IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 13, 2020, accepted April 17, 2020, date of publication April 22, 2020, date of current version May 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2989665

Multi-Attention-Based Capsule Network for Uyghur Personal Pronouns Resolution

QIMENG YANG[1], LONG YU[2], SHENGWEI TIAN[3], AND JINMIAO SONG[1]
[1]College of Information Science and Engineering, University of Xinjiang, Urumqi 830000, China
[2]Network Center, University of Xinjiang, Urumqi 830000, China
[3]School of Software, University of Xinjiang, Urumqi 830000, China

Corresponding author: Long Yu (yul@xju.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61563051, Grant 61662074, and Grant 61262064, in part by the Key Project of National Natural Science Foundation of China under Grant 61331011, and in part by the Xinjiang Uyghur Autonomous Region Scientific and Technological Personnel Training Project under Grant QN2016YX0051.

ABSTRACT Anaphora resolution of Uyghur is a challenging task because of complex language structure and limited corpus. We propose a multi-attention based capsule network model for Uyghur personal pronouns resolution, which can obtain the multi-layer and implicit semantic information effectively. Independently recurrent neural network (IndRNN) is applied in this model to achieve the interdependent features with long distance. Moreover, the capsule network can extract richer textual information to improve expression ability. Compared with the single attention-based model which combines Long Short-Term Memory (LSTM), the multi-attention based capsule network can capture multi-layer semantic information through a multi-attention mechanism without using any external parsing results. Experimental results on Uyghur dataset show that our approach surpasses the state-of-the-art models and gets the highest F-score of 83.85%. Meanwhile, our experimental results demonstrate the proposed method can effectively improve the performance of Uyghur personal pronouns resolution.

INDEX TERMS Capsule network, anaphora resolution, IndRNN, attention mechanism.

I. INTRODUCTION

Anaphora, as a special linguistic phenomenon, is pervasive in the expression of natural language. It is useful to simplify expression and maintain language coherence. Unambiguous interpretation of the anaphora part is conducive to machine analysis and text understanding. Personal pronouns resolution is the task of finding the correct antecedent for a given pronominal anaphor in a document [1]. Following shows an example of personal pronoun in Uyghur document, where personal pronouns are represented as "φ".

[زەھەر چەكلەش] ساقچىلىرى خەلق ئۈچۈن بىرىنچى سەپتە كۈرەش قىلىۋاتقان ، [ئۇلار] ھەر ۋاقىت خەتەر ئالدىدا ، [مەن] بۇنىڭدىن بەك تەسىرلەندىم.

(The [anti-drug police] are struggling for the people on the front line. [They]φ1 are always facing danger and [I]φ2 am deeply touched.)

A personal pronoun can be an anaphoric personal pronoun if it coreferes to one or more mentions in the associated text, or unanaphoric, if there are no such mentions. In this example, the first pronoun "φ1" is anaphoric and coreferes to the mention "زەھەر چەكلەش (anti-drug police)" while the personal pronoun "φ2" is unanaphoric. These mentions that contain the important information for interpreting the personal pronoun are called the antecedents [2].

Recent advances in deep learning models have shown superior performance in natural language processing tasks. Lu and Ng proposed an adversarial attention model for the task of multidimensional emotion regression [3]. Zhu *et al.* explored multi-channel graph neural network for entity alignment task [4]. Cao *et al.* introduced a multi-channel CNN based inner-attention for compound sentence relation classification [5].

Anaphora resolution is an important sub-task in natural language processing. In recent years, deep learning models for anaphora resolution have been widely investigated [6]–[11]. These methods concentrate on anaphoric pronoun resolution, applying numerous neural network models to pronoun-candidate antecedent prediction. Deep learning

The associate editor coordinating the review of this manuscript and approving it for publication was Lefei Zhang.

models have demonstrated their capabilities to learn vector-space semantics of pronouns and their pronoun-candidate antecedent, and substantially surpass classic models, obtaining state-of-the-art experiment results on the benchmark dataset.

Though these previous methods have achieved ideal performance, all of these studies are based on English or Chinese with large-scale corpus. However, the study of anaphora resolution in minority languages still remains a huge challenge for several reasons. For minority language research, both corpus annotation and entity recognition need to master multi-level grammar knowledge, semantic knowledge, and even corresponding language domain knowledge. In the current research stage of natural language processing (NLP), it is still difficult to acquire and learn this knowledge. Meanwhile, most machine learning methods overly rely on handcrafted features which require numerous manual design and extract, and it is time-consuming and cost-intensive. Though the problem is helped greatly by the proposal of deep learning in recent years, these neural network based approaches cannot encode and learn the word sequence dependencies and contexts efficiently. Moreover, the distance between the antecedent and the anaphora cannot be effectively identified in current research. For instance, given a sentence ''Because Yang is a scholar, Zhao respects him.'' with its candidate mention ''Yang'' and ''Zhao'', it is challenging to infer whether mention ''him'' is possible to be the antecedent of ''Yang''. In that case, the resolver may incorrectly predict ''Zhao'' to be the antecedent since ''Zhao'' is the nearest mention. Hence, a desirable model should be able to 1) take advantage cues of multi-layer semantic features to predict pronoun-candidate antecedent and 2) Analyze deep context semantics and mine word sequence dependencies and 3) identify the distance between personal pronouns and candidate antecedents

To achieve these goals, we propose a multi-attention based capsule network for personal pronoun anaphora resolution. On top of the neural network models [12]–[14], three main innovations are introduced that are capable of efficaciously leveraging multi-layer semantic features provided by personal pronouns and candidate antecedent, Mining word sequence dependencies, and identifying the distance between personal pronouns and candidate antecedent. The contributions of the paper are listed as follows:

✓ We propose a multi-attention mechanism to obtain multi-level semantic information of personal pronouns and candidate antecedents, which solves the problem of relying only on content-level features.

✓ The semantic of each sentence is represented by IndRNN, which can analyze deep context and mine word sequence dependencies.

✓ We propose a position recognition algorithm that allows the model to take full advantage of the positional information of each word in the text.

✓ The capsule network with multi-attention is devised to extract richer text information. It improves the text expression ability and acquires more important clues to improve anaphora resolution performance.

The rest of this paper is organized as follows. The next section outlines related work. Section 3 describes our multi-attention based capsule network for personal pronouns anaphora resolution. Section 4 presents our experiments, including the dataset description, hyperparameter setting, evaluation metrics, experiment results, and analysis. The Section 5 is about the conclusion and future work.

II. RELATED WORK

A. ANAPHORA RESOLUTION

The traditional anaphora resolution methods mainly focus on anaphora dictionary or machine learning approach. Soon *et al.* [15] applied a noun phrases anaphora resolution system based on decision tree to induction to two standard anaphora resolution data sets (MUC6, MUC7). It is the first time that anaphora resolution is viewed as a binary classification task: given a pair of referring expressions, the resolver has to determine whether they are anaphoric or unanaphoric. Their work is later improved by Ng and Cardie [16] in two aspects: first, the decision tree returns score instead of a hard-decision of anaphoric or not so that improved model is able to choose the ''best'' candidate on the left, as opposed the first in Soon *et al.* [15]; Second, Ng and Cardie [16] expand the feature sets of Soon *et al.* [15]. Yang *et al.* [17] proposed a competition learning model to anaphora resolution, which adopts a twin-candidate learning method. Such a model can give the competition criterion for pronoun-candidate antecedents reliably, and ensure that the most preferred candidate is chosen. The experimental results on (MUC-6 and MUC-7) data set show that the approach can surpass those based on the single-candidate method. These traditional methods identify the anaphora relationship effectively by constructing text grammar features, syntax features, and text content features. In addition, the anaphora resolution has been widely studied in many languages.

In recent years, deep learning techniques have been extensively studied [7], [18], [19] for anaphora resolution. Chen and Ng [20] introduced an unsupervised approach for this task. In this work, underlying their method is the novel idea of employing a model trained on manually resolved overt pronouns to resolve pronouns. Clark and Manning [21] apply reinforcement learning to directly optimize a neural mention-ranking model for anaphora resolution evaluation metrics, and experiment with two approaches: the reinforce policy gradient algorithm and a reward-rescaled max-margin objective.

The current methods of anaphora resolution can typically be categorized as (1) mention-ranking models, (2) entity-level models, (3) latent-tree models, (4) mention-pair classifiers. [22] The major difference between our model and previous techniques lies in the applying of multi-attention and capsule network. In this word, we propose a multi-attention based capsule network to obtain multi-level and richer text semantic

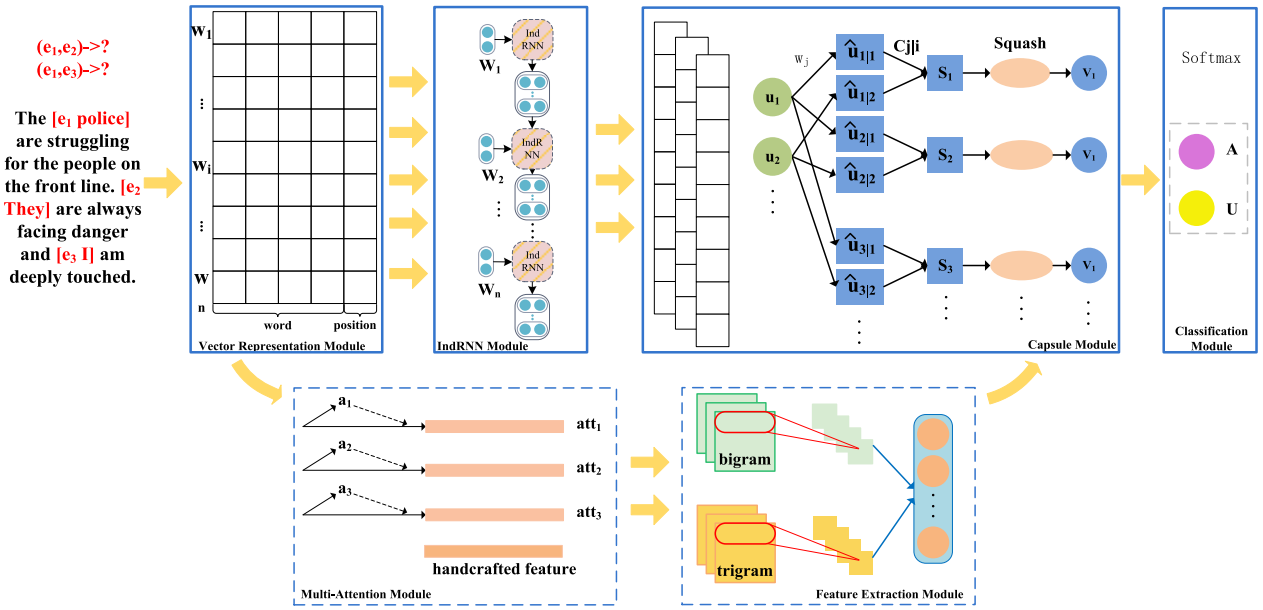


FIGURE 1. Multi-attention based capsule network for personal pronouns resolution.

information. Furthermore, we also design a position recognition algorithm, which can make effective use of position during model training.

B. LANGUAGE SPECIFIC ISSUES IN UYGHUR

Uyghur is a kind of agglutinative language, which has various forms and grammatical forms. It expresses different grammatical functions by suffixing different affixes at the end of words. For instance, given a sentence "تۇرقۇربان (Kurban feels)", and its affixes "تۇر (feels)" is connected to the name "قۇربان (Kurban)", it is expressing the meaning of "Kurban feels".

Uyghur personal pronouns are divided into first person pronouns, such as "مەن(I)" second person pronouns, such as "سىز(you)" third person pronouns, such as "ئۇ (he, she, it)". The biggest difference between the Uyghur personal pronouns and Chinese (or English) is that the Uyghur third person pronoun has no gender concept. For example, the English third person singular has "he/she/it", while in Uyghur, the third person can represent male, women and objects, so the third person is more extensive than the first person and second person, and the anaphora phenomenon is more frequent.

The characteristics of Uyghur personal pronouns are mainly influenced by the "grid" grammar. The "grid" grammar is a special form of language. The form of "grid" is different, and the additional "suffix" is different. The "grid" grammar reflects the syntactic function of noun phrases in sentences, has independence in grammatical form, and has stability in grammatical sense. It is one of the important linguistic features of Uyghur personal pronouns anaphora resolution. The "grid" grammar includes ten forms such as subject, genre, and directional.

III. MULTI-ATTENTION BASED CAPSULE NETWORK FOR PERSONAL PRONOUNS RESOLUTION

We propose a multi-attention based capsule network for personal pronouns resolution and the structure is shown in Figure.1. It consists of six modules: Vector Representation Module, Multi-Attention Module, IndRNN Module, Feature Extraction Module, Capsule Module and Classification Module. Assuming that the input sentence is $s = \{w_1, b_1, w_3 \ldots a_i \ldots w_n\}$, the goal of this model is to predict the anaphora relationship of antecedent b_1 and pronoun a_i, which will be Anaphoric (A) or Unanaphoric (U).

A. VECTOR REPRESENTATION MODULE

Represent each word as a multi-dimensional distributed vector [23]. For each sentence s_i, we use pretrained word embeddings to map each word token onto the d_w-dimensional space. Ultimately, the vector representation module encodes the sentence representation as $S = [w_1 \ldots w_i \ldots w_n] \epsilon R^{n \times d}$, where $w_i = [x_{i1} \ldots x_{ij} \ldots x_{id}]$ corresponds to the word vector of the word w_i in the sentence. We propose a bi-directional scanning position algorithm as the combinations of the relative distances from the personal pronouns to pronoun-candidate antecedent and encode these distances in d_w-dimensional vectors.

B. INDRNN MODULE

The Independently Recurrent Neural Network (IndRNN) was first proposed by Li *et al.* [24]. It's a variant of Recurrent Neural Networks (RNN). IndRNN solves the gradient disappearance and gradient explosion problems of traditional RNN and also learns long-term dependencies, especially in text processing. We make use of IndRNN to deeply learn the semantic meaning of a sentence and the long-term

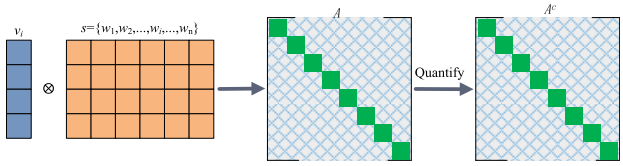


FIGURE 2. The operation of multi-attention mechanism.

dependencies of words quickly. We concatenate the current memory cell hidden state vector h_t of IndRNN as the output vector $h_t = [\vec{h_t}] \epsilon R^B$ at time t, where B denotes the dimensionality of IndRNN.

C. MULTI-ATTENTION MODULE

The multi-attention mechanism enables the model to focus on the target words with different feature information during the training process, and learn more hidden information of the word, so as to better identify the candidate antecedent. For the sentence $s = \{w_1, w_2, w_3 \ldots w_i \ldots w_n\}$, we extract the word vector of word w_i as the attention matrix, and perform an inner product operation between the attention matrix and the word vector of sentence s to obtain the attention feature matrix C^T. As shown in Figure 2, where C^T is a diagonal matrix.

$$C_{i,i} = innerproduct(v_i, x_i) \tag{1}$$

$$C^T_{i,i} = \frac{\exp(C_{i,i})}{\sum_{j=1}^{n} \exp(C_{j,j})} \tag{2}$$

Finally, we use the attention feature matrix C and the original word vector to obtain the input matrix of the model.

$$z^c_i = x_i \oplus C^T_{i,i} \tag{3}$$

$$z^c_i = x_i \cdot C^T_{i,i} \tag{4}$$

Both methods can be used to calculate the input matrix. In our experiment, we use Equation (3) to calculate the input matrix.

In this work, we propose three types of attention mechanisms to construct the model, namely, word vector attention mechanism, distance attention mechanism and part-of-speech (POS) attention mechanism. In addition, the above two attention mechanisms are calculated in the same way as the word vector attention mechanism.

1) WORD VECTOR ATTENTION MECHANISM

The attention mechanism allows the model to focus on key information during the training process to achieve better classification results. For the personal pronouns anaphora resolution task, the text content level information is most important. Analysis of anaphor and candidate antecedent semantic information from multiple aspects can improve classification performance.

We propose a word vector attention mechanism for Uyghur personal pronouns resolution task. For the sentence $s = \{w_1, w_2, w_3 \ldots w_i \ldots w_n\}$, the w_i word vector is extracted as the word vector attention matrix, and then the word vector attention matrix and the word vector matrix are operated to obtain the word vector attention feature matrix C^T. The operations are shown in Equation (5)-(6).

$$e^t_i = f_{ATT}(z_{t-1}, w_i, \{a^{t-1}_j\}^M_{j=1}) \tag{5}$$

$$C^T_{i,i} = \frac{\exp(e^t_i)}{\sum_{j=1}^{M} \exp(e^t_i)} \tag{6}$$

The matrix C^T indicates the importance of each word, which can be reflected by the score, so the attention feature matrix C^T can be rewritten into Equation (7).

$$C^T_{i,i} = \varphi\left(\{w_i\}^M_{i=1}, \{a^t_i\}^M_{i=1}\right) = \sum_{i=1}^{M} a_i w_i \tag{7}$$

The model input matrix can be obtained by using the attention feature matrix C^T and the word vector matrix w_i. The operations are shown in Equation (8).

$$Input^t_i = x_i \oplus C^T_{i,i} \tag{8}$$

where $\oplus$ represents the splicing operation, our method constructs the model input matrix by using the attention feature matrix and the original word vector splicing operation.

2) PART-OF-SPEECH ATTENTION MECHANISM

The content-level information of the anaphora chain is the key to the anaphora resolution. However, in the case where the word segmentation error and the low coverage of the antecedent in the data set, this method of relying only on the content-level information will reduce the performance of the experiment. To solve this problem, we propose an attention mechanism based on POS, which combines the word vector attention mechanism as the input of the network. We re-label the POS of each word in the text, which allows the model to learn the contact information between the anaphor and the candidate antecedents.

For the sentence $s = \{w_1, w_2, w_3 \ldots w_i \ldots w_n\}$, we re-label each word in the sentence, as shown in Figure 3. The result of the annotation is a combination of words and POS. For a sentence of length n, the result of the annotation is as shown in Equation (9), where w_i is the ith word, c_i is the POS, and $\oplus$ is the splicing operation.

$$z_{1,n} = w_1 \oplus c_1, w_2 \oplus c_2, w_3 \oplus c_3 \ldots w_n \oplus c_n \tag{9}$$

For the case where the antecedent is a noun phrase, since the noun phrase contains multiple words, the processing is different. In this case, we extract the word vector attention matrix of all words in the noun phrase, and obtain the POS attention feature matrix according to the Equation (10):

$$Z_i = \alpha \times \frac{\sum_{i=1}^{n} f_{ATT}(w_i \oplus c_i)}{n} \tag{10}$$

α is the noun phrase weight coefficient, which can be set manually or automatically during the model training process.

Like the word vector, we map each POS to a multi-dimensional continuous value vector called the POS vector $R^{K \times V}$, where K and V indicate the dictionary size and dimension of the POS vector respectively. We extract the POS vector of anaphor and candidate antecedents, and finally obtain the POS attention matrix according to Equation (5)(6).

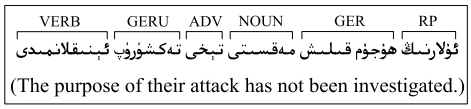


FIGURE 3. Example of POS tagging.

3) POSITION ATTENTION MECHANISM

The position of the anaphor and the candidate-antecedents hides important information, which provides key linguistic cues for anaphora resolution. We generally believe that the closer the distance between the candidate-antecedents and the anaphor, the greater the probability that there is an anaphora relationship. For instance, given a sentence "As a squad leader, Zeng Zhiqiang helps classmates and selfless dedication. We respect him very much." with its candidate mentions "we" and "him", it is challenging to infer whether mention "him" is possible to be the anaphor if it is considered separately. In that case, the resolver may incorrectly predict "we" to be the antecedent since "we" is the nearest mention. To solve this problem, we propose a bidirectional scanning algorithm to calculate the positional relationship between personal pronouns and candidate antecedents, as shown in table 1.

TABLE 1. Position recognition algorithm.

Input: a sentence of length n $s = \{w_1, w_2, w_3 \ldots a_i \ldots w_n\}$
Output: Set of positional values between candidate antecedent and personal pronoun L

1. The value of the position of the anaphor is set to 0, the value of the position of all candidate antecedents is set to n, where n is the length of the sentence;
2. Define the work pointer p to scan forward from the position of the personal pronoun a_i;
   Begin
   for epoch = 1…i **do**
   If object is not an antecedent, then $w_i \rightarrow L$,
   If object is an antecedent, calculate the distance according to the formula φ:
   $$\varphi = \frac{\sum_{i=1}^{len(L)} \alpha \times f_{ATT} L_i}{len(L)} \times min\{5, num(punc)\}$$
   If object is a punctuation mark, then $punc \rightarrow L$
   Set the set L to null: $L \rightarrow$Null
   end for
   End procedure

D. FEATURE EXTRACTION MODULE

We design a double-layer parallel convolutional neural network to extract and represent the anaphora chain features.

The purpose of feature extraction module is to extract semantic features of the anaphora chain, each convolution and pooling kernel corresponds to a certain part of feature and the feature mappings can be obtained after convolution and pooling operations. The convolution is operated on each attention and handcrafted ATT = $\{att_1, \ldots, att_t, \ldots, att_n\}$, which is the output of the previous multi-attention module, by Equation (11):

$$S = f\,(W * ATT + b) \tag{11}$$

where W and b represent the weight matrix and bias of the network respectively. Meanwhile, K-Max pooling is used to select the top-K value of each filter to represent the semantic information. The value of K is set to $\lfloor (len - f_s + 1)/4 \rfloor$, where len is the dimension of words and f_s is the convolution filter size.

After the feature extraction module operation, the feature vector dimension is significantly reduced, and important information is reserved.

E. CAPSULE MODULE

Capsule Network is proposed by Sabour *et al.* [25], Hinton *et al* [26]. Compared with CNN, it replaces the scalar-output feature detectors with vector-output capsules and has the ability to save additional information such as position and thickness. We combine the capsule network and IndRNN model to implement the anaphora resolution. Capsule network can extract abundant content-level information, and also anaphora chain position, syntactic and semantic structure can be encoded effectively. It improves the feature expression ability and acquires more important clues further.

In the capsule network, the activation function squash preserves the direction of the input vector and compresses the modulus of the input vector to (0,1). The output v_j is shown in Equation (12).

$$v_j = \frac{||s_j||^2}{1 + ||s_j||^2} \frac{s_j}{||s_j||} \tag{12}$$

Here v_j is the vector output of capsule j and s_j is the total input vector.

The first layer of the capsule network is a convolution layer whose activation function is ReLU. Except for the first layer capsule, the total input s_j of the capsule is the weighted sum of all prediction vectors $\hat{u}_{j|i}$, which is the output ui through the lower capsule and the weight matrix W_{ij}. The operations are shown in Equation (13)(14).

$$s_j = \sum\nolimits_i c_{ij} \hat{u}_{j|i} \tag{13}$$

$$\hat{u}_{j|i} = W_{ij} u_i \tag{14}$$

where c_{ij} is the coupling coefficient determined during the dynamic routing process, indicating the weight between each lower layer capsule and its corresponding high-level capsule. For each capsule i, the sum of all weights c_{ij} is 1. c_{ij} is determined by the softmax function in the dynamic routing algorithm. The operation is shown in Equation (15).

$$c_{ij} = \frac{exp\,(b_{ij})}{\sum_k exp\,(b_{ik})} \tag{15}$$

where b_{ij} is the logarithmic probability of capsule i and capsule j, which is used to update c_{ij} and initialize it to 0.

TABLE 2. Statistics on the training and test dataset.

word	translation	ingredient	anaphora number
چۈنكى	Because		
قۇربان	Kurban	antecedent	1
ئەنگلىيە	is		
هازىرقى	contemporary		
مەكتەپ	scholar		
ئەركىن	Elken	anaphor	2
هۆرمەتلەيدۇ	respects		
ئۇنى	him	PP	1

TABLE 3. Statistics on the training and test dataset.

	#Documents	#Number
Training	341	38487
Test	86	6084

During route iteration, b_{ij} will be continuously updated, as shown in Equation (16).

$$b_{ij} = b_{ij} + \hat{u}_{j|i} \cdot v_j \tag{16}$$

F. CLASSIFICATION MODULE

For anaphora resolution, we calculate the length of the vector v_j which represents the probability of each anaphora chain. Finally, we choose the anaphora chain with the largest v_j value as the result of the anaphora resolution.

IV. EXPERIMENTS

A. DATASET

Most of the current research on the anaphora resolution is based on Chinese and English. There are few studies on minority languages such as Uyghur, and there is no public corpus.

For the above problems, we collect and screen Uyghur raw data and perform dataset annotation under the guidance of natural language processing experts. For the following sentences, the anaphora information is annotated as shown in Table 2.

چۈنكى قۇربان ئەنگلىيە هازىرقى زامان ئالىي مەكتەپ قىلغۇچىلارنىڭ
بىرى ، شۇڭا ئەركىن ئۇنى هۆرمەتلەيدۇ

(Because Kurban is a contemporary scholar, Elken respects him.)

We annotated the antecedent and anaphor in the sentence and their anaphora number. There is a anaphora relationship with the same anaphora number. In addition, for each word, we annotated its part-of-speech, semantic category, named entity type, singular and plural, syntax structure, semantic role, gender and ''grid'' grammar type.

In this experiment, a total of 427 annotation data were used, and 44571 data instances were extracted. The training data and test data statistics are shown in Table 3:

B. EVALUATION MEASURES

Following previous work [6], [8], [9], [18] on anaphora resolution, metrics employed to evaluate our model are: precision, recall, and F-score (F). We report the performance for each hyperparameter except as the overall result.

TABLE 4. Hyperparameter setting.

#Parameter	#Parameter Description	#Value
t	training epochs	80
b	batch	128
d	dropout rate	0.5
l	learning rate	0.005
o	optimizer	adam

TABLE 5. The effectiveness of each module.

Model	Prec.	Reca.	F-Scor.
IRCC	81.36	78.84	80.08
CMAC	83.17	77.76	80.37
IMACN	83.24	80.12	81.65
IMAC	84.51	81.33	82.89
Our model	86.63	81.25	83.85

C. IMPLEMENTATION DETAILS

We randomly initialize the parameters and minimize the objective function using Adagrad algorithm [27].

Moreover, applying dropout regularization [28] to each layer during the experiment can effectively accelerate model training and prevent overfitting. Based on the preliminary experiment, the experimental hyperparameters of this paper are shown in Table 4.

D. EXPERIMENT RESULTS

We propose four comparative experiments to verify the performance of our model: (1) the effectiveness of each module; (2) performance comparison of different attention mechanism models; (3) the impact of handcrafted features and position recognition algorithm on model performance; (4) experiment results of different models; (5) model performance in different personal pronouns.

To highlight the strengths and weaknesses of our model, we provide both quantitative and qualitative analyses [29]. As shown in Table 4, We verify the performance of single module rather than the ensembled model. The four models we propose are as follows:

1) IRCC: consists of IndRNN, Capsule network and CNN.
2) CMAC: Consists of Capsule network, Multi-attention and CNN.
3) IMACN: Consists of IndRNN, Multi-attention and CNN.
4) IMAC: Consists of IndRNN, Multi-attention and Capsule network.
5) Our model: Consists of IndRNN, Multi-attention, Capsule network and CNN.

As can be seen from Table 5, our model surpasses all baselines and achieves the best performance. More specifically, for the ''Overall'' results, our model obtains a considerable improvement by 0.96% in F-score over the best baseline, which demonstrates the efficiency of the proposed technique.

TABLE 6. Performance comparison of different attention mechanism models.

Model	Prec.	Reca.	F-Scor.
SATT-WV	83.14	77.26	80.09
SATT-POS	84.86	77.63	81.08
SATT-POSI	79.29	75.01	77.09
Our model	86.63	81.25	83.85

It can be seen from the analysis of the experimental results that the proposed model can effectively deal with the task of anaphora resolution. Multi-attention can obtain deeper feature information from three aspects, which makes up for the lack of single attention mechanism to pay attention to content-level information. The introduction of IndRNN and Capsule can identify the long-term dependencies of words and capture the features of word by the dynamic routing algorithm respectively. Meanwhile, the use of convolution and pooling can further extract high-dimensional features and reduce model complexity. In all words, all these suggest that each module can improve the overall performance of the model, which demonstrates the efficiency of the proposed technique.

In Table 6, we compare the results of our model with single-attention models in the Uyghur resolution dataset. Where SATT-WV, SATT-POS and SATT-POSI represent single word vector attention, single POS attention and single position attention. Our multi-attention model surpasses all baselines.

More specifically, it can be seen from the experimental results that our model obtains a considerable improvement by 2.77% in F-score over the best baseline. Compared with the single-attention mechanism, model with word vector attention (SATT-WV), part-of-speech attention (POS), and SATT-POSI (position attention) can acquire semantic information at multiple levels. In addition, our model can obtain deeper text feature information without external knowledge such as dependency parsing to effectively identify anaphora relationship.

In order to verify the effectiveness of handcrafted features and position recognition algorithm, we removed the handcrafted features and position recognition algorithm for further experiments. The experimental results are shown in Table 7.

The experimental results show that the removal of handcrafted, including only the position feature ($V_{position}$), its F-score is reduced by 14.42% compared to our model. This shows that the anaphora resolution task relies more on the representation of words in terms of rules and knowledge. The experimental results show that the performance of the model without handcrafted features is significantly reduced, which proves that the introduction of handcrafted features plays a key role in improving the performance of anaphora resolution. Compared to our model, the F-score of the removal of the position feature ($V_{handcrafted}$) is reduced by 3.76%. This shows that the position recognition algorithm can accurately calculate the distance of the anaphora chain and identify the importance of different words.

TABLE 7. The impact of handcrafted features and position recognition algorithm on model performance.

Feature type	Prec.	Reca.	F-Scor.
$V_{position}$	73.78	65.57	69.43
$V_{handcrafted}$	82.45	77.87	80.09
Our model	86.63	81.25	83.85

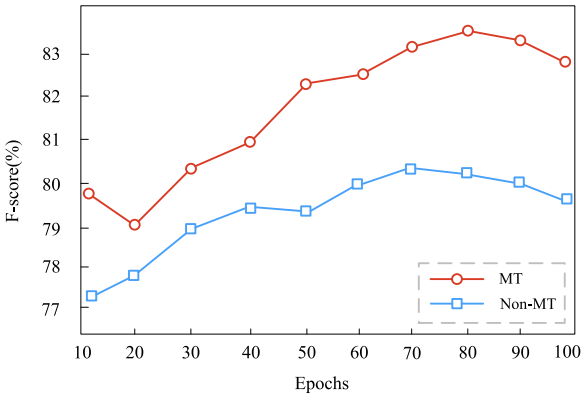


FIGURE 4. Experiment results of different models.

Ideally, our model learns multi-level semantic information from personal pronouns and candidate antecedents, which solves the problem of relying only on content-level features. Moreover, on purpose of better illustrating the effectiveness of the proposed multi-attention method, we run a set of experiments with different settings. Specifically, we compare the model with (MT) and without (Non-MT) the proposed multi-attention using different training iterations.

Figure 4 shows the performance of our model with and without multi-attention. We can see from the figure that our model with multi-attention achieves better performance than the model without this all across the board. With the help of multi-attention, our model learns to extract semantic information from multiple levels. It enriches the expressiveness of features, which effectively overcomes the shortcomings of focusing only on content-level information.

In order to explore the differences in personal pronouns resolution performance of personal pronouns subclasses, we conducted a series of experiments on three types of personal pronouns. In particular, we compare the performance of models in first-person pronouns, second-person pronouns, and third-person pronouns. using different iterations. For all these experiments, we retain the rest of the model unchanged.

As can be seen from Figure 5, the third-person model has the highest performance and the F-score reaches 83.4%. This is because in the Uyghur, the third person has a rich use scene, which can corefere to humans, objects, etc., and the third-person Uyghur pronoun has no gender and can corefere to male or female; but the second person except for certain specific use environments most of them appear in the text dialogue, and the usage scene is relatively simple compared to third-person. The third person is most widely distributed in the Uyghur resolution dataset, which makes the model

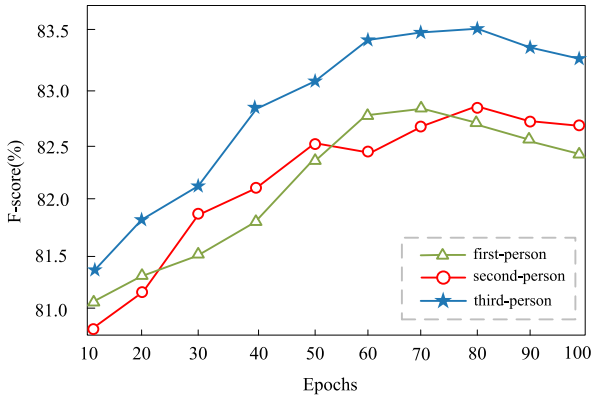


FIGURE 5. **Model performance in different personal pronouns.**

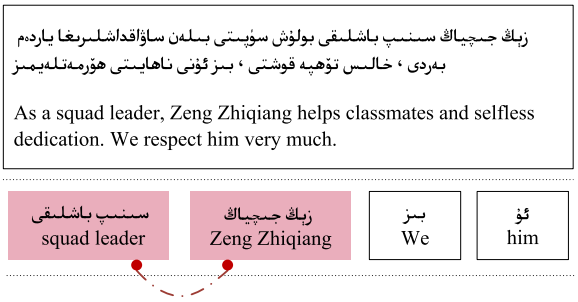


FIGURE 6. **Example of case study.**

have enough feature vectors to train and obtain deep semantic information. Therefore, compared with the first person and the second person, our model achieves the best performance in the third person.

E. CASE STUDY

Lastly, we show a case to illustrate the effectiveness of our proposed model, as is shown in Figure 6. In this case, we can see that our model correctly predict mentions "زېڭ جىچياڭ/Zeng Zhiqiang" and "سىنىپ باشلىقى/squad leader" as the antecedents of the personal pronoun "ئۇ/him". This case demonstrates the efficiency of our model.

V. CONCLUSION

We introduce a multi-attention based capsule network for personal pronouns resolution. Multi-attention can obtain deeper feature information from three aspects, which makes up for the lack of single attention mechanism to pay attention to content-level information. Meanwhile, the capsule network can capture the features of anaphor and candidate-antecedents by the dynamic routing algorithm iteratively, and also the semantic information within the context is retained by the learning model. Experimental results on Uyghur dataset show that our approach surpasses the state-of-the-art models and gets the highest F-score of 83.85%.

In the future, we plan to apply our model to other related anaphora resolution tasks, such as noun phrases resolution and zero pronoun resolution. Furthermore, we will explore more auxiliary neural networks to enforce our model for better performance.

REFERENCES

[1] D. Li, T. Miller, and W. Schuler, "A pronoun anaphora resolution system based on factorial hidden Markov models," in *Proc. Assoc. Comput. Linguistics*, 2011, pp. 1169–1178.

[2] Q. Yin, Y. Zhang, W.-N. Zhang, T. Liu, and W. Y. Wang, "Deep reinforcement learning for chinese zero pronoun resolution," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, vol. 1, 2018, pp. 569–578.

[3] J. Lu and V. Ng, "Joint learning for event coreference resolution," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, vol. 1, 2017, pp. 90–101.

[4] S. Zhu, S. Li, and G. Zhou, "Adversarial attention modeling for multi-dimensional emotion regression," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 471–480.

[5] Y. Cao, Z. Liu, C. Li, Z. Liu, J. Li, and T.-S. Chua, "Multi-channel graph neural network for entity alignment," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 1452–1461.

[6] K. Sun, Y. Li, D. Deng, and Y. Li, "Multi-channel CNN based inner-attention for compound sentence relation classification," *IEEE Access*, vol. 7, pp. 141801–141809, Oct. 2019.

[7] J.-L. Wu and W.-Y. Ma, "A deep learning framework for coreference resolution based on convolutional neural network," in *Proc. IEEE 11th Int. Conf. Semantic Comput. (ICSC)*, 2017, pp. 61–64.

[8] G. Veena, D. Gupta, A. N. Daniel, and S. Roshny, "A learning method for coreference resolution using semantic role labeling features," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2017, pp. 67–72.

[9] B. Nitoń, P. Morawiecki, and M. Ogrodniczuk, "Deep neural networks for coreference resolution for polish," in *Proc. 11th Int. Conf. Lang. Resour. Eval. (LREC)*, 2018, pp. 395–400.

[10] J. Plu, "Sanaphor++: Combining deep neural networks with semantics for coreference resolution," in *Proc. 11th Int. Conf. Lang. Resour. Eval. (LREC)*, 2018, pp. 412–417.

[11] I. Haponchyk and A. Moschitti, "A practical perspective on latent structured prediction for coreference resolution," in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics*, vol. 2, 2017, pp. 143–149.

[12] C. Zhang, Y. Li, N. Du, W. Fan, and P. Yu, "Joint slot filling and intent detection via capsule neural networks," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, Jul. 2019, pp. 5259–5267.

[13] Z. Chen and T. Qian, "Transfer capsule network for aspect level sentiment classification," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, Jul. 2019, pp. 547–556.

[14] Y. Dong, Y. Fu, L. Wang, Y. Chen, Y. Dong, and J. Li, "A sentiment analysis method of capsule network based on BiLSTM," *IEEE Access*, vol. 8, pp. 37014–37020, Mar. 2020.

[15] W. M. Soon, H. T. Ng, and D. C. Y. Lim, "A machine learning approach to coreference resolution of noun phrases," *Comput. Linguistics*, vol. 27, no. 4, pp. 521–544, Dec. 2001.

[16] V. Ng and C. Cardie, "Improving machine learning approaches to coreference resolution," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguistics (ACL)*, 2001, pp. 104–111.

[17] X. Yang, G. Zhou, J. Su, and C. L. Tan, "Coreference resolution using competition learning approach," in *Proc. 41st Annu. Meeting Assoc. Comput. Linguistics (ACL)*, 2003, pp. 176–183.

[18] K. Lee, L. He, M. Lewis, and L. Zettlemoyer, "End-to-end neural coreference resolution," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 188–197.

[19] A. Marasovic, L. Born, J. Opitz, and A. Frank, "A mention-ranking model for abstract anaphora resolution," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 221–232.

[20] C. Chen and V. Ng, "Chinese zero pronoun resolution: An unsupervised approach combining ranking and integer linear programming," *Springer Verlag*, vol. 36, no. 5, pp. 823–834, 2014.

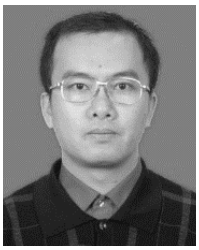
[21] K. Clark and C. D. Manning, "Deep reinforcement learning for mention-ranking coreference models," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 2256–2262.

[22] M. Poesio, Y. Grishina, V. Kolhatkar, N. Moosavi, I. Roesiger, A. Roussel, F. Simonjetz, A. Uma, O. Uryupina, J. Yu, and H. Zinsmeister, "Anaphora resolution with the ARRAU corpus," in *Proc. 1st Workshop Comput. Models Reference, Anaphora Coreference*, 2018, pp. 11–22.

[23] F. Luo, L. Zhang, B. Du, and L. Zhang, "Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, early access, Jan. 27, 2020, doi: 10.1109/TGRS.2020.2963848.
[24] S. Li, W. Li, C. Cook, C. Zhu, and Y. Gao, "Independently recurrent neural network (IndRNN): Building a longer and deeper RNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5457–5466.
[25] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3856–3866.
[26] G. Hinton, S. Sabour, and N. Frosst, "Matrix capsules with EM routing," in *Proc. 6th Int. Conf. Learn. Represent. (ICLR)*, Vancouver, BC, Canada, Feb. 2018, pp. 1–29.
[27] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, pp. 2121–2159, Feb. 2011.
[28] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, *arXiv:1207.0580*. [Online]. Available: http://arxiv.org/abs/1207.0580
[29] F. Luo, L. Zhang, X. Zhou, T. Guo, Y. Cheng, and T. Yin, "Sparse-adaptive hypergraph discriminant analysis for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, early access, Sep. 10, 2019, doi: 10.1109/LGRS.2019.2936652.

QIMENG YANG received the M.S. degree in software engineering from Xinjiang University, in 2019, where he is currently pursuing the Ph.D. degree in computer science and technology. His research interests include natural language processing and sentiment analysis.

LONG YU received the M.S. degree in computer science and technology from Xinjiang University, in 2008. She is currently a Professor with the Xinjiang University of Technology. Her research interests include intelligence computing, information security, and natural language processing.



SHENGWEI TIAN received the Ph.D. degree in computer science and technology from the Xinjiang University, in 2010. He is currently a Professor with the Xinjiang University of Technology. His research interests include intelligence computing, image processing, and natural language processing.

JINMIAO SONG received the M.S. degree in computer technology from North Minzu University in 2014. He is currently pursuing the Ph.D. degree in computer science and technology with the University of Xinjiang. His research interest includes bioinformatics.