Abstract

In online arguments, identifying how users construct their arguments to persuade others is important in order to understand a persuasive strategy directly. However, existing research lacks empirical investigations on highly semantic aspects of elementary units (EUs), such as propositions for a persuasive online argument. Therefore, this paper focuses on a pilot study, revealing a persuasion strategy using EUs. Our contributions are as follows: (1) annotating five types of EUs in a persuasive forum, the so-called ChangeMyView, (2) revealing both intuitive and non-intuitive strategic insights for the persuasion by analyzing 4612 annotated EUs, and (3) proposing baseline neural models that identify the EU boundary and type. Our observations imply that EUs definitively characterize online persuasion strategies.

The annotated dataset, annotation guideline, and implementation of the neural model are available in public.\footnote{Dataset and annotation guideline: http://katfuji.lab.tuat.ac.jp/nlp_datasets/. Implementation: https://github.com/EdoFrank/emnlp2019_morio_egawa}

1 Introduction

Persuasive discourses are common in social media. Recently, the automatic identification of such persuasive discourse is receiving attention in the field of computational linguistics (Habernal and Gurevych, 2016b; Tan et al., 2016; Habernal and Gurevych, 2016a; Persing and Ng, 2017; Musi et al., 2018; Hidey and McKeown, 2018; Ji et al., 2018; Durmus and Cardie, 2019; Gleize et al., 2019). Existing studies have examined persuasive arguments with regard to dynamics or lexical features (Tan et al., 2016), claim and premise analysis (Musi and Aakhus, 2018), and concessions (Musi et al., 2018). However, to the best of our knowledge, fine-grained analyses with an elementary unit (EU) semantic, such as propositions, are still in their infancy, although we estimate that EUs are important in obtaining the persuasion strategy. Also, existing studies do not extract token-level EUs via machine learning, which would be a basic step for an automatic understanding of the strategy.

Therefore, this paper shows the fundamental role of EUs in a persuasive forum by annotating five types of token-level EUs (i.e., Testimony, Fact, Value, Policy, and Rhetorical Statement) in a persuasive forum and also provides a baseline neural model to identify the EUs automatically.\footnote{Although we use the partial data of the previous study (Egawa et al., 2019), this work is distinct from the previous one because we focus on the novel analysis of EUs and a novel task to predict EUs. See Section 4 and 5.}

In this study, we use the effective dataset of ChangeMyView (Tan et al., 2016), which is a subreddit where original post (OP) users provide a controversial view in the title to change their
Table 1: Corpus comparison related to our study.

<table>
<thead>
<tr>
<th>source</th>
<th>#unit level</th>
<th>IAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ours</td>
<td>CMV</td>
<td>4612 token</td>
</tr>
<tr>
<td>Al Khatib et al.</td>
<td>editorial</td>
<td>14313 token</td>
</tr>
<tr>
<td>Hidey et al.</td>
<td>CMV</td>
<td>2615 token</td>
</tr>
<tr>
<td>Park et al.</td>
<td>eRulemaking</td>
<td>4899 clause</td>
</tr>
</tbody>
</table>

perspective through opponent users’ arguments. OP users award a delta (Δ) point when their view is changed. Figure 1 presents an overview of ChangeMyView where the positive post is an awarded post and the negative post is a non-awarded one.

Our analyses to find out the role of EUs in ChangeMyView show both intuitive and non-intuitive phenomena (Section 4.) We also inject a structural feature into the neural model to improve the performance; the results show that EUs have characteristic positional roles (Section 5.)

2 Related Work

EUs have been considered in investigating the characteristics of argumentation in an argument mining discipline. Park et al. (2015); Park and Cardie (2018) deemed Testimony to be a type of objective evidence using personal state or experience. Al Khatib et al. (2016) proposed the argument model for analyzing the strategy of a news editorial. The model considers token-level discourse units of six different types. The study that is most closely related to our research is (Hidey et al., 2017) because it annotated semantic types of claims and premises in ChangeMyView; however, the authors did not take objective propositions such as facts into consideration.

Table 1 summarizes the comparison in terms of size, annotation granularity level, and inter-annotator agreement (IAA) of the related corpora. The table shows that our corpus is middle sized but has sufficient granularity and the reasonable IAA reliability.

3 Annotating EUs

Typology of EUs

This study considers five types of EUs. For the introduction of the scheme, the motivation is derived from our expectation that we can feature persuasive arguments by considering personal experience, facts (Park and Cardie, 2018) and rhetoric (Blankenship and Craig, 2006). The five types of EUs are defined as follows:

1. **Testimony** is an objective proposition related to the author’s personal state or experience such as the following: *I do not have children.*

2. **Fact** is a proposition describing objective facts that can be verified using objective evidence and therefore captures the evidential facts in persuasions: *Empire Theatres in Canada has a “Reel Babies” showing for certain movies.*

3. **Value** is a proposition that refers to subjective value judgments without providing a statement on what should be done: *it is absolutely terrifying.*

4. **Policy** offers a specific course of action to be taken or what should be done: *intelligent students should be able to see that.*

5. **Rhetorical Statement** implicitly states the subjective value judgment by expressing figurative phrases, emotions, or rhetorical questions: *does it physically hurt men to be raped by women (as in PIV sex)?*

Annotation Process

We extracted 115 ChangeMyView threads from the train set of Tan et al. (Tan et al., 2016) through a simple random sampling. Each thread contains a triple of OP, positive, and negative. Therefore, we use 345 posts for the annotation.

EUs are annotated by 19 non-native annotators who are proficient in English, excluding the authors. Three annotators independently annotated 87 threads; the remaining 28 threads were annotated by 8 experts who were selected from the 19 annotators. Each annotator was asked to read the annotation guideline prior to the actual annotation. We also held several meetings to train the annotators and re-annotated the erroneous annotations when required. Given that the annotators are non-native speakers, the posts are translated into their language, and the translated documents are only used as a reference for the annotators. For the 83 threads, a gold standard is established by merging three posts using a majority vote.

We consider the token-level annotation (Stab and Gurevych, 2017), rather than the clause level one, to extract accurate minimal EU boundary and to remove irrelevant boundaries such as *for example* and *therefore.*

---

3 More precisely, 419 threads were randomly extracted from the train data range 2013/01/01–2015/05/07, and the older threads were actually used due to time constraints.
It is also intuitive that Rhetorik-skipped breakfast. For many years, I’ve regularly experiences, i.e., OP: positive authors are likely to assert their view with ex-
Rhetorical Statements in OP vs. reply. Therefore, difference exists in Testimony in OP vs. reply and
Mann–Whitney U test is employed. A significant positive and negative case in Table 2. Although Haber-
that this is because users are persuaded by effective indications of EUs, rather than simply present-
Table 2: Significance of the proportion of EU frequency of OP vs. reply and positive vs. negative. Sign-
ificant at \( p < 0.01 \) if **.

Table 3: Significance of the EU position between a positive and negative post. Significant at \( p < 0.05 \) if *.

**Positional Role of EUs is Characteristic**

To examine the positional role of EUs as a persua-
sion strategy, we investigate the position with a
one-dimensional axis, that is, a normalized position
wherein the beginning of a post is 0.0 and the
end is 1.0. Figure 3 shows the resulting histogram
of the positions for each EU type. It presents the
characteristic distributions: Fact and Testimony are
possibly located at the beginning of the post, e.g.,
positive: *Eating first thing in the morning stabilizes
blood sugar levels, which regulates appetite and
energy*, indicating the persuasion strategy to state
solid facts prior to claiming assertion. Moreover, a
Policy appeared at the bottom, indicating that what
should be done is regarded as a conclusion, e.g.,
negative: *Just keep in mind that eating breakfast may
increase your ability to concentrate, even if you think you are doing ok right now*. Value is the type
that dominates in most posts, and it appears in uniform distribution.

The statistical tests were performed to deter-
mine the existence of a difference between a posi-
tive and negative in terms of the position. In
general, given that addressing the entire positions
simultaneously could simplify the property, we
set the window size to 0.5 and investigated three
ranges as follows: beginning (0.00 – 0.50), mid-
dle (0.25 – 0.75), and end (0.50 – 1.00). For
each window, we employed the Mann–Whitney U
test and Kolmogorov-Smirnov (KS) significance
for the average positions in threads which have
at least five propositions, as described in Table 3.
5 Neural EU Extraction to Automatically Obtain the Strategy

In practice, we have to extract EUs computationally to obtain the strategy automatically. We consider the EU extraction problem as a sequence tagging problem in analogy using a related argument mining work (Eger et al., 2017; Habernal and Gurevych, 2017). The sequence tagging problem involves labeling each token in an input document. In each time step $i$, the label $y_i \in \mathcal{Y}$, where each token is associated with this label, is predicted as follows: $\mathcal{Y} = \{(b, t) \mid b \in \{B, I, O\}, t \in \{F, T, P, R, V, N\}\}$, where $b$ denotes the unit identification whether the current token is non-EU (O), begins (B), or continues (I) an EU. In addition, $t$ is the unit classification, i.e., Non-EU (N), Fact (F), Testimony (T), Policy (P), Rhetorical Statement (R), and Value (V).

The Multi-task Learning Approach

We employed neural multi-task learning (MTL) models for the sequence tagging problem because several existing argument mining studies achieved advanced performance when the MTL models were used (Schulz et al., 2018). Figure 4 shows the overview of the proposed model. This model comprises two modules. (i) The bi-directional LSTM (BiLSTM) module encodes the input tokens in a given document. (ii) The module for conditional random fields (CRFs) for each subtask (unit identification and classification) is applied to execute label classification, denoted by the layers toward the top of Figure 4. We call our model BiLSTM-CRF (BLC).

(i) Module of the BiLSTM

In our model, LSTMs were used to encode the input tokens. We employed the BiLSTM that uses two LSTMs to represent the forward and backward contexts in a sentence. Given an input token representation $x_i$ (i.e., a word embedding vector) at time $i$, we obtained the context-aware hidden state $h_i = LSTM_F(x_{1:i}) \circ LSTM_B(x_{n-i})$, where $n$ denotes the length of the input tokens.

(ii) Module of the CRFs

This output module discriminates the BIO tag and EU type using the hidden representations generated by the BiLSTM. When a CRF layer was utilized, the past and future tags to predict the present tag are used efficiently. Therefore, the CRF layer is effective for solving a non-independent sequence tagging problem, including BIO tagging, i.e., the label after “B” is obviously “I.” Note that we provided an MTL model that shares all the trainable parameters of the BiLSTM module but does not share any parameter of the CRF module for each task.

Experimental Setting

The train and test threads were randomly divided per a ratio of 8 : 2, where 30% of the train set is considered to be a development set. Herein, we set up two cases: post-level, a case in which an entire post is provided for the input and sentence-level, a case in which each sentence is given. The motivation is that the EUs cannot cross over a sentence. However, the sentence-level case would lose structural information. To remedy this problem, we focus on two structural features: whether or not the sentence is in OP post and the normalized position of the sentence in a post.

Each system was trained for 50 iterations with random seeds and the model that exhibited the optimal performance using the development set was selected (Schulz et al., 2018). The hyperparameters of our network are 100-dim GloVe embeddings (Pennington et al., 2014), the hidden layer size 200 for the BiLSTMs, the dropout rate of 0.5, and the optimizer is Adam (Kingma and Ba, 2014).

Results: Overall Performance

Table 4 shows that our proposed sentence-level BLC with structural information performs best regarding macro F1 in either boundary identification or unit classification tasks. The presence of
structural features provides an excellent boost in classifying unit types. For the post-level BLC, the model yields a better score in predicting non-EU parts because post-level discrimination can capture the entire post and thus identify irrelevant boundaries, such as supplementary notes, in posts.

Results: Analysis of Structural Features

Figure 5 shows the performance improvement by structural features in a distinct experiment from Table 4. In general, we see our structural features perform better for EUs that have characteristic positional distributions such as Policy and Testimony as shown in Figure 3. Interestingly, the structural features did not perform well in Fact classification for OPs and positive posts. The result implies Facts in the OPs and positive posts would be solid and can be predicted without structural information.

6 Conclusion

This paper demonstrated empirical work on the five types of elementary units (EUs) in online arguments. To find out the role of EUs, we annotated 4612 EUs and showed both intuitive and non-intuitive results. We also proposed baseline neural models to discriminate EUs. Experimental results showed that positional roles of some EUs are essential in detecting EUs. The annotated corpus and neural models will be applicable for future persuasion reasoning or evaluation of persuasiveness.

Acknowledgments

This work was supported by JST CREST (JPMJCR15E1) and JST AIP-PRISM (JPMJCR18ZL), Japan.

Table 4: Performance of the average F1. Max score if bold and significant differences with p < 0.05 if *.

<table>
<thead>
<tr>
<th>BLC variant</th>
<th>unit identification</th>
<th>unit classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>I</td>
</tr>
<tr>
<td>sentence-level</td>
<td>75.4</td>
<td>92.8</td>
</tr>
<tr>
<td>structural</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sentence-level +</td>
<td>75.6</td>
<td>92.8</td>
</tr>
<tr>
<td>structural</td>
<td></td>
<td></td>
</tr>
<tr>
<td>post-level</td>
<td>67.8</td>
<td>92.9</td>
</tr>
</tbody>
</table>

Figure 5: Performance improvement (F1 of a unit classification) by adding the structural features.

References


