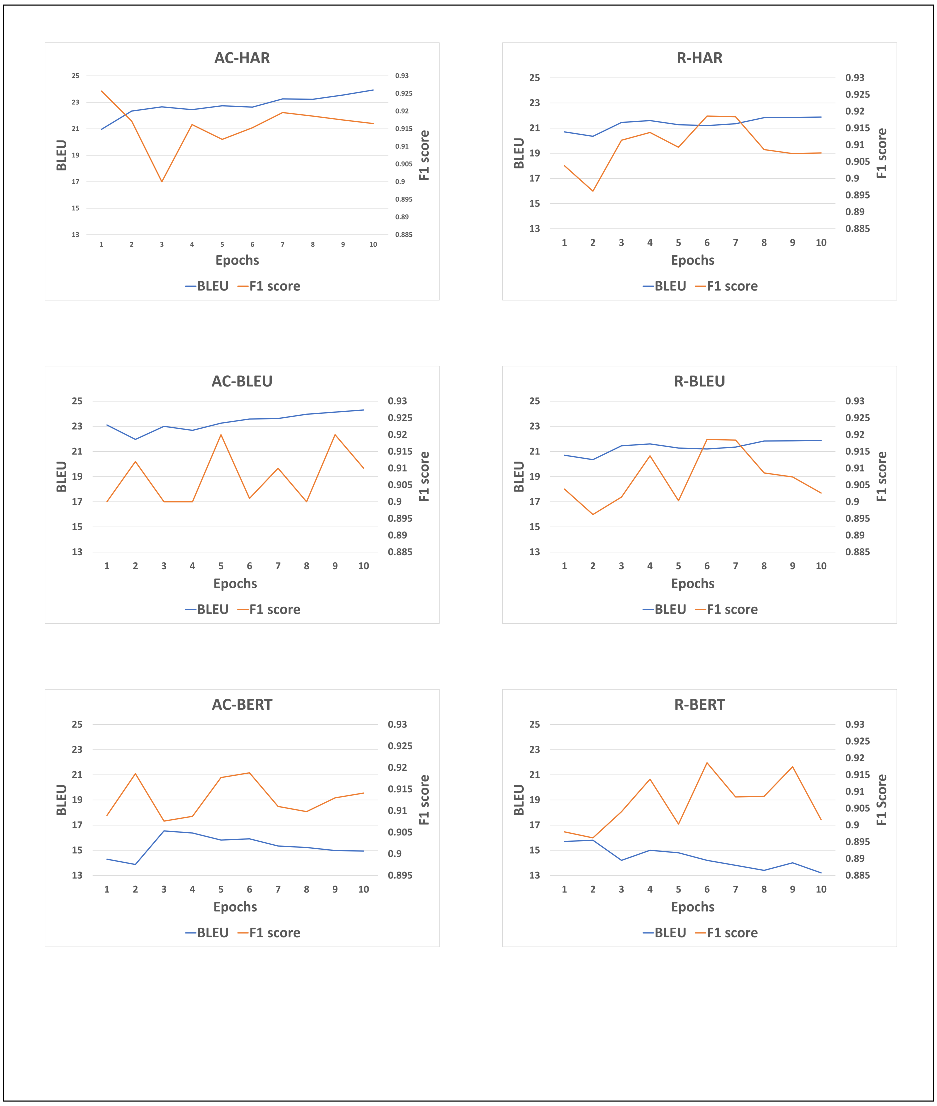

Fig. 3. The RL devset scores in terms of two metrics (BLEU and $F_1$ score) for actor-critic (left) and REINFORCE (right) across 10 epochs using (i) harmonic mean as a reward, (ii) SBLEU as a reward, and (iii) only classifier feedback signal as a reward.

source-side of the RL testset. By comparing rows (i), (ii), (iii) vs. (v), we observe that the best result is obtained with AC-BERT (92.17) followed by MO (91.19) and GeNMT (88.51), which is better than the language-specific classifier score (71.00). This finding for Hindi–English is similar to Reference [47] and suggests that sentiment analysis of the translated Hindi texts using an English (a resource-rich) language sentiment analysis system is a more viable choice than doing sentiment analysis of the source Hindi text directly. This is a preferred alternative even if the MT system is a vanilla NMT, i.e., GeNMT. We see from results in Tables 6 and 7 that the translation from a controlled NMT (i.e., MO, AC-HAR and AC-BERT) improves the performance of the classifier further suggesting they learned to preserve sentiment of the source text.

However, for Italian–English tweets classification, our observation is different from claims of Reference [47]. The language-specific classifier score (cf. Table 6, last column) suggests that despite an observed absolute improvement of +4.77 points for MO Reinforce and +2.27 points for AC-BERT, respectively (row (ii), (iii)), over the referenced GeNMT (53.15), the target (English) language $F_1$ scores could not surpass source classifier score (68.00). One possible reason could be the nature of in-domain task data (noisy tweets) for Italian–English that significantly varies from the referenced GeNMT training data, which is contrary to the case for Hindi–English SC training data (sentences majorly drawn from the ILCI corpus). As a result, the source itself being noisy might have negatively affected the translation quality. Please note, our approach still improves the classification score over GeNMT (55.42). While comparing MO Reinforce with AC-BERT for Hindi–English and Italian–English (column (iv) and (v)), we found that AC-BERT shows small improvement (+0.98 point) over MO Reinforce in the multi-class classification where as MO Reinforce performs better with the absolute improvement of +2.5 points in the binary-class sentiment classification over AC-BERT.

Now, we discuss the performance of the NMT systems (Hindi–English). Unlike Reference [47], our AC-HAR substantially improves the GeNMT (cf. Table 5, column (iii)). We obtain +4.57 points absolute gain in BLEU score over GeNMT and +7.86 points over MO Reinforce. This gain is statistically significant with 95% confidence level ($p$<0.05). This indicates that our reinforced NMT (AC-HAR) systems tend to preserve the source-side non-emotional semantic content in the target translation too. However, when comparing AC-HAR with continued training, the latter is observed to have achieved the highest improvement (+6.89 points) over our first baseline and +2.32 points over our proposed AC-HAR. However, as discussed previously, AC-HAR performs better than Continued training for the sentiment preservation.

To understand the class biases (toward neutral, see Section 3.1) of GeNMT as claimed by Lohar et al. [30] and Mohammad et al. [37] vs. other NMT(s), we show in Table 7 the class-wise sentiment distribution (in percentage) of the RL testset (tourism). Additionally, Table 8 shows the percentage change in labels to other classes, when the translations are obtained from different NMT(s). Table 7, row (a) presents the true sentiment distributions as per the manual annotation of the RL testset. Row (b) refers to the distribution as per the predictions of e-uBERT, when applied on the gold RL testset. Rows (c) to (e) of Table 7 represent the distributions, when e-uBERT is applied to the translations of RL testset through GeNMT, Continued, MO, and AC-HAR, respectively.

We see a rise in the number of neutral classes (33.95%) when the e-uBERT is applied to the translations obtained by the GeNMT model as compared to case, when it is applied to the gold RL testset (30.86%). Results from Table 8, row (a) similarly suggest that more label changes occur from other class to the neutral class. For example, 2.3% of label changes to neutral from positive class, and 2.5% changes from negative to the neutral.

We clearly see that the baseline NMT system, GeNMT, has a bias toward the neutral class followed by the positive class, and our finding is corroborated by the other studies [31, 37]. As expected, neutral class bias is not observed when the classifier is applied to the translations of AC-HAR.

### 5.2 Ablation Study

In this section, we investigate the contribution of two key components, viz. (i) harmonic reward and (ii) critic as a average reward estimator to stabilize the actor training of the proposed model.

*5.2.1 Reward Utility.* To contemplate the harmonic reward contribution in the proposed RL framework, we train two more models keeping the same pre-trained critic unchanged, and removing each of the reward components at a time. Accordingly, we train three different models in total from the same critic as a common starting point with the following three different rewards: (i) classifier-based reward, (ii) SBLEU, and (iii) the harmonic-mean of (i) and (ii).

For each model, we obtain RL devset evaluation scores ($F_1$ and BLEU) for the 10 RL epochs each, which were recorded for all the three reward formulations and shown in Figure 3. As can be seen from the left side of Figure 3, the harmonic reward (which, in fact, represents AC-HAR) shows more consistent improvements in both the metrics than its individual counterparts, which, unlike harmonic reward, show improvements in the targeted-metric that the models take into account for optimization.

Interestingly, an increase in BLEU does not lead to an improvement in $F_1$ score (see middle graph of Figure 3). An increase in BLEU is analogous to the improvement in the quality of translation, and this is a well-accepted belief in the MT community. In our case, the increase in BLEU does not lead to the expected improvement in $F_1$ score, which indicates the MT system's inability to translate the sentiment bearing expressions in comparison to the non-sentiment bearing expressions. This observation is analogous to our observation discussed in Section 5.1, where an MT system inspired by the MLE objective with the highest recorded BLEU score does not perform at par with our proposed AC-HAR in sentiment preservation.

One would expect that correct translations of the opinionated expressions would result in an improved sentiment classifier. However, this was not also the fact in our case as we see that the correct translations of sentiment bearing expressions have not resulted in an improved classifier's performance. In sum, there was not much observed correlation between the BLEU and $F_1$ score. This may be because BLEU is calculated based on the percentage of overlapping *n*-grams. However, these overlapping *n*-grams contain not only the content words but also the functional words. We further study this by carrying out a manual evaluation, and discuss this in Section 5.3, where we see that an increase in the BLEU score may have a co-relation with the preservation of the functional words.

*5.2.2 Critic Utility.* To analyse the critic utility, we train three variations of REINFORCE [50] using the three rewards described above. Similarly, we record the BLEU scores and $F_1$ scores on the development set (RL devset), and plot them in the graph with respect to epochs, which are shown in the right side of Figure 3.

On comparing the left and right side graphs of Figure 3, we see that the critic indeed helps the model to better identify and utilise task-specific rewards at hand in all the three rewards formulation. For example, observe the left and right sides of the first and the middle graphs of Figure 3, which shows a clear upward trend for the targeted metrics. This entails that the use of critic guides the actor model toward optimizing the target metrics.

### 5.3 Human Evaluation

To examine, why an improvement in BLEU does not result in an improved $F_1$ score, we carry out further analysis. First, we define two models, the harmonic-mean as a reward and SBLEU as a reward for our AC (actor-critic) model, which are referred to as Model A and Model B, respectively. Let us refer A to be the set of translations obtained by Model A (harmonic reward) and correctly classified by the sentiment classifier; and B as the set of translations obtained by Model B (SBLEU reward) and correctly classified by the sentiment classifier.

We take examples from set, $A - B$ (i.e., classification examples in which Model A are right but Model B is wrong) and, vice versa. We show some of the translations of $A - B$ and $B - A$ in Tables 9 and 10, respectively. On evaluating set $A - B$, the evaluator suggested that Model A, unlike Model B, is more likely to preserve the preferred variant of sentiment in the translation. It justifies the observed gain in performance of the classifier when it is fed with the translations produced by

Table 9. Samples where Model A Performs Better

| | | |
|---|---|---|
| Gold Output | pos | not just kids, it attracts adults as well. |
| Model A | pos | it is not the kids, the adults also attract. |
| Model B | neu | this is not the kids, the adults. |
| Gold Output | neu | regular bus and taxi services are available to go to tattapani. |
| Model A | neu | there is a regular bus and taxi for the water. |
| Model B | neg | regular bus and taxi service is available to burn. |
| Gold Output | pos | wow. its kinda amazing experience. can't forget. mughal regime architecture and sprawling lawns. simply no words. |
| Model A | pos | wow. it is a wonderful experience. can not forget. can not forget. the mughal rule and the vast lot. just no term. |
| Model B | neg | wow. it is a wonderful experience. no. can. no. can. the mughal rule and the big lolan. just no term. |
| Gold Output | pos | because of being situated at a height of 2110 feet from sea level the season of summer in dehradun is also pleasant. |
| Model A | pos | the dehradun is also a pleasant weather in dehradun because of the height of 2110 ft in the height of 2110 ft in the 21st. |
| Model B | neu | the dehradun is also the season of the weather in the height of 2110 feet from the sea sea, the weather of thunder in dehradun. |
| Gold Output | pos | the temple is really good and the people are so helpful and caring. |
| Model A | pos | the temple is really good and people are helpful and care. |
| Model B | neg | the temple is really good and people are helpful and careless. |
| Gold Output | neg | i found this temple hugely depressing. the day i visited it the elephant was encased in a tiny chamber deep inside the temple ( not outside at all ). |
| Model A | neg | i found this temple very disappointing. the day i was seen was tied to the temple, a small room in the elephant. |
| Model B | pos | i found this temple very fruitfully. the day i was seen it was seen inside the elephant, a small room (which was not the outside of the outside). |

Model A. The evaluator also looked at the set $B - A$, and suggested that, in most cases, model A is robust to preserve the sentiment-bearing expressions in translation. The miss-classification occurs mainly due to the appearance of some words that are the indicators of opposite sentiments in the output (cf. Table 10) or due to the error of the classifier.

The human evaluation results also favour the hypothesis (cf. Section 5.2.1) that the improvement in BLEU score might have come mainly due to the preservation of the non-content words. Further, Model B is more susceptible to the missed translation of sentiment bearing expression which justifies the dropped $F_1$ score.

Table 10. Samples where Model B Performs Better

| | | |
|---|---|---|
| Gold Output | neg | the sarovar was full of dirt. |
| Model A | neu | the lake was full of gandgar. |
| Model B | neg | the sarovar was full of dirt. |
| Gold Output | neu | iconic places in delhi to visit during the holiday. one should visit this place for sure. |
| Model A | pos | the prestigious place of delhi for the sake of leave during the leave. nobody should travel in this place. |
| Model B | neu | the prestigious place in delhi for the sake of leave is also to travel in this place. |
| Gold Output | pos | nature at its best. close to nature. had a great time with family. |
| Model A | neg | nature is the best form. there is not good time with the family.very good time with the family. |
| Model B | pos | nature is the best form. there is very good time with the family. very good time in the family. |
| Gold Output | pos | great experience. never miss it. |
| Model A | neg | the excellent experience. it is not going. |
| Model B | pos | the excellent experience is not known. |
| Gold Output | pos | the immense beauty of this mountain has made discoverers and pilgrims believe that they must see a glimpse of this heaven. |
| Model A | neg | the boundless beauty of this mountain has assured the loss and pilgrimage to the pilgrims and pilgrims that they must see a glimpse of this paradise. |
| Model B | pos | the boundless beauty of this mountain is the confidence and pilgrims and pilgrims who have assured that they see a glimpse of this paradise. |

In this context, BLEU is an $n$-gram precision-based metric and deals with the functional or content bearing expressions equally likely in its calculation. In other words, no additional penalty is given even if the MT system omits the important information in translation, e.g., sentiment bearing expressions. In our case, we obtain a moderate gain on BLEU for Model B and observe no improvement in the classifier performance. Similar observation was made in Section 5.1 about Continued training where despite the highest observed BLEU score, the performance of sentiment model was not at par with the AC-HAR. This indicates that the gain on BLEU does not always correlate with the gain in $F_1$ score.

## 6 CONCLUSION

In this work, we have made use of a state-of-the-art reinforcement learning technique to fine-tune a global-attention-based NMT system [4] in such a way that it learns to (i) generate translations that are best suited for a downstream classification (sentiment) task and (ii) keep the source-side semantics preserved in the translation.

For this, we use a popular policy gradient method (actor-critic) that makes use of our novel reward function that operates by taking a weighted harmonic mean of two individual rewards: (i) sentence-level BLEU that provides a measure in the form of translation quality of a sampled translation given the reference translation and (ii) a function that performs element-wise dot product between a predicted sentiment distribution and the ground-truth sentiment distribution.

This general-purpose reward function let the NMT system preserves the (a) underlying sentiment and (b) non-emotional semantic content intact in the translation. We have empirically demonstrated that this learning strategy brings about an improved NMT system, which, in turn, elevates the performance of the sentiment classifier. In particular, for Hindi–English, this improved the performance of the sentiment classifier by 3.48 points in terms of $F_1$ score and 4.57 points in BLEU over the generic model.

In this regard, applying the Hindi sentiment classifier on the source-side (Hindi) sentences resulted in an inferior performance. However, for translation direction Italian–English, we find that the language-specific classifier performs better. In sum, performing sentiment analysis of the translated Hindi text to English, through an automated MT system, is a more viable choice than performing sentiment analysis of the original Hindi text. This is a preferred alternative even if the MT system is a vanilla NMT baseline.

Moreover, to the best of our knowledge, there is no freely and readily available gold standard for evaluating the sentiment preservation in MT. Due to the unavailability of such resources, we manually created a polarity labelled balanced bilingual corpus for English–Hindi.[13] This serves our gold standard evaluation test set for studying the sentiment preservation in MT.

We performed a systematic ablation study to validate the effectiveness of our reward formulation and critic network in the learning framework. This study showed that our novel harmonic reward with critic as an average reward estimator has a significant impact on this learning task. More specifically, our reward formulation based on a sentence-level BLEU function and a weak reward signal from the classifier is more robust to weigh actions taken under a policy training than their individual counterparts, which when used in collaboration (i.e., as a weighted harmonic reward) seem to complement and help the learner (NMT) to consistently improve rewards in both the tasks, viz. sentiment and semantic content preservation.

Our human evaluators endorsed the findings of our automatic evaluation process in the sense that our best-performing MT system learns to preserve the sentiment-bearing expressions in the target translation. The evaluators also indicated that the BLEU improvement does not necessarily signify the fact that content words are better preserved in automatic translation. In many cases, the improvement in the BLEU score may be observed due to mainly the preservation of functional words from the source to the target translation.

Although the underlying learning principles of the RNN MT model [4] and Transformer [49] are quite similar despite many differences, in the future, we intend to test our proposed strategy on Transformer too. We also aim to investigate exploring different reward functions for semantic transfer such as METEOR [12], which, unlike precision-based metrics such as BLEU, is robust in terms of capturing lexical variations of items.

## REFERENCES

[1] Haithem Afli, Sorcha Maguire, and Andy Way. 2017. Sentiment translation for low resourced languages: Experiments on Irish general election tweets. In *Proceedings of the 18th International Conference on Computational Linguistics and Intelligent Text Processing*.

[2] Sweta Agrawal and Marine Carpuat. 2019. Controlling text complexity in neural machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP'19)*. 1549–1564. https://doi.org/10.18653/v1/D19-1166

[3] Dzmitry Bahdanau, Philemon Brakel, Kelvin Xu, Anirudh Goyal, Ryan Lowe, Joelle Pineau, Aaron Courville, and Yoshua Bengio. 2017. An actor-critic algorithm for sequence prediction. In *Proceedings of the 5th International Conference on Learning Representations*.

[13] Contact authors for the dataset.

[4] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *Proceedings of the International Conference on Learning Representations*.

[5] Alexandra Balahur and Marco Turchi. 2012. Multilingual sentiment analysis using machine translation? In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*. Association for Computational Linguistics, 52–60. Retrieved from https://www.aclweb.org/anthology/W12-3709.

[6] Alexandra Balahur, Marco Turchi, Ralf Steinberger, José Manuel Perea Ortega, Guillaume Jacquet, Dilek Küçük, Vanni Zavarella, and Adil El Ghali. 2014. Resource creation and evaluation for multilingual sentiment analysis in social media texts. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC'14)*. Citeseer, 4265–4269.

[7] A. R. Balamurali, Mitesh M. Khapra, and Pushpak Bhattacharyya. 2013. Lost in translation: Viability of machine translation for cross language sentiment analysis. In *Proceedings of the 14th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing'13)*. 38–49. https://doi.org/10.1007/978-3-642-37256-8_4

[8] Carmen Banea, Rada Mihalcea, Janyce Wiebe, and Samer Hassan. 2008. Multilingual subjectivity analysis using machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 127–135.

[9] P. Basile, F. Cutugno, M. Nissim, V. Patti, and R. Sprugnoli. 2016. EVALITA 2016: Overview of the 5th evaluation campaign of natural language processing and speech tools for italian. In *Proceedings of the CEUR Workshop*. 1749.

[10] Boxing Chen and Colin Cherry. 2014. A systematic comparison of smoothing techniques for sentence-level bleu. In *Proceedings of the Association for Computational Linguistic*. 362–367.

[11] Boxing Chen and Xiaodan Zhu. 2014. Bilingual sentiment consistency for statistical machine translation. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*. 607–615.

[12] Michael Denkowski and Alon Lavie. 2011. Meteor 1.3: Automatic metric for reliable optimization and evaluation of machine translation systems. In *Proceedings of the 6th Workshop on Statistical Machine Translation*. 85–91.

[13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[14] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 4171–4186. https://doi.org/10.18653/v1/N19-1423

[15] Marie Escribe. 2019. Human evaluation of neural machine translation: The case of deep learning. In *Proceedings of the 2nd Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT'19)*. 36.

[16] Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychol. Bull.* 76, 5 (1971), 378.

[17] Markus Freitag and Yaser Al-Onaizan. 2016. Fast domain adaptation for neural machine translation. Retrieved from https://arXiv:1612.06897.

[18] Markus Freitag, Isaac Caswell, and Scott Roy. 2019. APE at scale and its implications on MT evaluation biases. In *Proceedings of the 4th Conference on Machine Translation*. 34–44.

[19] Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford* 1, 12 (2009).

[20] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Comput.* 9, 8 (1997), 1735–1780.

[21] Hiroshi Kanayama, Tetsuya Nasukawa, and Hideo Watanabe. 2004. Deeper sentiment analysis using machine translation technology. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING'04)*. 494–500. Retrieved from https://www.aclweb.org/anthology/C04-1071.

[22] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proceedings of the International Conference on Learning Representations (ICLR'15)*.

[23] Philipp Koehn. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 388–395.

[24] Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*. 177–180.

[25] Anoop Kunchukuttan, Pratik Mehta, and Pushpak Bhattacharyya. 2018. The IIT Bombay English-Hindi parallel corpus. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC'18)*. European Language Resources Association, 3473–3476.

[26] Tsz Kin Lam, Julia Kreutzer, and Stefan Riezler. 2018. A reinforcement learning approach to interactive-predictive neural machine translation. In *Proceedings of the European Association for Machine Translation conference*. 169–178.

[27] Tsz Kin Lam, Shigehiko Schamoni, and Stefan Riezler. 2019. Interactive-predictive neural machine translation through reinforcement and imitation. In *Proceedings of Machine Translation Summit*. 96–106.

[28] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. Retrieved from https://arXiv:1910.13461.

[29] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A robustly optimized BERT pretraining approach. Retrieved from https://arXiv:1907.11692.

[30] Pintu Lohar, Haithem Afli, and Andy Way. 2017. Maintaining sentiment polarity in translation of user-generated content. *Prague Bull. Math. Linguist.* 108, 1 (2017), 73–84.

[31] Pintu Lohar, Haithem Afli, and Andy Way. 2018. Balancing translation quality and sentiment preservation. In *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas.* 81–88.

[32] Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective approaches to attention-based neural machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing.* 1412–1421. https://doi.org/10.18653/v1/D15-1166

[33] Evgeny Matusov. 2019. The challenges of using neural machine translation for literature. In *Proceedings of the Qualities of Literary Machine Translation.* 10–19.

[34] Paul Michel and Graham Neubig. 2018. Extreme adaptation for personalized neural machine translation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics.* Association for Computational Linguistics, 312–318. Retrieved from https://www.aclweb.org/anthology/P18-2050.

[35] Hideki Mima, Osamu Furuse, and Hitoshi Iida. 1997. Improving performance of transfer-driven machine translation with extra-linguistic informatioon from context, situation, and environment. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'97).* 983–989.

[36] Shachar Mirkin, Scott Nowson, Caroline Brun, and Julien Perez. 2015. Motivating personality-aware machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing.* 1102–1108. https://doi.org/10.18653/v1/D15-1130

[37] Saif M. Mohammad, Mohammad Salameh, and Svetlana Kiritchenko. 2016. How translation alters sentiment. *J. Artific. Intell. Res.* 55 (2016), 95–130.

[38] Khanh Nguyen, Hal Daumé III, and Jordan Boyd-Graber. 2017. Reinforcement learning for bandit neural machine translation with simulated human feedback. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing.* 1464–1474. https://doi.org/10.18653/v1/D17-1153

[39] Xing Niu, Marianna Martindale, and Marine Carpuat. 2017. A study of style in machine translation: Controlling the formality of machine translation output. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing.* 2814–2819.

[40] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on Association for Computational Linguistics.* 311–318.

[41] Alberto Poncelas, Pintu Lohar, James Hadley, and Andy Way. 2020. The impact of indirect machine translation on sentiment classification. In *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas.* Association for Machine Translation in the Americas, 78–88.

[42] Ella Rabinovich, Raj Nath Patel, Shachar Mirkin, Lucia Specia, and Shuly Wintner. 2017. Personalized machine translation: Preserving original author traits. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics.* 1074–1084.

[43] Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Controlling politeness in neural machine translation via side constraints. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* 35–40. https://doi.org/10.18653/v1/N16-1005

[44] Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Controlling politeness in neural machine translation via side constraints. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* 35–40.

[45] Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics.* 1715–1725. https://doi.org/10.18653/v1/P16-1162

[46] Gabriel Stanovsky, Noah A. Smith, and Luke Zettlemoyer. 2019. Evaluating gender bias in machine translation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.* 1679–1684. https://doi.org/10.18653/v1/P19-1164

[47] Amirhossein Tebbifakhr, Luisa Bentivogli, Matteo Negri, and Marco Turchi. 2019. Machine translation for machines: The sentiment classification use case. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP'19).* 1368–1374. https://doi.org/10.18653/v1/D19-1140

[48] Eva Vanmassenhove, Christian Hardmeier, and Andy Way. 2018. Getting gender right in neural machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 3003–3008. https://doi.org/10.18653/v1/D18-1334

[49] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the Conference on Advances in Neural Information Processing Systems*. 5998–6008.

[50] Ronald J. Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Mach. Learn.* 8, 3–4 (May 1992), 229–256. https://doi.org/10.1007/BF00992696

[51] Shuly Wintner, Shachar Mirkin, Lucia Specia, Ella Rabinovich, and Raj Nath Patel. 2017. Personalized machine translation: Preserving original author traits. In *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL'17)*. 1074–1084.

[52] Lijun Wu, Fei Tian, Tao Qin, Jianhuang Lai, and Tie-Yan Liu. 2018. A study of reinforcement learning for neural machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 3612–3621. https://doi.org/10.18653/v1/D18-1397

[53] Jingjing Xu, Xu Sun, Qi Zeng, Xiaodong Zhang, Xuancheng Ren, Houfeng Wang, and Wenjie Li. 2018. Unpaired sentiment-to-sentiment translation: A cycled reinforcement learning approach. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. 979–988. https://doi.org/10.18653/v1/P18-1090

Received July 2020; revised November 2020; accepted February 2021