



Research Article

Learning High-Order Semantic Representation for Intent Classification and Slot Filling on Low-Resource Language via Hypergraph

Xianglong Qi,¹ Yang Gao,² Ruibin Wang,² Minghua Zhao,³ Shengjia Cui ,³ and Mohsen Mortazavi ^{4,5}

¹Liaoning Huading Technology Co Ltd, Shenyang, Liaoning 110167, China

²Digital China Information Service Company Ltd, Beijing 100085, China

³Baidu Co Ltd, Beijing 100085, China

⁴Department of Computer Science, Islamic Azad University, Mahshahr, Iran

⁵Department of Computer Education and Instructional Technologies, Eastern Mediterranean University (EMU), Famagusta 99628, Cyprus

Correspondence should be addressed to Shengjia Cui; shengjia_cui@163.com and Mohsen Mortazavi; mohsen.mortazavi.edu@gmail.com

Received 12 June 2022; Accepted 22 August 2022; Published 16 September 2022

Academic Editor: Junwei Ma

Copyright © 2022 Xianglong Qi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Representation of language is the first and critical task for Natural Language Understanding (NLU) in a dialogue system. Pretraining, embedding model, and fine-tuning for intent classification and slot-filling are popular and well-performing approaches but are time consuming and inefficient for low-resource languages. Concretely, the out-of-vocabulary and transferring to different languages are two tough challenges for multilingual pretrained and cross-lingual transferring models. Furthermore, quality-proved parallel data are necessary for the current frameworks. Stepping over these challenges, different from the existing solutions, we propose a novel approach, the Hypergraph Transfer Encoding Network “HGTransEnNet”. The proposed model leverages off-the-shelf high-quality pretrained word embedding models of resource-rich languages to learn the high-order semantic representation of low-resource languages in a transductive clustering manner of hypergraph modeling, which does not need parallel data. The experiments show that the representations learned by “HGTransEnNet” for low-resource language are more effective than the state-of-the-art language models, which are pretrained on a large-scale multilingual or monolingual corpus, in intent classification and slot-filling tasks on Indonesian and English datasets.

1. Introduction

The pretrained language models, such as ELMo [1], BERT [2], RoBERTa [3], and XLNet [4], play vital roles in modern neural NLP systems, which learn a widely applicable and informative representation of words and sentences [5–7]. With the optimization of high-quality semantic representation, the performance of models for most of the downstream tasks such as text generation [8] or text classification [9, 10] is upsurging. Recently, the multilingual-BERT [11] and Bilingual Generative Transformer (BGT) [12] for low-resource languages draw attention in both research literature

and industry. However, pretraining a specific and reliable word embedding model from scratch for the low-resource language requires large-scale corpus and expensive computing costs. Meanwhile, it is unwise and redundant to pay so many efforts for each low-resource language as there are hundreds of low-resource languages in the world. Consequently, to our knowledge, the popular strategies for embedding low-resource languages are composed of two branches: (i) utilizing the multilingual pretrained word embedding model [2, 11] and fine-tuning with annotated training data directly; (ii) cross-lingual transferring [13] based on multilingual embedding pretrained model from a

resource-rich language in some designed methods, such as aligning vectors [14, 15] or mixing codes [16].

The blue circles denote the resource-rich natural language sentences (i.e., English corpus), each of which has its encoded representation from the pretrained embedding model. The green circles denote the low-resource natural language sentences (i.e., Indonesian corpus). By learning the high-order semantic representation from hypergraph, the representation of the Indonesian corpus is generated.

Under these two solutions, we also have to face and overcome corresponding challenges. For the first challenge, despite the large scale of multilingual pretrained models, many words of a low-resource language are still not included in the vocabulary, which leads to the out-of-vocabulary problem. Fine-tuning the multilingual pretrained model is an approach to update and adapt for the relevant data. However, it is technically challenging, as the hyper-parameters picking for the fitting needs to be carried out carefully [17]. Otherwise, this will cause either losing valuable information learned from the origin multilingual embedding model or merging the new corpus into the embedding latent space poorly. And lots of irrelevant and useless language embedding will also sparse the word representation model [11]. The second challenge is from the cross-lingual branch. They not only suffer obstacles of the multilingual pretrained model but also accumulate more representing loss during training or fine-tuning [17]. And they rely heavily on large amounts of a parallel corpus, which is expensive to collect and hard to control quality [13].

In this study, we address these above problems via a Hypergraph based Transfer Encoding Network (HGTransEnNet), which allows learning high-order representation of semantic information (rather than fine word-level representation) for low-resource language benefiting from high-quality monolingual pretrained word embedding model in a transductive clustering manner. The proposed HGTransEnNet is built upon a cross-lingual hypergraph structure as illustrated in Figure 1, which is fed with a bilingual but nonparallel corpus. Hypergraph structure takes advantage of collecting knowledge and explores and learns the co-relationship of high-order semantic representation shared between the low-resource language and resource-rich language within the same domain. We conduct extensive experiments on the annotated Indonesian dialogue dataset and the English dialogue dataset (MultiWOZ [18]). Our approach achieves better performance than existing methods on all of the domains in terms of intent classification and slot-filling tasks. And we investigate the model's performance on a different scale of feeding training data, rare domains, out-of-vocabulary, and other languages by abundant comparison experiments. This study is mainly divided into four sections. In the abstract, we briefly describe the main issues to be addressed, the structure of our proposed framework, and the experimental results. Second, in the Introduction, we discuss two challenges that exist in the field of low-resource language representation learning, as well as how our model addresses these challenges. We divide related work into three subsections (Low-resource Language, Spoken Language

Understanding, and Preliminary on Hypergraph Learning) to introduce our study on the basis of a solid theoretical foundation. Third, we disassemble and explain the structure of the proposed model in detail in the Hypergraph Transfer Encoding Network. Finally, in the Experiment, we first introduce datasets and compared methods and metrics. Then, we use figures and tables to demonstrate the effectiveness and robustness of our proposed HGTransEnNet. The main contributions of this study are summarized as follows:

- (i) We propose a hypergraph-based framework for representing high-order semantic information by transferring and learning from resource-rich language data.
- (ii) Our framework is not only capable of effectively solving embedding for low-resource languages but also has the potential ability to overcome the out-of-vocabulary problem for intent classification and slot-filling tasks.
- (iii) The proposed method outperforms state-of-the-art related methods in intent classification and slot-filling tasks on the Indonesian dialogue dataset (IDWOZ), which will be released as one of the resource contributions, as well as other widely adopted multilingual dialogue datasets (Multilingual WOZ 2.0, Multilingual NLU).

2. Related Works

2.1. Low-Resource Language. Many works make efforts on representing low-resource languages [19–21]. Cross-lingual transfer learning has become a popular topic aiming to discover the underlying co-relationships between the source and target languages. Reference [20] proposes to integrate English syntactic knowledge into a state-of-the-art model and shows that it is reasonable to leverage English knowledge to improve low-resource language understanding. Reference [22] conducts the cross-lingual word embedding mapping using zero supervision signals. Reference [14] proposes a self-learning framework in a small size of word dictionary to learn the mapping between source and target word embeddings. Reference [13] utilizes multilingual embeddings obtained from training Machine Translation systems [15] in Thai and Spanish. Reference [23] investigates to align the cross-lingual sentence-level representations by leveraging the large monolingual and bilingual corpus and achieves state-of-the-art performance in several cross-lingual tasks. In line with these methods, encoding semantic information directly within the same cross-lingual latent space could avoid semantic misunderstanding. But relying on aligned parallel sentence pairs can suffer from noise and imperfect alignments [16]. What is more, it is quite challenging to collect enough high-quality bilingual parallel corpus with fine-labeled annotation. To our knowledge, our approach is designed with a high-order structure-hypergraph modeling, to overcome these training and collection problems. Simply with the help of easy-obtained monolingual corpus and the off-the-shelf pretrained language

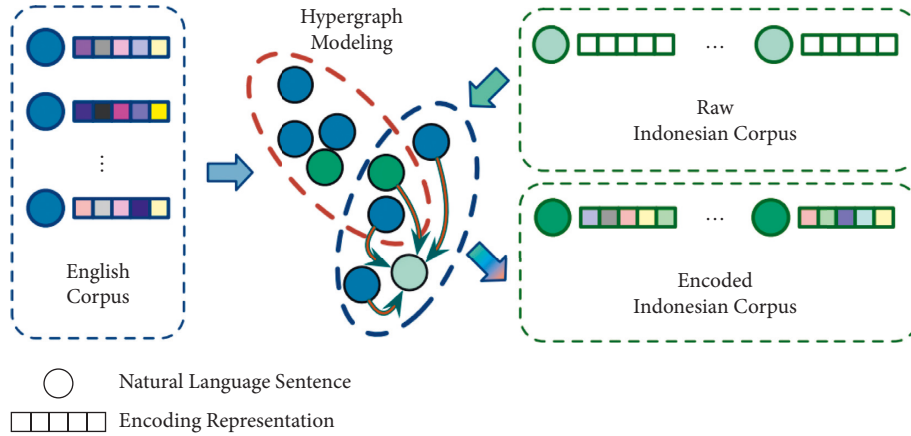


FIGURE 1: Illustration of the achievement for our framework.

model, our proposed framework is capable of achieving comparable and reliable performance on the semantic classification task (i.e., intent classification) on a low-resource language dataset.

2.2. Spoken Language Understanding. There are several complex components in dialogue systems, mainly separated into three parts: “Natural Language Understanding (NLU)” [24–28] (a.k.a., Spoken Language Understanding, SLU), “Dialogue Management (DM)” [29, 30], and “Natural Language Generation (NLG)” [31, 32]. Some trials on end-to-end modeling [31, 33–37] have also been considered. The difference among languages mainly reflects on the first part and the third part [38, 39], the former of which is the foremost challenge to tackle in this work. In recent years, brief concepts have been extracted efficiently from the growing data on the Internet. A method called Swarm Intelligence (SI) is widely used in the construction of automatic text summary frameworks. Moreover, several NLU models based on SI perform well in both single-document and multi-document summarization [40, 41]. Research in SLU fields has not only been applied to dialogue models but has gradually expanded to the field of chatbots. Kabiljo et al. [42] propose an ADA (Academic Digital Assistant) chatbot supported by natural language understanding to deal with the impact of COVID-19 [43] on the education system. Matic et al. [44] thoroughly investigate the structure of common chatbots and introduced corresponding meta-models. This study also designs mapping rules between common natural language understanding models, which can be used to make chatbot architecture more flexible. Gupta et al. [45] propose a novel health care chatbot based on the RASA framework, which can initially predict the disease of the patient and give certain treatment suggestions through dialogue without any hassle. SLU typically involves identifying the intent and extracting semantic constituents from the natural language query, two tasks that are often referred to as intent detection and slot filling [28]. Usually, the performance on these two tasks can efficiently embody the quality of semantic representation for understanding spoken languages. Therefore, we mainly demonstrate the ability of

our framework by conducting experiments on these two tasks in this work.

2.3. Preliminary on Hypergraph Learning. Hypergraph learning has been widely applied in many tasks, such as identifying nonrandom structure in structural connectivity of the cortical microcircuits [46], identifying high-order brain connectome biomarkers for disease diagnosis [47], and studying the co-relationships between functional and structural connectome data [48]. Hypergraph learning was first introduced in [49], in which each node represents one case, each hyperedge captures the correlation between each pair of nodes, and the learning process is conducted on a hypergraph as a propagation process. By this method, the transductive inference on hypergraph aims to minimize the label differences between vertices that are connected by more and stronger hyperedges. Then, the hypergraph learning is conducted as a label propagation process on the hypergraph to obtain the label projection matrix [50], or as a spectral clustering [51]. Other applications of hypergraph learning include video object segmentation [52], images ranking [53], and landmark retrieval [54]. Hypergraph learning has the advantage of modeling high-order correlation modeling, but the mining and learning of the co-relation among different languages for semantic understanding on the hypergraph have not been well investigated.

3. Hypergraph Transfer Encoding Network

In this section, we introduce the detailed structure of our proposed Hypergraph Transfer Encoding Network (HGTransEnNet), as shown in Figure 2. In the first stage, the encoding hypergraph is constructed from the resource-rich language dataset and low-resource language data, which includes the initial vertex feature matrix (denoted as X_0) and the hypergraph incidence matrix (denoted as H). Then, the second stage learns the semantic representation for low-resource language sentences from the pretrained resource-rich language model by the designed hypergraph encoding convolutional layers (denoted as “HGEnConv”) in a transductive learning manner. Finally, we obtain the

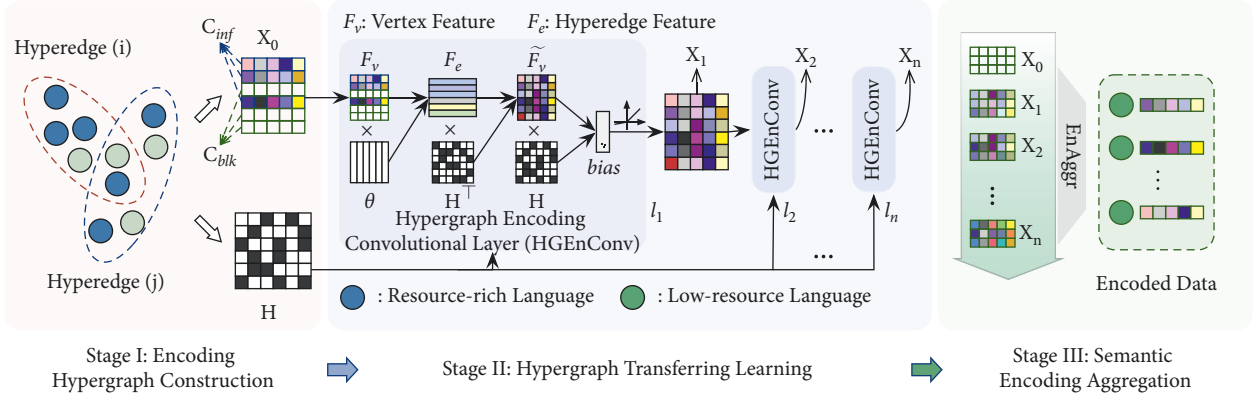


FIGURE 2: Illustration of the proposed framework—hypergraph transfer encoding network (HGTransEnNet), consisting of three stages, i.e., encoding hypergraph construction, hypergraph transferring learning and semantic encoding aggregation. X_0 and H denote the initialized encoded feature matrix and the hypergraph incidence matrix, respectively. The “HGEnConv” layers are capable of merging and extracting high-order semantic representation in a transductive clustering manner. So that the low-resource language data, represented by blank feature vectors (i.e., C_{blk}) could be generated and updated with the transferring of representation from resource-rich language sentences (i.e., C_{inf}). The “EnAggr” module is designed for aggregating the high-order encoding into the final semantic representation.

encoded features of low-resource language for further intent classification and slot-filling tasks. Next, we introduce each individual step of the proposed framework detailed furthermore.

3.1. Encoding Hypergraph Construction. Same as a fundamental hypergraph, our encoding hypergraph is defined as $G = \langle V, E \rangle$, where V and E denote a set of vertices and a set of hyperedges respectively. Each hyperedge is assigned with a weight by the diagonal matrix W . We let each vertex denote a semantic feature of sentence, including resource-rich language (i.e., English) and low-resource language (i.e., Indonesian). The crucial components of constructing stage in HGTransEnNet are the sentence vertex feature matrix $X \in \mathbb{R}^N \times C$ and the incidence matrix $H \in \mathbb{R}^N \times E$, where the N denotes the number of vertices $|V|$, the C denotes the dimension of the vertex feature, and the E denotes the number of hyperedges $|E|$. As shown in Figure 2, we firstly group the English data and the Indonesian data based on the same intent class ($y_i \in Y$) within the same domain, since the same semantic classification could share a similar combination of latent space patterns. In this grouping manner, the sets of hyperedges are generated, which are denoted as the blue oval dotted frame or the red oval dotted frame in Figure 2. The initial vertex feature matrix $X_0 \in \mathbb{R}^N \times C$ is formed by $N = N_{en} + N_{id}$ sentences, i.e., the sum of N_{en} English data and N_{id} Indonesian data, whose structure can be formulated as:

$$X_0 = \begin{bmatrix} [\dots n_i \dots]^T & [\dots n_j \dots]^T \\ \vdots & \vdots \\ C_{inf \in \mathbb{R}^{N_{en} \times C}} & C_{blk \in \mathbb{R}^{N_{id} \times C}} \end{bmatrix}^T, \quad (1)$$

where n_i and C_{inf} denote the pretrained encoded feature vector for English sentence and the informative features matrix, respectively. N_{en} and N_{id} denote the number of English sentences and Indonesian sentences, respectively. n_j and C_{blk} stand for the blank-semantic feature vector for the target Indonesian sentence and the target Indonesian features matrix, respectively. Note that the informative encoded

features matrix of English sentences C_{inf} are encoded by the 5 pretrained English-BERT-Base-based models [2, 55], provided by the popular repository bert-as-service1. The features matrix of Indonesian sentences C_{blk} are initialized randomly within a word bank but share the same dimension $n_j \in \mathbb{R}^{1 \times C}$, which are generated and updated by several layers of hypergraph encoding convolutional layers. In our approach, to leverage as much as informative encoding from high-quality pretrained English-BERT, we set up the correlation between each English vertex and Indonesian vertex in the incidence matrix H , based on the same classification of the data (e.g., the same intention), denoted as the “blue” oval wireframe and the “red” oval wireframe. Our incidence matrix H is calculated by $|V| \times |E|$ and the entries are defined in Eq. (2):

$$H = \begin{cases} 1, & n_i \in y_j, \\ 0, & n_j \notin y_j, \end{cases} \quad (n_i \in \mathcal{V}, n_j \in E). \quad (2)$$

If a vertex n_i is connected by a hyperedge y_j , then the value of the corresponding element of the incidence matrix $H(i, j)$ is 1, otherwise it has the value 0. The degree of a vertex $v \in V$ is defined as (v) , and the degree of a hyperedge is defined as (e) . The diagonal matrices of the hyperedge degrees and the vertex degrees, denoting as D_e and D_v , respectively, could be generated as shown in:

$$\begin{cases} D_e = [\delta(e_i)], & i \in [1, E], \\ [\delta(e_i)] = \sum_{v_j \in \mathcal{V}} h(v_j, e_i), \\ D_v = [d(v_j)], & j \in [1, N], \\ d(v_j) = \sum_{v_j \in e_i, e_i \in E} \mathcal{W}(e_i) h(v_j, e_i), \\ W = [\mathcal{W}(e_i)], \end{cases} \quad (3)$$

where $W \in \mathbb{R}^{1 \times |E|}$ denotes the weight matrix of hyperedges. $w(e)$ denotes the weight of each hyperedge; we here set it to

1, i.e., ratio (e)=1, all of the co-relationships between sentences share static same weight.

3.2. Hypergraph Transferring Learning. The vertex feature matrix $F_v \in RN \times C$, as denoted in Figure 2, is composed of semantic informative vectors $C_{inf} \in RN_{en} \times C$ and semantic-blank initialized vectors $C_{blk} \in RN_{id} \times C$, representing the English sentences and the Indonesian sentences, respectively. The hyperedge is the bridge of transferring and learning the representation for the low-resource language from the existing pretrained model. The hyperedge feature matrix $F_e \in RN \times E$ is the product of the vertex feature matrix F_v and the learnable parameter matrix $\Theta(l)$. Then, the transfer operation by the incidence matrix H makes the blank vectors C_{blk} filled and updated from the C . Since there is no unique mathematical definition of translation on the hypergraph from the spatial perspective, we take the widely adopted classical spectral hypergraph convolution operation [56] as the base structure of the hypergraph encoding convolutional layer (HGENconv(\cdot)). The hypergraph Laplacian L , i.e., the normalized positive semi-definite Laplacian matrix of the resulting hypergraph, is obtained by:

$$L = I - D_v^{-1/2} H W D_e^{-1} H^T D_v^{-1/2}, \quad (4)$$

where $I \in RN \times N$ is an identity matrix. H is the incidence matrix, W is the weight matrix for hyperedges. Therefore, the convolutional operation, i.e., HGENconv(\cdot) of HGTransEnNet, can be formulated in:

$$\begin{cases} X^{(i+1)} = \sigma((I - L)X^{(i)\Theta(i)}), \\ \sigma = \text{LeakyReLU}(0), \end{cases} \quad (5)$$

where the $X(l) \in RN \times C$ is a signal with N vertexes fed in a layer l . And the $X(l+1)$ is the output of a layer l . The σ denotes the nonlinear activation function like $\text{LeakyReLU}(\cdot)$. $\Theta(l)$ denotes a learnable parameter in the layer l . Finally, by adding the fine-tuned parameters bias, the HGENconv(\cdot) finishes fusing and generating the high-order representation for the Indonesian sentences.

3.3. Semantic Encoding Aggregation. After a few layers (i.e., k) of transferring hypergraph convolutional operation HGENconv(\cdot), a set of output feature maps are obtained $X_{in} = \{X_1, X_2, \dots, X_k\} \in RN \times kC$, denoting different scale of transferring. We design an encoding aggregation function EnAggr(\cdot) to control the rate of transferring, shown in

$$\text{EnAggr}(X_{in}, \text{ratio}) = \left\{ \begin{matrix} X_1, \dots, X_k \\ \text{ratio} * k \end{matrix} \right\} \longrightarrow \widehat{X}_{in} \in \mathbb{R}^{N_{id} \times kC}. \quad (6)$$

ratio represents the proportion of encoding aggregation. The higher the transfer ratio, the more the coupling scale, which means that the Indonesian representation enjoys more fusion. The representation of Indonesian sentences X_{blk} is extracted from the global representation matrix X_{in} . We treat different feature channels (a.k.a., attributes) as different semantic factors that represent and affect the final

intent classification and slot-filling tasks. Considering that the feature channels of each sentence are used to describe different attributes with the same dimension latent embedding space, we select the most remarkable or the average value for each attribute via a column-wise aggregating operation (i.e., AttrAggr(\cdot)) to reserve each attribute information as much as possible, which can be defined as:

$$\text{AttrAggr} \left(\begin{matrix} X' \\ \mathbb{R}^{a \times C} \end{matrix} \right) = \frac{1}{a} \cdot \left[\begin{matrix} \sum_{i=1}^a X'[i][1] \\ \dots \\ \sum_{i=1}^a X'[i][C] \end{matrix} \right]^T \longrightarrow X'' \in \mathbb{R}^{1 \times C}. \quad (7)$$

AttrAggr(\cdot) is the column-wise aggregation function that can be max-pooling, mean-pooling, etc. a indicates the number of aggregated attributes. The final encoded representation for Indonesian data $X_{out} \in RN_{id} \times C$ is aggregated by the following algorithm 1. And then, if the task needs word-level representations (e.g., slot filling), the embeddings for each word can be extracted, as shown in:

$$X_{out}^{(i)} = [\mathcal{W}_{emb}^1, \mathcal{W}_{emb}^2, \dots, \mathcal{W}_{emb}^M] \in \mathbb{R}^{1 \times C}, \quad (8)$$

where M denotes the number of tokens contained in each sentence. Finally, we feed the output of our framework forward to the BiLSTM with the attention mechanism and CRF layer [57] to train the intent classification and slot-filling model further. Note that in this work, we mainly focus on the representation of language, the BiLSTM-CRF model could be exchanged for other related classified models; therefore, we left these bunch of extending experiments as our future works.

4. Experiments

4.1. Datasets and Evaluations. We take the English dataset MultiWOZ [18] as the resource-rich language corpus and our Indonesian dataset named ID-WOZ, as the low-resource language corpus. In terms of collection and annotation, we adopt the Wizard-of-OZ [58] dialogue-collecting approach, which has been shown to be effective for obtaining a high-quality corpus at relatively low costs and with a small-time effort. Following the success of MultiWOZ [18], we conduct a large-scale corpus of natural human-human conversations on a similar scale. Based on the given templates for various domains, users and wizards generate conversations using heuristic-based rules to prevent the overflow of information. We design and develop a collection-annotation pipeline platform with a user-friendly structure for building the dataset. At the stage of annotation, we divided the number of well-trained annotators (i.e., 80 local people, 70 of whom spoken ID as their native language, 10 of whom were bilingual citizens, plus 2 main organizers) into two groups to produce dialogue and annotation. A quarter of annotators (i.e., 20) are trained following the guidance we provide to play the wizard role. After collecting 1 k dialogues initially

```

Input:  $X_{in} = \{X_1, X_2, \dots, X_k\}$ ,  $ratio \in [0, 1]$ , pooling
Output:  $X_{out} = [C_{blk}^{(1)}, \dots, C_{blk}^{(j)}]^T \in \mathbb{R}^{N \times C}$ 
(1)  $X_{out}, X_{blk} \leftarrow \text{InitializeTensor}(\Phi)$ 
(2)  $\bar{X}_{in} \leftarrow \text{EnAggr}(X_{in}, ratio)$ 
(3) for  $X \in X$  in do
   $X_{blk(i)} = [C_{blk}^{(1)}, \dots, C_{blk}^{(j)}]^T \leftarrow \text{Extract}(X_i)$ 
(5)  $X_{blk} \leftarrow \text{VerticalStack}(X_{blk}, X_{blk(i)})$ 
(6) end for
(7)  $X_{dbl} \leftarrow \text{AttrAggr}(X_{blk}, \text{pooling})$ 
(8)  $X_{out} \leftarrow \text{Transpose}(X_{dbl})$ 
(9) return  $X_{out}$ 
(5)  $X_{blk} \leftarrow \text{VerticalStack}(X_{blk}, X_{blk(i)})$ 
(6) end for
(7)  $X_{dbl} \leftarrow \text{AttrAggr}(X_{blk}, \text{pooling})$ 
(8)  $X_{out} \leftarrow \text{Transpose}(X_{dbl})$ 
(9) return  $X_{out}$ 

```

ALGORITHM 1: Semantic encoding aggregation.

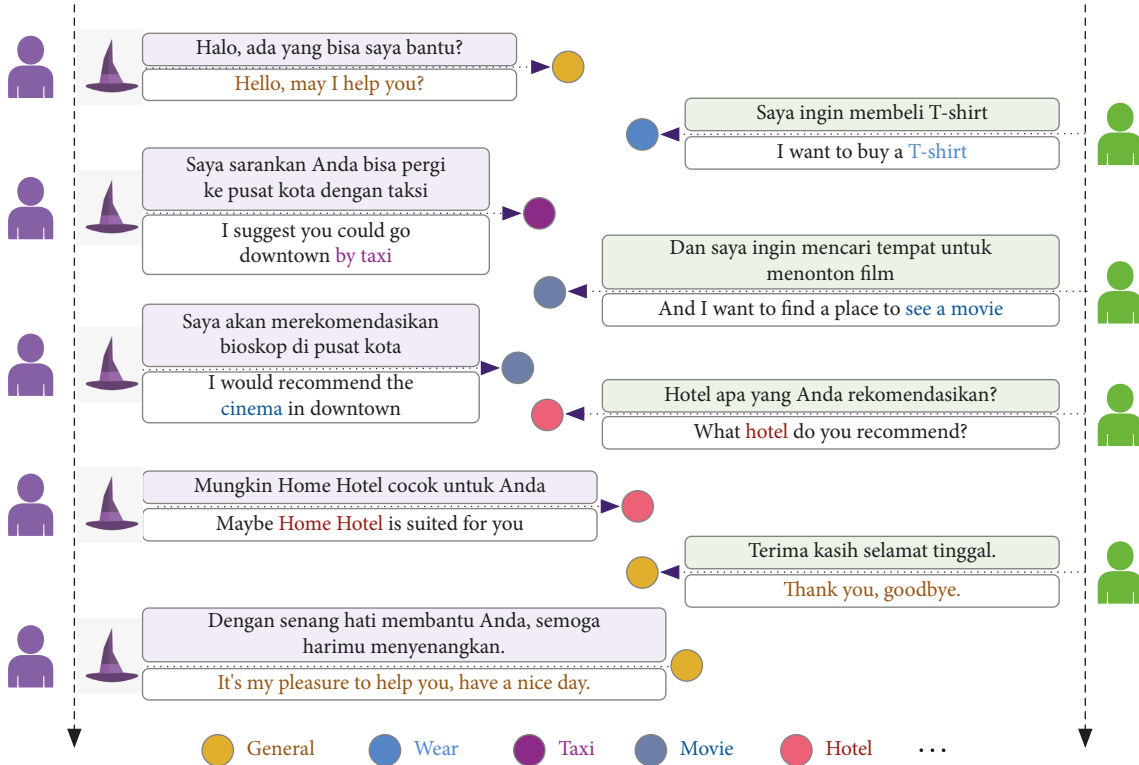


FIGURE 3: The example of our collected task-oriented dialogue dataset.

(about one week), while the collecting conversation is still ongoing, the second group of annotators (i.e., 62) joins in work toward the detailed full-labeled corpus, including domains, actions, intents, and slots. A brief example is shown in Figure 3. It consists of nine domains, namely plane, taxi, wear, restaurant, movie, hotel, attraction, hospital, and police. And we organize a group of annotators to label the corpus including actions, slots, and intents. As the hospital and police domains in MultiWOZ contain very few dialogues (5% of total dialogues) and only appear in the training

dataset, we choose to ignore them in our main experiments, following [59]. Therefore, we only adopt four domains restaurant, hotel, taxi, and attraction shared by MultiWOZ and ID-WOZ datasets in our main experiments. Statistics of them are shown in Table 1. We use the F1 score as the evaluation metric, which is the harmonic mean of precise (P) and recall (R) and is widely adopted in the intent classification and slot-filling tasks. Given a set of training data and corresponding testing data, we split the training data into 5 folds. In our implementation, five-fold cross-validation is

TABLE 1: Statistics for the total number (#.) of sentences, intents, and slots in four domains on the ID-WOZ dataset and the English MultiWOZ dataset.

Datasets	Data	#. Sentences	#. Intent	#. Slots
MultiWOZ	Restaurant	62, 703	41, 177	28, 351
ID-WOZ	Restaurant	28, 095	22, 312	5, 809
MultiWOZ	Hotel	64, 284	42, 434	25, 985
ID-WOZ	Hotel	30, 865	24, 694	8, 720
MultiWOZ	Taxi	48, 080	28, 976	7, 160
ID-WOZ	Taxi	28, 178	22, 168	6, 038
MultiWOZ	Attraction	55, 186	34, 053	21, 004
ID-WOZ	Attraction	36, 523	29, 513	9, 198

employed to investigate the optimal parameter setting within training datasets. To verify the stability of the proposed method, we run the experiments ten times for each set of parameter settings and compare their mean performance.

4.2. Compared Methods and Implementation. We first unify the sentence length based on the longest sentence by padding, i.e., each sentence containing M tokens ($M = 64$). And the following related approaches for word embedding are compared with our network:

- (i) Random Initialization. In this simple method, we set the dimension of word embedding the same with EnglishBERT-Base-cased (i.e., $W_{emb} \in R1 \times 768$), and randomly generate the word embedding bank.
- (ii) Machine Translation (MT). In order to embed reasonably, under the lacking Indonesian pre-trained model situation, we compare with the machine translation preprocessing method. And then we utilize the English-BERT-Basecased (i.e., $W_{emb} \in R1 \times 768$) to encode the corpus. Note that mostly the last two layers of BERT are used as the embedding output.
- (iii) Multilingual-BERT (ML-BERT) [2]. Released on the popular repository–Multilingual-BERT. It contains 104 languages, 12 layers, 768 hidden nodes, 12 heads, and 110M parameters. So that the Indonesian corpus can be represented directly by this pretrained model (i.e., $W_{emb} \in R1 \times 768$).
- (iv) Indonesia-fastText [60, 61]. This work released pretrained word vector models for 157 languages based on each monolingual corpus, including Indonesia. They pretrained it on the Common Crawl and Wikipedia using fastText. This Indonesian word vector model is trained using CBOW with position-weights, in dimension 300 (i.e., $W_{emb} \in R1 \times 300$), with character n-grams of length 5, a window of size 5 and 10 negatives.
- (v) Cross-lingual Transfer [14]. There are multiple ideas of cross-lingual transfer in recent years, we reproduced several of them and reported performance of this practical and reliable method. This method is capable of learning the Indonesian word embeddings notwithstanding still requiring a few bilingual

data, which we will also release (i.e., $W_{emb} \in R1 \times 768$).

- (vi) Indonesian-BERT (ID-BERT). To compare with the state-of-the-art embedding method, we pretrain a specific BERT for Indonesian from scratch with about 3.3 billion tokens from Indonesian websites’ document-level corpus, which covers news reports, research assay, daily articles, and other text genres. The size of its vocabulary is 0.9M, which is much larger than Multilingual-BERT (0.12M). We believe that this size of the vocabulary is sufficient to cover most situations. The training takes one week using Google Cloud TPU v3_8; the Indonesian-BERT-Base (Cased, $L = 12$, $H = 768$, $A = 12$) is eventually obtained (i.e., $W_{emb} \in R1 \times 768$).

The left user plays the wizard role pretending the assistant chatbot, and the right green plays the customer. One dialogue may contain several different domains.

After embedding words by the above models as well as our proposed network, each sentence is represented by a word embedding tensor, whose dimension is $RM \times W$. Besides feeding the original word embeddings of sentences to our network, we also follow the sentence encoder method [5] to experiment. We adopt the average pooling of the word embedding feature map into a 1-dimensional vector $R1 \times W$ and overlay the BiLSTM and CRF layers as the base model [28] to finish the task.

Compared with pretrained ML-BERT and the ID-BERT, our framework is capable of classifying the intention and slots of sentences more accurately, especially in several complex classes, such as hotel_name, area, destination, request_area, and inform_departure.

4.3. Results and Discussion. The comparison results of all the methods are summarized in Table 2. Based on these quantitative results, we have the following analysis.

4.3.1. Intent Classification: Our Proposed Framework. HGTransEnNet outperforms others on this task across all of the boards in the F1 score, achieving 87.21%, 85.03%, 91.26%, 91.44%, and 87.69%, 85.23%, 91.38%, and 91.59% for restaurant, hotel, taxi, attraction domains on the BiLSTM with attention mechanism and BiLSTM with attention mechanism and CRF, respectively. Compared with the nearest baseline models, our method averagely achieves a gain of 0.98% ($pvalue = 1.371 * 10^{-2}$), 0.91% ($p - value = 3.075 * 10^{-3}$), 1.58% ($p - value = 7.864 * 10^{-3}$), and 0.45% ($p - value = 2.609 * 10^{-2}$) compared with the nearest baseline models, cross-lingual, Multilingual-BERT, Indonesian-fastText, and Indonesian-BERT, respectively.

4.3.2. Slot Filling. Our model is also capable of outperforming other methods on slot filling. The accuracy reaches 76.94%, 77.28%, 87.58%, 88.22% and 77.08%, 77.86%, 87.47%, 89.01% on F1 score for restaurant, hotel, taxi, attraction on the BiLSTM-Attention and BiLSTM-

TABLE 2: Experimental comparison of different methods on the four main domains of ID-WOZ for intent classification and slot-filling tasks. ("†" denotes the significance testing, taking HGTransEnNet as the base model, all of p value < 0.05).

Main domains (F1)		Restaurant		Hotel		Taxi		Attraction	
Embeddings	Base model	Intent	Slots	Intent	Slots	Intent	Slots	Intent	Slots
Random initialization	BiLSTM-attention	85.48	74.36	80.73	73.49	89.15	85.22	89.64	86.26
Machine translation	BiLSTM-attention	85.86	73.43	81.82	73.74	89.34	84.17	89.92	86.12
Cross-lingual [14]	BiLSTM-attention	86.59	76.09	84.32	73.92	89.97	86.31	90.73	87.03
Multilingual-BERT [2]	BiLSTM-attention	86.44	75.95	83.92	74.13	90.09	86.28	90.51	87.22
Indonesia-fastText [60]	BiLSTM-attention	85.88	75.27	83.17	74.09	88.92	85.08	90.02	86.88
ID-BERT	BiLSTM-attention	87.03	76.32	84.29	75.51	90.38	86.33	91.37	87.49
HGTransEnNet	BiLSTM-attention	87.21 [†]	76.94 [†]	85.03 [†]	77.28 [†]	91.26 [†]	87.58 [†]	91.44 [†]	88.22 [†]
Random initialization	BiLSTM-attention-CRF	85.86	74.72	81.49	73.52	89.82	85.59	90.03	86.62
Machine translation	BiLSTM-attention-CRF	85.79	73.58	82.04	73.81	89.54	85.31	90.23	86.43
Cross-lingual [14]	BiLSTM-attention-CRF	86.34	76.27	84.51	74.09	90.33	86.45	90.22	87.34
Multilingual-BERT [2]	BiLSTM-attention-CRF	86.35	76.42	84.68	73.73	90.61	86.77	90.98	87.52
Indonesia-fastText [60]	BiLSTM-attention-CRF	85.97	75.38	83.72	74.68	89.37	85.35	90.79	86.41
ID-BERT	BiLSTM-attention-CRF	87.38	76.64	84.57	75.48	90.79	87.04	91.44	88.65
HGTransEnNet	BiLSTM-attention-CRF	87.69 [†]	77.08 [†]	85.23 [†]	77.86 [†]	91.38 [†]	87.47 [†]	91.59 [†]	89.01 [†]

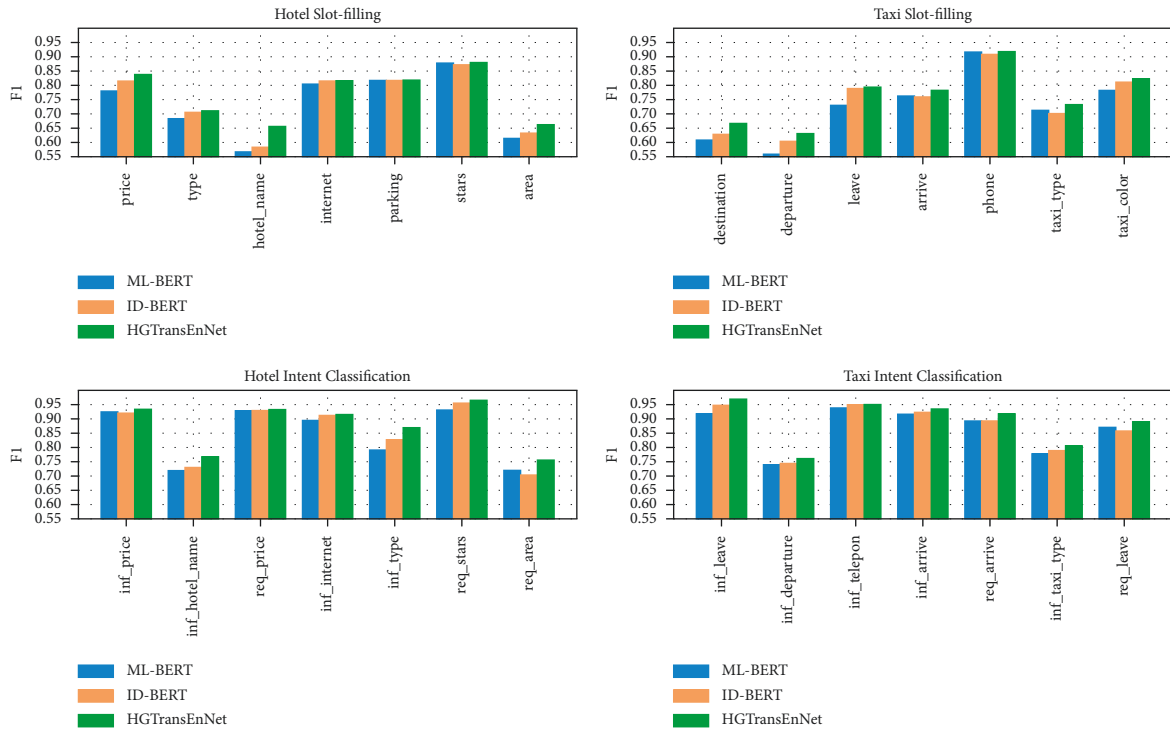


FIGURE 4: Illustration of three methods' performance on the intent classification and slot-filling tasks.

Attention-CRF, respectively. Because our model mainly focuses on the general high-order semantic representation and the slot filling is a kind of more relying on the word-level subtle task. Our method does not achieve relatively high performance, gaining of 1.74% (p - value = $1.715 * 10^{-2}$), 1.68% (p - value = $1.047 * 10^{-2}$), 2.29% (p - value = $5.913 * 10^{-3}$), and 1.00% (p - value = $4.183 * 10^{-3}$) compared with the nearest baseline models, cross-lingual, multilingual-BERT, Indonesian-fastText, Indonesian-BERT, respectively.

4.3.3. *Validation and Analysis.* As shown in Figure 4, we select the top three methods for intent classification and slot-

filling tasks in hotel and taxi domains and draw the bar chart to analyze qualitatively. For intent classification, the model relies on understanding the general semantic information of the given sentences and classifying it into different classes. We can see that `inform_type`, `request_area` intents in hotel domain and the `request_arrive`, `request_leave` intents in taxi domain are higher than the other two methods, because they are obscure and hard to detect. We analyze and draw the conclusion that the specific ID-BERT achieves slightly higher performance than our approach in the precision (P) evaluation metric (1.04%). Because it owns the richest background knowledge of Indonesian. Therefore, it also proves that our model has slightly weaker learning ability without

TABLE 3: Experimental comparison of different methods on the five rare domains of ID-WOZ for intent classification and slot-filling tasks. (“†” denotes the significance testing, taking HGTransEnNet as the base model, all of p value <0.05).

Rare domains (F1)	Plane		Police		Movie		Hospital		Wear	
	Intent	Slots	Intent	Slots	Intent	Slots	Intent	Slots	Intent	Slots
BiLSTM-attention-CRF										
Random initialization	89.47	70.87	86.41	69.47	87.73	74.63	90.63	74.13	89.84	68.93
Machine translation	89.08	72.66	87.89	70.26	88.72	75.79	91.29	75.43	90.90	69.74
Multilingual-BERT [2]	90.09	74.39	89.16	70.89	89.47	74.28	92.43	76.23	92.06	72.69
Indonesia-fastText [60]	89.80	73.44	89.02	70.17	88.93	74.17	92.32	75.16	90.83	70.11
ID-BERT	90.43	75.95	90.35	70.89	89.68	77.30	92.55	76.43	92.19	72.75
HGTransEnNet	91.71 [†]	75.86	90.39 [†]	71.08 [†]	90.31 [†]	76.97	92.89 [†]	77.31 [†]	92.78 [†]	71.80

the support of large-scale datasets. However, we perform better than others in the recall (R) metric about 2.17%, which reflects our model can learn the high-order general semantic representation from the English language model in the transductive clustering manner. The slot-filling task requires the model to detect the value of slots and recognize the class of slots at the same time. In this procession, representation for every word is critical. Our proposed framework HGTransEnNet outperforms others on this task across all of the boards in the F1 score. For the complex types of slots, like hotel_name, area and destination, departure in hotel and taxi domains, respectively, our model has the ability to leverage the language semantic knowledge by the grouping manner and outperforms others (Tables 3 and 4).

4.4. Analysis on Training Data Amount. One of our proposed approach’s strengths is to learn more informative semantic representations for low-resource language. The main limitation of most low-resource languages is lacking high-quality annotated collected corpus. Therefore, to investigate the performance of the model within different amounts of training data, we conduct a series of experiments incrementally. As shown in Figure 5, we feed the model annotated data batch by batch, i.e., 1 k, 2 k, 4 k, 8 k, 16 k, and full-scale. We here select restaurant and taxi domains as examples. The statistics line chart is shown in Figure 5, where the two leftmost sub-graphs denote intent classification and the two rightmost sub-graphs denote slot filling. And we compare three approaches to stand for three popular strategies, namely “machine translation”, “cross-lingual” [14], and “HGTransEnNet”(our introduced method), denoted as blue, yellow, and green lines, respectively. As shown in Figure 5, the red line represents the benchmark performance from ID-BERT pretrained by us former. Based on the quantitative results shown in Figure 5, we have the following observations:

The red line denotes the performance of pretrained ID-BERT, regarded as the reliable benchmark. The green line stands for our proposed HGTransEnNet, showing its strength over other compared methods.

- (1) For machine translation method, the main issue is the quality of translation. We conduct the BLEU [62] test for the entire MultiWOZ, and the performance of translation is 28.46 (BLEU-5) on 30 k sentences. However, during the translation of

TABLE 4: Statistics for the total number (#.) of sentences, intents and slots in five rare domains on ID-WOZ.

Datasets	Data	#. Sentences	#. Intent	#. Slots
ID-WOZ	Plane	30, 538	22, 701	9, 323
ID-WOZ	Police	17, 825	10, 868	3, 446
ID-WOZ	Movie	26, 577	18, 382	6, 105
ID-WOZ	Wear	26, 092	18, 953	10, 197
ID-WOZ	Hospital	26, 491	17, 076	5, 881

dialogue messages, one incorrect word could cause complete misunderstanding. We pick several examples and show them in Figure 6. The top half of the examples stand for the mistakes implied in the slot-filling task. Apparently, the translation method suffers a few mistakes when accounting for the “name”, “area”, “address”, etc., which are quite challenging problems to solve. And the below part shows that a tiny translation mistake will cause the wrong result of the entire intention classification for the sentence and will lead to totally different progress of the dialogue. We can see that with the help of pretrained English-BERT, the performance of the translation method has the ability to get close to the benchmark, but cannot reach better.

- (2) When the scale of fed annotated low-resource language data gets larger, the strength of cross-lingual becomes more obvious. It is capable of avoiding misunderstanding caused by translation and mitigating the shrink effect of the English corpus, which makes it achieve the best performance and even better than the benchmark performance of ID-BERT, when the Indonesian data reach around 16 k for restaurant and taxi domains.
- (3) The accuracy reaches 86.44%, 76.00%, 89.45%, and 85.21% on F1 scores in intent classification and slot-filling tasks, respectively.
- (4) Based on Figure 5, we can draw the conclusion that our method is capable of getting close to the benchmark performance with the less collected corpus. Enjoying the designed representation transferring and aggregation modules, our network manages to perform better than compared related approaches stably and reaches higher achievement in the same scale of training data. However, we can

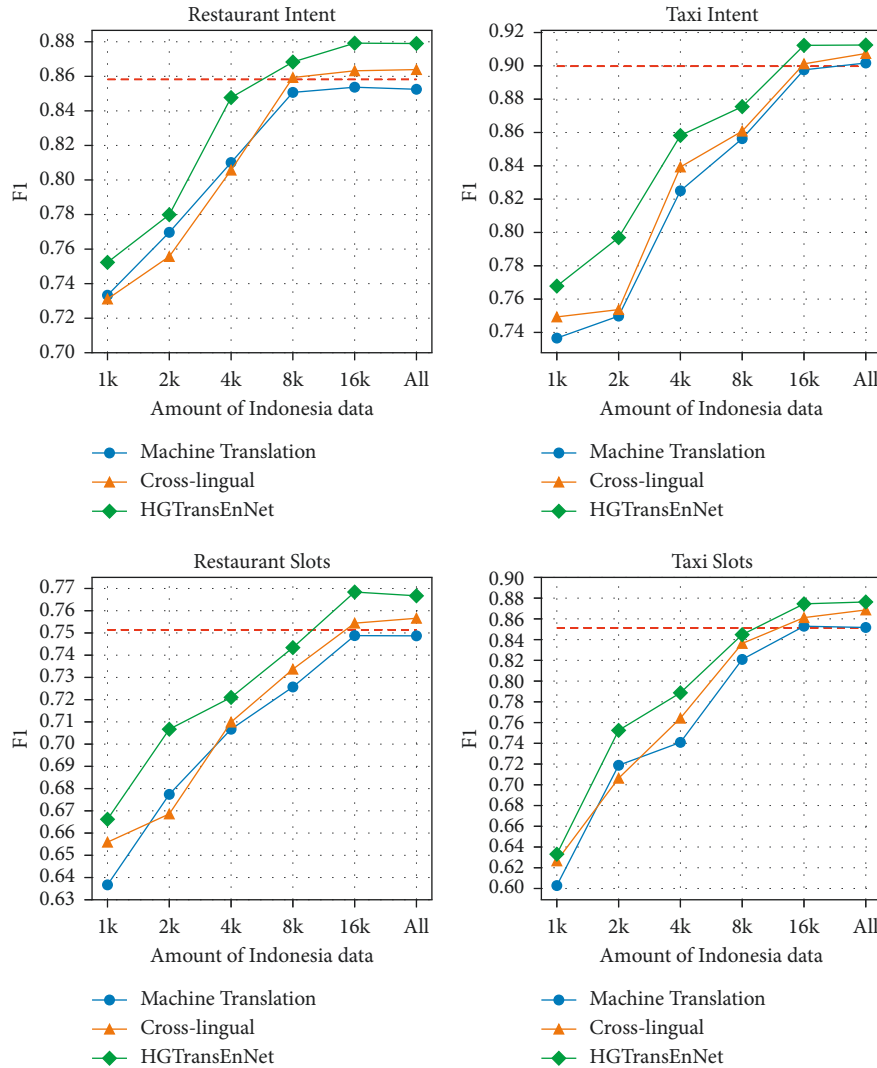


FIGURE 5: Illustration of three methods’ performance on the intent classification and slot-filling tasks within the different scales of feeding training data.

observe that both the traditional pretrained model and our proposed hypergraph model achieve poor performance when the amount of data is extremely small. The main reason is that the length of the dialogue data is short, and the hypergraph structure barely uses sufficient contextual correlation to understand semantic information.

4.5. Performance on Rare Domains. We also conduct extensive experiments on other rare domains (i.e., plane, police, movie, hospital, and wear), which reflect the local cultural background in Indonesia. The MultiWOZ dataset contains the police and hospital domains but the scale is small. The other three domains are special in our collected ID-WOZ. We use our model trained by main domains and large-scale data to generate the sentence encoding and fine-tune for rare domains furthermore. In the plane, movie, and wear domains, the specific ID-BERT achieves a slightly higher performance in the slot-filling task. Because in these

domains, the training data are not only relatively limited but also lack the English corpus in the same domain. But still, generally speaking, our approach shows its ability in the small-scale data situation and outperforms others across all of the board in intent classification and most in slot filling. Overall, the results on rare domains reflect that our approach is capable of transferring the semantic representation for Indonesian and outperforms the pretrained embedding model ID-BERT.

4.6. Ablation Study for Adjacency. Different from the incidence matrix, the adjacency matrix is another popular solution for associating the nodes. In the broad sense, this approach can also be viewed as the graph convolutional neural network [63] when the incidence matrix H connects each pair of sentences and becomes the symmetric matrix. To compare with GCN, we also conduct a comparison experiment in all of the domains. In most domains, its performance is close to the random initialization approach,

Christ	's	College	is	located	in	the	centre	at	saint	Andrew	's	street	.
B-name	I-name	I-name	O	O	O	O	B-area	O	B-addr	I-addr	I-addr	I-addr	O
Chunks: {'name': 'Christ's College', 'area': 'centre', 'addr': 'saint Andrew 's street'}													
Christ	's	College	terletak	di	pusat	di	jalan	suci	dan	suci	.		
O	O	O	O	O	B-area	O	B-addr	I-addr	O	B-addr	O		
Translated Chunks: {'name': 'Perguruan Tinggi Kristus', 'area': 'pusat', 'addr': 'jalan suci orang suci'}													
Christ	's	College	terletak	di	pusat	jalan	saint	Andrew	's	.			
B-name	I-name	I-name	O	O	B-area	B-addr	I-addr	I-addr	I-addr	O			
Human Annotated Chunks: {'name': 'Christ's College', 'area': 'pusat', 'addr': 'jalan saint Andrew 's'}													
Source Sentence	mas, itu taksinya adanya warna apa aja ya?										request_taxi_color		
True Meaning	sir, what colour do you have for that taxi?										request_taxi_color		
Machine Translation	Bro, what's the color of the cab?										request_taxi_color		
Source Sentence	tiket pesawatnya kalo beli besok berapaan ya?										request_price		
True Meaning	how much is the airplane ticket if i buy it tomorrow?										request_price		
Machine Translation	What are the plane tickets for tomorrow?										request_type		
Source Sentence	kalo mau pesen tiket, lewat mana mas pesennya?										request_ticket		
True Meaning	if i want to order the ticket, how do i order it?										request_ticket		
Machine Translation	if you want to order a ticket, where do you order it?										request_location		

FIGURE 6: Examples of translation mistakes. The upper half shows the loss in the slot-filling task. And the lower denotes a little error when the model tackle of intent classification will cause significant misunderstanding.

or even worse. So we report a few results of this section in Table 5, which do not have much meaning to discuss in this implementation strategy. The result demonstrates that lacking the ability to group and learn high-order information will make the model perform poorly and lose skills.

4.7. *Analysis on Out-of-Vocabulary.* To verify how our approach performs in the out-of-vocabulary situation, we furthermore conduct a validation experiment, which is reported in Tables 6 and 7. When the language embedding model encounters some unfamiliar words, we expect the reasonable solution is to group the coming new words into some existing semantically similar groups. Therefore, we take the Indonesian words as the out-of-vocabulary words of the English-BERT-Base-Cased model. We believe that different vertex does not matter in different syntax or in different languages, as long as they share the same group. There are general semantic latent attributes that represent the same classification features. We know that the main purpose of the pretrained language model is to embed the natural language words or sentences into a specific latent space. The words or sentences having similar semantic or syntactic information share similar latent mapping and have close spatial distances. In this analysis scale, the achievement of our proposed HGTransEnNet has the ability to imitate a similar sophisticated embedding latent space. We select the top 1k frequent words from the corpus by TF-IDF [65] and calculate the Euclidean distance between them and corresponding Indonesian word embeddings. From beginning to end, we never feed the model parallel corpus, and even though each word is embedded to a complex dimension vector ($W_{emb} \in R^{1 \times 768}$), the mean distance is 11.6327. (The distance of most synonyms in English within the English-BERT model is around 9.8722.) And the scatter plot demonstrates an obviously clustering effect of sharing similar semantic words.

TABLE 5: Experimental comparison of GCN and HGTransEnNet on the four main domains of ID-WOZ for intent classification and slot-filling tasks.

Task	Intent classification		Slot filling	
	GCN	HGTransEnNet	GCN	HGTransEnNet
Restaurant	40.88	87.69	30.56	77.08
Hotel	46.01	85.23	35.91	77.86
Taxi	47.56	91.38	40.15	87.47
Attraction	37.86	91.59	27.66	89.01

This proves that in the latent space, our model is capable of embedding the language into a similar semantic cluster.

4.8. *Performance on Other Languages.* Besides the exploration above, we also conduct several thorough experiments on the performance of approaches in other languages.

4.8.1. *Multilingual Datasets.* We briefly introduced the experiment datasets here. Multilingual WOZ 2.0 [66] is expanded from the restaurant WOZ dataset by including two more languages, 1200 dialogues of each, i.e., German and Italian. Following the settings of [64], we use 600 dialogues for training, 200 for validation, and 400 for testing. The corpus contains four slots types: food, price, area, and request. And the multilingual task-oriented natural language understanding dialogue dataset (denoted as ‘‘Multilingual NLU’’) proposed by [67] contains English, Spanish and Thai, across three domains (alarm, reminder, and weather). It contains 12 intent types and 11 slot types.

4.8.2. *Compared Methods.* We adopt two more related methods in this experiment besides those introduced in Section 4.2.

TABLE 6: Distance between English and synonyms in Indonesian.

English	Indonesian	Distance	English	Indonesian	Distance	English	Indonesian	Distance
Address	Alamat	9.8347	City	Kota	11.8735	Start	Bintang	8.8871
Location	Lokasi	10.6342	Type	Tipe	10.0988	Road	Jalan	10.4652
Price	Harga	9.9535	Time	Waktu	10.3028	Room	Kamar	10.8014
Departure	Keberangkatan	11.3498	Parking	Parkir	9.9913	Cheap	Murah	10.2911
Arrive	Tiba	11.8612	Hour	Jam	10.2833	Phone	Telepon	9.7668

TABLE 7: Distance between English synonyms within the English-BERT model.

English synonyms	Distance	English synonyms	Distance	English synonyms	Distance	
Arrive	Reach	9.8753	Complete	Finish	9.9882	
Pick	Take	9.7968	Like	Adore	10.1749	
Location	Position	9.8681	Street	Avenue	10.2311	
				Divide	Separate	10.7683
				Happen	Occur	10.3149
				Attain	Obtain	10.2579

TABLE 8: Experimental comparison of different methods on several multilingual datasets for intent classification and slot-filling tasks. (“†” denotes the significance testing, taking HGTransEnNet as the base model, all of p value <0.05).

Settings	Datasets	Multilingual WOZ 2.0 [65]				Multilingual NLU [66]					
		Language		German		Italian		Spanish		Thai	
Embeddings	Base model	Request	Slots	Request	Slots	Request	Slots	Request	Slots	Request	Slots
Random initialization	BiLSTM-attention	80.14	61.37	81.32	64.25	82.01	68.23	67.91	24.33		
Machine translation	BiLSTM-attention	81.33	60.02	82.84	64.03	82.23	67.48	68.30	23.71		
Multilingual-BERT [2]	BiLSTM-attention	83.03	62.43	83.12	65.34	84.18	70.82	70.23	25.92		
Cross-lingual transfer [13]	BiLSTM-attention	83.91	62.95	83.44	65.88	84.03	70.21	70.46	25.61		
Cross-lingual code-mix [16]	BiLSTM-attention	84.19	65.28	83.27	66.26	84.14	71.88	70.13	26.82		
MLT + Multilingual-BERT [64]	BiLSTM-attention	85.22	67.48	83.58	68.08	84.22	73.59	70.91	27.14		
HGTransEnNet	BiLSTM-attention	87.18 [†]	68.01 [†]	84.91 [†]	69.47 [†]	86.56 [†]	74.84 [†]	72.14 [†]	27.44 [†]		
Random initialization	BiLSTM-attention-CRF	80.32	62.14	81.47	64.82	82.47	69.66	68.10	24.47		
Machine translation	BiLSTM-attention-CRF	81.53	62.22	82.81	64.26	82.44	69.63	68.58	24.81		
Multilingual-BERT [2]	BiLSTM-attention-CRF	83.24	63.07	83.04	65.84	84.19	70.76	70.87	26.11		
Cross-lingual transfer [13]	BiLSTM-attention-CRF	83.78	64.33	83.31	68.61	84.73	71.32	70.88	26.59		
Cross-lingual code-mix [16]	BiLSTM-attention-CRF	84.17	65.82	83.98	69.04	84.68	73.81	70.91	26.90		
MLT + Multilingual-BERT [64]	BiLSTM-attention-CRF	85.41	67.57	84.04	69.78	85.03	74.16	71.00	27.42		
HGTransEnNet	BiLSTM-attention-CRF	87.37 [†]	69.21 [†]	85.17 [†]	71.04 [†]	86.42 [†]	75.52 [†]	72.77 [†]	27.79 [†]		

- (i) Zhang et al. [16] proposed a code-mixing approach to tackle the cross-lingual dependency parsing task. By adopting the code-mixing transfer method, it is capable of leveraging syntactic knowledge to transfer to the target language. Therefore, we here utilize this transferring idea to implement the multilingual dialogue NLU tasks as one of the compared methods, with the Multilingual-BERT as the embedding pre-trained model.
- (ii) Liu et al. [67] designed a zero-shot adaptation method for a cross-lingual task-oriented dialogue system, noted as “Attention-Informed Mixed-Language Training (MLT)”. It leverages a few task-related parallel word pairs generated by the attention layer from the trained English model and existing bilingual dictionaries. We here implement it with multilingual-BERT as the embedding pretrained model.

4.8.3. *Analysis on Other Languages.* From the results summarized in Table 8, we have the following observations: considering the difference among these languages, grammar, syntactic, cultural background, language family, etc., our proposed transferring representation method outperforms others across all of the board. Note that though the quality of the annotated dataset also affects all of the model’s performance, the kind of language still plays a more critical role in the cross-lingual intent classification and slot-filling tasks. The domain, quality, and data scale of Multilingual WOZ 2.0 and Multilingual NLU are distinctive, but all of the models perform reliably in German, Italian, and Spanish. Based on the trained model, we analyze the Euclidean Distance of different languages in detail. (1) Since within our transferring method, the model can embed multiple languages corpora into one single latent space, which makes it possible to compare their latent distance. As shown in Figure 7, we can intuitively see that Spanish is the closest language to English. Italian, German, and

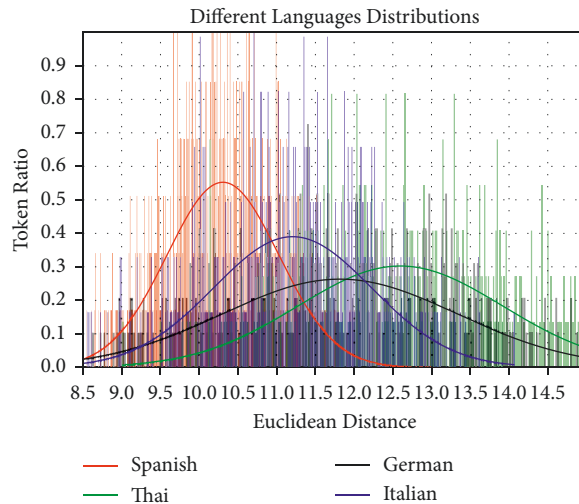


FIGURE 7: Euclidean distance between different languages. We calculate the Euclidean distance of every token in each language and its corresponding English tokens in latent space. And the visualization of their Gaussian distributions shows the distinctive difference between each language intuitively. Spanish is the closest language with English and Thai is the farthest.

EN	Shabu Ghin is a Japanese restaurant that serves the All You Can Eat
EN	Shabu Shabu menu with a choice of Special Beef , Premium Beef ,
EN	Wagyu Beef and Predmium Wagyu Beef .
ID	Shabu Ghin merupakan salah satu restoran Jepang yang menyajikan
ID	menu All you can eat Shabu Shabu dengan pilihan Special beef ,
ID	Premium Beef , Wagyu beef dan Predmium Wagyu Beef .
EN	What is the weather forecast for Tallahassee , Florida next week ?
ES	Cuál es el pronóstico del tiempo para Tallahassee , Florida la
ES	próxima semana ?
DE	Wie ist die Wettervorhersage für Tallahassee , Florida nchste Woche ?

FIGURE 8: Visualization of the model’s attention in different languages. The trained model can allocate weights to each token based on semantic features.

Thai are getting farther gradually. This conclusion is also reflected in Table 8. All of the methods in Spanish, Italian, and German outperform those in Thai. (2) The visualization [68] of our model attention encoding result in different languages is shown in Figure 8. The above half of Figure 8 comes from our collected dataset, i.e., English and Indonesian, and shows a quite reasonable attention bias. The model can allocate more weights on several more domain-specific slots, such as “restaurant name”, “address” or “food name” either in English or Indonesian. The below half shows the representation of the model for other languages, in which the attention weights also make sense. For instance, though the model may make a few mistakes for slot filling, more weights are focused on the related keywords like “time”, and “location”. We can see that our approach has the ability to capture semantic and syntactic information in different languages.

5. Conclusion and Future Work

This study presents a Hypergraph Transfer Encoding Network for the tasks of intent classification and slot filling on

low-resource language, in which the encoding hypergraph is constructed from both the low-resource language dataset and the high-resource language dataset. The semantic representation of low-resource language is generated by the well-designed hypergraph encoding convolutional layers (HGenConv). It is achieved by learning the high-order semantic representation in a transductive clustering manner from the pretrained resource-rich language model. In addition, we construct a well-annotated Indonesian dataset named ID-WOZ, which consists of multiple domains, to fairly evaluate baselines and our proposed HGTransEnNet. Experiments on MultiWOZ and ID-WOZ demonstrate the superior performance of our model to state-of-the-art neural models on intention classification and slot-filling tasks. And our method can also facilitate the exploration of the out-of-vocabulary problem in the semantic representing scale. As stated before, representation learning for the low-resource language is still a highly data-dependent task. Traditional pretraining models and cross-lingual models rely heavily on large amounts of the parallel corpus or multi-language datasets. Future work will consider zero-shot learning,

attention mechanism, and high-order relationships in small sample data to encode embeddings for low-resource languages. We will also explore its capability in other tasks in future works.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declare that there are no conflicts of interest.

References

- [1] E. P. Matthew, N. Mark, I. Mohit, and G. Matt, "Deep contextualized word representations," 2018, <https://arxiv.org/abs/1802.05365>.
- [2] D. Jacob, C. Ming-Wei, L. Kenton, and T. Kristina, "Bert: pre-training of deep bidirectional transformers for language understanding," 2018, <https://arxiv.org/abs/1810.04805>.
- [3] L. Yinhan, O. Myle, G. Naman et al., "A robustly optimized bert pretraining approach," 2019.
- [4] Y. Zhilin, D. Zihang, Y. Yiming, C. Jaime, S. Ruslan, and V. L. Quoc, "Xlnet: generalized autoregressive pretraining for language understanding," 2019, <https://arxiv.org/abs/1906.08237>.
- [5] R. Nils and G. Iryna, "Sentence-bert: sentence embeddings using siamese bert-networks," 2019, <https://arxiv.org/abs/1908.10084>.
- [6] C. Alexis, K. Douwe, S. Holger, B. Loic, and B. Antoine, "Supervised learning of universal sentence representations from natural language inference data," 2017, <https://arxiv.org/abs/1705.02364>.
- [7] D. Cer, Y. Yang, S.-yi Kong et al., "Universal Sentence Encoder," 2018, <https://arxiv.org/abs/1803.11175>.
- [8] S. Golovanov, R. Kurbanov, S. Nikolenko, K. Truskovskiy, T. Alexander, and T. Wolf, "Large-scale Transfer Learning for Natural Language Generation," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, July 2019.
- [9] H. Peng, J. Li, Q. Gong, S. Wang, and L. He, "Hierarchical Taxonomy-Aware and Attentional Graph Capsule Rcnn's for Large-Scale Multi-Label Text Classification," 2019, <https://arxiv.org/abs/1906.04898>.
- [10] D. Singh Sachan, M. Zaheer, and R. Salakhutdinov, "Revisiting Lstm Networks for Semi-supervised Text Classification via Mixed Objective Function," in *Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence*, Palo Alto, California USA, March 2019.
- [11] T. Pires, E. Schlinger, and D. Garrette, "How Multilingual Is Multilingual Bert?," 2019, <https://arxiv.org/abs/1906.01502>.
- [12] W. John, N. Graham, and T. Berg-Kirkpatrick, "A Bilingual Generative Transformer for Semantic Sentence Embedding," 2019, <https://arxiv.org/abs/1911.03895>.
- [13] T. Schuster, O. Ram, R. Barzilay, and A. Globerson, "Cross-lingual Alignment of Contextual Word Embeddings, with Applications to Zero-Shot Dependency Parsing," 2019, <https://arxiv.org/abs/1902.09492>.
- [14] M. Artetxe, G. Labaka, and E. Agirre, "Learning Bilingual Word Embeddings with (Almost) No Bilingual Data," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada, July 2017.
- [15] B. McCann, B. James, C. Xiong, and R. Socher, "Learned in Translation: Contextualized Word Vectors," *NeurIPS*, vol. 30, 2017.
- [16] M. Zhang, Y. Zhang, and G. Fu, "Cross-lingual Dependency Parsing Using Code-Mixed Treebank," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, Hong Kong, China, November 2019.
- [17] P. Bojanowski, O. Celebi, T. Mikolov, E. Grave, and J. Armand, "Updating Pre-trained Word Vectors and Text Classifiers Using Monolingual Alignment," 2019, <https://arxiv.org/abs/1910.06241>.
- [18] P. Budzianowski, T.-H. Wen, and Bo-H. Tseng, "Inigo Casanueva, and Etc. Ultes. Multiwoz-A Large-Scale Multi-Domain Wizard-Of-Oz Dataset for Task-Oriented Dialogue Modelling," 2018, <https://arxiv.org/abs/1810.00278>.
- [19] J. Guo, W. Che, D. Yarowsky, H. Wang, and T. Liu, "Cross-lingual Dependency Parsing Based on Distributed Representations," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Beijing China, June 2015.
- [20] H. Wang, Y. Zhang, G.Y. L. Chan, J. Yang, and H. L. Chieu, "Universal Dependencies Parsing for Colloquial Singaporean English," 2017, <https://arxiv.org/abs/1705.06463>.
- [21] W. Ammar, M. George, M. Ballesteros, C. Dyer, and N. A. Smith, *Many Languages, One Parser*, China, 2016.
- [22] A. Conneau, G. Lample, M. A. Ranzato, L. Denoyer, and H. Jégou, "Word translation without parallel data," 2017, <https://arxiv.org/abs/1710.04087>.
- [23] G. Lample and A. Conneau, "Cross-lingual Language Model Pretraining," 2019, <https://arxiv.org/abs/1901.07291>.
- [24] M. Henderson, B. Thomson, and S. Young, "Deep neural network approach for the dialog state tracking challenge," in *Proceedings of the SIGDIAL 2013 Conference*, pp. 467–471, Cambridge, U.K, October 2013.
- [25] P. Xu and R. Sarikaya, "Convolutional neural network based triangular crf for joint intent detection and slot filling," in *Proceedings of the 2013 Ieee Workshop on Automatic Speech Recognition and Understanding*, pp. 78–83, Olomouc, Czech Republic, December 2013.
- [26] N. Mrkšić, D. O Séaghdha, T.-H. Wen, B. Thomson, and S. Young, "Neural belief tracker: data-driven dialogue state tracking," 2016, <https://arxiv.org/abs/1606.03777>.
- [27] X. Zhang and H. Wang, "A joint model of intent determination and slot filling for spoken language understanding," *IJCAI*, vol. 16, pp. 2993–2999, 2016.
- [28] B. Liu and I. Lane, "Attention-based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling," 2016, <https://arxiv.org/abs/1609.01454>.
- [29] M. Gašić and S. Young, "Gaussian processes for pomdp-based dialogue manager optimization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 1, pp. 28–40, 2014.
- [30] C. Tegho, P. Budzianowski, and M. Gašić, "Benchmarking uncertainty estimates with deep reinforcement learning for dialogue policy optimisation," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AB, Canada, April 2018.

- [31] T.-H. Wen, D. Vandyke, N. Mrksic et al., “A network-based end-to-end trainable task-oriented dialogue system,” 2016, <https://arxiv.org/abs/1604.04562>.
- [32] C. Kiddon, L. Zettlemoyer, and Y. Choi, “Globally coherent text generation with neural checklist models,” in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 329–339, Washington, June 2016.
- [33] T. Zhao and M. Eskenazi, “Towards End-To-End Learning for Dialog State Tracking and Management Using Deep Reinforcement Learning,” 2016, <https://arxiv.org/abs/1606.02560>.
- [34] M. Eric and C. D. Manning, “Key-value retrieval networks for task-oriented dialogue,” 2017, <https://arxiv.org/abs/1705.05414>.
- [35] T. Young, E. Cambria, I. Chaturvedi, H. Zhou, S. Biswas, and M. Huang, “Augmenting end-to-end dialogue systems with commonsense knowledge,” in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, Palo Alto, CA, USA, March 2018.
- [36] B. Liu, G. Tur, D. Hakkani-Tur, P. Shah, and L. Heck, “Dialogue Learning with Human Teaching and Feedback in End-To-End Trainable Task-Oriented Dialogue Systems,” 2018, <https://arxiv.org/abs/1804.06512>.
- [37] X. Li, Yu Wang, S. Sun, S. Panda, J. Liu, and J. Gao, “Microsoft dialogue challenge: building end-to-end task-completion dialogue systems,” 2018, <https://arxiv.org/abs/1807.11125>.
- [38] S. Young, M. Gašić, B. Thomson, and J. D. Williams, “Pomdp-based statistical spoken dialog systems: a review,” *Proceedings of the IEEE*, vol. 101, no. 5, pp. 1160–1179, 2013.
- [39] I. V. Serban, R. Lowe, P. Henderson, L. Charlin, and J. Pineau, “A survey of available corpora for building data-driven dialogue systems,” 2015, <https://arxiv.org/abs/1512.05742>.
- [40] M. A. Mosa, A. S. Anwar, and A. Hamouda, “A survey of multiple types of text summarization with their satellite contents based on swarm intelligence optimization algorithms,” *Knowledge-Based Systems*, vol. 163, pp. 518–532, 2019.
- [41] T. Bezdan, C. Stoean, A. A. Naamany et al., “Hybrid fruit-fly optimization algorithm with k-means for text document clustering,” *Mathematics*, vol. 9, no. 16, p. 1929, 2021.
- [42] M. Kabiljo, M. Vidas-Bujanja, R. Matić, and M. Zivković, “Education system in the republic of Serbia under covid-19 conditions: chatbot-academic digital assistant of the belgrade business and arts academy of applied studies,” *KNOWLEDGE-International Journal*, vol. 43, no. 1, pp. 25–30, 2020.
- [43] M. Zivkovic, C. Stoean, A. Petrovic, N. Bacanin, I. Strumberger, and T. Zivkovic, “A novel method for covid-19 pandemic information fake news detection based on the arithmetic optimization algorithm,” in *Proceedings of the 2021 23rd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, pp. 259–266, IEEE, Timisoara, Romania, December 2021.
- [44] R. Matic, M. Kabiljo, M. Zivkovic, and M. Cabarkapa, “Extensible chatbot architecture using metamodels of natural language understanding,” *Electronics*, vol. 10, no. 18, p. 2300, 2021.
- [45] J. Gupta, V. Singh, and I. Kumar, “Florence-a health care chatbot,” in *Proceedings of the 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 504–508, IEEE, China, June 2021.
- [46] P. Dotko, K. Hess, R. Levi et al., “Topological Analysis of the Connectome of Digital Reconstructions of Neural Microcircuits,” 2016, <https://arxiv.org/abs/1601.01580>.
- [47] C. Zu, Y. Gao, B. Munsell et al., “Identifying high order brain connectome biomarkers via learning on hypergraph,” *International Workshop on Machine Learning in Medical Imaging*, vol. 9, 2016.
- [48] C. M. Brent, G. Wu, Y. Gao, N. Desisto, and M. Styner, “Identifying relationships in functional and structural connectome data using a hypergraph learning method,” *International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 17, 2016.
- [49] D. Zhou, J. Huang, and B. Schölkopf, “Learning with hypergraphs: clustering, classification, and embedding,” *Advances in Neural Information Processing Systems*, vol. 23, pp. 1601–1608, 2007.
- [50] M. Liu, J. Zhang, P. T. Yap, and D. Shen, “View-aligned hypergraph learning for Alzheimer’s disease diagnosis with incomplete multi-modality data,” *Medical Image Analysis*, vol. 36, pp. 123–134, 2017.
- [51] Li Pan and O. Milenkovic, “Inhomogeneous hypergraph clustering with applications,” *Advances in Neural Information Processing Systems*, vol. 13, pp. 2308–2318, 2017.
- [52] Y. Huang, Q. Liu, and D. Metaxas, “Video object segmentation by hypergraph cut,” in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, p. 1738, IEEE, Miami, FL, USA, June 2009.
- [53] Y. Huang, Q. Liu, S. Zhang, and D. N. Metaxas, “Image retrieval via probabilistic hypergraph ranking,” in *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3376–3383, IEEE, San Francisco, CA, USA, June 2010.
- [54] L. Zhu, J. Shen, H. Jin, R. Zheng, and L. Xie, “Content-based visual landmark search via multimodal hypergraph learning,” *IEEE Transactions on Cybernetics*, vol. 45, no. 12, pp. 2756–2769, 2015.
- [55] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, “Bertscore: Evaluating Text Generation with Bert,” 2019, <https://arxiv.org/abs/1904.09675>.
- [56] Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao, “Hypergraph Neural Networks,” AAAI, NY City, 2019.
- [57] Z. Huang, W. Xu, and K. Yu, “Bidirectional lstm-crf models for sequence tagging,” 2015, <https://arxiv.org/abs/1508.01991>.
- [58] J. F. Kelley, “An Iterative Design Methodology for User-Friendly Natural Language Office Information Applications,” *ACM TOIS*, vol. 2, 1984.
- [59] C.-S. Wu, A. Madotto, E. Hosseini-Asl, C. Xiong, R. Socher, and P. Fung, *Transferable Multi-Domain State Generator for Task-Oriented Dialogue Systems*, <https://arxiv.org/abs/1905.08743>, 2019.
- [60] J. Armand, E. Grave, P. Bojanowski, M. Douze, H. Jégou, and T. Mikolov, “Fasttext. Zip: Compressing Text Classification Models,” 2016, <https://arxiv.org/abs/1612.03651>.
- [61] E. Grave, P. Bojanowski, P. Gupta, J. Armand, and T. Mikolov, “Learning Word Vectors for 157 Languages,” 2018, <https://arxiv.org/abs/1802.06893>.
- [62] P. Kishore, 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*, pp. 311–318, Association for Computational Linguistics, Philadelphia, June 2002.
- [63] N. Thomas, *Kipf and Max Welling. Semi-supervised Classification with Graph Convolutional Networks* ICLR, Beijing, 2017.
- [64] Z. Liu, G. I. Winata, Z. Lin, P. Xu, and P. Fung, “Attention-informed mixed-language training for zero-shot cross-lingual task-oriented dialogue systems,” 2019, <https://arxiv.org/abs/1911.09273>.

- [65] J. Ramos and J. Chen, *Using Tf-Idf to Determine Word Relevance in Document Queries* CML, Piscataway, NJ, 2003.
- [66] N. Mrkšić, I. Vulić, D. Ó Séaghdha et al., “Semantic specialization of distributional word vector spaces using monolingual and cross-lingual constraints,” *Transactions of the association for Computational Linguistics*, vol. 5, pp. 309–324, 2017.
- [67] S. Schuster, S. Gupta, R. Shah, and M. Lewis, “Cross-lingual Transfer Learning for Multilingual Task Oriented Dialog,” 2018, <https://arxiv.org/abs/1810.13327>.
- [68] J. Yang and Y. Zhang, “Ncrf++: an open-source neural sequence labeling toolkit,” in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Melbourne, Australia, June 2018.