2022 7th International Conference on Computational Intelligence and Applications (ICCIA) | 978-1-6654-9584-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCIA55271.2022.9828451

Intent-Aware Long Short-Term Memory for Intelligent Training of Clinical Handover

Xiang Zhang[§]
Department of Computing
The Hong Kong Polytechnic University
Hong Kong, China
doris.x.zhang@connect.polyu.hk

Bruce X. B. Yu[§]
Department of Computing
The Hong Kong Polytechnic University
Hong Kong, China
bruce.xb.yu@connect.polyu.hk

Yan Liu
Department of Computing
The Hong Kong Polytechnic University
Hong Kong, China
csyliu@comp.polyu.edu.hk

George Wing-Yiu Ng
Multi-disciplinary Simulation and Skills Centre
Queen Elizabeth Hospital
Hong Kong, China
georgeng@ha.org.hk

Nam-Hung Chia
Multi-disciplinary Simulation and Skills Centre
Queen Elizabeth Hospital
Hong Kong, China
cnhz01@ha.org.hk

Eric Hang-Kwong So
Multi-disciplinary Simulation and Skills Centre
Queen Elizabeth Hospital
Hong Kong, China
sohke@ha.org.hk

Sze-Sze So
Multi-disciplinary Simulation and Skills Centre
Queen Elizabeth Hospital
Hong Kong, China
sss083@ha.org.hk

Victor Kai-Lam Cheung
Multi-disciplinary Simulation and Skills Centre
Queen Elizabeth Hospital
Hong Kong, China
ckl414@ha.org.hk

***Abstract*—Clinical handover is a crucial yet high-risk communication event in the provision of safe patient care. However, training standardized clinical handover in real-world scenarios often requires huge labor cost. To tackle with this issue, we propose a computer-aided method for delivering intelligent training of clinical handover at a low labor cost. Specifically, we formulate it as a continuous intent detection task that provides timely feedback during a simulated clinical handover conversation. Towards this goal, we collaborate with experts from a local hospital to collect a clinical handover dataset on real-world handover scenarios. According to the sequential nature of the handover conversation, we further propose the Intent-Aware Long Short-Term Memory (IA-LSTM) model that yields superior performance to baseline methods. Our work shows promise for the computer-aided training of clinical handover in hospitals and can encourage researchers in natural language processing to develop methods on standardized communication.**

Keywords*—intent detection, conversational system, clinical handover, standardized communication training

I. Introduction

Clinical handover refers to the transfer of information about a patient's state and care plan from one clinician to the next [1]. It is an essential part in the whole train of patient care and any missteps could result in severe consequences, such as delayed treatment, medication errors, and even mortality [2]. The situation can become even more complicated and perilous during an epidemic outbreak, due to understaffed hospitals and panic psychology [3]. To help clinical staff avoid omitting vital information, standardized clinical communication frameworks have been developed. ISBAR (Identify-Situation-Background-Assessment-Recommendation), a standardized communication framework recommended by the World Health Organization [4], has been shown of great potential to improve the transparency and accuracy of inter-professional and non-face-to-face handover in hospitals [1]. This framework provides a systematic approach to clinical handover by breaking it down into a series of important elements, namely "Identify," "Situation," "Background," "Assessment," and "Recommendation". Table I provides the description of five key elements in the ISBAR Framework. Each element can be viewed as a distinct intent for delivering pertinent clinical information, thereby ensuring the integrity of the clinical handover process.

TABLE I. Description of the ISBAR Communication Framework

Element	Description
Identify	Identify yourself, the patient and verify the receiver.
Situation	Clarify the problem or reason for contact.
Background	Briefly summarize patient's previous history relevant to the current problem.
Assessment	Share the latest clinical assessment, investigation, and your interpretation of the current situation.
Recommendation	Ask for advice or intervention; state your expectation.

However, ISBAR has not been well applied in real-world clinical handover scenarios due to the lack of effective training. According to our interviews with clinical staff, there are two major reasons for this phenomenon: First, experienced senior doctors are usually too occupied to practice communication with junior doctors; Second, practices between junior clinical staff generally lack fidelity, and they seldom receive timely and accurate feedback. Although clinical staff have been instructed in standard communication procedures in the classroom, without sufficient practice, they may fail to deliver a complete handover, especially in emergent situations.

This work was supported by Innovation and Technology Fund (ITS/110/19 from the Innovation and Technology Commission of Hong Kong.

[§] These authors contributed equally to this work.

Recently, computer-aided method has shown its promise in automating some training processes such as social skills training [5] and hotline counsellor training [6], which provides lifelike practice with low labor cost. Meanwhile, Natural Language Processing (NLP) techniques have led to transformative advances in clinical informatics research [7]. Therefore, we are motivated to use computer-aided methods to facilitate low-labor-cost and intelligent clinical communication training. Specifically, in a simulated clinical scenario, the clinician practice standardized handover through a conversational system. And this system should be able to detect the intents of clinical handover based on the ISBAR framework and give timely feedback to the clinician. Towards this goal, the problem is formulated as continuously detecting the sentence-level intent given all of the conversation history so far.

With the advancement of Deep Learning (DL), neural networks have been widely applied in intent detection and achieve great performance improvements [8]. However, there remains some challenges to directly apply existing intent detection models on standardized clinical communication. First, these models capitalize on large amounts of labelled data [9], and scaling existing intent detectors to new target domains is vulnerable to negative transfer due to the large disparity in data characteristics [10]. Second, unlike general intents, intents derived from the ISBAR framework are sequentially interrelated. This emphasizes the critical nature of considering sequence information when detecting intents in clinical communication, which is not the case with the majority of existing models.

To address the aforementioned challenges, we propose a model called Intent-aware Long Shot-Term Memory (IA-LSTM) to incorporate the sequential feature in ISBAR standardized communication. Experiments on clinical handovers in real-world scenarios demonstrate that our IA-LSTM significantly outperforms baseline models. The proposed intent-aware mechanism can be further applied to other DL models, substantially boosting their performance.

II. Related Work

Intent detection has been studied for long, where classical machine learning approaches include support vector machine, K-nearest neighbours and decision tree [11]. With the success of DL, neural networks started to be widely used for this task [12]. For DL methods, text data is first represented using word embedding, which projects sparse word representations into low-dimensional, dense vector representations [13]. Typical word embedding algorithms include word2vec [14], GloVe [15], and FastText [16]. After word embedding, different neural networks can be applied for intent detection, such as convolutional neural networks [17] and recurrent neural networks (RNN) [18]. Long Short-Term Memory (LSTM) network [19], a popular variant of RNN, has shown the great power in modelling the temporal relationship of text and capturing long-term dependencies. It uses memory cell and gates to control the information flow and solves the gradients vanishing and exploding problems in vanilla RNN training [19]. Based on the LSTM structure, further developments have been made by adding bidirectional mechanism [20], attention mechanism [21], hierarchical structures [22] and convolutional layers [23].

Despite the superiority in handling sequential data, RNN-based models requires input data to be processed in order, which limits the speed of training. Transformer [24] solved this issue by using self-attention blocks solely - all tokens are processed at the same time and attention weights between them are calculated. In this way, Transformer facilitates more parallelization during training and enables training on larger datasets. It is now replacing older RNN models and leading to the development of large pretrained systems [25]. BERT (Bidirectional Encoder Representations from Transformers) [26], one of most popular pretrained models, has achieved state-of-the-art performance in many NLP tasks. The BERT-base model is constructed by 12 layers of transformer blocks with 100M parameters, and pretrained over 3.3 billion word corpus [26]. It then can be used for a wide range of NLP tasks simply by fine-tuning on the specific task without architecture modifications.

In addition to the development of neural network structures, significant progress has been made on intent detection by integrating extra information, e.g., learning with external knowledge [27], and taking slot-filling as joint tasks [28]. As discussed in Section I, in standardized clinical handover, sequential information plays an important role in understanding the intent of the current sentence. Therefore it can serve as the extra information for intent detection. How to model the sequential information varies with different classification tasks. For example, Wu et al. [29] constructed propagation graph to model the sequence of message spreading for online rumor detection; Zhou and Li [30] used headings and sentence locations as the sequential information for medical paper section identification.

In a clinical handover conversation, contents are organized under the guidance of the ISBAR framework. Specifically, the ISBAR standard communication framework divides clinical handover into five intents (i.e. Identify, Situation, Background, Assessment, and Recommendation), connecting in the specific order. Thus, the order of intents can be used to represent the sequential information in a clinical handover conversation. Unlike previous work with explicit sequential information [29], [30], the true order of intents are unknown in our task, and we only have the model's predicted ones. Hence, we can use the order of detected intents to simulate the sequential information, which unfolds as intent detection progresses.

III. Method

In this section, we formulate the problem of intent detection on an ongoing clinical handover and introduce our proposed IA-LSTM model in detail.

A. Problem Formulation

Given a conversation with N sentences from the clinician's side, we denote it as $\mathcal{D} = \{(s^{(n)}, y^{(n)}) \mid n \in \mathbb{Z}, 1 \leq n \leq N\}$, where $s^{(n)}$ is the n-th sentence and $y^{(n)}$ is the corresponding intent label represented in one-hot encoding. We further denote the n-th sentence as a sequence of word embeddings $s^{(n)} = (w_1, \dots, w_t, \dots, w_T)$, where T is the number of words in $s^{(n)}$, and $w_t \in \mathbb{R}^M$ is a M-dimensional word embedding of the t-th word. In the ongoing setting, we only have the first n sentences of the conversation when $s^{(n)}$ is given out. Thus, we formulate the problem as an objective of learning a model G for intent detection on a subset of the conversation, which can be written as

$$\hat{\boldsymbol{y}}^{(n)} = G\left(\boldsymbol{s}^{(n)}, \Theta\right) \quad (1)$$

where Θ is the parameters of the model G, and $\hat{\boldsymbol{y}}^{(n)} = \left(\hat{y}^{(1)}, \dots, \hat{y}^{(n)}\right)$ is the model predictions given the input sentences $\boldsymbol{s}^{(n)} = \left(s^{(1)}, \dots, s^{(n)}\right)$ in the conversation $\mathcal{D}$.

A general procedure of applying DL models to intent detection consists of three steps: preprocessing, vectorization, and classification, as illustrated in Fig. 1. The first step preprocesses the raw input sentence and to obtain a sequence of tokens. The second step vectorizes each token by word embedding. Taking embedded vectors as input, the third step classifies the intent with DL models such as RNN and CNN. In the ongoing setting, sentences are given in a consecutive manner in a clinical handover conversation, corresponding to a chain of intents.

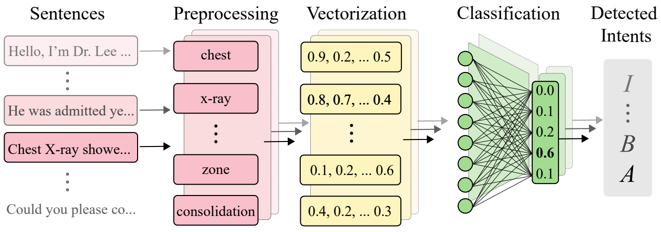

Fig. 1. Procedures of intent detection, including preporcessing, vectorization, and classification. Under the ongoing setting, we predict the intent vector of the first n sentences in the clinical handover.

B. Intent-aware LSTM

1) Sentence Representation: Given the vectorized word embeddings $s^{(n)} = (w_1, \dots, w_t, \dots, w_T)$, many DL models could be adopted to learn representations of the input sentence. It remains controversial which DL model can perform better especially on relatively small datasets [31]. For the ease of understanding, we adopt a basic LSTM [19] as the backbone model. The LSTM unit uses three different gates to regulate the information flow from previous steps to the current step: an input gate, an output gate, and a forget gate. At each time step $t \in [1, \dots, T]$, for its corresponding embedding w_t, LSTM calculates its current hidden state output vector h_t based on a memory cell c_t and an output gate g^o as

$$\begin{aligned} g^o &= \sigma(W^o h_{t-1} + I^o w_t) \\ h_t &= tanh\,(g^o \odot c_t) \end{aligned} \quad (2)$$

where W^o and I^o are weight and projection matrices, respectively. σ represents the logistic sigmoid function, and $\odot$ is the element-wise multiplication. While the memory cell c_t is calculated with three gates that can be defined as

$$\begin{aligned} g^c &= \sigma(W^c h_{t-1} + I^c w_t) \\ g^f &= \sigma(W^f h_{t-1} + I^f w_t) \\ g^u &= \sigma(W^u h_{t-1} + I^u w_t) \\ c_t &= g^f \odot c_{t-1} + g^u \odot g^c \end{aligned} \quad (3)$$

where g^c, g^f, and g^u are the activation vectors of the cell state, output, and input gates, respectively; The recurrent weight matrices are denoted by W^c, W^f, and W^u; The projection matrices are represented as I^c, I^f, and I^u. For the input sequence $s^{(n)} = (w_1, \dots, w_t, \dots, w_T)$, we take the last hidden state h_T of the LSTM model as its sentence representation $\hat{s}^{(n)}$, which will be used for intent detection.

2) Intent-Aware Design: Fig. 2 shows the structure of our proposed IA-LSTM with an intent-aware mechanism that incorporates the preceding intents for the intent detection of the current input. Let $p^{(n)} \in \mathbb{R}^C$ denote the probability distribution vector of the intent information for the n-th sentence $s^{(n)} \in \mathbb{R}^{D \times T}$, where C is the number of intent labels. Given the current input sentence $s^{(n)}$, we have $\boldsymbol{p}^{(n-1)} = \left(p^{(n-k)}, \dots, p^{(n-1)}\right)$, $1 \leq k < n - 1$, indicating its detected preceding intents in the probability distribution format. Fig. 2 illustrates the model structure when $k = 1$ (i.e., $\boldsymbol{p}^{(n-1)} = p^{(n-1)}$). To calculate $p^{(n)}$, we first concatenate the latest previous intent probability distribution vector $p^{(n-1)}$ with the current sentence representation $\hat{s}^{(n)}$. Then, using a fully connected layer and a SoftMax layer, our IA-LSTM makes predictions based on the concatenated representation. The inference process can be written as

$$p^{(n)} = \text{SoftMax}\left(F\left(\left[\boldsymbol{p}^{(n-1)}, \hat{s}^{(n)}\right]\right)\right) \quad (4)$$

where F is the fully-connected layer. For the first sentence in conversation $\mathcal{D}$, we set its preceding intent information as a C-dimensional zero vector.

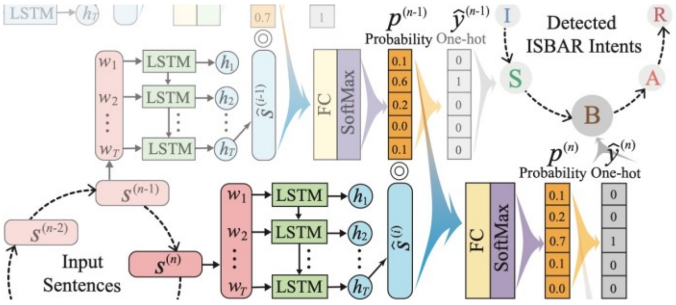

Fig. 2. Structure of IA-LSTM. The intent-aware mechanism is realized by concatenating intent information (i.e., $p^{(n-1)}$, the probability distribution after the SoftMax layer) of the preceding sentence $s^{(n-1)}$ with the current sentence representation $\hat{s}^{(n)}$ learned from the backbone model LSTM (i.e., the last hidden state h_T).

3) Optimization: Given the proposed IA-LSTM model G as defined in Equation 4, we use the cross-entropy loss to optimize it, which can be written as

$$\underset{\Theta}{\arg min} - \sum_{n=1}^{N} \sum_{c=1}^{C} y_c^{(n)} log\left(p_c^{(n)}\right) \quad (5)$$

IV. Experiments

In this section, we introduce the experimental setting, which includes the dataset, baseline methods, and hyper-parameters. We then present experiment results and discuss the effectiveness of IA-LSTM in intent detection and the generalizability of the intent-aware mechanism.

A. Experimental Setting

1) Dataset: In collaboration with Queen Elizabeth Hospital, we collected clinical handovers for both medical and surgical cases. During the handover process, a junior doctor reported the case to a senior doctor based on the related documents including medical records, notes, and various testing reports (e.g., hematology reports, chemical pathology reports, CT scans). To protect patients' privacy, we fabricated all personal information in the conversations. Finally, we gathered 100 recordings from the junior doctor's side, totalling 1895 sentences. Each sentence was labeled by the clinical expert based on the ISBAR framework. Here, we divided all sentences into training, validation, and test splits with a ratio of 6:2:2. The distribution of sentences of each intent is presented in Table II.

TABLE II. DATA SPLITS OF COLLECTED CLINICAL HANDOVERS.

Split Item	Total	I	S	B	A	R
#Training	1159	141	61	327	455	175
#Validation	366	40	18	91	151	66
#Test	370	49	18	95	156	52

2) Baselines: **LSTM** [19] is same as the backbone model used in our IA-LSTM. For LSTM and all its variants, we use 1 layer and set the hidden size as 16.

Bidirectional LSTM (BiLSTM) [20] uses two LSTMs taking the input in both forwards and backwards directions.

Attention-based LSTM (AttLSTM) [21] learns attention information from the embedding representation to guide the model in focusing on specific parts of the sentence for the classification task.

TextCNN [17] presents an implementation of CNN on NLP tasks which puts word embeddings into three separate convolutional layers and concatenate their output to a linear layer. Following the setting in the original paper, we use three kernel sizes (2, 3, 4) and each has 5 kernels.

Recurrent CNN (RCNN) [23] represents a sentence with a concatenation of the output of BiLSTM and the word embedding of GloVe [15].

Transformer [24] is a multi-head self-attention structure that has outperformed RNN/CNN based models on machine translation tasks with faster training speed. We used a 1-layer two-head encoder and average the encoder output layer before the fully connected layer.

BERT [26] is a deep bidirectional Transformer architecture pretrained over a 3.3 billion word corpus. It has shown state-of-the-art performance on many NLP tasks. Here we fine-tune on the BERT-base model and connect the output of the first token (the [CLS] token) to a fully connected layer.

Concatenate BiLSTM (CLSTM) [30] uses the nearby sentences processed by BiLSTM to aid classification of the current sentence. In our experiment, k sentences before the current sentence are used to model sentence-level interdependencies.

3) Hyper-parameters: We use glove.6B.50d [15] for embedding initialization (except for BERT), which is trained on Wikipedia 2014 and Gigaword5 with 6B tokens and 400K vocabularies. For Transformer and BERT, the sentence length is fixed to 32, while other models take inputs of variable lengths. We use a batch size of 16 to train BERT and 1 for other models. And we set k = 1 for CLSTM and IA-LSTM. During the training process, Adam optimizer is used for all the models. We set the learning rate to 1e-3, drop out rate 0.2. For parameters in the original BERT model, we set the learning rate to 1e-5.

For a fair comparison, we conduct five rounds of experiments for all models with five random seeds (1, 12, 123, 1234, and 12345) and record the test accuracy when each model achieves the best performance on the validation set within 50 epochs. We report the average test accuracy of five rounds for all models.

B. Results and Discussion

Since the dataset contains an unbalanced proportion of different classes, we report both *Accuracy* and *Macro F1-Score* as performance measures.

1) Effectiveness on Intent Detection: Table III shows a comparison of baseline models and the proposed IA-LSTM on the collected clinical handovers. Among all baseline models, BERT achieves the highest accuracy, 84.86%. And CLSTM improves BiLSTM by considering the interdependencies of nearby sentences, reaching an accuracy of 83.68%. Representing the sequential information by intent labels, our IA-LSTM outperforms all baselines with noticeable improvements: our model's performance surpasses the state-of-the-art BERT with an enhanced accuracy of 3.57%. IA-LSTM also significantly improves the results of its backbone model LSTM (from 81.84% to 88.43%), demonstrating the effectiveness of the intent-aware mechanism. With an accuracy of 88.43%, our model can feasibly be deployed to a standardized clinical communication training system.

TABLE III. PERFORMANCE OF BASELINES AND IA-LSTM (%).

Model	Accuracy	F1-Score
LSTM [19]	81.84	77.47
BiLSTM [20]	79.78	74.47
AttLSTM [21]	81.62	78.04
TextCNN [17]	79.90	74.14
RCNN [23]	82.86	78.51
Transformer [24]	78.92	73.39
BERT [26]	84.86	81.09
CLSTM [30]	83.68	84.36
Our IA-LSTM	**88.43**	**85.76**

Discussion: how does LSTM benefit from the intent-aware design? To further investigate how IA-LSTM improves the intent detection in each class, we provide the confusion matrices for LSTM and IA-LSTM (see Fig. 3). These matrices show that all detection accuracy along the diagonal are enhanced, with the detection accuracy of intents A and S achieving more noticeable improvements. This outcome is consistent with fact that intents

S and A are closely related and difficult to distinguish. Our intent-aware design enables the model to look at the preceding sequences and infer the difference.

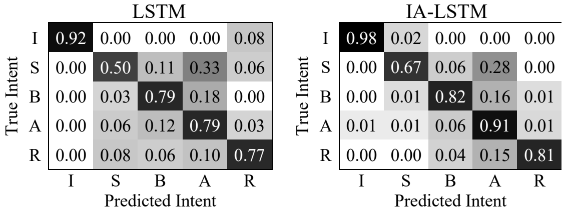

Fig. 3. Confusion matrices of LSTM and IA-LSTM on clinical handovers.

Table IV displays two consecutive sentences in a conversation which is classified wrongly by LSTM. Without knowing the intent of the previous sentence, LSTM classifies the target sentence as S (Situation). Indeed, this sentence could serve as a summary of the patient's situation or a reason for calling. However, based on the preceding sentence, it is clear that the subsequent sentence presents an assessment. With the intent information, IA-LSTM is able to classify the target sentence as A (Assessment).

TABLE IV. Two Consecutive Sentences In a Conversation.

	Sentence	Intent
(Previous)	And I have revealed the CT scan ...	A
(Target)	So I think the problem is that the patient suffered from an acute gangrenous appendicitis and probably and very likely perforation.	A

We have discussed the performance of IA-LSTM when $k = 1$, which outperforms all baseline models (see Table III). To further validate the effectiveness of our IA-LSTM, we perform ablation on the value of k – the number of preceding intents incorporated in the model. Fig. 4 shows the performance of IA-LSTM when choosing different values of k. With different values of k, the model maintains stable performance that is significantly better than when no preceding intents are included (Accuracy 81.84%, F1-Score 77.47%). It is observed that increasing the value of k improves performance further, with the highest accuracy and F1-Score being 90.91% and 90.51% when $k = 4$. This demonstrates the effectiveness and potential of the intent-aware design once again. From the sharp improvement before $k = 4$ and the slight decline after, we can learn that the most recent preceding intents are more important in understanding the current sentence, whereas the intents further away provide limited information, which is consistent with common human conversation.

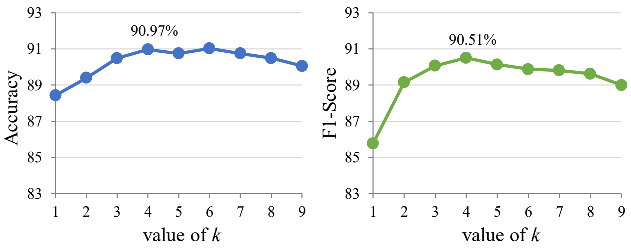

Fig. 4. The performance of IA-LSTM with different k values (%). The nodes on the lines indicate the average results of implementations with 5 different seeds. The shade areas indicate the upper and lower bands for the results.

2) Generalization to other DL Models: Given the effectiveness of our intent-aware design, we further expand it to other baseline models. Most DL models for text classification contain a fully connected layer as the final layer to predict the label. We denote the input of the final layer as a sentence representation $\hat{s}$. Based on the same idea in IA-LSTM, we extend our intent-aware design to general DL models by concatenating the intent vector with $\hat{s}$ and passing it to a fully connected neural network. For the ease of comparison, we set k to 1 for all the expanded implementations.

TABLE V. Performances of Baselines with and without the Intent-Aware Design (%).

Model	**w/o Intent-aware**		**w/ Intent-aware (*k*=1)**	
	Accuracy	***F1-Score***	***Accuracy***	***F1-Score***
LSTM	81.84	77.47	88.41 (↑ **6.61**)	85.81 (↑ **8.31**)
BiLSTM	79.78	74.47	88.41 (↑ **8.61**)	86.61 (↑ **12.11**)
AttLSTM	81.62	78.04	90.11 (↑ **8.41**)	87.91 (↑ **9.91**)
TextCNN	79.90	74.14	88.41 (↑ **8.51**)	86.71 (↑ **12.61**)
RCNN	82.86	78.51	90.31 (↑ **7.51**)	88.81 (↑ **10.31**)
Transformer	78.92	73.39	88.11 (↑ **9.21**)	85.81 (↑ **12.41**)
BERT	84.86	81.09	88.21 (↑ **3.31**)	86.41 (↑ **5.31**)

Table V shows a comparison of baseline models and their intent-aware versions. Notably, our approach to incorporating proceeding intent information is robust to different model structures and model sizes. All models' performance improve greatly, with RCNN achieving the best accuracy of 90.31%. This general improvement reflects strong correlations between sentences, consistent with our observation.

Discussion: why is the improvement of BERT not that significant? The BERT-base model used in this paper contains 110M parameters, which is almost a thousand times larger than the other baselines. To expand our intent-aware design to BERT, we concatenate the intent vector p with $\hat{s}$ generated by BERT (the output of the [CLS] token) and pass it to a fully connected layer. It is worth noting that different models may have different dimensions of $\hat{s}$, but the same dimension of p (i.e., 5). For other baseline models, the dimension of $\hat{s}$ is between 16 to 50: LSTM has $\hat{s}$ of dimension 16, BiLSTM 32, AttLSTM 16, TextCNN 15, RCNN 16, and Transformer 50. However, BERT generates a $\hat{s}$ of dimension 768, much larger than other baseline models. When confronted with this dominant vector size (768 vs. 5), the intent-aware design still improves the accuracy of vanilla BERT by 3.3%, which again confirms the effectiveness and generalization ability of our intent-aware mechanism.

V. Conclusion

In conclusion, this paper facilitates intelligent training of standardized clinical handover by solving the continuous intent detection problem. We propose a novel intent-aware algorithm IA-LSTM based on the sequential feature of standardized clinical communication. Extensive experiments and comparisons on clinical handovers have verified the effectiveness and generalization ability of our intent-aware design. Our work lays a foundation for the deployment of clinical communication training systems. This initial attempt of applying NLP techniques to the clinical communication training will promote the development of communication training systems and inspire researchers in the intent detection domain.

Acknowledgment

This work was supported by Innovation and Technology Fund (ITS/110/19) from the Innovation and Technology Commission of Hong Kong.

References

[1] S. Marshall, J. Harrison, and B. Flanagan, “The teaching of a structured tool improves the clarity and content of interprofessional clinical communication,” *BMJ Quality & Safety*, vol. 18, no. 2, pp. 137–140, 2009.

[2] M. Leonard, S. Graham, and D. Bonacum, “The human factor: the critical importance of effective teamwork and communication in providing safe care,” *BMJ Quality & Safety*, vol. 13, no. suppl 1, pp. i85–i90, 2004.

[3] K.M.Chow,B.M.Law,M.S.Ng,D.N.Chan,W.K.So,C.L.Wong, and C. W. Chan, “A review of psychological issues among patients and healthcare staff during two major coronavirus disease outbreaks in china: Contributory factors and management strategies,” *International journal of environmental research and public health*, vol. 17, no. 18, p. 6673, 2020.

[4] W. P. Safety and W. H. Organization, “Patient safety curriculum guide: Multi-professional edition,” 2011.

[5] H. Tanaka, H. Negoro, H. Iwasaka, and S. Nakamura, “Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders,” *PloS one*, vol. 12, no. 8, p. e0182151, 2017.

[6] L. K. Fryer, K. Nakao, and A. Thompson, “Chatbot learning partners: Connecting learning experiences, interest and competence,” *Computers in Human Behavior*, vol. 93, pp. 279–289, 2019.

[7] S.Velupillai,H.Suominen,M.Liakata,A.Roberts,A.D.Shah,K.Morley, D. Osborn, J. Hayes, R. Stewart, J. Downs, *et al.*, “Using clinical natural language processing for health outcomes research: overview and actionable suggestions for future advances,” *Journal of biomedical informatics*, vol. 88, pp. 11–19, 2018.

[8] J. Liu, Y. Li, and M. Lin, “Review of intent detection methods in the human-machine dialogue system,” in *Journal of Physics: Conference Series*, vol. 1267, p. 012059, IOP Publishing, 2019.

[9] H. S. Bhathiya and U. Thayasivam, “Meta learning for few-shot joint intent detection and slot-filling,” in *Proceedings of the 2020 5th International Conference on Machine Learning Technologies*, pp. 86–92, 2020.

[10] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, “A comprehensive survey on transfer learning,” *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.

[11] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, “Text classification algorithms: A survey,” *Information*, vol. 10, no. 4, p. 150, 2019.

[12] H. Weld, X. Huang, S. Long, J. Poon, and S. C. Han, “A survey of joint intent detection and slot-filling models in natural language understanding,” *arXiv preprint arXiv:2101.08091*, 2021.

[13] T. Mikolov, E´. Grave, P. Bojanowski, C. Puhrsch, and A. Joulin, “Advances in pre-training distributed word representations,” in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

[14] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in *Advances in neural information processing systems*, pp. 3111– 3119, 2013.

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

[16] A.Joulin,E.Grave,P.Bojanowski,M.Douze,H.Je´gou,andT.Mikolov, “Fasttext. zip: Compressing text classification models,” *arXiv preprint arXiv:1612.03651*, 2016.

[17] Y.ZhangandB.C.Wallace,“Asensitivityanalysisof(andpractitioners’ guide to) convolutional neural networks for sentence classification,” in *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 253–263, 2017.

[18] A. Bhargava, A. Celikyilmaz, D. Hakkani-Tu¨r, and R. Sarikaya, “Easy contextual intent prediction and slot detection,” in *2013 ieee international conference on acoustics, speech and signal processing*, pp. 8337– 8341, IEEE, 2013.

[19] S. Hochreiter, J. urgen Schmidhuber, and C. Elvezia, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[20] A. Graves, A.-r. Mohamed, and G. Hinton, “Speech recognition with deep recurrent neural networks,” in *IEEE international conference on acoustics, speech and signal processing*, pp. 6645–6649, Ieee, 2013.

[21] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” in *International Conference on Learning Representations*, 2015.

[22] X.NiuandY.Hou,“Hierarchicalattentionblstmformodelingsentences and documents,” in *International Conference on Neural Information Processing*, pp. 167–177, Springer, 2017.

[23] S. Lai, L. Xu, K. Liu, and J. Zhao, “Recurrent convolutional neural networks for text classification,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, 2015.

[24] A.Vaswani,N.Shazeer,N.Parmar,J.Uszkoreit,L.Jones,A.N.Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *NIPS*, 2017.

[25] T. Wolf, J. Chaumond, L. Debut, V. Sanh, C. Delangue, A. Moi, P. Cistac, M. Funtowicz, J. Davison, S. Shleifer, *et al.*, “Transformers: State-of-the-art natural language processing,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 38–45, 2020.

[26] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 4171–4186, 2019.

[27] W. Wang, S. Hosseini, A. H. Awadallah, P. N. Bennett, and C. Quirk, “Context-aware intent identification in email conversations,” in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 585–594, 2019.

[28] Y.-B. Kim, S. Lee, and K. Stratos, “Onenet: Joint domain, intent, slot prediction for spoken language understanding,” in *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 547–553, IEEE, 2017.

[29] Z. Wu, D. Pi, J. Chen, M. Xie, and J. Cao, “Rumor detection based on propagation graph neural network with attention mechanism,” *Expert systems with applications*, vol. 158, p. 113595, 2020.

[30] S. Zhou and X. Li, “Feature engineering vs. deep learning for paper section identification: Toward applications in chinese medical literature,” *Information Processing & Management*, vol. 57, no. 3, p. 102206, 2020.

[31] A. Ezen-Can, “A comparison of lstm and bert for small corpus,” *arXiv preprint arXiv:2009.05451*, 2020.