A Bootstrap Method for Automatic Rule Acquisition on Emotion Cause Extraction

Shuntaro Yada
Graduate School of Education
The University of Tokyo
Tokyo, Japan
shuntaroy@p.u-tokyo.ac.jp

Kazushi Ikeda and Keiichiro Hoashi
KDDI Research, Inc.
Saitama, Japan
{kz-ikeda,hoashi}@kddi-research.jp

Kyo Kageura
Graduate School of Education
The University of Tokyo
Tokyo, Japan
kyo@p.u-tokyo.ac.jp

***Abstract*—Emotion cause extraction is one of the promising research topics in sentiment analysis, but has not been well-investigated so far. This task enables us to obtain useful information for sentiment classification and possibly to gain further insights about human emotion as well. This paper proposes a bootstrapping technique to automatically acquire conjunctive phrases as textual cue patterns for emotion cause extraction. The proposed method first gathers emotion causes via manually given cue phrases. It then acquires new conjunctive phrases from emotion phrases that contain similar emotion causes to previously gathered ones. In existing studies, the cost for creating comprehensive cue phrase rules for building emotion cause corpora was high because of their dependencies both on languages and on textual natures. The contribution of our method is its ability to automatically create the corpora from just a few cue phrases as seeds. Our method can expand cue phrases at low cost and acquire a large number of emotion causes of the promising quality compared to human annotations.**

***Index Terms*—sentiment analysis; emotion cause; bootstrapping method; rule base; cost reduction**

I. Introduction

Sentiment analysis is an important research field in text mining [1]–[3]. *Emotion cause* (EC) extraction is one of the promising yet not well-investigated research topics in sentiment analysis. EC, also referred as emotion cause events or cause events, can be defined to "events that evoke an expressed emotion". As pointed out in [4], most psychological theories regard EC as an integral part of emotion [5]; in cognitive linguistics it is held that the cause of an emotion should be an event itself [6]. Constructing EC corpora through EC extraction enables the following applications. First, utilising EC as a source of external knowledge improves the quality of emotion estimation tasks [7], [8]. For example, while traditional lexicon-based emotion estimation may mark some sentences that contain no explicit emotion words as *neutral* (no emotion), an estimator can infer sentiment from EC events inside the sentences. The corpus also makes it possible to analyse human sentiment expressed in texts in a more comprehensive way. If we can construct a sufficiently large corpus, we would be able to find a new insight about human sentiment via chronological or geometric analysis of the corpus.

Existing work on EC extraction mostly proposes or adapts rule-based methods such as utilising cue phrases that appear frequently in emotion-bearing sentences containing ECs (e.g. "because"). Rule-based methods are language- and domain-dependent, and need to be adapted manually for addressing new languages or new domains, as in [8]. Adapting detection rules for another language or class of texts requires a careful linguistic analysis of the target language and target texts. Some studies apply supervised machine learning (ML) methods to this task. These methods require annotated EC corpora for training, which can be costly if the corpora are to be created by hand. As we show in Sec. IV-B, the portion of sentences containing an informative type of EC is relatively small even in a corpus consisting of emotional sentences. Furthermore, it is difficult to manually annotate ECs consistently. Besides, in application, ECs are linked to corresponding emotions evoked by the cause events, so that their corpora have to be recreated depending on adopted emotion classification theories. These difficulties might hinder the progress of EC extraction work.

This research proposes a bootstrapping method to automatically acquire cue phrases in order to construct EC corpora more efficiently. Our method iteratively acquires both new cue phrases and ECs by exploring a corpus, starting from a few instances of cue phrases. To the best of our knowledge, this is the first effort to apply a bootstrapping technique to EC extraction. Evaluation experiments for Japanese EC extraction have shown the following:

- Our method can increase the variety of cue phrases, and the quantity of ECs as well, at low cost;
- It also improves recall scores while maintaining precision in comparison to a rule-based baseline; and
- Its extraction errors are more predictable and treatable than those of ML methods.

The rest of this paper is organised as follows. Sec. II summarises previous work on EC extraction and bootstrap techniques. In Sec. III, we describe details of our method. Sec. IV explains the settings of four experiments we conducted. Sec. V shows the results, which are analysed in Sec. VI. Finally, we conclude this paper in Sec. VII.

II. Related Work

We briefly summarise (A) existing studies that deal with EC from the perspective of text mining and (B) work on bootstrapping techniques which we adopted.

A. *Emotion Cause*

While much work has been carried out that deals with sentiment analysis, events that cause emotion, or ECs, have not attracted much attention. Existing work on ECs can roughly be divided into two types: extracting or identifying ECs, and applying ECs to sentiment analysis. The former tends to focus on the method of extracting ECs from texts that bear sentiments, while the latter aims to use ECs to identify emotion more precisely.

1) Emotion Cause Extraction: A series of pioneering studies on EC extraction was conducted in [4], [9], [10] where ECs were automatically extracted from Chinese texts. The first work [9] created an EC corpus by manual annotation using a Chinese balanced corpus, and analysed it in terms of the way how ECs occur within texts. According to this study, in Chinese texts, expressions that represent EC tend to be surrounded by cue phrases such as causative verbs, perception verbs, prepositions, and conjunctions, and to occur before emotion words. Based on this observation, succeeding studies attempted to extract EC from the corpus by a rule-based method [4] or by a combination of rules and statistical learning [10].

A supervised learning method using Support Vector Machine was proposed in [11], which classifies events into EC events and unrelated events. Besides, new rules for EC extraction ware created in [12] based on OCC model [13] designed for computational emotion processing. Both work dealt with Chinese texts, while a sequential labelling with Conditional Random Field (CRF) was applied to EC extraction from an English corpus [14] which predicts spans and boundaries of ECs in the corpus.

2) Emotion Cause Application: EC was at first applied to sentiment processing in Japanese [7]. The method first gathers EC events from large web corpora by a rule-based method which utilises cue conjunctive phrases and emotion words. Next, in the sentiment estimation phase, the method searches the EC candidates which are similar to the gathered ECs in input utterances.

Another application on sentiment estimation was done in Chinese microblog posts [8]. It tuned and adapted the rule set developed in [9] for microblog texts, which are known for short and informal characteristics, because the original rules were based on the Chinese balanced corpus which mainly consists of printed news articles and books.

Except for [7] and [8], the studies introduced above were conducted with small corpora, consisting of less than 10,000 sentences. This is partly because manually creating EC corpora costs high. Note that these corpora are indispensable for both types of research, irrespective of whether they use rule- or ML-based methods. In rule-based methods, cue phrases around ECs are important, but to find them requires a careful linguistic analysis on corpora. These rules are dependent on languages and types of texts. Corpus creation for ML will also take resources because EC annotation is sometimes difficult even for humans. Furthermore, ECs are paired with their evoking emotion classes when they are investigated by social or psychological analyses. This causes corpus re-creation on an emotion-classification-theory basis due to the existence of several theories such as [5], [15].

B. *Bootstrap Technique*

Bootstrap technique is, in general, to automatically expand a desired type of data through iterated steps of acquiring textual patterns that appear with the type of data and data itself using acquired patterns, starting from a few instances of data called *seeds*. The method is applied to a wide variety of text mining tasks, especially for knowledge acquisition. The DIPRE algorithm was proposed for collecting pairs of information such as book titles and their authors from web texts [16]. Patterns of answers for question answering systems were gathered in [17]. A bootstrap method was also found useful for expanding entries of dictionaries [18].

As mentioned above, it is known that ECs occur with cue phrases in texts. Besides, both rule-based and ML-based methods focus cue phrases. By regarding cue phrases as textual *patterns*, bootstrapping methods can be applied to the extraction of ECs as well as their cue phrases.

III. Proposed Method

In this section, we first define the form of ECs we target. We then describe our proposed method.

A. *Targeted Emotion Cause*

The definition of EC is well discussed in [9] where ECs were defined as “explicitly stated events that trigger emotions expressed in texts”. In addition to this, we further limit its range to *ECs appearing in the same sentence*[1] of corresponding emotions. Although we recognise EC events are sometimes stated outside of a sentence which contains an expression of emotion triggered by those events, we decide to deal with such complex type of ECs as the next step. Given that the research in EC extraction is still in its beginning stage as summarised in Sec. II-A, we focus here on improving the EC extraction for the limited range of ECs.

From the grammatical view point, EC events are categorised into two types: verbal events and nominal events [4][2]. A Verbal events is *a clause (a group of words containing a subject and predicate verb) describing events of the cause of emotion*, whereas nominal events are its counterpart of *a nominal phrase*. We target *verbal events* here. For example, we collect “my birthday is celebrated” (a verbal event) from a sentence “I am happy that *my birthday is celebrated*”, but we do not catch “his word” (a nominal event) of a sentence “I’m sad on *his words*”.

The reason why we here focus on verbal events is because they are more informative as expressions of events. In contrast, nominal events tend to have less information about events than that in verbal events, i.e. some arguments could be

[1]We treat direct speech sentences within a sentence as a part of the sentence due to dependency parsing consistency.

[2]This categorisation is based on Chinese, but through corpus analysis we confirmed that they can also be applied to Japanese.

dropped in nominal events and the relations of subjects and predicates could be ambiguous. Considering that EC corpora would be applied to psychological or sociological analyses, it is desirable at least at the first stage that EC events have sufficient information about subject, predicate, object, and other arguments.

As for the position of ECs in the sentences, we consider only the *nearest* EC for an expressed emotion within a sentence as well as [9]. It is known that some ECs become multi-stepped and exist outside of an emotion bearing sentence through a demonstrative in the sentence. However, these complex cases bring another, possibly theoretical, issue of defining what sort of language expressions over what range can potentially bear events relevant to emotions. We safely put this issue aside for a future work and take a pragmatic standpoint.

In summary, we formulate the targeted EC as *the nearest clause describing events of the cause of emotion that is dependent on emotion words in an emotion-bearing sentence*.

B. Bootstrapping Method

We introduce here a bootstrapping way to acquire both ECs and their cue phrases iteratively. Cue phrases we targeting are *sequences of function words that are connected to emotion words in the sense of dependency*. In the sentence "I am happy *that* my birthday is cerebrated" the word "that" between the emotion word and the EC is the cue phrase. Note that the definition of function words constituting cue phrases should be adaptable corresponding to the formality of target corpora as described in III-B1.

Our method is based on the following idea: when an EC event of an emotion found in a sentence appears in another sentence bearing the same emotion, cue phrases between the EC and the emotion in both two sentences have the same conjunctive function. If we encounter another sentence "I am happy *because* my birthday is cerebrated", for instance, we can extract a new instance of cue phrases "because", starting from the original cue phrase "that". Our proposed method consists of two steps: (1) to extract ECs from a corpus by using known cue phrases (and emotion words), and (2) to acquire new cue phrases by searching another sentences that contain the same pairs of EC events and emotion words. By iterating these two steps, both ECs and their cue phrases will be augmented.

In step (1), i.e. to extract ECs by looking for cue phrases, it is preferable to use linguistic dependency information of sentences to remove noise. In step (2), i.e. acquisition of new cue phrases, EC events need to be dependent on emotion words or phrases as well. In addition, it is expected that searching common pairs of EC and corresponding emotions using exact matching may result no outcome, due to sparseness of data. We thus presume that judging by similarities of abstracted senses of sentences is more suitable than exact matching.

Taking into account these observations, we can more rigidly state the steps (1) and (2) as follows:

Seeds: A set of initial instances of cue phrases and emotion words (phrases);

Input: A set of (dependency-parsed) sentences that contain emotion words;

While: Until no new cue phrases is acquired from the input sentences:

Step (1): **Learning phase:** Collect nearest clauses as EC events among clauses that are dependent on emotion words and contain known cue phrases;

Step (2): **Acquisition phase:** Find clauses that are dependent on emotion words as EC candidates but do not contain known cue phrases, and then acquire functional word sequences between ECs and emotion words as cue phrases if the clauses are similar enough to previously learned ECs

This overall procedure is depicted in Fig. 1

This method has two threshold parameters s and t that will control the quality of gathered ECs and its cue phrases. The parameter s is the similarity threshold to judge whether to acquire cue phrases in step (2) or not. That is, if the semantic similarity between a learned EC and a found candidate clause is greater than the value, we regard these two sentences are sufficiently similar. The parameter t is the frequency threshold of cue phrases among clauses acquired in step (2). Even when a high similarity score is obtained for an EC candidate, but if the cue phrase of the clause appeared only once in acquired clauses, the cue phrase may be unsuitable to use for the next learning phase.

We further describe components of our proposed method in detail that need modifications depending on target languages and relying on emotion classification theories. In this research, we target Japanese, or one of the languages where few EC studies have been done but a set of cue phrases has been made available [7]. Since our focal point is to show effectiveness of our method, we implement these components in a simple manner instead of applying state-of-the-art techniques of related fields.

1) Definition of cue phrases: In step (2), the acquisition phase, a rule of cue phrases as sequences of words and part-of-speech types is required. Cue phrases mainly consists of functional words, while certain verbs and nouns sometimes constitute them [9]. In this research, under the Japanese grammar, we define the rule as sequences of part-of-speeches other than nouns, verbs, and adjectives, so that the sequences mainly consist of auxiliary verbs, particles, and symbols[3]. However, we also include some functional nouns that form noun clause.

2) Seeds of cue phrases: In the first learning phase, cue phrase seeds are used for storing ECs. The seeds can be selected based on a preliminary corpus analysis and a primitive observation, whereas they should follow the above definition of cue phrases. We chose eight Japanese cue phrases adopted in [7][4].

[3]We suppose symbols can be also key cue phrase component especially in informal texts.

[4]It used following Japanese connectives: ので, から, ため, て, のは, のが, ことは, ことが.

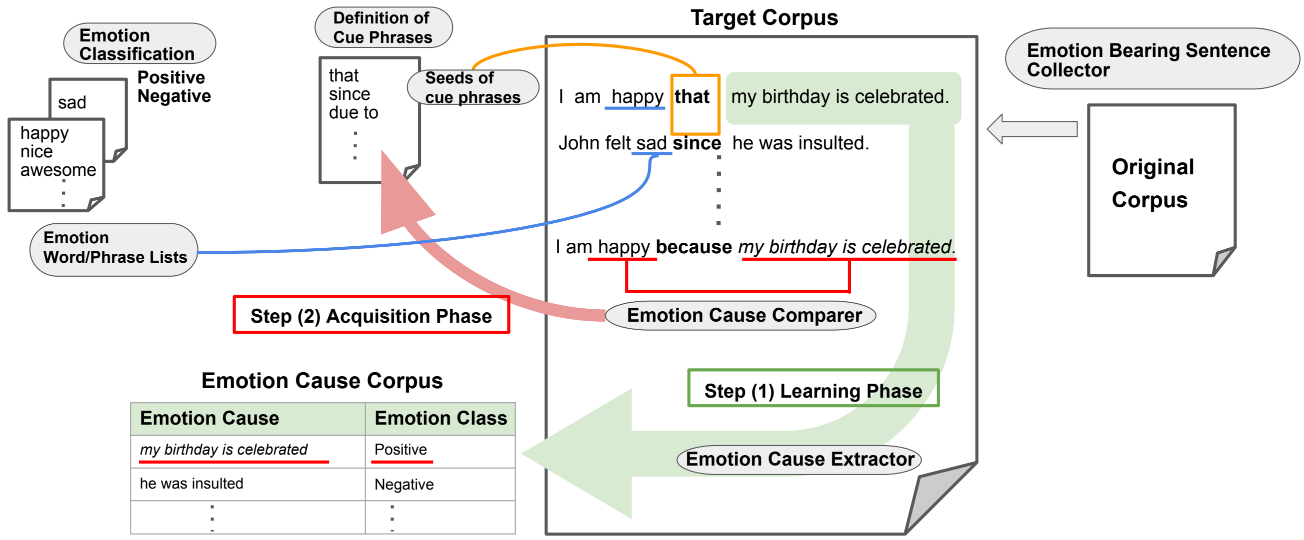

Fig. 1. The procedure of the proposed method

3) Emotion classification: In the learning phases, the proposed method stores ECs with corresponding emotion classes for later use in EC comparer (Sec. III-B7). Related studies we saw in Sec. II-A adopt different emotion classifications (e.g. Ekman and Plutchik, or more simply, binary and binary with neutral). That means the emotion classification to be adopted is determined by the purpose of creating EC corpora. For simplicity, this research, only considers binary emotion classes: positive and negative.

4) Emotion word/phrase lists: Our method elicits emotions from sentences based on the lists of emotion words as the input data. It is a well known approach to use information gain for the feature selection in text categorisation [19]. We adapted a simple way reported in [20] which finds characteristic words of a particular class among several classes of texts by an AIC-based feature selection algorithm. We first extracted characteristic words of positive and negative tweets by the algorithm from randomly collected 9,006 Japanese tweets which was manually labelled with positive (38%), negative (32%), and neutral (30%). We then manually selected 26 emotional adjectives of less semantic ambiguity[5] from top 10% scored words. In particular, positive words contain 10 words meaning *happy, nice, cute, funny, interesting,* and *awesome*, while negative words include 16 words meaning *sad, lonely, scare, hard, painful, tired, yucky,* and *hateable*. Note that we include different notations of words of the same meaning due to Japanese character sets (i.e. Hiragana, Katakana, and Kanji).

5) Emotion bearing sentence collector: This component gathers sentences that contain the selected emotion words or phrases from corpora. We simply adopted exact pattern matching of the emotion words, which can meet the minimum requirement. Note that, for instance, if an emotion word/phrase list contains some complex idiomatic phrases, this module may require dependency parsing.

6) Emotion cause extractor: The proposed method searches for and extracts ECs among clauses that are dependent on emotion words or phrases. The EC extractor first applies dependency parsing to emotion bearing sentences[6], then resolves dependencies of emotion words or phrases, and finally extracts ECs into a specified format. It requires dependency parser which can process a target language. In this research, we adopt a Japanese dependency parser ASA [21]. The dependency resolving part needs to identify whether a targeted clause contains cue phrases or not, which can be achieved by pattern matching. The extraction part outputs triplets of an extracted EC, a cue phrase (or its candidate, in the acquisition phase), and an evoked emotion word/phrase.

7) Emotion cause comparer: In acquisition phases, this component judges whether or not an EC candidate of an input emotion bearing sentence is similar enough to learned ECs. In order to judge the similarity, the way of semantic representation of clauses (sentences) has to be defined. While several techniques are available to represent the meaning of a sentence, we form numeric vectors with a fixed dimension produced by word embeddings of the skipgram algorithm [22] and calculate similarities of sentences by cosine similarity. This EC vector is composed by ordered concatenation of word vectors of a predicate and its arguments, which are inferred from dependency parsing information processed by the EC extractor. If the predicate and its arguments consist of more than a word, their vectors are constructed by the mean of component words' vectors. We defined the dimension of each word as 300, and the numbers of arguments we take into

[5]We consider to what extent a word is used for describing emotional states.

[6]As written in "Input:" item of the method's overview, we can alternatively process this dependency parsing on a whole input sentences before the main two step iteration.

account as 10[7]. The dimension for EC, therefore, is 3,300.

IV. Experiment

A. Experimental Setting

We design three experiments to investigate and evaluate the proposed method. Our aim is to acquire a large size of ECs and cue phrases with reasonable quality at low cost. We first investigate effects of parameters and seeds on the quantities of ECs and cue phrases in **Experiment 1**. From their definition, we can say that two parameters (s, t) control quantities of learned ECs and acquired cue phrases while seed cue phrases determine required steps to collect a certain amount of outcomes. For the parameter investigation, we choose the value of s from 0.7, 0.8, and 0.9, and that of t from 2, 10, and 100. In these 9 ($= 3 \times 3$) combinations, we fix the size of seed cue phrase to eight. On the other hand, for investigation of the seed effect, we compare results between the setting of all eight entities and that of four entities, with parameters fixed to the middle combination ($s = 0.8$ and $t = 10$) as a representative.

Second, in **Experiment 2**, we apply our method to different classes of corpora and examine the quantity of the output as a simulation of practical situations. Previous research targeted various corpora such as general books, news papers, web pages, and user-generated texts. If our method can work for these textual classes, it has generality in terms of textual domains. To conduct this experiment, we have created three corpora, which will be described in the next subsection (IV-B).

For quality evaluation, we compare the output of our method with human-annotated data. Using the annotated data as a gold standard, we calculate precision and recall of extracted ECs in **Experiment 3**. Especially for EC extraction, we also evaluate results with baselines defined in related work, the details of which are shown in Sec. IV-C. In order to calculate precision and recall, we compare the tokens labelled as EC by human and the methods. As tokens labelled by human annotation are true tokens, precision equals to the ratio of tokens labelled by both human and the methods divided by all tokens labelled as true by the methods. Similarly, recall is calculated as the ratio of tokens labelled by both human and the methods divided by all tokens labelled by human.

B. Data Set

In order to investigate whether our method has corpus-independent generality in Experiment 2, we created three corpora: **newspaper** articles[8], **web news** articles[9], and **Q&A site** texts[10]. We sampled 100,000 sentences that contain emotion words defined in Sec. III-B4 for each corpus, to make all the same size (henceforth, *emotion sentence corpora*).

[7]We use shallow cases in inferring predicate argument structures instead of deep cases (semantic roles) because we put importance on reliability at this stage.

[8]A Japanese newspaper source "Mainichi Shimbun" published by The Mainichi Newspapers Co., Ltd.

[9]Online news articles collected from 40 Japanese news web sites.

[10]Japanese Yahoo! Answers, one of the most popular Q&A site in Japan. http://chiebukuro.yahoo.co.jp

TABLE I
THE MEASUREMENT OF AGREEMENT BETWEEN TWO ANNOTATORS.

Annotation label	News paper	Web news	Q&A site
Beginning of EC	0.63	0.56	0.59
End of EC	0.74	0.66	0.69
Cue phrases*	0.77	0.68	0.69
Nominal EC	0.51	0.50	0.46

* Cue phrase agreement is calculated by exact match of strings.

For Experiment 3, we further sampled 1,000 sentences from each emotion sentence corpora, and asked two annotators to label spans of EC events separately. The detailed guideline for labelling each sentence is as follows: (1) find EC candidates corresponding to each emotion words in a sentence; (2) among the candidates, choose the nearest one from the emotion word; (3) label the span of the nearest EC counted by chunks; and (4) extract cue phrases between the EC and the emotion word. In addition, for our information, we asked annotators to label EC which appear to be nominal events that we excluded from the focus of this research (see Sec. III-A).

We evaluated the agreement of two annotators by measuring Cohen's κ (kappa) coefficient [23] in order to assess whether the annotation data can be applied to Experiment 3. The result is shown in Table I. According to the common interpretation of the Cohen's κ [24], the annotations of EC spans and cue phrases mostly achieve the "substantially good ($\kappa \geq 0.6$)" agreements for all sample sets. However, the fact that the agreements of nominal ECs are relatively low implies the interpretation and the judgement of nominal events are difficult even for humans. Two annotators report that they used 104 hours and 107.5 hours respectively, from which we can confirm the human annotation of EC requires high cost.

Table II shows the numbers of ECs and cue phrases labelled by one of the annotators, for each annotation datum. Since we asked annotators to label ECs per corresponding emotion words, the total numbers are greater than the sentence size (1,000). In this table, ECs are broken down into the aforementioned two grammatical types: verbal and nominal ECs. Verbal events with cue phrases are separately counted too. This result indicates more than a half of emotional sentences bring ECs, but the informative type of them (verbal events) are quite small. However, most verbal ECs appear to occur with cue phrases, which confirms that cue phrase is an important key for EC extraction.

C. Baseline

In Experiment 3, we evaluate the quantity and quality of acquired ECs. We set two baseline methods corresponding to two major techniques in this field, i.e., rule-based and ML-based methods.

1) Rule-based baseline: Among rule-based methods introduced in Sec. II-A, we chose [7] which targets Japanese texts. They extract ECs by matching patterns of combination with emotion words and cue phrases. This method is identical to executing the first learning phase of our method.

TABLE II
THE NUMBERS OF ANNOTATED ECS AND CUE PHRASES FOR EACH CORRESPONDING EMOTION WORD.

	News paper	Web news	Q&A site
With ECs	720	714	599
- Verbal events	272	341	320
· with cue phrases	241	299	280
- Nominal events	448	373	279
No ECs	311	355	450
Total emotion words	1,031	1,069	1,049

2) ML-based baseline: Following [14], we chose CRF-based sequential labelling, and adapted it for Japanese with a simple and standard feature set: one-hot word vectors, part-of-speech tags, emotion word labels, seed cue phrase labels. This method considers two tokens surrounding each focal token. Through five-fold cross-validation, we report the result produced by the best parameters of CRF ranked by F-measure, or the harmonic mean of precision and recall.

V. RESULT

In **Experiment 1**, we applied our method to the news paper corpus for the parameters investigation, while all the corpora were used for the seeds investigation. The results are shown in two tables, Table III and Table IV respectively. From Table III, increasing the values of two parameters cause smaller sizes of acquired cue phrases and learned ECs. According to Table IV, changing the size of cue phrases has small effect on the number of obtained candidates.

Fig. 2 and 3 report the result of **Experiment 2**. They show both numbers of cue phrases and ECs were increasing with cycles among all corpora. We can see they increase in a similar manner, although the extracted ECs of Q&A site were smaller than the others.

The result of **Experiment 3** is shown in Table V. In comparison to the rule-based baseline, our method improved recall value when its iteration works. It is notable that, for the web news corpus, our method can maintain the precision value. Except for the newspaper corpus, precision of our method is higher than the ML-based baseline although the baseline outperformed our method in recall.

The numbers of steps of the proposed method in Experiment 3 are smaller compared to those in Experiment 1, because we only used manually annotated 1,000 sentences to obtain emotion cues in Experiment 3 whereas we used 100,000 sentences in Experiment 1.

VI. ANALYSIS

A. Experiment 1

In Experiment 1, we have found that both parameters can control the quantity of cue phrases and ECs, and also they may coordinate the quality of outcomes too. We could not conduct this parameter investigation on annotated data, because the data are too small to apply our method according to the result of Experiment 3. However, we can reasonably conclude that

TABLE III
EXPERIMENT 1: THE EFFECTS OF PARAMETER COMBINATIONS ON ECS AND CUE PHRASE ACQUISITION.

s	t	Cue phrases	ECs	Steps
0.7	2	194	19,691	3
	10	59	19,832	3
	100	18	17,186	3
0.8	2	127	19,631	3
	10	42	17,820	4
	100	16	16,703	3
0.9	2	78	19,540	3
	10	25	17,335	3
	100	8	7,327	1

TABLE IV
EXPERIMENT 1: THE EFFECTS OF THE NUMBERS OF SEED CUE PHRASES.

Seed size	Corpora	Cue phrases	ECs	Steps
8	News paper	42	17,820	4
	Web news	37	18,130	3
	Q&A site	39	12,066	3
4	News paper	36	17,594	4
	Web news	32	17,879	3
	Q&A site	36	12,340	3

the obtained outcomes can achieve higher quality as the values of the parameters increase, since infrequently appeared cue phrases are excluded as t increases and increasing s means to keep high permissible logical relations between ECs and corresponding emotion classes connected by acquired cue phrases.

We have also found the our method needs only a few cue phrases as seeds. This means it is applicable to the situation where little or no EC analysis has done in a particular language or genre of corpus, because it allows us to start an EC research even from a primitive observation of a small cause event sample to pick up just a few cue phrases.

B. Experiment 2

From the result of Experiment 2, we can conclude that our method can be applicable to any textual domain with almost no modification. This also helps us to make an initial commitment to EC extraction of any specific purpose. We can estimate from the annotation data (Table II) that Q&A site seems to contain smaller ECs than the others. That can be one of the reasons the relatively small size of ECs is extracted from the Q&A site corpus, since this corpus mainly consists of informal texts, which are difficult to parse correctly.

C. Experiment 3

Through Experiment 3, we have found that the proposed method can expand previous rule-based techniques by improving recall scores. The result that our method did not iterate in the Q&A corpus suggests that this size of the corpus ($n = 1000$) is unsuitable to it, but we do not think this point as a drawback since a large volume of unlabelled textual data can easily be obtained.

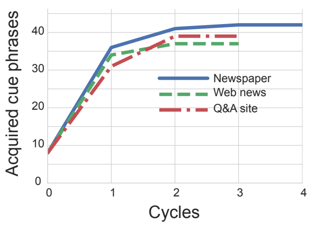

Fig. 2. Experiment 2: The numbers of acquired cue phrases with cycle steps

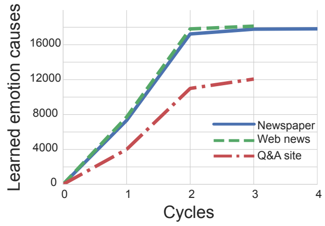

Fig. 3. Experiment 2: The numbers of acquired EC with cycle steps

TABLE V
EXPERIMENT 3: PRECISION AND RECALL VALUES OF EXTRACTED ECS FROM ANNOTATED SAMPLES COMPARING WITH BASELINE METHODS.

Method	Corpora	Precision	Recall	Steps
Proposed method	News paper	0.46	0.28	2
	Web news	0.75	0.21	2
	Q&A site	0.64	0.07	1
Rule-based baseline	News paper	0.65	0.21	-
	Web news	0.76	0.18	-
	Q&A site	0.64	0.07	-
ML-based baseline*	News paper	0.72	0.53	-
	Web news	0.68	0.54	-
	Q&A site	0.57	0.36	-

Here we report error analyses of false positives and false negatives produced by our method. First, we shall mention the negatively outstanding precision score on the news paper corpus. The precision drop means that the proposed method produced many false positives at through the iteration of cycles. This is due to an acquisition of irrelevant cue phrases at the second cycle. In fact, it acquired Japanese period and comma as cue phrases. Remember that we included signs and symbols into the definition of cue phrases (see the footnote of Sec. III-B1). As a result of this, it learned some normal dependent clauses that state general events instead of ECs.

Another factor is that some sentences contain sub-sentences inside direct speech quotations in the corpus because of our decision to keep whole dependency structures of such sentences (see footnote 1). The dependency parser we adopted treats such sub-sentences as normal clauses. This sometimes causes that a sub-sentence is wrongly parsed as a clause dependent on an emotion word. Emotion bearing sentences often appears in interview articles[11] in this kind of corpus, or news paper articles. These are the reasons our method acquired the period as a cue phrase and learned previous sentences occurred before nominal ECs, while almost all (96%) of the false positives were originally annotated as nominal events.

> A woman who experienced the Nankai earthquakes said "At that moment, *my mother carried me on her back and ran*. **Tsunami** is scary."

[11]For example, there exists numeral short or long interviews with sports players, entertainers, artists, professionals, bureaucrats, politicians, and witnesses of some events.

In this example, our method learned the sub-sentence with *italics* whereas an annotator labelled "**Tsunami**" as a nominal EC of emotion word "scary". Part of these types of false positive instances could be included in general interpretation of ECs, although we excluded them by dealing with the nearest EC from emotion words. At this stage, we put them aside the range of valid ECs, although they might be included in the future progress of this research area.

The above false positive problems also occurred on the other corpora. We note that false annotations of nominal ECs partly affect scores among all corpus settings. In addition to them, we here describe other corpus-specific false positives. In the web news corpus, we found some named entities that represent an EC and a corresponding emotion word such as a book title *Is it fun to be adult?*. That is because web news tend to publish articles that target specific and casual products more frequently than news paper. The Q&A corpus has question-formed sentences that sometimes ask, for instance, whether an EC uncommon for a questioner evoke an emotion or not. The annotators we asked labelled some of them as non EC. These cases suggest again that EC annotation is a difficult task even for humans.

Secondly, we summarise possible causes of false negatives based on a simple observation. They are categorised into two types: our simple implementation of the components and the complexity of real-world ECs. As for the first one, we implemented each component of the proposed method in simple ways. This brings errors especially in parsing steps, or mainly in EC extractor and EC comparer. We use a general dictionary for the dependency parser we adopted for EC extractor, which caused miss parsing such as splitting failure of noun or verb phrases, and labelling failure of cue phrases due to irrelevant named entity entries in the dictionary. However, these problems can be solved by tuning the parser for target corpora, and this is easily achieved when and where linguistic resources for general NLP tasks are sufficiently available. We do not further describe these points because they strongly depend on the available techniques in target languages.

On the other hand, in the real world, ECs appears in complex ways. Some sentences contain several ECs in a parallel or multi-stepped manner, while some ECs are wrapped by direct/indirect speeches or into noun clauses. Our method sometimes appears to elicit part of the parallel or multi-stepped ECs, but this is also related to the definition of their range. The latter "wrap" case can be solved by available dependency parsers or predicate argument structure analysers of given languages, which similarly depends on the progress of target languages' fundamental NLP research.

From a practical point of view, we conclude that precision and recall can be controlled by the definition of cue phrases and preprocessing on an input corpus. Moreover, our method holds the common merit of rule-based methods, that is, to predict error types is relatively easier than ML methods. Even when ML methods perform well in experimental situations, we will have less idea of failure patterns in general. At the end of this section, we note that our method is also applicable more easily in terms of time consumption since it does not require 100 hours to execute the process.

VII. Conclusion

This paper proposed a bootstrapping technique to automatically acquire conjunctive phrases as textual cue patterns for emotion cause (EC) extraction. Our method consists of language-independent specifications of components, and enables us to acquire ECs by extending cue phrases at low cost. While two major methods, rule-based and ML-based methods, have been proposed in related work, our method has promising possibilities to improve recall compared to rule-based methods and to output results which can be more predictable and easy to post-process in comparison with those of ML methods. In other words, our method contributes to the further development of this emerging research field.

The proposed method iterates the following two phases: (1) to gather ECs via manually given cue phrases, and then (2) to acquire new conjunctive phrases from emotion phrases that contain similar ECs to previously gathered ones. We conducted three experiments to investigate (1) the effects of its parameters and the size of seed cue phrases, (2) its corpus independent applicability and (3) its performance with standard methods previously proposed in the EC extraction field. The experimental results revealed that with the tuned parameters, our method can obtain various cue phrases and improves the recall of ECs keeping its precision in comparison with a rule-based baseline.

Our method still has room for improving precision and recall scores. For the next step, we plan to conduct further error analyses. In addition, we plan to expend our method by implementing a mechanism which enables the automatic coordination of its parameters and reports the degree of confidence in acquiring cue phrases or learning ECs. In order to achieve this, we seek appropriate definitions of evaluation functions for the parameters and outputs. Furthermore, considering the possibility that the relation between an EC and its evoking emotion can be regarded as a particular type of discourse relationships, we intend to adapt recent methods of discourse relation recognition [25].

Acknowledgement

This work was partly supported by JSPS KAKENHI Grant Number JP 16K12542.

References

[1] B. Liu and L. Zhang, "A Survey of Opinion Mining and Sentiment Analysis," in *Mining Text Data SE - 13*, C. C. Aggarwal and C. Zhai, Eds. Springer US, 2012, pp. 415–463.

[2] G. Paltoglou, "Sentiment Analysis in Social Media," in *Online Collective Action SE - 1*, ser. Lecture Notes in Social Networks, N. Agarwal, M. Lim, and R. T. Wigand, Eds. Springer Vienna, 2014, pp. 3–17.

[3] A. Yadollahi, A. G. Shahraki, and O. R. Zaiane, "Current State of Text Sentiment Analysis from Opinion to Emotion Mining," *ACM Computing Surveys*, vol. 50, no. 2, pp. 1–33, 2017.

[4] S. Y. M. Lee, C. Ying, and C.-R. Huang, "Emotion Cause Events: Corpus Construction and Analysis," *Proceedings of The Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, pp. 19–21, 2010.

[5] R. Plutchik, *The Emotions: Facts, Theories, and a New Model*. New York Random House, 1962.

[6] L. Talmy, *Towards a Cognitive Semantics, Volume II. Typology and Process in Concept Structuring*, 2000, vol. 2.

[7] R. Tokuhisa, K. Inui, and Y. Matsumoto, "Emotion Classification Using Massive Examples Extracted from the Web," in *Proceedings of the 22nd International Conference on Computational Linguistics*, ser. COLING '08. Stroudsburg, PA, USA: Association for Computational Linguistics, 2008, pp. 881–888.

[8] W. Li and H. Xu, "Text Based Emotion Classification using Emotion Cause Extraction," *Expert Systems with Applications*, vol. 41, no. 4 PART 2, pp. 1742–1749, 2014.

[9] S. Y. M. Lee, Y. Chen, and C. R. Huang, "A Text-Driven Rule-Based System for Emotion Cause Detection," *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, no. June, pp. 45–53, 2010.

[10] Y. Chen, S. Y. M. Lee, S. Li, and C.-R. Huang, "Emotion Cause Detection With Linguistic Constructions," *Proceedings of the 23rd International Conference on Computational Linguistics*, no. August, pp. 179–187, 2010.

[11] L. Gui, R. Xu, Q. Lu, D. Wu, and Y. Zhou, "Emotion Cause Extraction, A Challenging Task with Corpus Construction," in *Proceedings of Chinese National Conference on Social Media Processing*. Springer Singapore, 2016, pp. 98–109.

[12] K. Gao, H. Xu, and J. Wang, "A Rule-Based Approach to Emotion Cause Detection for Chinese Micro-Blogs," *Expert Systems with Applications*, vol. 42, no. 9, pp. 4517–4528, 2015.

[13] B. N. Colby, A. Ortony, G. L. Clore, and A. Collins, "The Cognitive Structure of Emotions." *Contemporary Sociology*, vol. 18, no. 6, p. 957, 1989.

[14] D. Ghazi, D. Inkpen, and S. Szpakowicz, "Detecting Emotion Stimuli in Emotion-Bearing Sentences," in *16th International Conference on Intelligent Text Processing and Computational Linguistics*. Springer International Publishing, 2015, pp. 152–165.

[15] P. Ekman, "Expression and the Nature of Emotion," in *Approaches to Emotion*, 1984, pp. 319–344.

[16] S. Brin, "Extracting Patterns and Relations from the World Wide Web," in *Selected Papers from the International Workshop on The World Wide Web and Databases*, ser. WebDB '98. London, UK, UK: Springer-Verlag, 1999, pp. 172–183.

[17] D. Ravichandran and E. Hovy, "Learning Surface Text Patterns for a Question Answering System," in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ser. ACL '02. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002, pp. 41–47.

[18] T. McIntosh and L. Yencken, "Relation Guided Bootstrapping of Semantic Lexicons," *Proceeding HLT '11 Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*, pp. 266–270, 2011.

[19] C. Lee and G. G. Lee, "Information Gain and Divergence-Based Feature Selection for Machine Learning-Based Text Categorization," *Information Processing and Management*, vol. 42, no. 1, pp. 155–165, 2006.

[20] K. Ikeda, G. Hattori, C. Ono, H. Asoh, and T. Higashino, "Twitter User Profiling based on Text and Community Mining for Market Analysis," *Knowledge-Based Systems*, vol. 51, pp. 35–47, 2013.

[21] K. Takeuchi, S. Tsuchiyama, M. Moriya, Y. Moriyasu, and K. Satoh, "Verb Sense Disambiguation Based on Thesaurus of Predicate-Argument Structure: An Evaluation of Thesaurus of Predicate-Argument Structure for Japanese Verbs," in *Proceedings of the International Conference on Knowledge Engineering and Ontology Development*, 2011, pp. 208–213.

[22] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," *arXiv preprint arXiv:1607.04606*, 2016.

[23] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[24] R. Artstein and M. Poesio, "Inter-Coder Agreement for Computational Linguistics," *Computational Linguistics*, vol. 34, no. 4, pp. 555–596, 2008.

[25] J. Li, R. Li, and E. Hovy, "Recursive Deep Models for Discourse Parsing," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, 2014, pp. 2061–2069.