Journal of Intelligent & Fuzzy Systems xx (20xx) x–xx
DOI:10.3233/JIFS-191690
IOS Press



# Topic modeling in short-text using non-negative matrix factorization based on deep reinforcement learning

Zeinab Shahbazi and Yung-Cheol Byun*
*Department of Computer Engineering, Jeju National University, Jejusi, Jeju Special Self-Governing Provience, Korea*



**Abstract**. Topic modeling for short texts is a challenging and interesting problem in the machine learning and knowledge discovery domains. Nowadays, millions of documents published on the internet from various sources. Internet websites are full of various topics and information, but there is a lot of similarity between topics, contents, and total quality of sources, which causes data repetition and gives the user the same information. Another issue is data sparsity and ambiguity because the length of the short text is limited, which causes unsatisfactory results and give irrelevant results to end-users. All these mentioned issues in short texts made an interesting topic for researchers to use machine learning and knowledge discovery techniques to discover underlying topics from a massive amount of data. In this paper, we propose a combination of deep reinforcement learning (RL) and semantics-assisted non-negative matrix factorization model to extract meaningful and underlying topics from short document contents. The main objective of this work is to reduce the problem of repetitive information and data sparsity in short texts to help the users to get meaningful and relevant contents. Furthermore, our propose model reviews an issue of the Seq2Seq approach based on the reinforcement learning perspective and provides a combination of reinforcement learning and SeaNMF formulation using the block coordinate descent algorithm. Moreover, we compare different real-world datasets by using numerical calculation and present a couple of state-of-art models to get better performance on short text document topic modeling. Based on experimental results and comparative analysis, our propose model outperforms the state of art techniques in terms of short document topic modeling.

Keywords: Topic modeling, knowledge discovery, short text, non-negative matrix factorization, machine learning

## 1. Introduction

In the past two decades, topic modeling for short types of text documents becomes an effective area in machine learning and natural language processing. Every day a major quantity of short text documents distributed in social media and another source, e.g., emails, search queries, image tags, advertisements, headlines, status messages, etc. These sources provide valuable data to extract the user's topic interests and to drive better decisions. Hence, the automatic identification and extraction of valuable topics from short text data is a fundamental problem in a wide range of applications, such as context analysis [1], and text mining and classification [2]. Nowadays, knowledge discovery becomes a fundamental and challenging task that achieves a lot of attention from researches all over the world [1–5]. The length of the short text is limited, which causes data sparsity and ambiguity. The sparsity and noisy data in short text documents are major problems in knowledge discovery from a large corpus and make the process of analyzing underlying knowledge difficult. There are

*Corresponding author. Yung-Cheol Byun, Department of Computer Engineering, Jeju National University, Jejusi 63243, Jeju Special Self-Governing Provience, Korea. E-mail: Byunyung cheoll@gmail.com.

ISSN 1064-1246/20/$35.00 

different topic identification techniques available to uncover and unearth content from short texts.

Topic identification techniques are used to discover semantically meaningful and valuable topics from a massive amount of data [6–8]. These techniques are capable of identifying underlying hidden patterns (topics) that explain and identify the similarity between documents. Another aspect of topic modeling is to consider purification of the document clustering by unsupervised machine learning strategy. This process is opposite to document clustering, in topic modeling procedure, many topics can occur in the individual document, but frequent topics among the document have more process in the training set. Frequently, topic modeling divided into two groups, i.e., the first group known as non-negative matrix factorization (NMF) [9], and the second group known as latent Dirichlet allocation (LDA) [6]. NMF topic model is a matrix base model that immediately reads the topics by break-down the document terms into a matrix that presents bag-of-words into text corpus with low-rank matrix factors. A low-rank matrix is a machine learning algorithm technique and a way to understand the recommendation system process. e.g., it means the given matrix $Z$ where $Z_i$, $X$ demonstrates the score which teacher $A$ gives to student $B$. LDA is a statistical measure for identifying and extracting hidden topics that occur in a set of documents.

To show the prominent performance of NMF-based model, clustering and dimension reduction techniques are used to get a high-dimensional dataset [10–12]. Regular documents primarily consist of consistent document structure and compact data contents. This document class exhibits the XML view of relational data. The data range of those documents is in the range of within 40 and 60 %. Irregular documents mainly consist of materials that have intense, complicated, and improper structures. This document class primarily focuses on the evaluation of the capableness of compressing improper structural data of XML documents. In the case of the regular type of documents, the conventional topic models come with great achievement, but this process in case of short text document collection is not working well. The reason for this problem getting high dimension in a short document set is because of not having enough information in a short type of text, extracting meaningful keywords is very difficult to extract [1, 3]. There are various methods presented to investigate topics modeling from short text messages in the recent past to overcome this challenge, and the current methods are used to combine short text documents into the pseudotype of the document and detect words transpire [3, 4, 13, 14]. But the topics generated by using this strategy is biased based on pseudo-texts, which cause exploratory.

In particular, the probability combination of most of the incoherent short texts into pseudo-texts is high. Another type of challenge in short texts is the lack of words or the limited length of the text. The internal semantic relationship between words proposed to extract information through short text documents, which based on this strategy, the problem of lack of words in the text is quite settled. The main reason to propose this strategy is the fact of semantic information of words in document which efficiently obtain along with neural network system, which rely on word embedding techniques, e.g., word2vec [15] and Glove [16].

Different word embedding techniques are presented [17–19] to discovering and underlying hidden topics from short texts based on the usage of words from various sources, e.g., word embedding techniques depends on Google News and Wikipedia. Based on word semantic differences between Wikipedia and short text documents causes noise in topics. Therefore, word embedding can be an effective process in modeling short text topics because it contains a huge number of words with similar semantic properties, which is very efficient in achieving better performance to cluster topic models. Another way of raising the performance of topic models is skip-gram with negative sampling (SGNS). This methodology is useful to obtain the relevance between words using a small type of sliding window approach [15, 20]. The SGNS is very useful because each window can present an individual document. In that case, word-context semantic relationships will be taken by SGNS inefficient way. The relationship between words can behold in another form of a word occurring.

This system potentially dominates the difficulty of data sparsity. The existing studies [21, 22], present that the SGNS algorithm is equal to factorize the relationship between matrix terms. Afterward, we enhance some questions about this technique: 1) Is it possible to transform the problem of matrix factorization to a non-negative matrix factorization problem? 2) Is there any way to combine these outputs into the conventional NMF for the term-document matrix? 3) Is this model suitable to detect underlying topics form short texts?

In this paper, we propose a novel semantic-assisted model based on RL and SeaNMF for short-text topic

modeling, which is used to extract topics from a large corpus. This work uses and analyzes four different real-world datasets of short text documents, such as Article headlines, Question Answers, Microblogs, and News. The following NMF-based topic modeling technique is used to discover underlying topics from a large volume of data. We analyzed and investigated the semantic relationship between keywords and document information using skip-gram, which shows the efficiency of disclosing the semantic relation of words. The reinforcement learning system applied to get the reward function from the algorithm in training set for input-specific RL and compare it in the test set with specific RL. Experimental results show the significance of the proposed model as compared with existing methods, e.g., document classification and topic coherence accuracy. Finally, we performed a comparative analysis of the proposed model with state of the art techniques. By comparing different topics keywords and analyze their network, the proposed system shows that the extracted topics from the benchmark datasets are meaningful, and extracted keywords using this model are more semantically associated. Therefore, the proposed RL+SeaNMF is an effective topic model for short texts.

The remaining of this paper is categorized into five sections, which are the following: Section 2 explains the related work about topic modeling in short texts based on the majority of state-of-the-art topic models. Section 3 presents the detailed system methodology, design architecture, and overall transaction process of the proposed system. Section 4, elaborates on the implementations and results of the proposed system, and section 5 concludes the paper.

## 2. Literature review

In this section, we will enlighten on some of the existing studies based on topic modeling. This section is categorized into three-part, i.e., topic modeling for short text, topic modeling for short text using Non-negative Matrix Factorization, and sequence to sequence for topic modeling (Seq2Seq). At the end of this section, we will discuss the novelty of our proposed model.

### *2.1. Topic modeling for short text*

Traditionally topic modeling algorithms, e.g., LDA and PLSA, are used to find the hidden and meaningful topic from a set of document corpus using a pattern like a word co-occurrence. These approaches have been used in various tasks and work well for normal text. Nevertheless, there are many shortcomings in traditional topic models approaches in terms of data sparsity, and it occurred when the size of the document is too short. Many studies are focus on and improve the quality of finding a topic in the short text to overcome the problem of data sparsity. For instance, in [23] author proposed a system that is used to find the hidden topic from the large external data source, e.g., Wikipedia. The presented framework aim is to find hidden patterns from the spare and short text and web segments. Similarly, in [24], Jin et al. presented an approach to cluster short text in order to find the topically related text from long auxiliary text data. The proposed Dual latent Dirichlet Allocation (DLDA) approach combined the topics parameters from the long text and short text and removed the data inconsistency between the datasets. These studies use a large set of data in order or improve the quality of topic representation. Numerous ingenious approaches extract latent topics from topic modeling using a large set of data by integrating a short text document. For example, in the study [25], use an LDA model to produce short texts by the similar user. The main aim of this study is to find influential users from social networking and blogging sites, e.g., twitter. Likewise, Davison and Hong et al. [3] develop two separate aggregated short text approaches that contain each word in the large set of data vocabulary, including text authors. In [26], author design, a scheme to produce pseudo-documents for LDA data preprocessing using tweet pooling. In another study [27] presented an approach based on word co-occurrence named WNTM to improve the sparsity and imbalance issues to generate pseudo-documents in word network. The develop scheme greatly improved the semantic compactness of data without introducing space and time complexity. Author Zuo et al. [14] also develop a scheme to generate pseudo-documents using the topic model. The presented approach uses the short text by influencing fewer pseudo documents to conjecture the latent pseudo documents. Kou et al. [28] presented a multifeatured probabilistic graphical approach to combined short text into long text in order to improve the search task in social networking sites. The develop MFPGM is also used to overcome the problem of semantic sparsity. Also, Jin et al. [24] propose an approach that is used to generate the pseudo documents from the short text that contains URLs. Afterward the short text is used for combining auxiliary information, e.g., named entities, hash tags, and

locations etc. [1, 3, 25, 29]. Kou et al. [30] analyzed a temporal and spatial characteristic which lead to develop an approach named as social network short text semantic modeling (STTM). The STTM is used to resolve the problem of semantic sparsity for short text in social network and improve the temporal and spatial characteristic.

Nowadays, many researcher and scientist used word co-occurrence information in order to improve the topic modeling techniques for short text. For instance, Yan et al. [5] design an approach used to generate word pair co-occurrence in the document. The develop technique biterm topic model (BTM) collectively used information of large set of data biterms which further used in for topic distribution. Moreover the biterms is also used to resolve the problem of document sparsity, but not consider the order of words in the algorithm processing. Lin et al. [31] proposed a dual-sparse topic scheme to investigate the problem of data sparsity in word and topic within the document. The design approach used to decouple the sparsity and improve the document in term of topic-word and document-topic distribution. Similarly, Quan et al. [13] presented self-aggregated approach for shot text. The designed self-aggregated based topic model (SATM) generate pseudo-document from the short text snippet sampled. The generated topic and the process of aggeneration is process using reinforcement techniques so that the aggregation is build based in topical affinity of texts. Nevertheless, the size of data will also increase the number of parameters for configuring a suitable extensive pseudo document. Temporarily inference procedure containing both the topic sampling and text aggregation consuming a large amount of time during processing. Bicalho et al. [32] highlighted a general approach for topic modeling by generating a corpus pseudo documents presentation from the actual documents. The proposed scheme used vector representation and word co-occurrence to create pseudo-documents. Afterward different topic modeling techniques are applied on pseudo-document for further analysis.

### 2.2. Non-negative matrix factorization for topic modeling

During the past few year researcher and data scientist have gained a lot of attention in the field of topic modeling. Different techniques like Non-Negative Matrix Factorization (NMF) are applied on topic modeling in order to improve the knowledge extraction process, data sparsity and repetition [10, 33, 34]. Although only few of studies focus on topic modeling for short text using NMF approach. For instance, Yan et al. [35] present asymmetric term correlation matrix based on NMF to extract topics from short text documents. The proposed system uses a quartic non-convex based loss function in its processing; however, the designed system is failed to work in terms of stability and reliability. To dominate this problem, they present the Sym-NMF model in [15, 17], but this model doesn't supply a good insight in topic modeling, and furthermore, we can't use this model to get document representation immediately. Similarly, in NMF, latent semantic space obtains the topics in a specific document cluster, and each text document proposes merged topics in document [9]. Automatic text segmentation plays an important role in this area. It helps to make instructive information based on document summaries that are the significant ideas of the original text documents. The main issue in automatic text summarization is how to calculate and extract the right information from the document. Tian et al. [36] proposed a SeaNMF approach integrated with local word-context correlation which is used to find topics for the short texts. The design approach use the block coordinate decent technique to calculate the proposed SeaNMF approach. Moreover, the develop system also equipped with sparse SeaNMF in order to achieve high interpretability.

### 2.3. Sequence to sequence for topic modeling

Seq2Seq is one of approach which is widely used in natural language processing for different purposes like machine translation [37–41], headline generation [41–43], text summarization [44, 45] and speech recognition [46–48]. Furthermore the Seq2Seq approach is divided into two parts which takes input and output of the document. The input section is a specified data sequence, and the output section is a sequence of the data module. Generally, Seq2Seq model training is based on the ground-truth sequence using the teacher-forcing technique where ground-truth is act as a teacher [49]. During the past few years many studies have been published by using Seq2Seq approach in order to enhance the complexity of topic modeling for short texts [50]. Rush et al. [39], seq2seq model based on encoder attention and neural network language model (NNLM) decoder based on the neural machine translation (NMT) for text summarization. Specifying document topics is one of the important step of document readability

Table 1
Critical analysis of sequence to sequence topic modeling

| Authors | Propose | Issues |
|---|---|---|
| Wei Li, Andrew McCallum [51] | Detecting word clusters which contain syntactic classes and semantic topics. | Task labeling due to words indeterminacy |
| Ilya Sutskever, Oriol Vinyals, Quoc V. Le [52] | Manufacture sequence learning approach to structure minimal hypothesis of the process | Sentence optimization problem between sentence and data origin |
| Martin Riedl, Chris Biemann [53] | LDA and TopicTiling optimization to improve the accuracy of segmentation | Segment numbers are not investigated. |
| Martin Riedl, Chris Biemann [54] | Improving the semantic information of topic models to get better performance in wordbased and TextTiling algorithms. | Sparsity issue which require different ways to smooth the data |
| Dan Fisher, Mark Kozdoba [55] | Using Full Dependence Mixture (FDM) to combine the general moment and second moment of data. | Hyper parameter doesn't contain the topic arrangement |

which usually comprised of three main approaches, i.e., `|total words|, |hiragana ratio|, and |word length|`.

Table 1 shows the critical analysis of state-of-art techniques based on short text document to develop the probabilistic method to paradigms for realizing short documents.

As stated above, these existing models are not adequately developed for topic modeling from short text documents and also have some overcoming in terms of lack of information, such as data sparsity and ambiguity. To the best knowledge of the author, there has been no semantic-based model to extract meaningful topics from short text documents using RL and SeaNMF techniques so far.

## 3. System architecture of proposed RL+SeaNMF model

The proposed paper comprised of three main modules, i.e., non-negative matrix factorization (NMF), Reinforcement learning, and sequence to sequence model (Seq2Seq). First, we explain the problem statement in sequence to sequence models and solutions based on our study. Second reinforcement learning steps in short text document topic generalization and some information about block coordinate decent methodology. Third, the NMF topic modeling system and SeaNMF model based on a block-coordinate algorithm to display the terms of short text documents. Figure 1 represents the proposed system architecture.

This figure propose the input data (context, word, and document) as $(X_i)$, $(J_i)$, $(Y_i)$. (G), $(R_c)$ and T is a vector representation of this proposed system, and each part of T represents a topic in the document.

The presented SeaNMF model [36] is able to extract the semantics out of short text data use as a basis of word-content, and word-document relationship, and the proposed objective function integrates the privilege of both the NMF model for topic modeling and the skip-gram model for obtaining word-context semantic relationships. To solve the problem of optimization, we used a block coordinate algorithm along with the SeaNMF model sparse version to realize further interpretability. Table 2 is used to summarizes frequently used notations

### 3.1. Sequence to sequence model (Seq2Seq)

Encoder decoder architecture applied in different types of natural language processing tasks, e.g., machine translation, dialogue generation, document summarization, and question answering. Seq2Seq model presented with various strategies, e.g., attention, copying, and coverage. The maximum result of the proposed cases focus on repetitive words and build a fixed-sized vocabulary based on words and tokens. Therefore, the efficiency of the model is based on vocabulary limitations. The Algorithm 1 describes a simple Seq2Seq model steps.

The Algorithm 1 consists of two phases, such as training and testing phases. In training phase, there are two groups of data, such as X and U. First, we run encoding on X to get the last encoded state $h_{T_e}$. After getting the last encoded state, we run the decod-

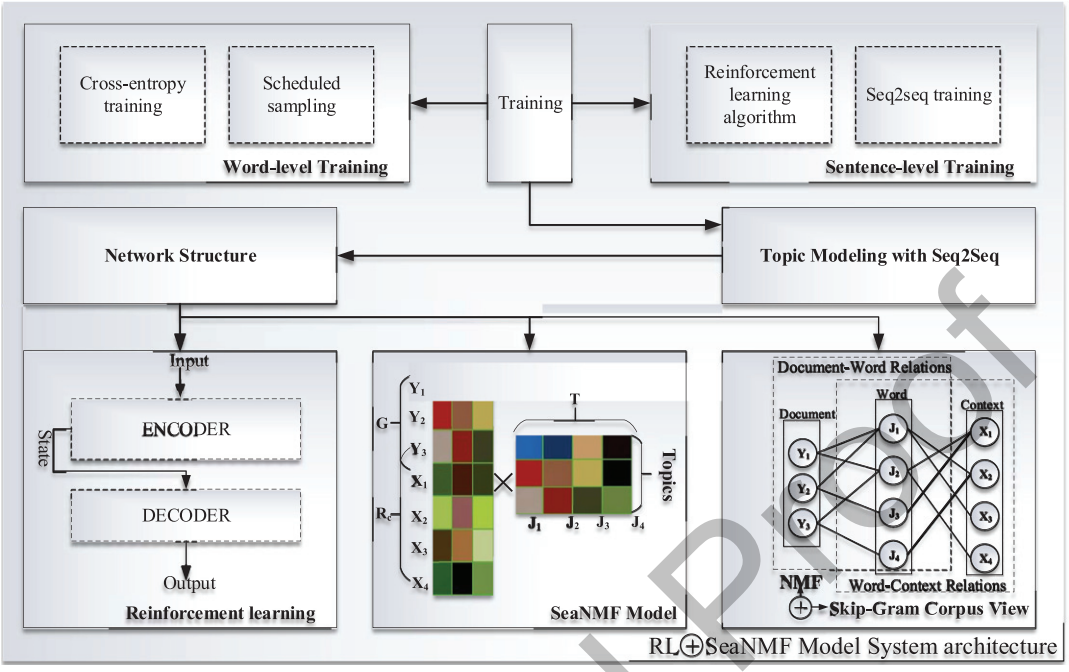


Fig. 1. Overview of the RL+ SeaNMF Model System architecture.

Table 2
Notation used in this paper

| Name | Description |
|---|---|
| Seq2Seq model parameters | |
| X | Length of input sequence $T_e$, X = $\{x_1, x_2, \ldots, x_{T_e}\}$ |
| Y | Ground-truth of output sequence T, U = $\{u_1, u_2, \ldots, u_T\}$ |
| $\hat{U}$ | Generated output length sequence T, $\hat{U} = \{\hat{u_1}, \hat{u_2}, \ldots, \hat{u_T}\}$ |
| $T_e$ | Encoder numbers of input sequence |
| T | Decoder number of output sequence |
| d | Input and output sequence representation size |
| B | shared vocabulary of input and output |
| $e_t$ | Hidden state encoder at time t. |
| $j_t$ | Hiden state decoder at time t. |
| $P_\theta$ | The seq2Seq model with parameter $\theta$. |
| Reinforcement Learning Parameters | |
| $r_t$ = r($s_t$, $u_t$) | The reward that the agent receives by taking action $u_t$ when the state of the environment is $j_t$. |
| $\hat{U}$ | Sets of actions that the agent is taking for a period of time T, $\hat{U} = \{\hat{u_1}, \hat{u_2}, \ldots, \hat{u_T}\}$ This is similar to the output that the Seq2Seq model is generating. |
| $\phi$ | The policy that the agent uses to take the action. |
| Q($j_t$, $u_t$) | The Q-value (under policy $\phi$) that shows the estimated reward of taking action $y_t$ when at state $j_t$. |
| SeaNMF Parameters | |
| B | word-document matrix. |
| J | semantic correlation matrix. |
| V | words Latent factor matrix. |
| $V_c$ | contexts Latent factor matrix. |
| E | Document Latent factor matrix. |
| $v_j$ | Words Vector representation $v_j$ . |
| $c_j$ | Context Vector representation $c_j$ . |
| R+ | Real number of Non-negative. |
| D | Document number in the corpus. |
| M | Distinct words Number in vocabulary. |

Algorithm 1
Seq2Seq model simple Training

**Input**: Input (X) as sequence and (U) as ground-truth
**Output**: Seq2Seq model Trained set
**Train Step**:
**for** <Group X and U> **do**
  Run encoding on X and get the last encoder state $h_{T_e}$.
  Run decoding by feeding $h_{T_e}$ to the first decoder and obtain the sampled output sequence $\hat{U}$
  Update the parameters of the model
**end for**
**Test Step:**
**for** <Group X and U> **do**
  sampling output based on training model $\hat{U}$
  Estimate the model by applying performance measure.
**end for**

ing process by feeding $h_{T_e}$ to the first decoder in order to get a sampled output sequence $\hat{U}$. Finally, we update the parameters to executes the next step. In the testing phase, the input samples are tested based on trained model $\hat{U}$. Finally, we evaluate the performance of the tested samples using different performance measures. The proposed Seq2Seq model in this paper is containing convolutional architecture, which is comprising words and topics, reinforcement learning system, and SeaNMF model. Ground-truth in this algorithm defined as $U$ which generates $\hat{U}$ as an output of model. In this paper, we used the following measures, such as CIDEr [56], BLEU [57], ROUGE [58], and METEOR, to evaluate the performance of the proposed model [59].

#### 3.1.1. Sequence to sequence model problem statement

The main problem of the current Seq2Seq model is, "cross-entropy minimization output," which has not to generate the acceptable result in running the network. Hence, applying cross-entropy (C) loss in the Seq2Seq procedure faces conformity in the operation process in training and testing set. Equation 1 defined to evaluate the loss (error) of output $\hat{u}_t$. Figure 2 describes the training process of the decoder, which has two inputs.

$$C_{IN} = -\sum_{t=1}^{T}(\log \pi_\theta(\hat{u}_t|\hat{u}_1 ... \hat{u}_{t-1}, j_t, j_{t-1}, X)) \quad (1)$$

Output state is defined as $j_{t-1}$, input of ground truth is defined as $u_t$. Current output state is defined as $j_t$ and next action calculation defined as $\hat{u}_t$. Although, in the testing set, all over the decoder process is based on generated action through the model distribution to anticipate the next action.

As shown in Fig. 2, the right side of the Figure is presenting the LSTM network, which is correlated with the encoder $e$ to show hidden state of encoder, and it has $T_e$ modules. $T_e$ is representing the length of the input sequence and the number of encoders. And the right side of the LSTM network is representing the correlate with decoder, and it has $T$ modules. $T$ is presenting the length of output sequence and number of decoders.

### 3.2. Reinforcement learning

Machine learning is an advance and robust area in artificial intelligence (AI), which Reinforcement learning is one of the tasks in this system. The reinforcement learning algorithm is evaluated based on Markov Decision Processes (MDPs). The agent-environment relationship in the reinforcement learning system is divided into three categories named "Agent," "Environment," and "State." The agent is described as software which produces high-order intentions, and it is also learner part of the

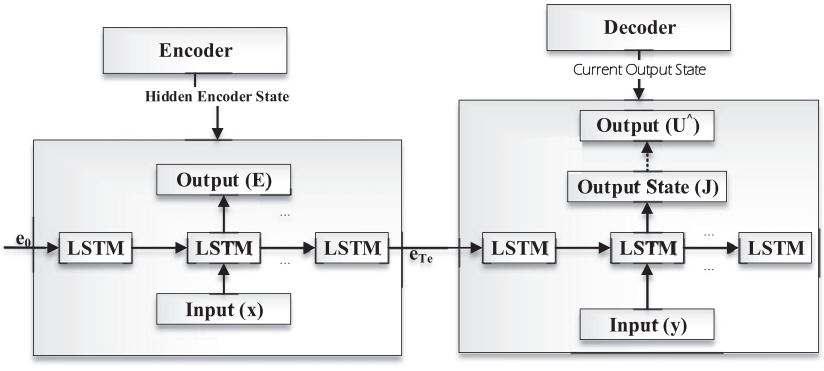


Fig. 2. Simple sequence to sequence model.

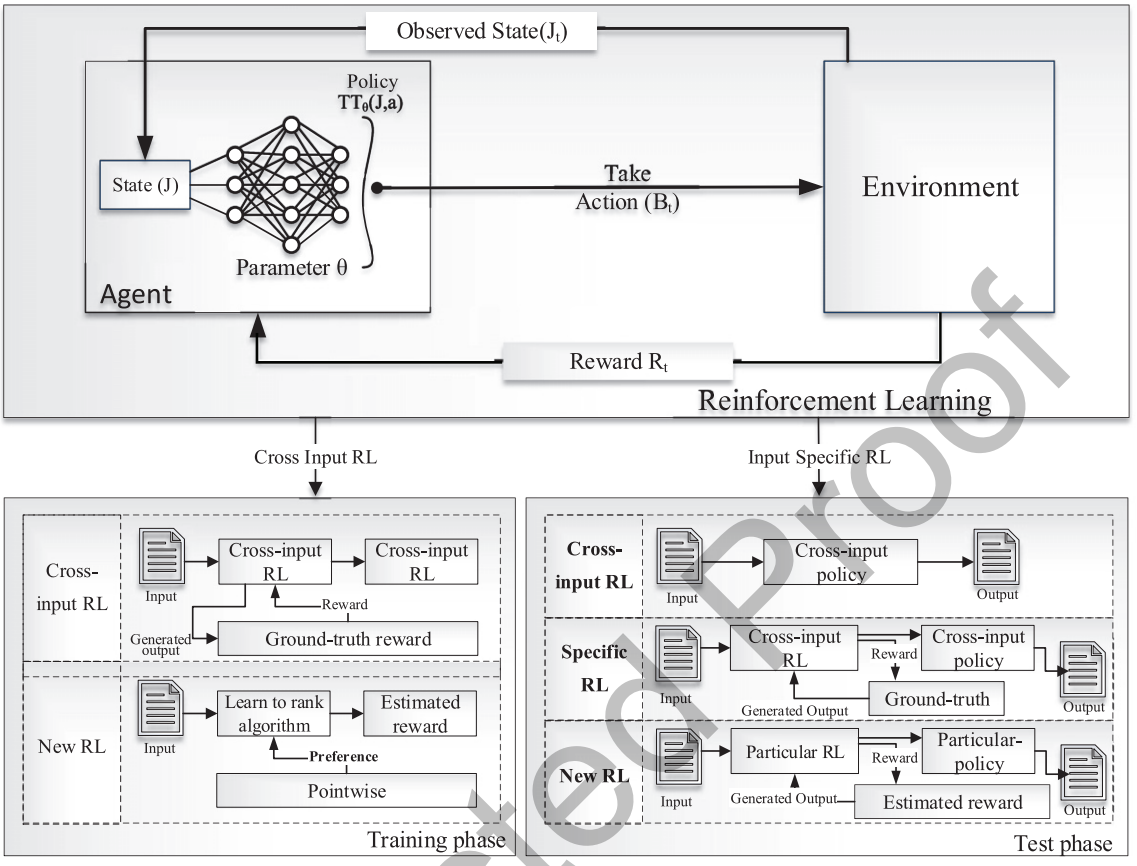


Fig. 3. Work flows of Reinforcement Learning model.

RL system. The environment represents the procedure problem, which should solve, make a real-world and also simulate the environment to communicate with the agent. The last category states which are the position of agent for a special pick of time in the environment. As a simple explanation, an agent performs an action the environment gives the agent reward and a new state where the agent reached by acting.

Markov Decision Processes is a tuple of (J, B, P, R, T), where J is a set of states, B is a set of actions. Transition function is defined as $P : J * B \rightarrow J$ where P(j, b) giving the next state after performing an action b in state j. The reward function is defined as $R : J * B \rightarrow R$ with R(j, b), which is giving the immediate reward for performing an action b in the state of j. Terminal states are given by $T \subseteq J$ that marks the end of an episode. Figure 3 shows RL total process in document similarity.

As presented in Fig. 3, the ground-truth of this architecture can supply by human or automatic metrics, which is generating similarity measurement between document output and reference output. Practically, the Seq2Seq model is applied to generate the function operation, and task-specification, e.g., text summarization is extracting the meaning of the function which denotes the next function summarization, while, in question answering the task, start and point of the action suppose to defined to get the document answer.

Through reinforcement learning, two output sequences generated with input sequence $x$. $\hat{u}$, which define as the first sequence capture by words that this process maximizes and sampling the output of probability distribution. Output of both $x$ and $\hat{u}$ sequences are a reward of the process, and it minimizes the loss of reinforcement based on Equation 2.

$$L_{RL} = -(r(u^j) - r(\hat{u})) \log p\theta(u^j) \qquad (2)$$

Equation 2 shows minimum reward ($r$) output of sequence $x$ and sequence $\hat{u}$.

#### 3.2.1. Reinforcement learning: Model-free solution technique

RL is essentially deal with how to capture the optimal policy when there is no accessible model. Through this issue, MDPs (Markov Decision Process) focus on estimation and defective information, which proposed earlier; it is a crucial problem statement in short text documents. To overcome this issue, "model-free" reinforcement learning presented, which is not depending on existing information and reward models. There is a lack of a model for this problem, which generates statistical knowledge to use the sample MDP model. Algorithm 2 presents the common RL based on online document resources.

Algorithm 2
General Reinforcement Learning algorithm for online short documents

**for** <Occurance> **do**
  $j \in J$ is marked as the start point.
  $t := 0$
  **while** select an action $b \in B_j$ **do**
    complete action $a$
    perceive the new state $\hat{s}$ and received reward $r$
    Update $\hat{T}$, $\hat{R}$, $\hat{Q}$ and $\hat{V}$
    Using experiment of $<j, b, r, \hat{j}>$
    Update the parameters of the model
    $j := \hat{j}$
  **end while** $<\hat{j}$ is a state goal>
**end for**

Algorithm 2 represents the short documents topic modeling issue by evaluating the minimum reward of the process.

### 3.3. SeaNMF model

SeaNMF model is a novel semantic-assisted NMF model which is proposed to unearth hidden topics from the short text document. This methodology combines the semantic information of document using word embedding dictionary in a training model, which this dictionary able to recover words relationship, hidden part of a document, etc.

#### 3.3.1. Model evaluation

One of the challenges in this paper is how to recommend words to the NMF system. Based on latent matrix $V \in (R_+^{M*F})$ (W components are all non-negative), non-negative limitations applied on words and vectors of document. Thus, $\overrightarrow{V} \in (R_+^{M*F})$ and $\overrightarrow{c} \in (R_+^{M*F})$ hold. To set the keyword $V_i \in V$ applied to the system based on $V_{i,:} = \overrightarrow{v_i}$. To detect the semantic information between words and document contents, $V_c(j,:) = \overrightarrow{c_j}$ for $c_j \in V$ matrix defined. To detect keywords and document content relationship (word-content) $J \approx VV_c^T$ matrix defined. The above equations defined as a matrix to show the relationship between words and contents. In this equations $V$ defined as latent factor matrix of words, $V_c$ defined as content matrix, $V_j$ represents vector representation of words and $c_j$ represents vector representation of contents.

#### 3.3.2. Optimization

Topic modeling on short text documents contains a different type of process that the one which used in this paper is, "each text is defined as a window in the skip-gram model."

- Skip-gram model is equal to the relative length of the text document.
- Number of short texts is equal to the number of windows.

Block coordinate descent algorithm used to define the term-document matrix $B$ for word representation based on bag-of-words and in next step evaluation of semantic correlation matrix $J$ presented for sampling the document contents. Equation 3 and 4 are representing the SeaNMF model procedure evaluation based on above mentioned explanation.

$$V_{(F)} \leftarrow \left[ V_F + \frac{(BE)_{(F)} + \alpha(JV_c)_{(F)}}{(E^T E)_{(F,F)} + \alpha(V_c^T V_c)_{(F,F)}} \right] - \left[ \frac{(VE^T E)_{(F)} - \alpha(V_T^c V_c)_{(F)}}{(E^T E)_{(F,F)} + \alpha(V_c^T V_c)_{(F,F)}} \right] \tag{3}$$

$$V_{c(F)} \leftarrow \left[ V_(cF) + \frac{(JV)_{(F)}}{(V^T V)_{(F,F)}} \right] - \left[ \frac{(V_c V^T V)_{(F)}}{(V^T V)_{(F,F)}} \right] \tag{4}$$

The block coordinate descent algorithm in the SeaNMF model is presented in Algorithm 3.

#### 3.3.3. Intuitive evaluation

Furthermore about SeaNMF model, Equation 5 and 6 processed from Equation 3 and 4 to represent the update procedure of topic modeling.

$$V_{(F)}^1 \leftarrow \left[ V_F + \frac{(BE)_{(F)}}{(E^T E)_{(F,F)}} - \frac{(VE^T E)_{(F)}}{(E^T E)_{(F,F)}} \right] \tag{5}$$

Algorithm 3
SeaNMF Algorithm for short text documents

**Input**: Matrix $B$ term-document

Matrix J semantic correlation;
Number of topics $F, \alpha$;
**Output**: $V$, $V_c$, $E$
**Initialize**: $V \geq 0$, $V_c \geq 0$, $E \geq 0$ random real numbers;
$t = 1$;

**while** <repeat> **do**
  **for** $<k = 1, K>$ **do**
    Compute $V^t_{(:,k)}$ by Equation 3
    Compute $V^t_{c(:,k)}$ by Equation 4
**end for**
$t = t + 1$
**end while** <Close>

$$V^2_{(F)} \leftarrow \left[ V_F + \frac{(JV_c)_{(F)}}{(V_c^T V_c)_{(F,F)}} - \frac{(VV_c^T V_c)_{(F)}}{(V_c^T V_c)_{(F,F)}} \right] \quad (6)$$

Equation 5 is proposed for standard NMF. It predicts the terms in short content into the same category using the term-document matrix. Furthermore, in Equation 6, words that have a similar meaning or hidden words with the same significance, share the common context keywords. Hence, using this model increase the topic's connections. e.g., in Fig. 1 words ($J_1 and J_4$) are in separate documents, but their connection is in $J_2$ based on the same keyword in both document contents. A simple explanation of SeaNMF model in short text is as below:

"Skip-gram model is equal to the relative length of the text document. e.g., "Harvard university of engineering" and "Oxford university of science" are two samples of short text documents. "Harvard" and "Oxford" are not occurring in the same position in the second sentence and also "engineering" and "science" are not occurring in the first sentence but the relationship between "Harvard, university" and "Oxford, university" is the meaning of standard NMF model."

Although, based on Equation 6, both sentences have a shared word "university," which is the meaning of the SeaNMF model. This model is to find the similarity between words and also hidden parts of document and relationship between vectors in short text documents to extract meaningful and valuable hidden topics, which is the main issue of this type of content because of lack of information.

### *3.4. Similarity measurement*

Similarity measurement is evaluating based on the word probability distribution and word relationships. Although topic similarity can be calculated based on word occurrence and average PMI. Similarity measurement divided into two categories named "Distribution of topic words" and "Semantic space of topic models."

Based on the distribution of topic words, Every topic is generated from a Dirichlet distribution of $V$ dimensions where $V$ is the size of a word. For every document, Generate a distribution over topics from a Dirichlet distribution of $T$ dimensions where $T$ is the number of topics in the corpus and For every word in the document, choose a topic according to the distribution generated and Choose a word according to the distribution corresponding to the chosen topic. So, each topic is a probability distribution over the words of the vocabulary.

Semantic space is representing by the topic model, which is applying to propose topics and words. Moreover, each topic is a probability of the words in the training set. Each corpus with a combination of words and documents shows the relationship of them in the topic modeling procedure.

## 4. Implementation and experimental result

### *4.1. Experimental setting*

In this section, we evaluate the performance of our model by substantial experiments on different short document datasets. Finally, Dataset information, metrics, baselines estimation, and present various sets of results. Although, comparison between proposed models and latent Dirichlet allocation (LDA), Non-negative Matrix Factorization (NMF), Pseudo-document-based Topic Model (PTM), and Dirichlet Multinomial Mixture model (GPUDMM) is presented.

### *4.2. Development and simulation environment*

The development environment of the proposed system is summarized in Table 3. All experiments and results of the system are carried out using Intel(R) Core(TM) i7-8700 CPU @3.20GHz 3.19 GHz processor with 32 GB memory. Moreover, Win-Python (v3.6.2), based on Jupyter notebook, is a prerequisite to develop the topic modeling environment, configure the model, and get the result detail information. Topic modeling on short text documents evaluates based on a combination of three different models in this area named "Sequence to sequence," "Reinforce-

Table 3
Development environment of proposed short text document topic modeling

| Component | Description |
|---|---|
| Programming language | WinPython–3.6.2 |
| Operating system | Windows 10 64bit |
| Browser | Google Chrome, opera |
| Library and framework | Jupyter notebook |
| IDE | Anaconda 2019 Release 3 |
| CPU | Intel(R) Core(TM) i7-8700 CPU @3.20GHz 3.19 GHz |
| Memory | 32GB |
| Topic modeling algorithm | Seq2Seq, Reinforcement Learning, SeaNMF |
| Distribution Modeling Algorithm | SeaNMF |
| Reinforcement Learning | Model Free |
| Seq2Seq | Encoder and Decoder |
| SeaNMF | Novel semantic-assisted NMF matrix |

ment Learning," and "SeaNMF model." Sequence to sequence model evaluated based on encoder and decoder architecture to solve the problem of cross-entropy in short texts. The reinforcement learning model evaluated based on a model-free technique to capture the optimal policy on short content, and finally, the SeaNMF matrix-based model evaluate extracted topics from limited information contents.

### 4.3. Dataset

The proposed system contains four real-world short text document datasets correlated with related applications, i.e., Article headlines, Question Answers, Microblogs and News. These categories divided into seven types of datasets that explain in detail below:

Tweet: This type of data is a good resource for short text document topic modeling, which accumulates and labeled by [60]. Out of a large amount of tweet data categories, 15 categories selected to apply in the proposed process, i.e., World, Regional, Art, Science, Health, News, Shopping, Reference, Society, Recreation, Computers, Business, Sport, Game, and Home. Each category contains almost 3000 separate tweets with a minimum of two keywords. Second is Collected GoogleNews dataset from the GitHub website. This type of short document contains 3 million English words based on 300-dimensional latent embedding applying the word2vec model. This data type used for the comparison method with GPUDMM [17]. Third is Yahoo.Ans which proposed dataset extracted from Yahoo Answers Manner Questions, version 2.04. Collected data is from 10 various categories related to Question subjects, and it also contains Financial Service, Diet, Fitness, etc. Forth is Tag.News that selected dataset is containing part of the TagMyNews dataset that includes tweets, news, and snippets. It divided into seven categories named by "Entertainment", "Health", "Sci and Tech", "Sport", "Business", "world" and "US". Each category contains a minimum of 25 keywords. Fifth is DBLP which is the computer science bibliography website. A collected dataset from this website is divided into four categories named "Database," "Machine Learning," "Information Retrieval," and "Data mining." Each category contains the title of conference papers.Sixth is Yahoo.CA which Proposed dataset extracted from Yahoo Answers Manner Questions, version 2. It contains Questions and a more suitable answer to that question. And finally ACM.IS data which is collected from Scientific Literature Dataset Data-verse Harvard university. it contains the published paper's information. Dataset information is summarized in Table 4.

### 4.4. Metrics estimation

To accurate the proposed methodology, Topic coherent and Document classification procedure applied in the topic modeling system. More detail information about each part is explained as below:

#### 4.4.1. Topic coherent

To evaluate the metrics in this process "point mutual information (PMI)" score used, which is to evaluate information theory and show all possible events occurring in text data. Given Equation 7 is used to calculate Topic coherent.

$$C_K = \frac{2}{M(M-1)} \sum_{1<=i<j<=M} \log \frac{p(v_i v_j)}{p(v_i v_j)} \quad (7)$$

Each topic is defined as $K$, and the Number of probability between words in a document is defined as $M$. $p(v_i v_j)$ is presenting the words occurring together in a similar document. To measure out the topic quality, the PMI score applied to the topic modeling process. Based on results of PMI in regular-sized text contents which it works well in that type, still, there are issues in short type of document, which means gold-standard topic may be assigned with a low PMI score.

To dominate this issue, First, we evaluate PMI in four categories of short text data. Next, in two other

Table 4
Dataset statistics

| Data Type | Documents | Terms | Aggregation (A) | Aggregation (B) | Document Length | Categories |
|---|---|---|---|---|---|---|
| Tweet | 54524 | 21381 | 1.3855% | 1.1824% | 8.84 | 15 |
| Yahoo.Ans | 51865 | 5445 | 1.2118% | 1.1184% | 5.41 | 10 |
| Tag.News | 39769 | 22636 | 2.3972% | 1.2471% | 29.25 | 7 |
| DBLP | 26112 | 3558 | 1.8714% | 1.3788% | 7.75 | 4 |
| Yahoo.CA | 41797 | 5445 | 6.1642% | 1.8865% | 53.72 | – |
| ACM.IS | 47413 | 3558 | 5.3778% | 2.1515% | 88.51 | – |

datasets (Yahoo.Ans and DBLP), which are arranged documents, based on an external corpus, PMI calculated. The comparison output of these datasets presents the effectiveness of the proposed methodology.

#### *4.4.2. Document classification*

Document classification is the further efficient topic modeling process used in the proposed system. In this process, classification is just used for a labeled dataset. To calculate the classification performance, five-fold validation applied to the process and, finally, the LIBLINEAR package used to figure out the quality of classification. F-measure ("Precision" and "Recall") used to qualify the classification performance.

### *4.5. Analysis methods*

To show the effectiveness of the presented method, we compare the efficiency of our method with other state-of-art techniques.

- Latent Dirichlet Allocation (LDA): LDA is a probabilistic-based topic modeling technique used to extract the topics from the rich source of information. LDA is used to extract topics from contents based on document keywords. Extracted words are the most relevant keywords which occurs in document related to special topics mentioned in content. There are different open-source libraries are available in python, such as Gensim [61], but it does not include NMF. LDA uses a probabilistic mechanism while NMF relies on linear algebra.
- Non-negative Matrix Factorization (NMF): NMF is an unsupervised learning task which used for reducing dimension and simulating clustering. In this paper, NMF is used to implement the block coordinate algorithm. NMF is a linear-algebraic model, that is used to reduce the high dimensional vectors into a low dimensional.

Table 5
Topic coherent outputs based on PMI

| | Tweet | Tag.News | DBLP | Yahoo.ANS |
|---|---|---|---|---|
| Latent Dirichlet Allocation | 1.3168 | 1.6159 | 0.1457 | 1.3168 |
| Non-negative Matrix Factorization | 1.9156 | 1.7525 | 0.1295 | 1.2415 |
| Pseudo-document-based Topic Model | 1.4856 | 1.7739 | 0.9616 | 1.2422 |
| GPUDMM | 0.1324 | 0.1862 | 0.3926 | 0.6819 |
| SeaNMF | 4.2588 | 3.7429 | 1.7248 | 1.8664 |
| Sparse SeaNMF | 4.2181 | 3.7164 | 1.7341 | 1.7192 |

Table 6
Topic coherent outputs based on Yahoo.CA and ACM.IS

| | Yahoo.ANS/ Yahoo.CA | DBLP/ ACM.IS |
|---|---|---|
| Latent Dirichlet Allocation | 0.7651 | 0.5393 |
| Non-negative Matrix Factorization | 0.6372 | 0.4737 |
| Pseudo-document-based Topic Model | 0.7615 | 0.5542 |
| GPUDMM | 0.4413 | -0.1261 |
| SeaNMF | 1.2115 | 0.7752 |
| Sparse SeaNMF | 1.1299 | 0.7558 |

- Pseudo-document-based Topic Model (PTM): PTM is to propose pseudo-documents to topic models in short text documents to extract the document topic.
- GPUDMM: Applied to extract topics based on the Dirichlet Multinomial Mixture model.

### *4.6. Experimental results*

#### *4.6.1. Topic coherent output results*

Table 5 represents the topic of coherent outputs via comparison with various methods and effectiveness of the SeaNMF model based on standard NMF in short document contents topic modeling. The comparison between LDA and PTM also shows the considerable prove that mentions the SeaNMF model can extract more coherent topics out of a document.

Table 6 represents the topic of coherent outputs via a comparison between Yahoo.CA and ACM.IS

dataset. Based on the mentioned reasons for PMI in topic coherent section, we evaluate topic coherent on external corpus because it has better performance on short text documents. The output result of coherent topics is presented in Table 6. Going deep into word correlation process, SeaNMF obtains more suitable topics out of short text documents.

As explained in Tables 5 and 6 by visualizing and comparing the words relationship , the main information of contents effects in extracting topics using Sparse SeaNMF model.

#### *4.6.2. Document classification output results*

Furthermore, Metrics estimation presented into two parts, which document classification is part of the proposed system. Table 7 shows the best results of our approach in short document dataset ("Tweets", "Yahoo.ANS", "Tag.News"). These outputs display that the proposed system is efficient in short content. Result comparison with LDA and NMF shows the presented SeaNMF model has better classification and also better performance results than other systems. The presented model output is summarized in Fig. 4.

A comparison between these systems represents that skip-gram has an effective role in the high performance of semantic documents, and also, the NMF model is not good enough comparing with LDA.

Another effective model in this process is GPUDMM, which accomplishes better than other baselines. The number of Tweets in various categories is almost the same and is more dependable. Although, the Tweet dataset passes the imbalanced class" issue. Table 7 shows the development of the SeaNMF model, which on average, 14 percent higher performance in overall the baselines which presents by precision (P), recall (R), F-measure (F).

It is observed that the better performance of the topic coherence does not indicate a better performance of document classification. The following Fig. 4 presents the performance of topic coherence and classification of documents of the SeaNMF model. The parameter alpha indicates the weighting factor, which used for factorizing the correlation matrix of semantic words. $\gamma$ is used as a smoothing factor, which used for the probability of sampling. It is evident in Fig. 4 that the F-score decreases as $\alpha$ increases. It is also noticed that there is no significant change in F-score with respect to $\alpha$ value. Therefore, our proposed SeaNMF-based model is stable for modeling topics for short text documents. The parameter f also plays a vital role in constructing a correlation matrix of semantic words. The parameter f is used to effects the data sparsity of a correlation matrix, which causes that the semantic words are less correlated. It is also observed from the graph that the value of the F-score is reduced when the parameter f increase. It is found that the F-score slightly increases and improves when a smoothing factor Y is increased. Hence, in the collection of short documents, there is a difference between a high coherent topic and a high-quality topic.

### *4.7. Combination of Seq2Seq and reinforcement learning model*

This section is presented the combination Seq2Seq model with reinforcement learning in every possible technique. The main goal of using these models is to overcome with the mismatching problem in document keywords and topics. Generally, this system is works base on the reward function for running object function, which explained earlier in the reinforcement learning section. The reinforcement learning algorithm is an effective system for improving the output of state-of-art in the Seq2Seq system. Although, there are a lot of progressive techniques, e.g. DQN, Actor-Critic models, and DDQN, which didn't apply to this task before. The main issue in this system is the problem of using Q-Learning and formative part of that in the Seq2Seq model. Text summarization can be an example of this problem. This system should evaluate words in each vocabulary even with a lower status of training. Related to this reason, the most focus on this area is a problematic approach, which is the REINFORCE algorithm that is a suitable task to train the Seq2Seq model. Table 8 presents "policy", "action", and "reward" function in Seq2Seq model.

### *4.8. Topics semantic analysis*

Generally, topic semantic analysis is to represent the effectiveness of the SeaNMF model based on extracting meaningful keywords. To discover topics out of short text document, specified top keywords based on standard NMF model compare with SeaNMF system output. Represented dataset extracted words depending on the word embedding system and describe the problem statement of topic extraction in short contents. Output covers the set of contents based on word similarity. The topic semantic analysis procedure is to make a group of sentences by using coherent topics which allocated to individ-

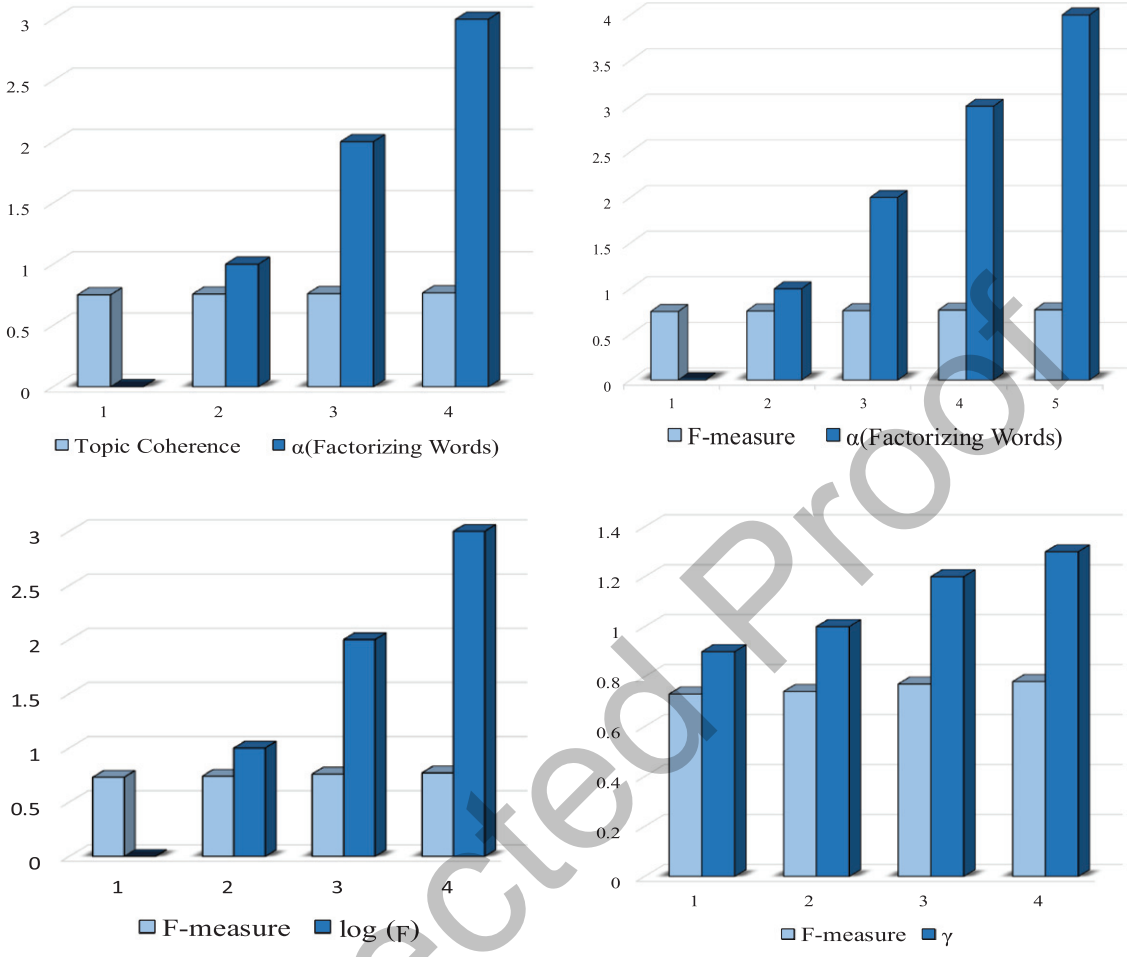


Fig. 4. Topic coherence and classification performance of SeaNMF model.

Table 7
Comparison method on various document classification

| | Tweet | | | Tag.News | | | Yahoo.ANS | | | DBLP | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P | R | F | P | R | F | P | R | F | P | R | F |
| LDA | 0.4938 | 0.4978 | 0.4957 | 0.8434 | 0.8295 | 0.8363 | 0.6131 | 0.6849 | 0.6469 | 0.7192 | 0.6184 | 0.6648 |
| NMF | 0.4788 | 0.4628 | 0.4704 | 0.7874 | 0.7482 | 0.7672 | 0.7414 | 0.6581 | 0.6972 | 0.7414 | 0.7337 | 0.7374 |
| PTM | 0.4152 | 0.4949 | 0.4524 | 0.8636 | 0.8417 | 0.8524 | 0.7411 | 0.7149 | 0.7277 | 0.7535 | 0.7478 | 0.7505 |
| GPUDMM | 0.4194 | 0.4177 | 0.4183 | 0.8954 | 0.8823 | 0.8887 | 0.6165 | 0.7419 | 0.6732 | 0.7780 | 0.7684 | 0.7731 |
| SeaNMF | 0.5759 | 0.5666 | 0.5712 | 0.8979 | 0.8897 | 0.8937 | 0.7677 | 0.7449 | 0.7560 | 0.7759 | 0.7663 | 0.7709 |
| Sparse SeaNMF | 0.5613 | 0.5679 | 0.5644 | 0.8915 | 0.8912 | 0.8913 | 0.7714 | 0.7471 | 0.7590 | 0.7811 | 0.7724 | 0.7766 |

ual sentences and by applying cross-entropy using document terms, system ables to extract meaningful topics.

For a start, "Yahoo.Ans" and "DBLP" data used in the training process and extracted topics out of process selected based on the PMI score. The higher score selected. By analyzing top keywords, most similar topics captured with the SeaNMF system selected.

To display semantically associated discovered topics in the SeaNMF model, we build a network, based on top keywords in each topic, which is representing in Fig. 5. Several words with great modulation in the corpus have less degree or in other meaning, less associated with other keywords.

Figure 5 shows the extracted topics using SeaNMF model which has minimum noise in words and extracted words are correlated together. Hence, pro-

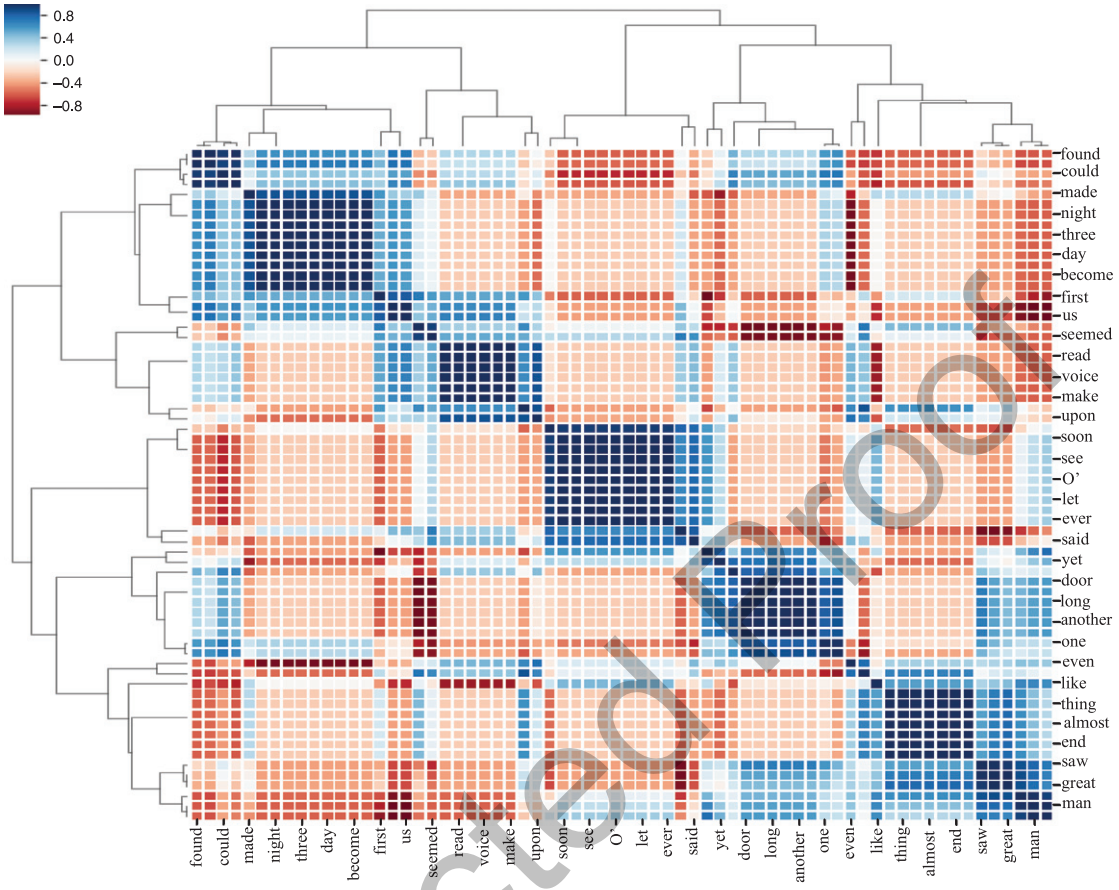


Fig. 5. Topic Semantic Analysis in SeaNMF model.

Table 8
Policy, action, reward in various Seq2Seq models

| Seq2Seq | Policy (P) | Action (A) | Reward(R) |
|---|---|---|---|
| T.Summarization<br>H.Generation<br>M.Translation<br>Q.Generation | Attention-based models,<br>pointer-generators, etc. | Summarize based on next token,<br>translate and headline | ROUGE, BLEU |
| Q&A | Seq2Seq | Vocabulary based answer selection<br>or selecting the<br>start and end index of the<br>answer in the input document | F-measure |
| I.Captioning<br>V.Captioning | Seq2Seq | Next token chosen as caption | CIDEr, SPICE, METEOR |
| S.Recognition | Seq2Seq | Next token chosen as speech | Connectionist Temporal<br>Classification (CTC)<br>BLEU |
| D.Generation | Seq2Seq | Generate Dialogue speech | Dialogue Length<br>Dialogue variety |

posed semantic analysis output display the presented model is able to discover meaningful topics from short text document. Further explanation shows, the probability of connection between words through dataset. As it shows the color bar in Figure, light colors have low chance to occur together in same sentence but dark colors represents the highest probability of words which occur together.

Figure 6 represents the topic visualization between words in the selected dataset. The number of topics

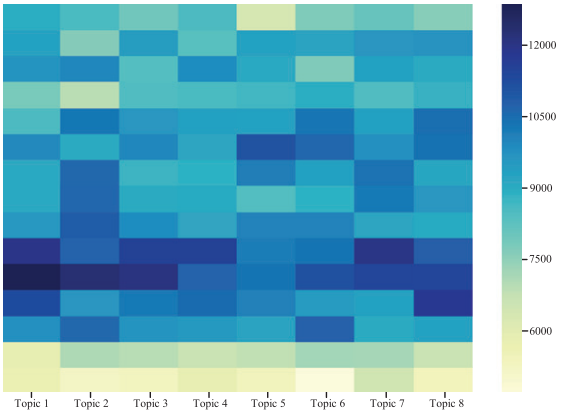


Fig. 6. Topic distribution in word corpus.

is manually selected as eight and Fig. 6 Visualize the NMF topic modeling output regarding the short text content matrix. Topics are including the detail information of the input document. Word embedding process applied to find the relationship between individual terms. Each colour shows the importance of the topic in the dataset and related contents similarity in which light colour present minimum probability of words relationship and dark colours show the maximum probability between topic terms.

## 5. Conclusion

In this paper, we present a combination of reinforcement learning Seq2Seq system using the SeaNMF model to improve the performance of topic modeling in short text documents. The proposed model leverages the word-context semantic correlations in training, which is effective to overcome the issue of short text documents, such as lack of information in meaningful topics extraction. The semantic relationship between keywords and document information is from skip-gram, which shows the efficiency for disclose semantic relationship of words. The reinforcement learning system applied to get the reward function from the algorithm in training set for input-specific RL and compare it in the test set with specific RL. It avoids the difficulty of the document extraction part. This work used a block coordinate descent algorithm to solve the SeaNMF model issue and compare the proposed model with other state-of-art models using real-world data. In this paper, we utilized accuracy as a standard measure to evaluate the performance of topic coherence and document classification. The accessible measurement shows that the proposed model performed better in terms of accuracy metric as compared to other models. Topic coherence and classification performance show the consistency and strength of the model. The proposed model results demonstrate that extracted hidden topics from short text data are meaningful, and terms are consistent with topics. The performance of the proposed model can be enhanced by using a graph-based word embedding technique. The graph embedding method overcome the limitations of sequential input methods which is used in the proposed system. In future we will use a graph-based word embedding technique to improve the performance of topics modeling from short text document using machine learning techniques.



## Acknowledgments

This research was supported by the 2020 scientific promotion program funded by Jeju National University.

## References

[1] W.X. Zhao, J. Jiang, J. Weng, J. He, E.-P. Lim, H. Yan and X. Li, Comparing twitter and traditional media using topic models. In *European conference on information retrieval*, (2011), pp. 338–349. Springer.

[2] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu and M. Demirbas, Short text classification in twitter to improve information filtering. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, (2010), pp. 841–842. ACM.

[3] L. Hong and B.D. Davison, Empirical study of topic modeling in twitter. In *Proceedings of the first workshop on social media analytics*, (2010), pp. 80–88. acm.

[4] Z. Wang and H. Wang, Understanding short texts. 2016.

[5] X. Yan, J. Guo, Y. Lan and X. Cheng. A biterm topic model for short texts. In *Proceedings of the* $22^{nd}$ *international conference on World Wide Web*, (2013), pp. 1445–1456. ACM.

[6] D.M. Blei, A.Y. Ng and M.I. Jordan, Latent dirichlet allocation, *Journal of Machine Learning Research* **3**(Jan) (2003), 993–1022.

[7] S. Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer and R. Harshman, Indexing by latent semantic analysis, *Journal of the American society for Information Science* **41**(6) (1990), 391–407.

[8] T. Hofmann, Probabilistic latent semantic indexing. In *ACM SIGIR Forum* **51** (2017), pp. 211–218. ACM.

[9] Daniel D. Lee and H. Sebastian Seung, Learning the parts of objects by non-negative matrix factorization, *Nature* **401**(6755) (1999), 788.

[10] J. Choo, C. Lee, Chandan K. Reddy and H. Park, Weakly supervised nonnegative matrix factorization for user-driven clustering, *Data Mining and Knowledge Discovery* **29**(6) (2015), 1598–1621.

[11] D. Kuang, J. Choo and H. Park, Nonnegative matrix factorization for interactive topic modeling and document clustering, In *Partitional Clustering Algorithms* (2015), pp. 215–243. Springer.

[12] D. Kuang, S. Yun and H. Park, Symnmf: nonnegative low-rank approximation of a similarity matrix for graph clustering, *Journal of Global Optimization* **62**(3) (2015), 545–574.

[13] X. Quan, C. Kit, Y. Ge and S.J. Pan, Short and sparse text topic modeling via self-aggregation. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[14] Y. Zuo, J. Wu, H. Zhang, H. Lin, F. Wang, K. Xu and H. Xiong, Topic modeling of short texts: A pseudodocument view. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, (2016), pp. 2105–2114. ACM.

[15] T. Mikolov, K. Chen, G. Corrado and J. Dean, Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.

[16] J. Pennington, R. Socher and C. Manning, Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, (2014), pp. 1532–1543.

[17] C. Li, H. Wang, Z. Zhang, A. Sun and Z. Ma, Topic modeling for short texts with auxiliary word embeddings. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval* (2016), pp. 165–174. ACM.

[18] V.K.R. Sridhar, Unsupervised topic modeling for short texts using distributed representations of words. In *Proceedings of the 1st workshop on vector space modeling for natural language processing*, (2015), pp. 192–200.

[19] G. Xun, V. Gopalakrishnan, F. Ma, Y. Li, J. Gao and A. Zhang, Topic discovery for short texts using word embeddings. In *2016 IEEE 16$^{th}$ international conference on data mining (ICDM)*, (2016), pp. 1299–1304. IEEE.

[20] T. Mikolov, I. Sutskever, K. Chen, Greg S. Corrado and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (2013), pp. 3111–3119.

[21] O. Levy and Y. Goldberg, Neural word embedding as implicit matrix factorization. In *Advances in neural information processing systems*, (2014), pp. 2177–2185.

[22] O. Levy, Y. Goldberg and I. Dagan, Improving distributional similarity with lessons learned from word embeddings, *Transactions of the Association for Computational Linguistics* **3** (2015), 211–225.

[23] X.-H. Phan, L-M. Nguyen and S. Horiguchi, Learning to classify short and sparse text & web with hidden topics from large-scale data collections. In *Proceedings of the 17th international conference on World Wide Web*, (2008), pp. 91–100.

[24] O. Jin, Nathan N. Liu, K. Zhao, Y. Yu and Q. Yang, Transferring topical knowledge from auxiliary long texts for short text clustering. In *Proceedings of the 20$^{th}$ ACM international conference on Information and knowledge management*, (2011), pp. 775–784.

[25] J. Weng, E-P. Lim, J. Jiang and Q. He, Twitterrank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference onWeb search and data mining*, (2010), pp. 261–270.

[26] R. Mehrotra, S. Sanner, W. Buntine and L. Xie, Improving lda topic models for microblogs via tweet pooling and automatic labeling. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, (2013), pp. 889–892.

[27] Y. Zuo, J. Zhao and K. Xu, Word network topic model: a simple but general solution for short and imbalanced texts, *Knowledge and Information Systems* **48**(2) (2016), 379–398.

[28] F. Kou, J. Du, C. Yang, Y. Shi, M. Liang, Z. Xue and H. Li, A multi-feature probabilistic graphical model for social network semantic search, *Neurocomputing* **336** (2019), 67–78.

[29] D. Ramage, D. Hall, R. Nallapati and Christopher D. Manning, Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora. *In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, (2009), pp. 248–256. Association for Computational Linguistics.

[30] F. Kou, J. Du, Z. Lin, M. Liang, H. Li, L. Shi and C. Yang, A semantic modeling method for social network short text based on spatial and temporal characteristics, *Journal of Computational Science* **28** (2018), 281–293.

[31] T. Lin, W. Tian, Q. Mei and H. Cheng, The dual-sparse topic model: mining focused topics and focused terms in short text. In *Proceedings of the 23rd international conference on World wide web*, (2014), pp. 539–550.

[32] P. Bicalho, M. Pita, G. Pedrosa, A. Lacerda and Gisele L. Pappa, A general framework to expand short text for topic modeling, *Information Sciences* **393** (2017), 66–81.

[33] J. Choo, C. Lee, Chandan K. Reddy and H. Park, Utopian: User-driven topic modeling based on interactive nonnegative matrix factorization, *IEEE Transactions on Visualization and Computer Graphics* **19**(12) (2013), 1992–2001.

[34] H. Kim, J. Choo, J. Kim, Chandan K. Reddy and H. Park, Simultaneous discovery of common and discriminative topics via joint nonnegative matrix factorization. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2015), pp. 567–576. ACM.

[35] X. Yan, J. Guo, S. Liu, X. Cheng and Y. Wang, Learning topics in short texts by non-negative matrix factorization on term correlation matrix. In *proceedings of the 2013 SIAM International Conference on Data Mining*, (2013), pp. 749–757. SIAM.

[36] T. Shi, K. Kang, J. Choo and Chandan K. Reddy, Short-text topic modeling via non-negative matrix factorization enriched with local word-context correlations. In *Proceedings of the 2018 World Wide Web Conference*, (2018), pp. 1105–1114.

[37] M.-T. Luong, H. Pham and Christopher D. Manning, Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025, 2015.

[38] Y. Wu, M. Schuster, Z. Chen, Quoc V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, et al., Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144, 2016.

[39] D. Bahdanau, K. Cho and Y. Bengio, Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473, 2014.

[40] S. Shen, Y. Cheng, Z. He, W. He, H. Wu, M. Sun and Y. Liu, Minimum risk training for neural machine translation. arXiv preprint arXiv:1512.02433, 2015.

[41] Alexander M. Rush, S. Chopra and J. Weston, A neural attention model for abstractive sentence summarization. arXiv preprint arXiv:1509.00685, 2015.

[42] S. Chopra, M. Auli and Alexander M. Rush, Abstractive sentence summarization with attentive recurrent neural networks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, (2016), pp. 93–98.

[43] S.-Q. Shen, Y.-K. Lin, C.-C. Tu, Y. Zhao, Z.-Y. Liu, M.-S. Sun, et al., Recent advances on neural headline generation, *Journal of Computer Science and Technology* **32**(4) (2017), 768–784.

[44] A. See, Peter J. Liu and Christopher D. Manning, Get to the point: Summarization with pointer-generator networks. arXiv preprint arXiv:1704.04368, 2017.

[45] R. Nallapati, B. Zhou, C. Gulcehre, B. Xiang, et al., Abstractive text summarization using sequence-tosequence rnns and beyond. arXiv preprint arXiv:1602.06023, 2016.

[46] D. Bahdanau, J. Chorowski, D. Serdyuk, P. Brakel and Y. Bengio, End-to-end attentionbased large vocabulary speech recognition. In *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, (2016), pp. 4945–4949. IEEE.

[47] A. Graves and N. Jaitly, Towards end-to-end speech recognition with recurrent neural networks. In *International conference on machine learning*, (2014), pp. 1764–1772.

[48] Y. Miao, M. Gowayyed and F. Metze, Eesen: End-to-end speech recognition using deep rnn models and wfst-based decoding. In *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, (2015), pp. 167–174. IEEE.

[49] S. Bengio, O. Vinyals, N. Jaitly and N. Shazeer, Scheduled sampling for sequence prediction with recurrent neural networks. In *Advances in Neural Information Processing Systems* (2015), pp. 1171–1179.

[50] M. Zamanian and P. Heydari, Readability of texts: State of the art, *Theory & Practice in Language Studies* **2**(1) 2012.

[51] W. Li and A. McCallum, Semi-supervised sequence modeling with syntactic topic models. In *AAAI* **5** (2005), pp. 813–818.

[52] I. Sutskever, O. Vinyals and Quoc V. Le, Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, (2014), pp. 3104–3112.

[53] M. Riedl and C. Biemann, Text segmentation with topic models, *Journal for Language Technology and Computational Linguistics* **27**(1) (2012), 47–69.

[54] M. Riedl and C. Biemann, How text segmentation algorithms gain from topic models. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, (2012), pp. 553–557. Association for Computational Linguistics.

[55] D. Fisher, M. Kozdoba and S. Mannor, Topic modeling via full dependence mixtures. arXiv preprint arXiv:1906.06181, 2019.

[56] R. Vedantam, C. Lawrence Zitnick and D. Parikh, Cider: Consensus-based image description evaluation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2015), pp. 4566–4575.

[57] K. Papineni, S. Roukos, T. Ward and W.-J. Zhu, Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, (2002), pp. 311–318. Association for Computational Linguistics.

[58] C.-Y. Lin, Looking for a few good metrics: Automatic summarization evaluation-how many samples are enough? In *NTCIR*, 2004.

[59] S. Banerjee and A. Lavie, Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, (2005), pp. 65–72.

[60] A. Zubiaga and H. Ji, Harnessing web page directories for large-scale classification of tweets. In *Proceedings of the 22nd international conference on world wide web*, (2013), pp. 225–226. ACM.

[61] R. ek and P. Sojka, Gensim—statistical semantics in python. Retrieved from genism. org, 2011.