Annotating and Extracting Synthesis Process of All-Solid-State Batteries from Scientific Literature

Fusataka Kuniyoshi[1,3]**, Kohei Makino**[2,3]**, Jun Ozawa**[1,3]**, Makoto Miwa**[2,3]

[1]Panasonic Corporation
[2]Toyota Technological Institute
[3]National Institute of Advanced Industrial Science and Technology (AIST)
{kuniyoshi.fusataka, kohei.makino, ozawa.jun, makoto.miwa}@aist.go.jp

Abstract

The synthesis process is essential for achieving computational experiment design in the field of inorganic materials chemistry. In this work, we present a novel corpus of the synthesis process for all-solid-state batteries and an automated machine reading system for extracting the synthesis processes buried in the scientific literature. We define the representation of the synthesis processes using flow graphs, and create a corpus from the experimental sections of 243 papers. The automated machine-reading system is developed by a deep learning-based sequence tagger and simple heuristic rule-based relation extractor. Our experimental results demonstrate that the sequence tagger with the optimal setting can detect the entities with a macro-averaged F1 score of 0.826, while the rule-based relation extractor can achieve high performance with a macro-averaged F1 score of 0.887.

Keywords: Synthesis process, Flow graph, Corpus annotation, Named entity recognition, Relation extraction

1. Introduction

With the rapid progress in the field of inorganic materials, such as the development of all-solid-state batteries (ASSBs) and solar cells, several materials researchers have noted the importance of reducing the overall discovery and development time by means of computational experiment design, using the knowledge of published scientific literature (Agrawal and Choudhary, 2016; Butler et al., 2018; Wei et al., 2019). To achieve this, automated machine reading systems that can comprehensively investigate the synthesis process buried in the scientific literature is necessary.

In the field of organic chemistry, Krallinger et al. (2015) proposed a corpus in which chemical substances, drug names, and their relations are structurally annotated in documents such as papers, patents, and medical documents, while composition names are provided in the abstracts of molecular biology papers. Linguistic resources are available in abundance, such as the GENIA corpus (Kim et al., 2003) of biomedical events on biomedical texts and the annotated corpus (Kulkarni et al., 2018) of liquid-phase experimental processes on biological papers. In biomedical text mining, the detection of semantic relations is actively researched as a central task (Miwa et al., 2012; Scaria et al., 2013; Berant et al., 2014; Rao et al., 2017; Rahul et al., 2017; Björne and Salakoski, 2018). However, the relations in biomedical text mining represent the cause and effect of a physical phenomenon among two or more biochemical reactions, which differs from the procedure of synthesizing materials.

In the field of inorganic chemistry, only several corpora have been proposed in recent years. Kononova et al. (2019) constructed a general-purpose corpus of material synthesis for inorganic material by aligning the phrases extracted by a trained sequence-tagging model. However, this corpus did not include relations between operations, and therefore, it was difficult to extract the step-by-step synthesis process. Mysore et al. (2019) created an annotated corpus with relations between operations for synthesis processes of general materials such as solar cell and thermoelectric materials. However, the synthesis processes of ASSBs are hardly included even though the operations, operation sequences, and conditions also have differences due to the characteristics of the synthesis process for each material category.

In this study, we took the first step towards developing a framework for extracting synthesis processes of ASSBs. We designed our annotation scheme to treat a synthesis process as a synthesis flow graph, and performed annotation on the experimental sections of 243 papers on the synthesis process of ASSBs. The reliability of our corpus was evaluated by calculating the inter-annotator agreements. We also propose an automatic synthesis process extraction framework for our corpus by combining a deep learning-based sequence tagger and simple heuristic rule-based relation extractor. A web application of our synthesis process extraction framework is available on our project page [1]. We hope that our work will aid in the challenging domain of scholarly text mining in inorganic materials science.

The contributions of our study are summarized as follows:

- We designed and built a novel corpus on synthesis processes of ASSBs named SynthASSBs, which annotates a synthesis process as a flow graph and consists of 243 papers.

- We propose an automatic synthesis process extraction framework by combining a deep learning-based sequence tagger and rule-based relation extractor. The sequence tagger with the best setting detects the entities with a macro-averaged F1 score of 0.826 and the rule-based relation extractor achieves high performance with a macro-averaged F1 score of 0.887 in macro F-score.

---

[1] http://synth-extractor-demo.fusataka-k.com/

The pure $Li_4Ti_5O_{12}$ material, denoted LTO, was obtained from Li_2CO_3 (99.99 %, Aladdin) and anatase TiO_2 (99.8 %; Aladdin) precursors, mixed, respectively, in a 4:5 molar ratio of Li:Ti. The precursors, dispersed in deionized water, were ball-milled for 4 h at a grinding speed of 350 rpm, and then calcined at 800 °C for 12 h after drying.

Figure 1: Example of a synthesis process. The underlined phrases relate to the material synthesis process.

2. Annotated Corpus

In this section, we present an overview of our annotation schema and annotated corpus, which we named the SynthASSBs corpus.

2.1. Synthesis Graph Representation

We used flow graphs to represent the step-by-step operations with their corresponding materials in the synthesis processes. Using the flow graphs, it was expected that we could represent links that are not explicitly mentioned in text. In the inorganic materials field, Kim et al. (2019) proposed the representation of the synthesis process using a flow graph and the definition of annotation labels in experimental paragraphs. In our annotation scheme, we followed their definition, with three improvements: (1) the property of an operation is treated as a single phrase, and not as a combination of numbers and units, (2) each label has been modified to capture the conditions necessary to synthesize the ASSB, and (3) a relation label for coreferent phrases is included to understand the anaphoric relations. A flow graph for the ASSB synthesis process is represented by a directed acyclic graph $G = (V, E)$, where V is a set of vertices and E is a set of edges. We provide an example section of the paper (Bai et al., 2016) in Figure 1, and the graph extracted from the sentences in the section in Figure 2.

2.2. Label Set

The labels contain vertices and edges for the synthesis graph representation in Section 2.2.1., and Section 2.2.2., respectively.

2.2.1. Vertices

The following vertex labels were defined to annotate spans of text, which correspond to vertices in the synthesis graph. The labels represent the materials, operations, and properties. For the material labels, we labeled all phrases that represent materials in the text, while operation and property labels were added only to those phrases related to the synthesis process. We segmented the roles of materials into categories. Moreover, we introduced multiple property types for analyzing the structure of the synthesis process. The defined labels and their examples mostly taken from Fig. 1 are explained in the following.

MATERIAL-START is a raw material used to synthesize the final material; for example, Li_2CO_3 or TiO_2.

MATERIAL-INTERMEDIUM indicates an intermediate material produced during the synthesis process; for example, "then, $LixMo_4O_6$ was obtained from a mixture of $InMo_4O_6$ and LiI."
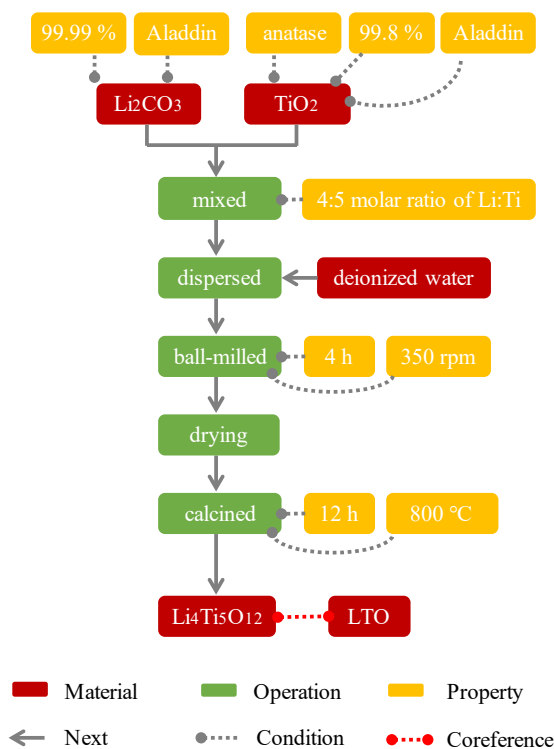

Figure 2: Example of the synthesis graph generated from Figure 1.

MATERIAL-FINAL represents the final material (or products) of the material synthesis process; for example, $Li_4Ti_5O_{12}$.

MATERIAL-SOLVENT is liquid that is used to dissolve substances and create solutions; for example, deionized water, ethanol, or methanol.

MATERIAL-OTHERS represents other materials that are not related to the synthesis process, such as compounds for thin films or catalysts; for example, "... and then purified with activated carbon and acid alumina."

OPERATION represents an individual action performed by the experimenters. It is often represented by verbs; for example, "... were ball-milled for 4 h ..."

PROPERTY-TIME represents a time condition associated with an operation; for example, "... were ball-milled for 4 h..."

PROPERTY-TEMP represents a temperature condition associated with an operation; for example, "... and then calcined at 800 °C ..."

PROPERTY-ROT indicates a rotational speed condition associated with an operation; for example, "... at a grinding speed of 350 rpm ..."

PROPERTY-PRESS represents a pressure condition associated with an operation; for example, "The powder was uniaxially cold pressed at 300 MPa."

PROPERTY-ATMOSPHERE represents an atmosphere condition associated with an operation; for example, "... was conducted in Ar atmosphere for 3 h."

PROPERTY-OTHERS represents other conditions associated with an operation or the manufacturer names and purity associated with a material; for example, "MgO (purity 99.999%)," "... pressed into pellets
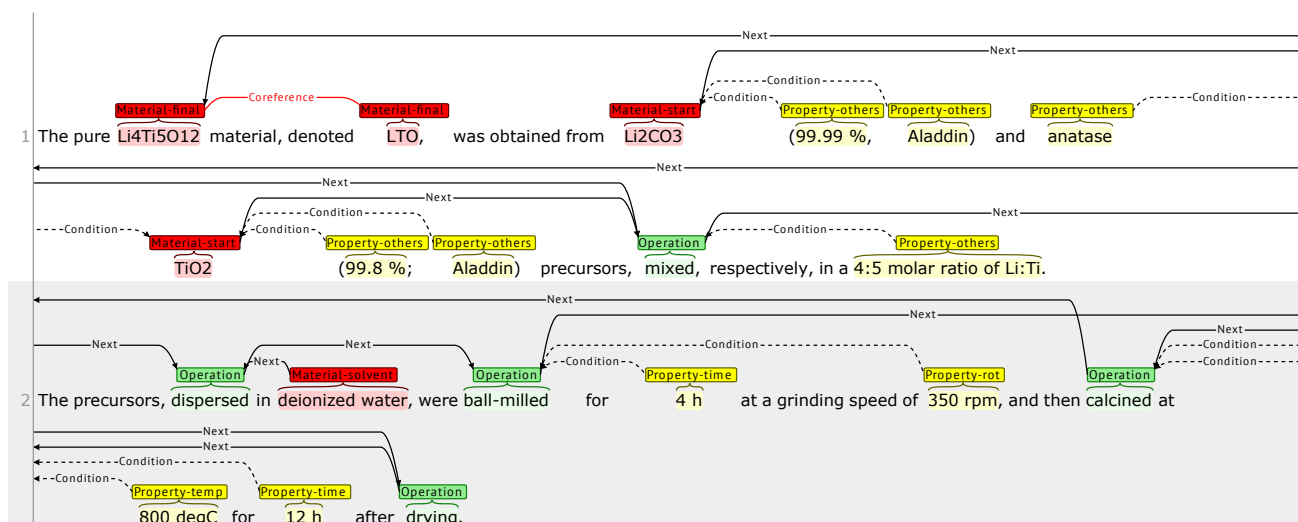
Figure 3: Screenshot of brat interface annotating synthesis process in Figure 1.

("10 mm diameter, 1 mm thick)," and "the starting materials in the 1/4 molar ratio."

2.2.2. Edges

We defined the following three edge labels, which represent the relations between vertices.

CONDITION indicates the conditions of an operation and properties of a raw material (for example, the temperature, time, and atmosphere) for performing an operation. This label is also used to express the relations between a raw material and its manufacturer name or purity.

NEXT represents the order of an operation sequence and indicates the input or output relations between a material and an operation.

COREFERENCE is a link that associates two or more phrases when these phrases refer to the same material.

3. Annotation Details and Evaluation

In this section, we explain the annotation details, including the text preparation, preprocessing, and annotation settings; thereafter we present the settings and results of the inter-annotator agreement experiments.

3.1. Annotation Details

We constructed a corpus including the experimental sections of 243 papers on material synthesis processes in the following manner.

We collected papers on experimental processes from online journals. To limit the annotation target to the ASSB, which is synthesized using the "solid phase method" or "liquid phase method". We set the search queries to identify papers containing "solid electrolyte" or "ionic conductivity", but not containing "poly", "SEI", and "solid electrolyte interphase" in the titles, abstracts, and keywords. The four experts in material science are involved in selecting target journals and keywords.

Thereafter, we manually selected 243 papers that were confirmed to include the synthesis process in the "Experimental", "Preparation" or "Method" sections, because synthesis processes often appear in these sections. We applied the

PDF Parser[2] to extract text from the downloaded PDF papers. We extracted the texts of the above sections, manually corrected several typos, and unified certain orthographical variants in composition formulae and quantitative expressions. For example, a "°C" was replaced with the token "degC".

Finally, we annotated the synthesis graph on the obtained texts. Three annotators, who were master's course students in materials science, were involved in the annotation. Annotator A tagged 77 papers, annotator B tagged 68 papers, and annotator C tagged 98 papers. Finally, one professional in materials science verified the annotations of the three student annotators and corrected the annotation errors. We used the brat annotation toolkit (Stenetorp et al., 2012) for manual annotation. Figure 3 illustrates an annotation interface by brat.

3.2. Inter-Annotator Agreement

The agreement calculations were based on whether the spans of the labels were precisely matched the three annotators in materials science on the spans by using 30 randomly selected synthesis processes from the SynthASSBs corpus. We calculated the agreements using Cohen's kappa. For each pair of two annotators selected from the three annotators A, B, and C, the agreement score was calculated by regarding the labels identified by one annotator as gold and the labels by the other annotator as the prediction, and the average of the scores in two directions was determined. For the vertices, we calculated two agreement scores: the agreement score of the spans and types (All), and the agreement score of the types on the spans that were annotated by both annotators (Type). For the edges, we also calculated two agreement scores on the vertices that were annotated by both annotators. One score was calculated by comparing the existence of edges and their types (All), while the other score was calculated by comparing the types on the edges that were annotated by both annotators (Type). The inter-annotator agreement results are presented in Table 1.

---

[2]https://github.com/allenai/science-parse

1943

	Vertices		Edges	
Annotators	All	Type	All	Type
A–B	0.637	1.000	0.705	0.990
B–C	0.667	1.000	0.671	0.991
A–C	0.608	1.000	0.651	0.990

Table 1: Inter-annotator agreement results using Cohen's kappa.

Item	Count
Documents	243
Sentences	2,877
Tokens	46,477
Entities	10,995
Vertex types	12
Edge types	3
Avg. sentences/document	12
Avg. tokens/document	191
Avg. entities/document	45

Table 2: SynthASSBs corpus statistics.

Vertex / Edge types	Count
MATERIAL	2,749
MATERIAL-START	1,319
MATERIAL-INTERMEDIUM	138
MATERIAL-FINAL	532
MATERIAL-SOLVENT	212
MATERIAL-OTHERS	548
OPERATION	1,680
PROPERTY	3,994
PROPERTY-TEMP	704
PROPERTY-TIME	642
PROPERTY-ROT	66
PROPERTY-PRESS	81
PROPERTY-ATMOSPHERE	275
PROPERTY-OTHERS	2,226
CONDITION	4,139
NEXT	3,018
COREFERENCE	759
TOTAL	23,082

Table 3: Statistics of vertices and edges annotated in SynthASSBs corpus.

We confirmed that the types (Type) of vertices and edges were almost perfectly matched among the annotators (both kappa coefficients were over 0.99) and the spans and types (All) of them were also substantially matched. This demonstrates that the annotation scheme of the vertices and edges was clear when selecting types. However, the kappa coefficients in the All settings were lower than those in the Type settings. This indicates that annotation ambiguity was caused when deciding which phrase should be involved in the synthesis process. We leave the improvements in the annotation guidelines to reduce this ambiguity problem for future work.

3.3. Statistics

Several key statistics of the SynthASSBs corpus, such as the number of documents, sentences, tokens, and entities, are summarized in Table 2. The number of vertices or edges per type is indicated in Table 3. In the statistics, we used scispaCy (Neumann et al., 2019) [3] to split sentences, perform tokenization and extract entities.

4. Synthesis Process Extraction

Our framework performed extraction of synthesis processes in a pipeline manner, using two modules: deep learning-based sequence taggers for extracting the phrases we defined as vertices, and a rule-based relation extractor (RE) for connecting the edges that were pairs of extracted phrases. As illustrated in Figure 4, our framework first performed sequence tagging (a) to extract the phrases related to the material synthesis process. Thereafter, the relations between entities were extracted by the rule-based RE (b).
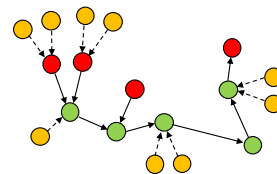
4.1. Sequence Tagging

To train the sequence-tagging model, we employed Bidirectional Long Short-Term Memory with Conditional

(a) Sequence Tagging   (b) Rule-based RE

Figure 4: Overview of synthesis process extraction. The red phrases and circles indicate terms related to materials, green indicates operations, and yellow indicates properties. The solid and broken arrows represent the next and condition edges, respectively.

Random Fields (Huang et al., 2015) as a sequence-tagging model to identify the spans of the vertices. We used six different base representations in the neural network-based sequence tagger: character-level embedding (CE) (Zhang et al., 2015); byte pair encoding (BPE) (Sennrich et al., 2016); word embeddings for inorganic material science Mat-WE (Kim et al., 2017) and mat2vec (Tshitoyan et al., 2019); Mat-ELMo (Kim et al., 2017), which is an embeddings from language models (ELMo) (Peters et al., 2018) model pretrained on materials science texts; and SciBERT (Beltagy et al., 2019), which is a bidirectional encoder representations from transformers (BERT) model (Devlin et al., 2019), pretrained on biomedical and computer science texts. These representations were fine-tuned during training on the sequence-tagging task.

4.2. Relation Extraction

We developed the following five rules using the training portion of the SynthASSBs corpus. The illustrations following the rule descriptions are used for visualization. The

circles used in the figures represent sequential tokens; the red, green, yellow, and white circles corresponds to MATERIAL, OPERATION, PROPERTY, and other words/phrases, respectively. A bounding box around circles represents a sentence. A solid arrow represents an edge of NEXT, while a broken arrow represents an edge of CONDITION.

Rule of OPERATION to OPERATION (O-O):
An OPERATION phrase is connected to the next OPERATION phrase in the same sentence or in the next sentences.

Figure 5: Illustration of O-O.

Rule of MATERIAL to OPERATION (M-O):
When an OPERATION phrase appears in brackets, a MATERIAL-START or MATERIAL-SOLVENT phrase before the left bracket is connected to the OPERATION. In the example sentence "Samples were prepared from H_3BO_3, AL_2O_3, SiO_2 and either Li_2CO_3 (dried at 200 degC)," the OPERATION phrase "dried" written in brackets is connected to its previous MATERIAL-START phrase, "Li_2CO_3", not "H_3BO_3", "AL_2O_3", and "SiO_2".

As for other MATERIAL-START or MATERIAL-SOLVENT phrases, we applied the following rules with ignoring the OPERATION phrases in brackets. A MATERIAL-START or MATERIAL-SOLVENT phrase is connected to its closest OPERATION phrase in a sentence. If two candidates exist within the same distance, the previous candidate is selected. If no OPERATION phrase exists in a sentence, the phrase is connected to the next-closest OPERATION phrase beyond the sentence boundaries.
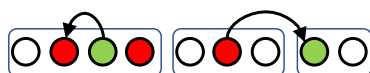
Figure 6: Illustration of M-O.

Rule of OPERATION to MATERIAL (O-M):
An OPERATION phrase that appears at the end of the operation sequence is connected to all MATERIAL-FINAL phrases in the text.

Figure 7: Illustration of O-M.

Rule of PROPERTY-OTHERS to OPERATION or MATERIAL (PO-OM):
When a PROPERTY-OTHERS phrase appears in brackets, the phrase is connected to the closet previous MATERIAL-START phrase. In the example phrase "TiO_2, GeO_2 and $NH_4H_2PO_4$ (purity 99.999 %)," "purity 99.999 %" is connected to its closest previous MATERIAL-START phrase, namely "$NH_4H_2PO_4$", and not "TiO_2" and "GeO_2".

A PROPERTY-OTHERS phrase is connected to the closest phrase of MATERIAL-START, MATERIAL-FINAL, MATERIAL-INTERMEDIUM, MATERIAL-SOLVENT, MATERIAL-OTHERS, or OPERATION. If two candidates exist with the same distance, the previous candidate is selected.
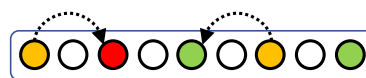
Figure 8: Illustration of PO-OM.

Rule of PROPERTY to OPERATION (P-O):
A PROPERTY-TIME, PROPERTY-TEMP, PROPERTY-ROT, PROPERTY-PRESS, or PROPERTY-ATMOSPHERE (that is, properties other than PROPERTY-OTHERS) phrase is connected to its closest previous OPERATION phrase in the sentence or before it.

Figure 9: Illustration of P-O.

5. Evaluation

5.1. Evaluation Settings

We evaluated the sequence tagger and rule-based RE individually. The sequence tagger was implemented using Flair (Akbik et al., 2019)[4], which is a multi-lingual, neural sequence-labeling framework for state-of-the-art natural language processing. When training the sequence tagger, we set the number of training epochs to 200, and used the default hyper-parameters of Flair.

The sequence tagger was evaluated using two settings of type sets. In the first setting, we extracted three coarse-grained distinct types of vertices in the flow graph: the MATERIAL, OPERATION, and PROPERTY vertices. In the second setting, we extracted all 12 fine-grained types of vertices in the flow graph: MATERIAL-START, MATERIAL-INTERMEDIUM, MATERIAL-FINAL, MATERIAL-SOLVENT, MATERIAL-OTHERS, OPERATION, PROPERTY-TIME, PROPERTY-TEMP, PROPERTY-ROT, PROPERTY-PRESS, PROPERTY-ATMOSPHERE, and PROPERTY-OTHERS.

We divided the SynthASSBs corpus into three subsets: 145 sections for training, 49 for development, and 49 for testing. We used an F1 score as the primary evaluation metric. We also report the macro-averaged F1 score of the three coarse-grained types (ALL) for the first setting and the micro-averaged F1 scores for the three coarse-grained types (MATERIAL, OPERATION, and PROPERTY) and the macro-averaged F1 score of these three types (ALL) for the second setting.

We also plot the changes in F1 score of the methods as the training set is increased in increments of 5% to answer the question about whether the corpus size is large enough to train the sequence tagging. The evaluation was performed on the fine-grained types and the scores were calculated

---

[4] https://github.com/zalandoresearch/flair

	MATERIAL			OPERATION			PROPERTY			ALL
Model	F1	P	R	F1	P	R	F1	P	R	F1
CE (Zhang et al., 2015)	0.686	0.644	0.733	0.741	0.779	0.708	0.571	0.673	0.496	0.666
BPE (Sennrich et al., 2016)	0.860	0.837	0.883	0.799	<u>0.785</u>	0.814	0.706	0.726	0.686	0.788
mat2vec (Tshitoyan et al., 2019)	0.841	0.826	0.858	0.804	0.742	0.877	0.697	0.727	0.668	0.781
Mat-WE (Kim et al., 2017)	0.834	0.816	0.854	0.797	0.769	0.827	0.702	0.703	<u>0.701</u>	0.778
Mat-ELMo (Kim et al., 2017)	**0.917**	**0.897**	**0.938**	<u>0.823</u>	0.768	**0.887**	**0.739**	**0.761**	**0.718**	**0.826**
SciBERT (Beltagy et al., 2019)	<u>0.879</u>	<u>0.866</u>	<u>0.893</u>	**0.839**	**0.798**	<u>0.884</u>	<u>0.709</u>	<u>0.749</u>	0.673	<u>0.809</u>

Table 4: F1 scores of sequence-labeling models with different base representations on development dataset. Macro-averaged F1 scores were calculated using all three coarse-grained types (ALL). The highest and second-highest for each metric are indicated in bold and underline, respectively.

on the development set. We show the micro-averaged F1 scores for the three coarse-grained types and the macro-averaged F1 score of the three types (ALL) for the plot.

For the rule-based RE, we used 145 sections (used for training in sequence tagging) in designing rules, whose details are shown in Sec. 4.2., and 98 sections (used as development and testing in sequence tagging) for evaluating the rules. To evaluate the RE, an F1 score based on an exact match was used as the primary evaluation metric. We used COREFERENCE relations in the evaluation: phrase pairs with COREFERENCE relations were treated as the same phrase in the RE evaluation. The performance of the rule-based RE was further analyzed in detail by evaluating the efficiency of the fine-grained labels in the entities as the ablation study, and by demonstrating the accuracy and coverage of each rule.

5.2. Sequence-Tagging Results

Table 4 summarizes the sequence-labeling results for extracting three coarse-grained vertex types over the six base representations as shown in Sec. 4.1.. The results show reasonably high performance, in which Mat-ELMo achieved the highest performances, with an F1 score of 0.917 on MATERIAL, and 0.826 on ALL, while SciBERT achieved the best score on OPERATION.

The performance of the sequence tagger with Mat-ELMo, evaluated on the fine-grained types, is presented in Table 5. Among the MATERIAL types, MATERIAL-START achieved the highest F1 score of 0.887. The F1 score of OPERATION was 0.821, which was higher than the average. Among the PROPERTY types, PROPERTY-TIME achieved the highest F1 score of 0.928. However, the F1 score of MATERIAL-INTERMEDIUM was 0.105 in the sequence tagging. This may be because it is difficult to extract MATERIAL-INTERMEDIUM without understanding the whole structure of the synthesis process.

Changes in F1 score according to training set size are presented in Figure 10. In this result, we observe that the curves of ALL remain almost flat after using around 20% of the training set is used. Therefore, we conclude that the size of the SynthASSBs corpus is large enough to train the sequence tagger. In detail, for MATERIAL, the F1 score gradually increases as the training set size increases because material phrases often include unknown terms. OPERATION's performance is flat after 5% of the training set is used because there are several types of OPERATION verbs

Types	F1	P	R
MATERIAL	0.661	0.692	0.665
MATERIAL-START	0.887	0.885	0.888
MATERIAL-INTERMEDIUM	0.105	0.286	0.065
MATERIAL-FINAL	0.675	0.591	0.786
MATERIAL-SOLVENT	0.793	0.852	0.742
MATERIAL-OTHERS	0.845	0.845	0.845
OPERATION	0.821	0.792	0.852
PROPERTY	0.780	0.778	0.784
PROPERTY-TEMP	0.842	0.806	0.880
PROPERTY-TIME	0.928	0.932	0.925
PROPERTY-ROT	0.889	0.857	0.923
PROPERTY-PRESS	0.605	0.619	0.591
PROPERTY-ATMOSPHERE	0.775	0.775	0.775
PROPERTY-OTHERS	0.641	0.676	0.609
ALL	0.754	0.754	0.767

Table 5: F1 scores of the sequence labeling models with Mat-ELMo on the development dataset by type. MATERIAL and PROPERTY indicate micro-averaged F1 scores of each fine-grained types, respectively. ALL is the macro-averaged scores of three coarse-grained types (i.e., MATERIAL, OPERATION, and PROPERTY in this table).

used in the synthesis process. Because PROPERTY is also steady state when 20% or more of the training set is used, it seems that the properties are described in a regular manner.

5.3. Relation Extraction Results

Table 6 displays the results of the rule-based model as well as the rule-based RE results obtained by the ablation tests. The high performance with a macro-averaged F1 score of 0.887 shows the effectiveness of the rules. To confirm the effectiveness of the fine-grained types or sub-labels, we compared the F1 score with three settings. In the first setting, we extracted the relations without using material sub-labels (– MATERIAL-∗), by applying the rule of M-O to all of the MATERIAL types and ignoring the rule of O-M. In the second setting, we extracted relations without using PROPERTY sub-labels (– PROPERTY-∗), by applying the rule of PO-OM to all of the PROPERTY types and ignoring the rule of P-O. The final setting was without either MATERIAL or PROPERTY sub-labels (– both sub-levels). According to the ablation tests, the F1 scores were improved by 7.8% on CONDITION and 11.1% on NEXT when applying
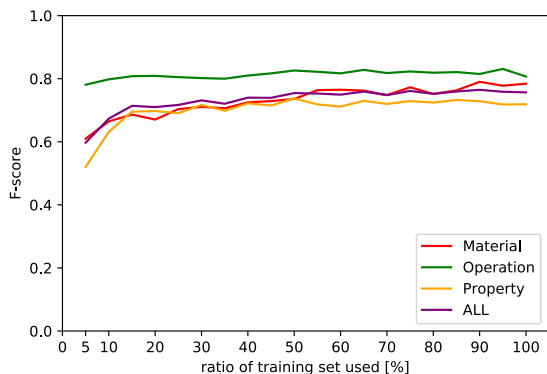
Figure 10: Changes in F1 score according to training set size which is increased in increments of 5 % until training set size reaches 145 sections. ALL shows the macro-averaged F1 score of the three coarse-grained types.

	CONDITION	NEXT	ALL
Rule-based RE	0.914	0.860	0.887
– MATERIAL-∗	0.914	0.749	0.832
– PROPERTY-∗	0.836	0.860	0.848
– both sub-levels	0.836	0.749	0.793

Table 6: F1 scores of rule-based system and ablation test results. Macro-averaged F1 scores were calculated using CONDITION and NEXT (ALL).

the sub-label rules.

To analyze the effects of the rules in further detail, the coverage and accuracy for each rule were determined, and these are presented in Table 7. By comparing the rule coverage and accuracy, it could be observed that the rules of PO-OM and P-O, which exhibited wide coverage and high accuracy (over 25% and 85%, respectively), contributed significantly to the extraction performance. This indicates that the rules for extracting the relation between the PROPERTY and MATERIAL or OPERATION successfully mimicked the manner of reading a paper. Although the coverage of the rule O-M was extremely low and the accuracy was relatively low (4.6% and 48.9%, respectively), this rule was essential for constructing the synthesis graph and could not be omitted.

6. Qualitative Evaluation

We present a thorough evaluation on a real-world scientific literature to demonstrate the efficacy of our framework. A prediction obtained by our framework and the synthesis graph are shown in Figure 11 and Figure 12, respectively. In this result, our framework could extract phrases related to material synthesis almost without error. In detail, the relations across the sentences were extracted without problems; for example, our framework created a NEXT edge between "mixed" written in the first sentence and "dispersed" written in the second sentence. Moreover, our framework succeeded in identifying the type of MATERIAL even if material written in an abbreviation form; for example, our framework could detect that "$Li_4Ti_5O_12$" and "LTO" are

Rule	Coverage	Accuracy
O-O	0.219	0.811
M-O	0.160	0.811
O-M	0.046	0.489
PO-OM	0.322	0.853
P-O	0.254	0.951

Table 7: Coverage and accuracy of our rules applied to training data.

MATERIAL-FINAL in the first sentence. However, the label type was wrong in "anatase" in the first sentence, and OPERATION connection between "calcined" and "drying" on the second line was different from that labeled by the annotator. This is because our rule-based RE could not understand the meaning of "after drying".

7. Error Analysis

We analyzed 135 errors in the sequence-tagging results. The over-detection errors constituted 49 cases, which were often PROPERTY-OTHERS types that were not directly related to the synthesis process; for example, vessel size or thickness, and milling machine properties. A total of 49 entities were missing and were often caused by PROPERTY types, were missing due to rare adverbs, adjectives, or units; for example, "naturally", "constant", "mm-thick", and, "micrometers".

In the RE, we identified two major problems when we analyzed the 129 errors. The first problem was caused by the definition of the distance, which used the number of words and ignored syntactic structures. For example, in the sentence "LiNO3 were weighed according to the stoichiometry of the Li3xLa2/3-xTiO3 and dissolved in ethylene," our distance-based rule predicted that "Li3xLa2/3-xTiO3" qualifies "dissolved" instead of "weighed". This type of problem included 73 errors. The second problem was complex operation sequences. Where two or more material synthesis processes were described in one document, there were cases in which a synthesis process indicated at the beginning was omitted in the second and subsequent explanations. In such cases, branching and merging of synthesis processes occurred. Our rules assumed that the operation sequence was described sequentially, so they could not identify these processes. This type of complex operation sequence caused 28 errors. One means of addressing the above problems is to incorporate additional rules; however, it is not realistic to create more rules manually, because the descriptions are sometimes ambiguous, without an understanding of the contents. We are considering developing a deep learning-based extractor that can take syntactic structures into account.

8. Related Work

Process extraction from procedure texts has been studied in a wide range of fields. Such studies include an effort to extract liquid mixing procedures from text (Long et al., 2016), an annotated corpus of photosynthesis and formation erosion processes (Dalvi et al., 2018), the extraction of response processes from guidance texts at the time of disaster
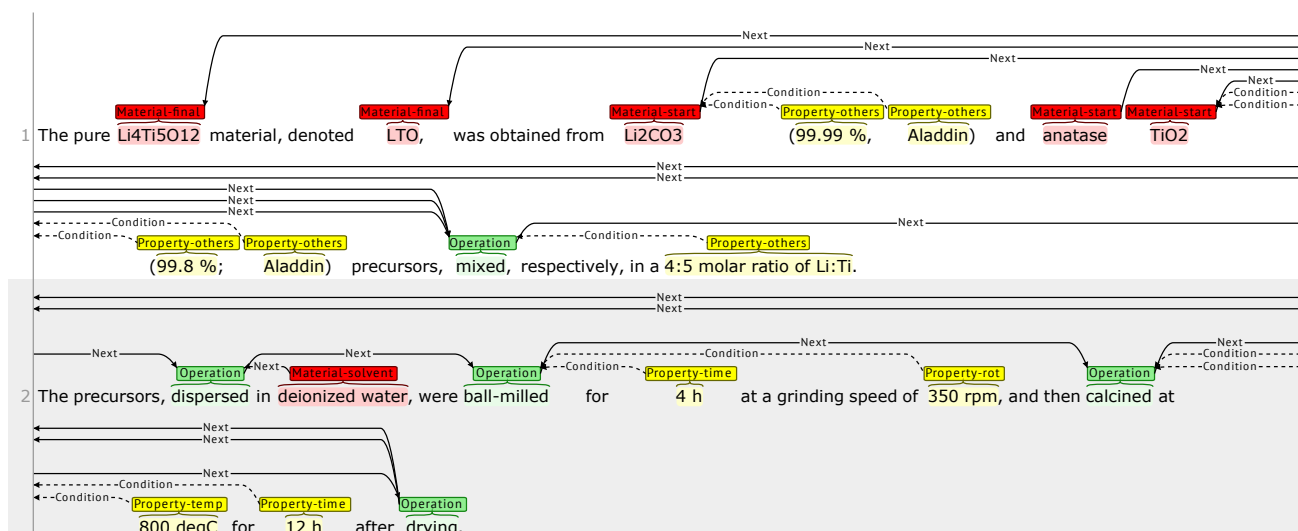
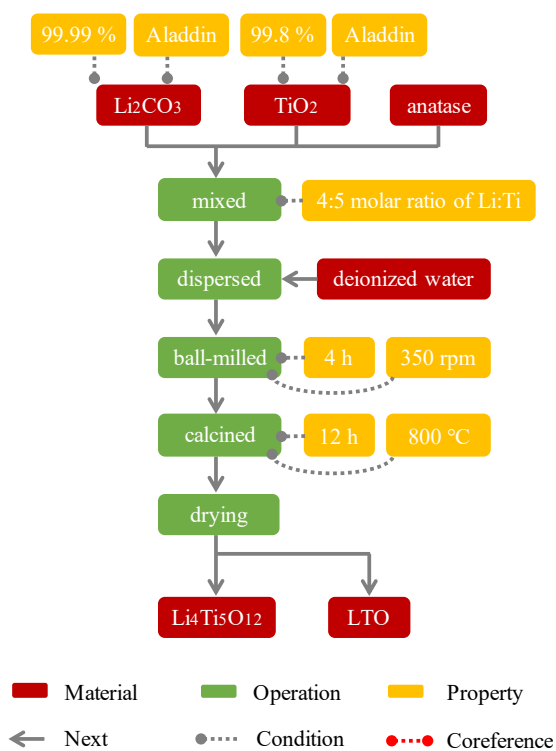Figure 11: Synthesis process extraction results from the text in Figure 1

Figure 12: Synthesis graph of the extracted synthesis process in Figure 11.

occurrence (Guo et al., 2018), and several attempts to structure and extract a series of cooking-related actions, such as baking and boiling, from cooking recipe sentences (Mori et al., 2014; Kiddon et al., 2015; Maeta et al., 2015; Abend et al., 2015). Numerous language resources exist in the organic chemistry field (Kim et al., 2003; Krallinger et al., 2015; Tsubaki et al., 2017; Kulkarni et al., 2018; Tanaka et al., 2018), which have been annotated with the experimental processes that appear in the papers. Moreover, an attempt has been made to extract processes by applying event extraction methods to realize machine-based text reading for biomedical papers (Miwa et al., 2012; Scaria et al., 2013; Berant et al., 2014; Rao et al., 2017; Rahul et al., 2017; Björne and Salakoski, 2018). In the inorganic chemistry field, several corpus are available for general-purpose materials (Mysore et al., 2019; Kononova et al., 2019), while some studies are underway to extract the synthesis process from papers (Mysore et al., 2017; Tamari et al., 2019); however, no corpus and extraction system exist for synthesizing ASSBs. Therefore, we have presented a domain specific corpus of the synthesis process for ASSBs, and an automated machine-reading system for extracting the synthesis processes buried in the scientific literature.

9. Conclusion

This study has addressed the problem of the lack of labeled data, which is a major bottleneck in developing ASSBs. We constructed the novel SynthASSBs corpus, consisting of the experimental sections of 243 papers. The corpus annotates synthesis graphs that represent the synthesis process of ASSBs in text. Moreover, we proposed an automatic synthesis process extraction framework using our corpus by combining a deep learning-based sequence tagger and rule-based relation extractor that mimics the experience in human reading. As a result, the sequence tagger with the best setting can detect the entities with a macro-averaged F1 score of 0.826. Furthermore, the rule-based RE demonstrates high performance with a macro-averaged F1 score of 0.887.

In future work, we will develop a deep learning-based relation extractor that incorporates syntactic information into the model to improve the extraction performance. We will also apply our extracting framework to existing papers, and, using the extracted abundant knowledge, we will build a computational synthesis design framework for discovering novel material.

10. Bibliographical References

Abend, O., Cohen, S. B., and Steedman, M. (2015). Lexical event ordering with an edge-factored model. In *Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (HLT-NAACL)*.

Agrawal, A. and Choudhary, A. (2016). Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science. *APL Materials*.

Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., and Vollgraf, R. (2019). Flair: An easy-to-use framework for state-of-the-art nlp. In *Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (HLT-NAACL Demonstrations)*.

Bai, X., Li, W., Wei, A., Li, X., Zhang, L., and Liu, Z. (2016). Preparation and electrochemical properties of mg2+ and f- co-doped li4ti5o12 anode material for use in the lithium-ion batteries. *Electrochimica Acta*.

Beltagy, I., Lo, K., and Cohan, A. (2019). Scibert: A pretrained language model for scientific text. In *International Joint Conference on Natural Language Processing (IJCNLP)*.

Berant, J., Srikumar, V., Chen, P.-C., Linden, A. V., Harding, B., Huang, B., Clark, P., and Manning, C. D. (2014). Modeling biological processes for reading comprehension. In *Empirical Methods in Natural Language Processing (EMNLP)*.

Björne, J. and Salakoski, T. (2018). Biomedical event extraction using convolutional neural networks and dependency parsing. In *Biomedical Natural Language Processing Workshop (BioNLP) in ACL 2018*.

Butler, K. T., Davies, D. W., Cartwright, H. M., Isayev, O., and Walsh, A. (2018). Machine learning for molecular and materials science. *Nature*.

Dalvi, B., Huang, L., Tandon, N., tau Yih, W., and Clark, P. E. (2018). Tracking state changes in procedural text: a challenge dataset and models for process paragraph comprehension. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

Guo, W., Zeng, Q., Duan, H., Yuan, G., Ni, W., and Liu, C. (2018). Automatic extraction of emergency response process models from chinese plans. *IEEE Access*.

Huang, Z., Xu, W. L., and Yu, K. (2015). Bidirectional lstm-crf models for sequence tagging. *In arXiv:1508.01991*.

Kiddon, C., Ponnuraj, G. T., Zettlemoyer, L. S., and Choi, Y. (2015). Mise en place: Unsupervised interpretation of instructional recipes. In *Empirical Methods in Natural Language Processing (EMNLP)*.

Kim, J.-D., Ohta, T., Tateisi, Y., and Tsujii, J. (2003). Genia corpus - a semantically annotated corpus for bio-textmining. *Bioinformatics*.

Kim, E. Y., Huang, K., Tomala, A., Matthews, S. L., Strubell, E., Saunders, A. R., McCallum, A. L., and Olivetti, E. (2017). Machine-learned and codified synthesis parameters of oxide materials. In *Scientific data*.

Kim, E., Huang, K., Kononova, O., Ceder, G., and Olivetti, E. (2019). Distilling a materials synthesis ontology. *Matter*.

Kononova, O., Huo, H., He, T., Rong, Z., Botari, T., Sun, W., Tshitoyan, V., and Ceder, G. (2019). Text-mined dataset of inorganic materials synthesis recipes. In *Scientific Data*.

Krallinger, M., Rabal, O., Leitner, F., Vazquez, M., and et al., D. S. (2015). The chemdner corpus of chemicals and drugs and its annotation principles. *Journal of Cheminformatics*.

Kulkarni, C., Xu, W., Ritter, A., and Machiraju, R. (2018). An annotated corpus for machine reading of instructions in wet lab protocols. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

Long, R., Pasupat, P., and Liang, P. (2016). Simpler context-dependent logical forms via model projections. In *Association for Computational Linguistics (ACL)*. Association for Computational Linguistics.

Maeta, H., Sasada, T., and Mori, S. (2015). A framework for procedural text understanding. In *International Conference on Parsing Technologies (IWPT)*.

Miwa, M., Thompson, P., and Ananiadou, S. (2012). Boosting automatic event extraction from the literature using domain adaptation and coreference resolution. *Bioinformatics*.

Mori, S., Maeta, H., Yamakata, Y., and Sasada, T. (2014). Flow graph corpus from recipe texts. In *International Conference on Language Resources and Evaluation (LREC)*.

Mysore, S., Kim, E., Strubell, E., Liu, A., Chang, H.-S., Kompella, S., Huang, K., McCallum, A., and Olivetti, E. (2017). Automatically extracting action graphs from materials science synthesis procedures. *Workshop on Machine Learning for Molecules and Materials in NeurIPS 2017*.

Mysore, S., Jensen, Z., Kim, E., Huang, K., Chang, H.-S., Strubell, E., Flanigan, J., McCallum, A., and Olivetti, E. (2019). The materials science procedural text corpus: Annotating materials synthesis procedures with shallow semantic structures. In *Proceedings of the 13th Linguistic Annotation Workshop (LAW) at ACL 2019*.

Neumann, M., King, D., Beltagy, I., and Ammar, W. (2019). Scispacy: Fast and robust models for biomedical natural language processing. In *Biomedical Natural Language Processing Workshop (BioNLP) in ACL 2019*.

Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. (2018). Deep contextualized word representations. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

Rahul, P. V. S. S., Sahu, S. K., and Anand, A. (2017). Biomedical event trigger identification using bidirectional recurrent neural network based models. In *Biomedical Natural Language Processing Workshop (BioNLP) in ACL 2017*.

Rao, S., Marcu, D., Knight, K., and Daumé, H. (2017). Biomedical event extraction using abstract meaning representation. In *Biomedical Natural Language Processing Workshop (BioNLP) in ACL 2017*.

Scaria, A. T., Berant, J., Wang, M., Clark, P., Lewis, J., Harding, B., and Manning, C. D. (2013). Learning biological processes with global constraints. In *Empirical Methods in Natural Language Processing (EMNLP)*.

Sennrich, R., Haddow, B., and Birch, A. (2016). Neural machine translation of rare words with subword units. In *Association for Computational Linguistics (ACL)*.

Stenetorp, P., Pyysalo, S., Topic, G., Ohta, T., Ananiadou, S., and Tsujii, J. (2012). brat: a web-based tool for nlp-assisted text annotation. In *European Chapter of the Association for Computational Linguistics (EACL)*.

Tamari, R., Shindo, H., Shahaf, D., and Matsumoto, Y. (2019). Playing by the book: An interactive game approach for action graph extraction from text. In *Workshop on extracting structured knowledge from scientific publications in NAACL 2019*.

Tanaka, K., Iwakura, T., Koyanagi, Y., Ikeda, N., Shindo, H., and Matsumoto, Y. (2018). Chemical compounds knowledge visualization with natural language processing and linked data. In *International Conference on Language Resources and Evaluation (LREC)*.

Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., Persson, K. A., Ceder, G., and Jain, A. (2019). Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature*.

Tsubaki, M., Shimbo, M., and Matsumoto, Y. (2017). Protein fold recognition with representation learning and long short-term memory. *IPSJ Transactions on Bioinformatics*.

Wei, J., Chu, X., Sun, X.-Y., Xu, K., Deng, H.-X., Chen, J., Wei, Z., and Lei, M. (2019). Machine learning in materials science. *InfoMat*.

Zhang, X., Zhao, J. J., and LeCun, Y. (2015). Character-level convolutional networks for text classification. In *Neural Information Processing Systems (NeurIPS)*.