*Research Article*

# Resolution of English Event Pronouns Based on Machine Learning

## Lu Qiu [1] and Wei Feng[2]

[1]*Institute of International Education, Jiangsu Vocational College of Electronics and Information, Huaian 223003, Jiangsu, China*
[2]*School of International Education, Jiangsu Vocational College of Finance and Economics, Huaian 223003, Jiangsu, China*

Correspondence should be addressed to Lu Qiu; 108157@jsei.edu.cn

As the basis of machine translation, anaphora aims to let the machine determine the entity or event to which the sentence refers by exploring the anaphora relationship between sentences. Prior to this, the research on anaphora resolution mainly focused on the resolution of entity anaphora. Through unremitting efforts, the elimination of entity-reference relationship has achieved great success, but the equally important event reference has been stagnant. This means that we can promote the development of machine translation by enhancing event reference. In this paper, a new method is proposed, which uses the latest machine learning algorithm to eliminate English event pronouns. Through feature extraction, data preprocessing, and the introduction of end-to-end double-loop neural network and attention mechanism, the network's ability to acquire contextual features is improved, and finally, the purpose of eliminating English event pronouns is achieved. In the experimental part, this paper also conducts training and testing on the latest data set KBP. It is found that the model algorithm proposed in this paper can perform the task of experimental setup well, and the value of 40.3% $F1$ is given under CONLL evaluation index. This proves that the model can understand semantic information very effectively and extract relevant information from the given semantic information.

## 1. Introduction

As a very important aspect of language in real life, reference is ubiquitous in various language expressions. It serves as a basic unit in an article. The main function is to replace objects and events, and referential dissolution refers to the process of finding pronouns based on the article information. Event reference is a common phenomenon in reference. It is often used in our daily expressions. However, in previous research, event reference is often handled together with entity reference. Event reference resolution has not attracted enough attention. In recent years, with the advancement of information extraction and text understanding research, the demand for event understanding has been growing, and the role of event reference resolution is highlighted. This topic aims to study the elimination of English event pronouns. We hope to improve the accuracy and efficiency of machine translation through the research on deference elimination and provide a new choice for the translation model of machine translation.

Reference parsing is an indispensable part of information extraction and text comprehension. Due to human habits of expression and the coherent nature of discourse, the phenomenon of reference is widespread. In order to accurately summarise and understand discourse, it is inevitable that entities and events scattered throughout the discourse are linked to the pronouns that point to them. Given the importance of referential relations in natural language processing, referential parsing was identified as one of the key techniques for information extraction at the Information Understanding Conference (MUC), funded on an ongoing basis by the US Defense Advanced Research Projects Agency, and a special body was established to evaluate referential digestion [1]. As a type of referent, event referents have a special meaning beyond that of traditional referential parsing. Firstly, in previous work, event referents and entity referents were often treated together, neglecting the independence and specificity of event referents and affecting the effectiveness of referential resolution. Secondly, as event reference disambiguation deals with the

relationship between pronouns and events, it is important to rationalise this relationship for the understanding of the text, which is beyond the scope of entity reference disambiguation [2]. Finally, event pronouns are ubiquitous. In the Onto Notes corpus, temporal pronouns account for 20% of all pronouns in the corpus, which illustrates the importance of event pronouns in life. Therefore, we must use machine learning techniques to eliminate event pronouns [3].

Fatemeh Mirzapour regards the whole event reference digestion process as a clustering task. By comparing each pair of events, the maximum entropy model is used to judge whether each active event is grouped with the previous event, and finally, the purpose of co-reference resolution is achieved [4]. Yang et al. proposed a method to construct language resources for event clustering and co-reference resolution. According to their definition, there are many types of events in each document, and the events in each type of event may be related to each other across documents. They aim to solve two problems: first, the relationship between which events are classified into the same type of events. The second is to use the unsupervised model for the first problem (realizing event clustering) [5]. Rudnev used the nonparametric Bayesian model to realize unsupervised event cross-reference resolution. However, they defined event co-reference as the relationship between events rather than the relationship between events and pronouns [6]. By summarizing relevant research, it can be found that their understanding of pronoun resolution is more optimization at the algorithm level and does not highlight the real purpose of pronoun resolution.

This paper proposes a research model of English event pronouns resolution based on machine learning methods. The starting point of the model is to utilize semantic information in the sentence, combined with the contextual content in the article. Then, multiple feature vectors are extracted and weight distribution is used to merge these features to eventually achieve the purpose of eliminating English events. After that, the present study also used a variety of evaluation indicators to conduct an effective evaluation of our model. Through analysis of the result, we put forward some effective conclusions and future research directions mainly focused on the solution of imbalanced samples and treatment. The English event pronoun resolution research model based on the machine learning method studied in this paper can effectively solve the problem of unclear pronoun reference in machine translation and can greatly improve the accuracy of translation.

## 2. Basic Concepts and Methods

### 2.1. Referential Digestion Concept.

Anaphora resolution is an important part of natural language processing. Anaphora resolution technology is used in information extraction. In the anaphoric resolution, the first candidate list must be constructed, and then, multiple selections are made from the candidates. Syntax-based resolution is an earlier approach that attempts to make full use of syntactic-level knowledge and apply it to subferential resolution in a heuristic manner. Therefore, the resolution of the referential relationship has

become an irreplaceable position in applications such as machine translation and question answering systems.

In general, we can divide the referential relationship into co-referential and anaphora according to the different ways of referring in the referential relationship. Co-reference refers to the fact that three linguistic units (nouns, pronouns, or noun phrases) point to the same objective existence. Even if it is out of context, it will not affect the change of its referential relationship. The anaphora refers to the anaphora in the current article pointing to an entity or event mentioned above. There is a semantic connection between them. This referential relationship depends on the contextual connection, and the antecedent referred to in the anaphora varies with the change in the context. For example, in the example in Figure 1, [Vice President Lu Xiulian] and [Vice President Lu] point to the same person, and the referential relationship belongs to the common referential relationship. The reference word "he" in Example 2 refers to the "Xiao Li" mentioned earlier. Moreover, "he" may also point to other characters as the context changes [7, 8].

Although there are certain similarities between the two referential relations of common referent and anaphora, the relationship between them is not inclusive. Anaphora is when the antecedent of the pronoun comes before the pronoun, and antithetical is when the antecedent of the pronoun comes after the pronoun. Because these two referential relations have different characteristics, the resolution of these two referential relations cannot be solved by a single method and language model. It needs to be treated differently according to its own characteristics. The current referential resolution work mainly focuses on the resolution of the common reference relationship [9].

In addition, from another point of view, reference can be divided into two categories according to the different objects: entity reference and event reference. As the name implies, entity reference refers to the antecedent that the anaphora points to, which is an objectively existing thing. The biggest difference between the two kinds of references is the difference in the reference object. For example, as shown in Figure 2, in Example 3, John Smith pointed to He, which is an objective person. As shown in Figure 2, the strong growth in Example 4 points to the [rose] event (in the Onto Notes corpus, the verbs in the sentence are usually marked as antecedent to indicate an event, so rose represents car sales growth event). The strong growth is a noun phrase, so we refer to this as a noun phrase event referential relationship [10]. For example, in Example 5, it, as a pronoun, points to the event do (do sports). Since it is a pronoun, we call this a pronoun event referential relationship [11]. The three instances in Figure 2 have a common feature, that is, the pronouns are converted for better translation.

### 2.2. Classic Event Referential Resolution Method

### 2.2.1. Event Pronoun Referential Resolution Based on Compound Kernel Function.

Composite kernel function, also called mixed kernel function, is to use two or more

Example 1 [Vice President Lu Xiulian] appointed the lawyers to formally request the publication of [Vice President Lu] who once called the media executives to disclose the scandal, and must apologize publicly by next Sunday, otherwise they will be told.

Example 2 [Xiao Li] participated in the spring dinner party organized by everyone today,This day happened to be [his] birthday. What surprised him was that his friends prepared a birthday cake for him, which impressed [his] at the same time.

FIGURE 1: Line of events.

Example 3 [John Smith] is The president of the bank. [He] loves his job.job
Example 4 Japan's domestic sales of cars, trucks and buses in October [rose] 18% from a year earlier to 500,004 units, a record for the month, the Japan Automobile Dealers' Association said. [The strong growth] followed year-to- year increases of 21% in August and 12% in September.
Example 5 I always [do] more sports in everyday morning because [it] can make you healthy.

FIGURE 2: Line of events.

kernel functions together to form a new kernel function. For the first time in 2010, machine learning was used to complete the referential elimination task. In the aspect of sample generation, the verbs in the sentence where each event pronoun is located and the first two sentences are used as antecedent candidates, and they form an example pair with anaphora. In terms of feature extraction, it adds some feature information according to the specificity of the event based on the partial features of the entity referential resolution such as (Surrounding Words and POS Tags) and (Co-occurrences of Surrounding Words). In addition, the method also compares the minimum expansion tree, simple expansion tree, and full expansion tree. Experiments show that the simple expansion tree achieves an $F1$ value of 47.2% while achieving a relatively moderate accuracy and recall rate. Because the sample generation method caused a serious imbalance in the proportion of positive and negative examples, the positive and negative examples were balanced by four methods such as down-sampling and adjusting the data hyperplane. Among them, adjusting the hyperplane method obtained the best system performance [12, 13].

### 2.2.2. Noun-Verb Anaphora Resolution.

This method studies the referential relationship composed of noun phrases and verbs to eliminate. Different from event pronouns, this method removes the verbs before and after the noun phrase as its antecedent candidate and its constituent instance pairs. In terms of feature selection, in addition to using the features in the entity digestion that is suitable for the event-referential resolution, it adds some features according to the meaning of the noun phrase in the context. Morphological features are proposed by using the noun phrase itself with certain content, Synonym features are expressed with the help of WordNet, the frequency of instance pairs appearing in the training corpus is used as a Fixed Pairings feature, and

according to the Named Entity, this feature will belong to the name and location of the person. Noun phrases such as institutions and others are excluded. In addition, it also uses the context similarity and part of the entity referral chain information as the features of noun phrase event referential resolution [14]. Many scholars have studied the noun anaphora, and through these methods, it is found that the system has the best effect when the minimum expansion tree is used in combination with planar features, and its system performance reaches 61.36%.

### 2.2.3. Event Reference Resolution Based on the Positive and Negative Balance Strategy.

The system balances the positive and negative cases by using the relevant content of the central theory, using the semantic role method to let Arg0 and Arg1 participate in calculating the semantic similarity of the two events, and compares it with the N-Window method [15]. The system uses the SVM classifier to experiment with the English expectations in Onto Notes 3.0. Experimental results show that the recall of the system has reached 58.67%. The system performance has reached 41.42%.

### 2.3. Application of Machine Learning in Referential Resolution.

For natural language processing tasks, due to the continuity of the language, feed-forward neural networks have limitations. In many cases, the words in the text are related to each other, and the context does not exist independently of each other. However, traditional feed-forward neural networks cannot model such position information well. All these have led to the result that the traditional method cannot effectively model language text information [16]. In order to better solve the above problems, researchers use recurrent neural network (RNN) to solve most natural language processing problems. Compared with the traditional feed-forward neural network, the biggest difference between the recurrent neural network is its unique cyclic structure. Using this cyclic representation of serialized information, it is possible to model the net listing of the variable-length input before and after the model [17]. Figure 3 is the basic structure and expanded structure diagram of the recurrent neural network. Each node (circle in the figure) represents a layer in the recurrent neural network. In the recurrent neural network, the output of each hidden layer is not only related to the input at the corresponding time but also affected by the output of the last hidden layer; this way of circular modeling can help the network effectively retain historical information [18].

A complete recurrent neural network is composed of an input layer (Output Layer), an output layer (Output Layer), and a hidden layer (Hidden Layer), which can map the serialized input $\{x_1, x_2, x_3 \ldots\}$ to the output sequence $\{y_1, y_2, y_3 \ldots\}$, and the corresponding output at each moment is $\{h_1, h_2, h_3 \ldots\}$. In the figure above, $U$ represents the parameter from the input layer to the hidden layer, and $V$ represents the hidden layer to the output layer parameters. In the recurrent neural network, the most important work is completed by these hidden layers. The output of each hidden layer is
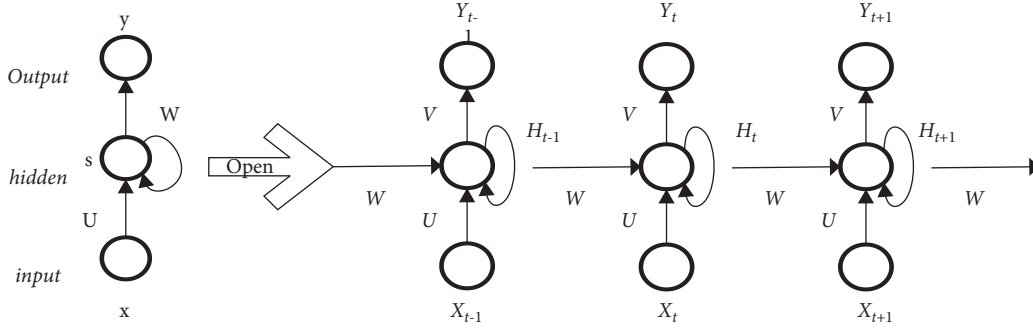
FIGURE 3: Recurrent neural network example.

passed to the output layer and is also passed back as the input of the next hidden layer. The specific definition and detailed calculation process of each part of the recurrent neural network are as follows:

(1) $x_t$ is the input at time $t$, and the text corresponds to the word vector of the word input at time $t$. For example, $x_0$ represents the word vector of the first word of the input text sequence, and $x_1$ is the word vector of the second word.

(2) $h_t$ is the output of the corresponding hidden layer at time $t$. $ht$ is calculated from the output $h_{t-1}$ of the previous hidden layer and the input $x_t$ at the moment. When calculating the first hidden layer, since there is no upper hidden layer output, a zero vector is generally used as the initial hidden layer direction during the calculation.

$$ht = f\left(Ux_t + Wh_{t-1}\right), \tag{1}$$

where $f$ is a nonlinear activation function, and the input vector is nonlinearly transformed and mapped to the output vector. Activation functions are very important for artificial neural network models to learn and understand very complex and nonlinear functions. The activation functions commonly used in RNN mainly include the following: Sigmoid, tanh, and ReLU [19, 20].

The traditional recurrent neural network processes the input at each moment and the output of the hidden layer at the previous moment indiscriminately, so as the length of the sequence increases, it will cause the loss of historical injury, and the gradient of the parameters will disappear during training (gradient vanishing). This problem means that in the process of modeling longer serialized text information, as the sequence grows, the network will gradually forget some of the original information [21].

To solve this problem, researchers have proposed a long-term memory neural network model (long short-term memory, LSTM). Later, LSTM was successfully applied to multiple natural language tasks, such as sentiment analysis, speech recognition, machine translation, and reading comprehension. The biggest difference between it and the traditional recurrent neural network is that the hidden

layer of LSTM is composed of units containing multiple gates. In the calculation process, the input $x_t$ needs to be passed to the three gate units, and the corresponding output is passed to the memory unit to obtain the hidden layer output ($h_t$) [22].

In the LSTM network, the update of historical information and hidden layer output update weights are controlled by three different gates, including the input gate (Input Gate), output gate (Output Gate), and forget gate (Forget Gate). Among them, the input gate is used to control the weight of the input information at each moment, the forget gate is used to control the weight of the information that needs to be retained by the hidden layer output at the previous moment, and the output gate is used to control the information weight of the hidden layer output at the corresponding moment [23]. Similarly, the input time sequence is defined as $\{x_1, x_2, x_3 \ldots\}$, the input corresponding to time $t$ is $x_t$, the memory unit information of LSTM is $C_t$, and the hidden layer output is $h_t$. The calculation and update process of LSTM are as follows:

$$\begin{aligned}
\tilde{C}_t &= \tanh\left(W^{(c)} \cdot [x_t; h_{t-1}] + b^{(c)}\right), \\
f_t &= \sigma\left(W^{(f)} \cdot [x_t; h_{t-1}; C_{t-1}] + b^{(f)}\right), \\
i_t &= \sigma\left(W^{(i)} \cdot [x_t; h_{t-1}; C_{t-1}] + b^{(i)}\right), \\
o_t &= \sigma\left(W^{(o)} \cdot [x_t; h_{t-1}; C_{t-1}] + b^{(o)}\right), \\
C_t &= i_t \otimes \tilde{C}_t + f_t \otimes C_{t-1}, \\
h_t &= o_t \otimes \tanh\left(C_t\right).
\end{aligned} \tag{2}$$

In these formulas, $\sigma$ represents a nonlinear transformation function. Through the control of input gate ($i_t$), forget gate ($f_t$), output gate ($o_t$), and unique memory unit design, LSTM can better update, store, and even "discard" longer sequence information. In extreme cases, if the value of the input gate is 0, it means that the input at the current time is completely ignored, and if the forget gate is 0, it means that the memory unit information of the previous time is completely forgotten ($c_{t-1}$).

On the basis of LSTM, subsequent researchers have successively proposed gate recurrent neural networks (GRU). The main difference is that GRU omits the memory unit and only needs to calculate two gates to control the flow of information. The specific calculation is as follows:

$$r_t = \text{sigmod}\left(W^{(x)} \cdot \left[x_t; h^{t-1}\right] + b^{(r)}\right),$$
$$z_t = \text{sigmod}\left(W^{(z)} \cdot \left[x_t; h_{t-1}\right] + b^{(z)}\right),$$
$$\widetilde{h}_t = \tanh\left(W^{(h)} \cdot \left[xt; r_t \otimes h_{t-1}\right] + b(h)\right), \quad (3)$$
$$h_t = z_t \otimes h_{t-1} + (1 - z_t) \otimes \widetilde{h}_t.$$

In the calculation process, through the control of different doors, the model also achieves the purpose of modeling long-distance sequences and has obtained better performance than LSTM on some tasks [24]. We have proposed a system to eliminate English event pronouns based on a recurrent neural network model. The English event pronoun elimination system based on the recurrent neural network model has high recognition efficiency for complex pronoun reference and can effectively identify different reference models.

### 2.4. Evaluation Index.

Due to the relatively few research work on the resolution of the event, there is no evaluation system specifically for this task. The entity generally refers to the evaluation method in the resolution. The main use is the MUC evaluation method. It includes recall rate $R$, precision Rate $P$, and $F$ value, where recall rate refers to the percentage of the number of objects that are correctly digested in the digestion result to the total number of objects that should be digested by the digestion system. Accuracy refers to the proportion of the correct quantity to the total quantity in use. $F$ is a mixed value of the two indicators of recall rate and accuracy rate. It can fully represent its performance, and $P$ and $R$ can get the maximum value. In general, the value is 1, so the $F$ value is also called the $F1$ value. In the evaluation model of this paper, the larger the $f$ value is, the better the effect of the algorithm is.

$$R = \frac{TP}{TP + FP},$$
$$P = \frac{TP}{TP + FN},$$
$$F = \frac{(1 + \beta)P \times R}{\beta \times P + R}, \quad (4)$$
$$F_1 = \frac{2 \times P \times R}{P + R}.$$

## 3. Experimental Data Set and Model

### 3.1. Experimental Corpus.

There are many types of corpus, and the main basis for determining the type is its research purpose and use, which can often be reflected in the principles and methods of corpus collection. Currently, research events refer to the data sets used in the resolution task as long as they include ACE2005, KBP2015, 2016, and ECB data sets. Among them, the KBP2015 and 2016 data sets are currently the largest, most extensive, and most complete data sets, so this article uses the KBP2015 and 2016 data sets. Some general overviews of the KBP2015 and

2016 data sets are shown in Table 1. The KBP database is mainly related to articles in the field of news (newswire, NW) and forums (discussion forums, DF), and each data has attributes such as event type (type); especially in KBP2016, the total number of event types decreased from 38 to 18, so this experiment only selected these 18 events for training and testing.

From Table 1, we can intuitively see that among the 3,929 events, 2441 events accounted for 63.75% of the events, which are not to be resolved. However, by reviewing the information, we find that this ratio is 20% in the field of physical reference. This shows that we have to face the imbalance of positive and negative samples.

### 3.2. English Event Pronoun Resolution Model Based on Machine Learning Method.

In this paper, the research model of English event pronouns based on the machine learning method consists of two parts, which are the extraction of multidimensional features and the classification of event expressions.

### 3.2.1. Extraction of Multidimensional Features.

Feature extraction refers to the extraction of new features from the original data. This "extraction" can be performed automatically using a certain algorithm. The purpose of extraction is to reduce multidimensional or related features to low dimensional in order to extract the main information or generate the target. First, this paper uses the trained word vectors to encode relevant features. After that, the bidirectional recurrent neural network (Bidirectional recurrent neural network, Bi-RNN) is used to preprocess the data of these word vectors so that he can learn contextual information. In addition, this paper also introduces an attention mechanism (Attention) to redistribute the weight of data features to make it grasp the key features. Finally, from the obtained features, the score function is used for scoring to select the highest scoring element for event description. The specific process is shown in Figure 4.

In the event-related research, the information of the trigger word is obviously more critical than other word information in the text, and it should be focused on. However, relying on the recurrent neural network alone cannot capture the important information in the text. Therefore, we add an attention mechanism to the model. After the redistribution of the attention mechanism, we sum the word vectors in each candidate element to obtain a vector $\widehat{x}_i$ and combine the output of Bi-LSTM and the feature vector as the final candidate element $\varphi(i)$ vector characterization, which $\varphi(i)$ represents the span of candidate elements as shown in the following equations:

$$a_t = \omega_a \bullet FFNN_a\left(x_t^0\right), \quad (5)$$

$$a_{i,t} = \frac{\exp(a_t)}{\sum_{k=\text{start}(i)}^{\text{end}(i)} \exp(a_k)}, \quad (6)$$

TABLE 1: Distribution of events in KPB corpus.

| Data set | | Doc | Sentence | Event | Event representation | The timetable of the referential relationship | Single event representation chain |
|---|---|---|---|---|---|---|---|
| Train | ALL | 509 | 16359 | 1770 | 10682 | 5511 | 5117 |
| | ZW | 178 | 3495 | 667 | 4323 | 2030 | 2293 |
| | DF | 331 | 12864 | 1103 | 6305 | 3481 | 2824 |
| Verification | ALL | 139 | 4712 | 485 | 2956 | 1519 | 1437 |
| | ZW | 49 | 910 | 159 | 1040 | 471 | 569 |
| | DF | 90 | 3802 | 326 | 1916 | 1048 | 868 |
| Test | ALL | 169 | 4499 | 451 | 3829 | 1388 | 2441 |
| | ZW | 85 | 1690 | 288 | 2319 | 903 | 1416 |
| | DF | 84 | 2809 | 163 | 1510 | 485 | 1025 |
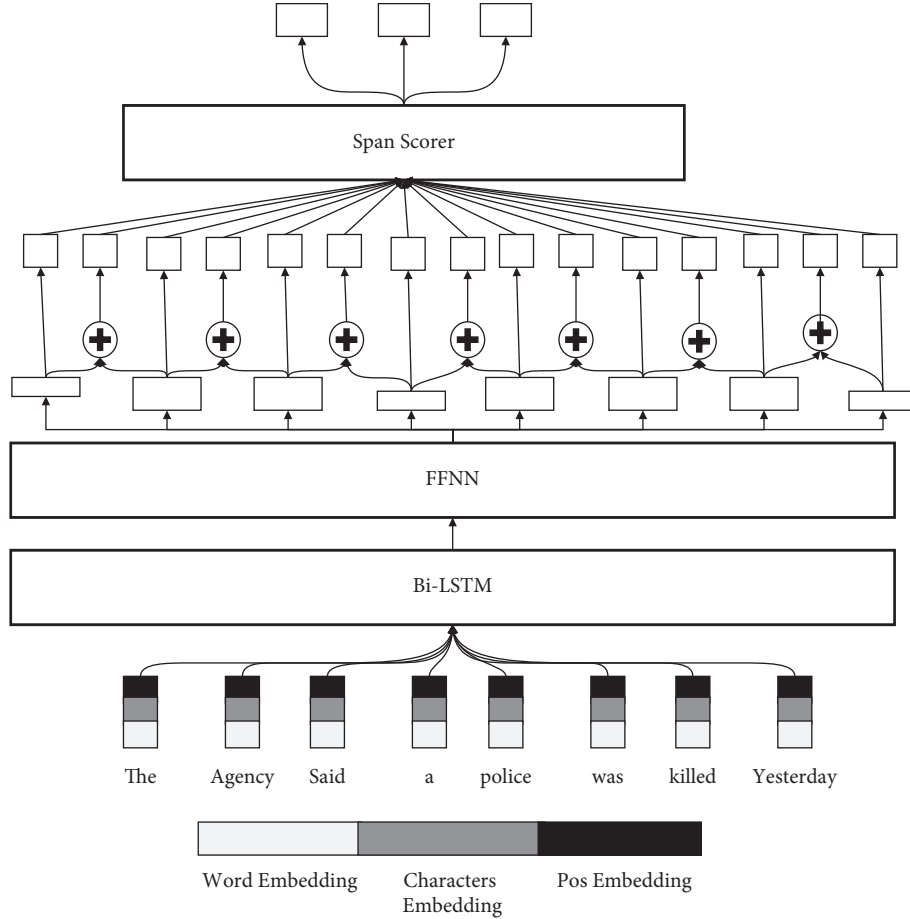


FIGURE 4: Candidate element extraction.

$$\widehat{x}_i = \sum_{t=\text{start}(i)}^{\text{end}(i)} a_{i,t} \bullet x_t, \tag{7}$$

$$hi = \left[ x^0_{\text{start}(i)}, x^0_{\text{end}(i)}, \widehat{x}_i, \varphi(i) \right]. \tag{8}$$

3.2.2. Classification of Event Description. As shown in Figure 5, the model scores $S_m(i)$ for each candidate element through a standard feed-forward neural network FFNNm and sets a threshold $\lambda$ to retain the candidate element with the highest score to participate in the final reference resolution task. We arrange the reserved candidate elements in the order of the text, and each candidate element $i$ must be paired with its antecedent $j$ to form an event pair, and then, each event pair is scored $S_m(i, j)$ ($1 \le j \le i - 1$) through the feed-forward neural network FFNNa. Among them, "." represents the dot product operation between vectors, and the feature vector ($i, j$) is composed of text type (NW, DF) vectors. The final learning goal of the model is to maximize the following marginal likelihood function, where GOLD ($i$) represents a standard set of antecedent elements as shown in the following equations:
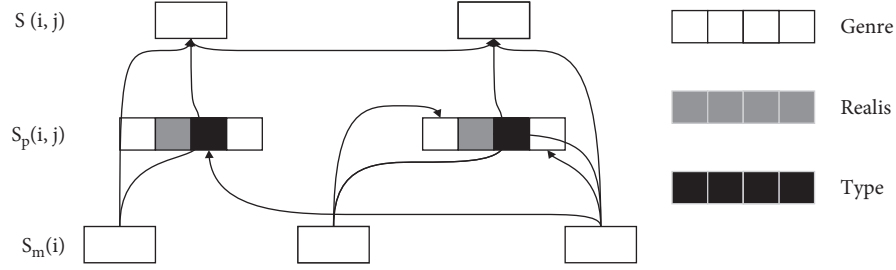
FIGURE 5: Event score.

$$S_m(i) = \omega \bullet \text{FFNN}_m(h_i), \tag{9}$$

$$S_p(i, j) = \omega_p \bullet \text{FFNN}_p\big([h_i, h_j, h_i \cdot h_j, \varphi(i, j)]\big), \tag{10}$$

$$S(i, j) = \begin{cases} 0, & j = \epsilon, \\ sm(j) + sm(i) + s(i, j), & j \neq \epsilon, \end{cases} \tag{11}$$

$$L\text{loss}(i) = \log \frac{\sum_{j' \in \text{GOLD}(i)} \exp(s(i, j'))}{\sum_{j = \epsilon, 0, \dots, i-1} \exp(s(i, j))}. \tag{12}$$

## 4. Experimental Results and Analysis

### 4.1. Balanced Sample Experiment

*4.1.1. Use Case Selection Experiment.* For the English event reference, the size of the data set is not large. When the data set is small, the data set is divided into a training set and a test set so that the variance between the test set and the training set will be very large, and the model training ability will change, which may eventually result in poor model performance. In order to solve this problem, we use a tenfold cross-validation method to process the data set. Table 2 and Figure 6 compare the results of using two different use case selection strategies.

As can be seen from Figure 6, the first method has certain defects. Some verbs that cannot be antecedents at all are also negatively paired with pronoun phrases, resulting in the number of negative examples being far greater than that of positive examples. The second method filters out these verbs, which basically has no effect on the number of positive examples, so it gets a good experimental result. In the following experiment, we will use the second method to select use cases.

*4.1.2. Undersampling Experiment.* In entity referential resolution, we can filter according to whether the anaphora and antecedent are consistent with their own words, numbers, and the semantic categories to which they belong so as to ensure that the proportion of positive and negative examples participating in machine learning is reasonable and can be trained to obtain high-performance classification. However, in the process of model pronoun resolution, the anaphora is a neutral pronoun such as this, that, or it, which contains very few words, numbers, and semantic information. In addition, the antecedent candidate is a description of an

event. Usually, we choose the driving verb of the event to represent it. It has a completely different semantic classification system from the anaphora (pronoun). We cannot tell if it is a word or a number. In this case, the ratio of positive and negative examples generated by our construction reaches $1 : 9$; even after filtering constraints of various rules, the ratio is still higher than $1 : 4$. Therefore, we illustrate through the negative sample undersampling experiment; as far as the event pronoun resolution task is concerned, the positive and negative examples can get better performance when the ratio is $1 : 1$. In the experiment, we sampled the negative examples. Figure 7 shows the performance of event pronoun digestion under different proportions of positive and negative examples when undersampling.

It can be seen from Figure 7 that as the ratio of negative examples to positive examples increases, the correct rate shows an upward trend, the recall rate shows a downward trend, and the recall rate decreases more, so the $F1$ value shows a downward trend. Therefore, unless otherwise specified, the experimental results given later in this article are the results of using the undersampling method to adjust the ratio of positive and negative examples by $1 : 1$.

*4.2. Machine Learning Experiment Setup Based on Neural Network.* In this paper, the model is trained and evaluated on the KBP data set. Among them, 648 of KBP2015 and 2016 are used as samples of the training set, and 169 of KBP2015 and 2016 are used as samples of the test set. At the same time, 139 are divided from the training set as cross-validation verification set. In order to better evaluate our model's ability to eliminate English event pronouns, we have added three evaluation indicators B3, CEAF, and BLANC on the basis of MUC. In order to make the model more stable, we also use weighted summation, and averaging CONLL and AVG-F are used to evaluate the final model. The specific calculation method is shown in the following equations:

$$\text{CoNLL} = \frac{\left(\text{MUC} + B^3 + \text{CEAF}_e\right)}{3}, \tag{13}$$

$$\text{AVG} - F = \frac{\left(\text{MUC} + B^3 + \text{CEAF}_e + \text{BLANC}\right)}{4}. \tag{14}$$

In this experiment, the word vector Word Embedding is the 300Vec word vector obtained by the trained model, the character vector Characters Embedding is the word vector obtained by inputting the character vector to the training of

TABLE 2: Comparison of two methods.

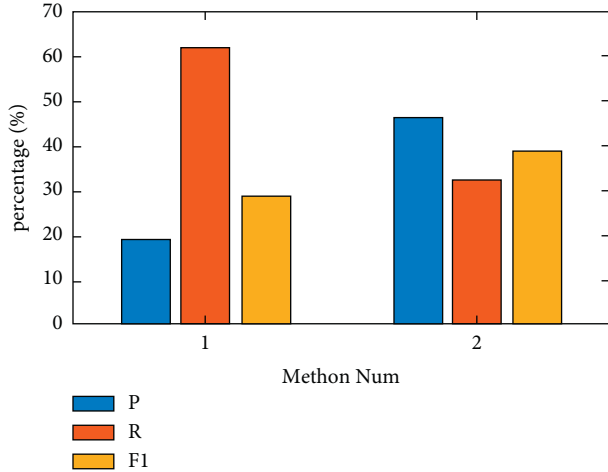|           | P (%) | R (%) | F1 (%) |
|-----------|-------|-------|--------|
| Method 1  | 19.01 | 62.33 | 28.78  |
| Method 2  | 46.75 | 32.48 | 38.98  |



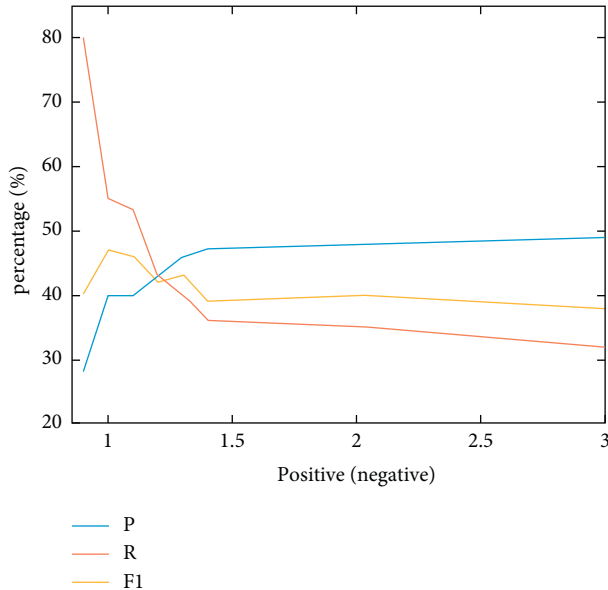FIGURE 6: Comparison of two methods.



FIGURE 7: The trend of experimental results with the change of the proportion of positive and negative cases.

the convolution network, and the final output vector dimension is 100Vec. In addition, we set the dimensions of features such as part of speech, text type, event type, and tense to 20 dimensions. After passing through the feed-forward neural network FFNNm, we screen all candidate elements and screen the threshold as shown in Figure 8.

It can be seen from the change of the above results that when the threshold $\lambda$ is set to 0.2, the model has the highest comprehensive score on CONLL and AVG-F. The specific experimental parameter configuration is shown in Table 3.
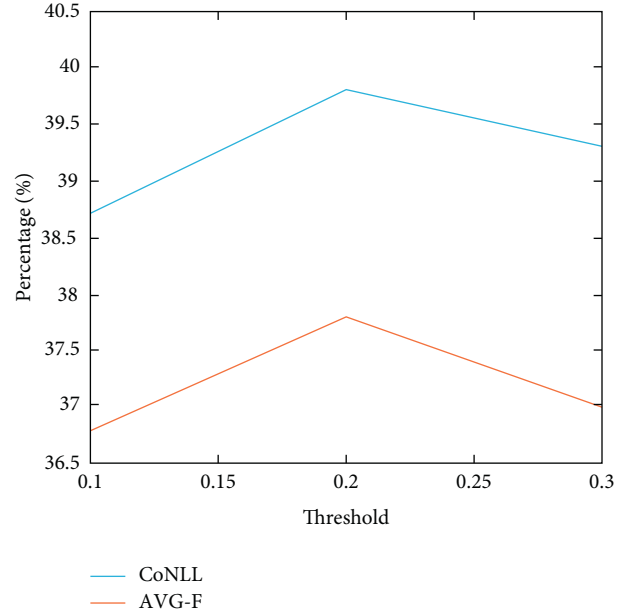


FIGURE 8: Comparison of threshold results.

TABLE 3: Experimental parameter configuration.

| Bi-LSTM | | Hidden dim: 200 |
|---------|---|---|
| | | Number layers: 1 |
| | | Dropout: 0.2 |
| FFNN | | Hidden dim: 200 |
| | | Number layers: 1 |
| | | Dropout: 0.2 |
| | | Activation: RELU |
| Word Embedding | | Embedding dim: 300 |
| Character Embedding | | Embedding dim: 100 |
| Features | Pos | |
| | Type | |
| | Realis | Embedding dim: 20 |
| | Genre | |
| Optimizer | | Adam |
| Learning rate | | 0.0001 |

The experimental results are shown in Table 4 and Figure 9. Our experiments used the same data configuration as the results reported in the current mainstream research papers.

From the experimental results in Figure 9, we can see that our neural network model has much higher performance under the two evaluation indicators of MUC and BLANC than the Lu model, which is 16.06% and 5.56% higher than that of the Lu model. Compared with the comprehensive evaluation score, our system has achieved better experimental results whether it is CONLL or AVG-F value.

In order to verify the influence of semantic information and context content on this task, we give three other sets of comparative experiments.

(1) "-Character": indicates that the character vector Character Embedding is deleted, and only the word

TABLE 4: Experimental results of event reference resolution.

|  | MUC | $B^3$ | CEAFe | BLANC | CONLL | AVG-F |
|---|---|---|---|---|---|---|
| Lu | 27.41 | 40.90 | 39 | 25 | 35.77 | 33.08 |
| Ours | 43.47 | 38.93 | 37.3 | 30.56 | 40.3 | 37.57 |
| -Character | 41.29 | 36.99 | 36.1 | 29.04 | 38.13 | 35.86 |
| RNN | 36.13 | 32.33 | 31.99 | 22.94 | 33.48 | 30.85 |
| -Attention | 43.81 | 38.18 | 35.82 | 30.46 | 39.27 | 37.07 |
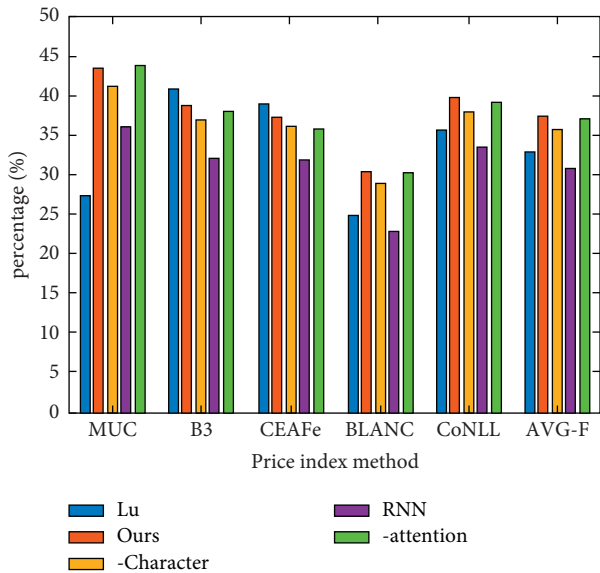


FIGURE 9: Experimental results of event reference resolution.

TABLE 5: Event extraction experiment results.

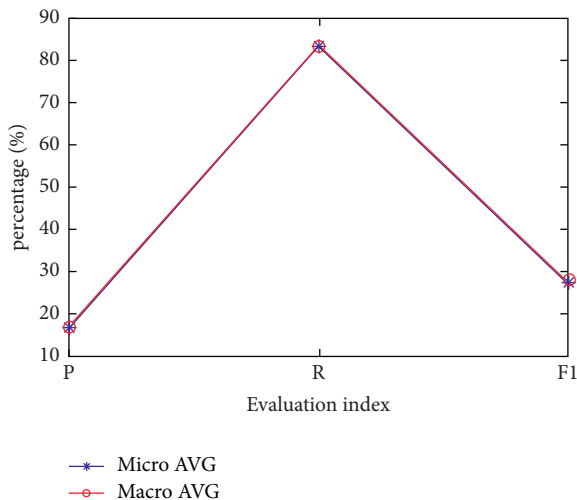|  | P | R | F1 |
|---|---|---|---|
| Micro AVG | 16.45 | 83.05 | 27.17 |
| Macro AVG | 16.87 | 83.12 | 27.98 |



FIGURE 10: Event extraction experiment results.

vector Word Embedding is retained. By comparing the experimental results, it can be seen that the results of CONLL and AVG-F decreased by 1.77% and 1.71%, respectively.

(2) RNN: we tried to replace the LSTM with RNN to evaluate the model. By observing the results of Table 4, it was found that the performance of using RNN encoding is very poor.

(3) "-attention": the evaluation results of the model after removing the attention mechanism; CONLL and AVG-F decreased by 0.63% and 0.5%, respectively.

In addition, we will output the results after Span Score, use the event evaluation toolkit provided by KBP2016 to score the reserved candidate elements, and calculate the number of correct events contained in these candidate elements. The results are shown in Table 5. It should be noted that the event syndication platform given in this article only focuses on the performance of the final event common reference.

By analyzing Table 5 and R and P in Figure 10, it can be seen that the model retains most of the correct results after the scoring operation while also including a large number of negative examples. For event referential resolution tasks, these negative examples have an impact on the final performance of the model. According to the statistics of the English corpus of KBP2016, it is found that this negative corpus accounts for 60% of the corpus, so our model still has a lot of room for improvement.

## 5. Conclusions

According to different objects, anaphora resolution can be divided into entity anaphora resolution and event anaphora resolution. As the attention of English event reference is far less than that of entity reference, event reference resolution used to be the same as entity reference resolution, and a unified resolution method was adopted. However, with the increase of information, information is often transmitted with the occurrence of events. With the increasing information extraction technology, the demand for event reference resolution is increasing day by day, and event pronoun resolution is of great significance.

This paper proposes a research platform for English event pronoun resolution based on machine learning. In this study, the latest machine learning algorithm is adopted to eliminate English event pronouns, and feature extraction, data preprocessing, and attention mechanism are introduced to improve the ability of the network to acquire contextual features. This paper also conducts training and testing on the latest data set KBP and finally achieves the purpose of eliminating English event pronouns.

At the same time, at the end of the paper, we pointed out the shortcomings of the model, that is, we could not keep the positive samples and filter out more negative samples. At the same time, this is also the most difficult point of this topic. In the future, we will seek to solve the problem of event reference elimination in negative samples so as to really improve the system performance.

## Data Availability

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

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

[1] A. N. Yurchenko, O. V. Fedorova, A. V. Kurganskii, and R. I. Machinskaya, "Event-related potentials in the brain on perception of referentially ambiguous Russian pronouns," *Neuroscience and Behavioral Physiology*, vol. 48, no. 1, pp. 101–108, 2017.

[2] X. Yang, X. Mao, M. Han, and C. LiGuo, "Electrophysiological evidence of encoding in self-referential effect," *NeuroReport*, vol. 30, no. 13, pp. 901–907, 2019.

[3] L. Vernich, R. Argus, and L. K. Merfeldienė, "Extending research on the influence of grammatical gender on object classification: a cross-linguistic study comparing Estonian, Italian and Lithuanian native speakers," *Eesti Rakenduslingvistika Ühingu aastaraamat. Estonian Papers in Applied Linguistics*, vol. 13, pp. 223–240, 2017.

[4] F. Mirzapour, "Differences in the use of hedges and first person pronouns in research articles of applied linguistics and chemistry[J]," *International Journal of Applied Linguistics & English Literature*, vol. 5, no. 6, pp. 166–173, 2016.

[5] Z. Yang, Z. Gong, and F. Kong, "Exploit comparable corpus to Chinese zero pronoun resolution[J]," *Acta Scientiarum Nauralium Universitaties Pekinensis*, vol. 53, no. 2, pp. 279–286, 2017.

[6] P. Rudnev, "Minimal pronouns, logophoricity and long-distance reflexivisation in Avar," *Studia Linguistica*, vol. 71, no. 1-2, pp. 154–177, 2017.

[7] C.-J. Lin, C. W. Pao, Y.-H. Chen, T. Liu, and H. H. Hsu, "Ellipsis and coreference resolution in a computerized virtual patient dialogue system," *Journal of Medical Systems*, vol. 40, no. 9, p. 206, 2016.

[8] J. A. Evans and A. Pedro, "Machine translation: mining text for social theory[J]," *Annual Review of Sociology*, vol. 42, no. 1, pp. 539–547, 2016.

[9] C. A. Bail, "Combining natural language processing and network analysis to examine how advocacy organizations stimulate conversation on social media," *Proceedings of the National Academy of Sciences*, vol. 113, no. 42, pp. 11823–11828, 2016.

[10] E. K. Chang, C. Y. Yu, R. Clarke et al., "Defining a patient population with cirrhosis," *Journal of Clinical Gastroenterology*, vol. 50, no. 10, pp. 889–894, 2016.

[11] C. Salvador, M. Joaquim, and A. Oliver, "Bilingual newsgroups in catalonia: a challenge for machine translation[J]," *Journal of Computer-Mediated Communication*, vol. 9, no. 1, p. 1, 2017.

[12] M. Rikters, M. Fishel, and O. Bojar, "Visualizing neural machine translation attention and confidence," *The Prague Bulletin of Mathematical Linguistics*, vol. 109, no. 1, pp. 39–50, 2017.

[13] E. Maldonado, E. Shihab, and N. Tsantalis, "Using natural language processing to automatically detect self-admitted technical debt," *IEEE Transactions on Software Engineering*, vol. 43, no. 11, pp. 1044–1062, 2017.

[14] A. Irvine and C. B, "End-to-end statistical machine translation with zero or small parallel texts," *Natural Language Engineering*, vol. 22, no. 4, pp. 517–548, 2016.

[15] O. Caglayan, M. G. Martínez, and A. Bardet, "NMTPY, A flexible toolkit for advanced neural machine translation systems," *The Prague Bulletin of Mathematical Linguistics*, vol. 109, no. 1, pp. 15–28, 2017.

[16] G. Cabanac, M. K. Chandrasekaran, I. Frommholz et al., "Report on the joint workshop on bibliometric-enhanced information retrieval and natural language processing for digital libraries (BIRNDL 2016)," *Acm Sigir Forum*, vol. 50, no. 2, pp. 36–43, 2016.

[17] D. Ataman, M. Negri, M. Turchi, and M Federico, "Linguistically motivated vocabulary reduction for neural machine translation from Turkish to English," *The Prague Bulletin of Mathematical Linguistics*, vol. 108, no. 1, pp. 331–342, 2017.

[18] A. Srivastava, G. Rehm, and F. Sasaki, "Improving machine translation through linked data," *The Prague Bulletin of Mathematical Linguistics*, vol. 108, no. 1, pp. 355–366, 2017.

[19] M. Turchi, M. Negri, M. A. Farajian, and M Federico, "Continuous learning from human post-edits for neural machine translation," *The Prague Bulletin of Mathematical Linguistics*, vol. 108, no. 1, pp. 233–244, 2017.

[20] P. Williams, R. Sennrich, M. Post, and P Koehn, "Syntax-based statistical machine translation," *Synthesis Lectures on Human Language Technologies*, vol. 9, no. 4, pp. 1–208, 2016.

[21] P. Wiriyathammabhum, D. Summers-Stay, C. Fermüller, and Y. Aloimonos, "Computer vision and natural language processing: recent approaches in multimedia and robotics," *ACM Computing Surveys*, vol. 49, no. 4, pp. 1–44, 2017.

[22] J. Biamonte, P. Wittek, and N. Pancotti, "Quantum machine learning," *Control Theory & Applications*, vol. 549, no. 7671, pp. 195–202, 2017.

[23] M. Feurer, A. Klein, K. Eggensperger, J. Springenberg, M. Blum, and F. Hutter, "Efficient and robust automated machine learning," *Ecological Informatics*, vol. 30, pp. 49–59, 2016.

[24] J.-X. Wang, J.-L. Wu, and H. Xiao, "Physics informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data," *Phys.rev.fluids*, vol. 2, no. 3, pp. 1–22, 2017.