Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 6753–6760
Marseille, 11–16 May 2020

SC-CoMIcs: A Superconductivity Corpus for Materials Informatics

Kyosuke Yamaguchi[1], **Ryoji Asahi**[2], **Yutaka Sasaki**[1]
[1]Toyota Technological Institute
2-12-1 Hisakata, Tenpaku-ku, Nagoya, 468-8511 Japan
[2]Toyota Central R&D Labs., Inc., 41-1, Yokomichi, Nagakute, Aichi 480-1192, Japan
[1]{sd19453,yutaka.sasaki}@toyota-ti.ac.jp, [2]rasahi@mosk.tytlabs.co.jp

Abstract

This paper describes a novel corpus tailored for the text mining of superconducting materials in *Materials Informatics (MI)*, named *SuperConductivety Corpus for Materials Informatics (SC-CoMIcs)*. Different from biomedical informatics, there exist very few corpora targeting *Materials Science and Engineering (MSE)*. Especially, there is no sizable corpus which can be used to assist the search of superconducting materials. A team of materials scientists and natural language processing experts jointly designed the annotation and constructed a corpus consisting of manually-annotated 1,000 MSE abstracts related to superconductivity. We conducted experiments on the corpus with a neural *Named Entity Recognition (NER)* tool. The experimental results show that NER performance over the corpus is around 77% in terms of micro-F1, which is comparable to human annotator agreement rates. Using the trained NER model, we automatically annotated 9,000 abstracts and created a term retrieval tool based on the term similarity. This tool can find superconductivity terms relevant to a query term within a specified Named Entity category, which demonstrates the power of our SC-CoMIcs, efficiently providing knowledge for Materials Informatics applications from rapidly expanding publications.

Keywords: Materials Informatics, Superconductivity Corpus, Named Entity Recognition, Materials Text Mining

1. Introduction

Recently, *Materials Informatics (MI)* is a hot topic in the field of *Material Science and Engineering (MSE)*. The reason behind this surge of interests is that a lot of materials scientists are trying to use machine learning to accelerate the search of new materials. To find new materials, scientists need to select base materials, adequate doping, and process conditions to synthesize the materials. This is a very costly, time-consuming process. Traditionally, the experience and intuition of researchers often led to unexpected discovery of new useful materials. MI is studied with a great expectation to change the paradigm in such a long process of materials discovery. There has been, however, a bottleneck, *i.e.*, existence of few datasets for MI to work effectively (Ramprasad et al., 2017).

The search of superconducting materials is one of the most challenging issues in MSE. Even after historical discovery of high-Tc cuprate superconductors (Bednorz and Müller, 1986), materials with a higher transition temperature, higher current flow, and higher processability are required for practical applications (Malozemoff et al., 2005; Foltyn et al., 2007). In materials exploration, experts often consult thousands of published and new articles for getting knowledge to decide a research direction. This can be one of the factors that hinder the efficient search of superconducting materials.

In this paper, we propose a new corpus named *SuperConductivety Corpus for Materials Informatics (SC-CoMIcs)* tailaored for the superconductivity domain, evaluate the corpus using state-of-the-art Natural Language Processing (NLP), and introduce a term search tool which efficiently provides knowledge hidden in a large amount of literature. To this end, we here particularly focus on construction of *Named Entity Recognition (NER)* model based on manually-annotated corpus, which categorizes entity useful for materials scientists and thus provides new ideas in the search of new superconducting materials.

2. Annotation Design

A team of materials scientists and natural language processing experts jointly designed annotations. Term categories are carefully considered to match the requirements from the domain experts. As a preparation for annotations, we initially referred to a keyword list[1] of letter journal *scripta materialia*, and the domain experts updated it for enrichment. The original keyword list has five categories: Synthesis/Processing, Characterization, Material Type, Property/Phenomena, and Theory/Computer Simulation/Modeling. We modified this categorization and defined the following seven categories as summarized in Table 1:

`Characterization:` The `Characterization` category lists the terms of characterization methods such as X-ray diffraction (XRD) and scanning electron microscope (SEM). The information is useful in MSE in relation to `Element` and `Property`.

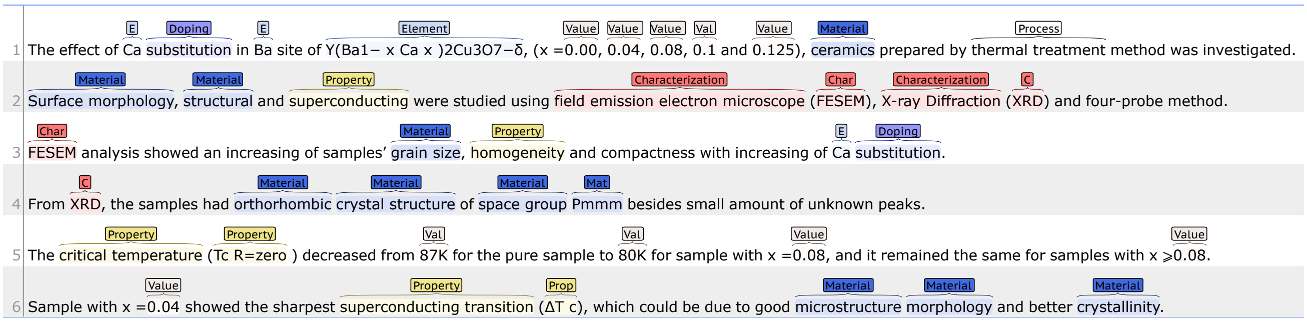
`Process:` The `Process` category lists the terms of Synthesis/Processing such as the sol-gel method for film samples, calcination for bulk samples, and AC/DC sputtering for thin-film preparation. This information is hardly obtained from theoretical simulations such as density functional theory, thus providing unique database useful for materials scientists.

`Property:` The `Property` category lists the terms of Property/Phenomena and their Theoretical representations such as electrical conductivity, mechanical hardness, electron-phonon coupling, and Fermi surface. Property and Theory including Simulation and Modeling can belong to each different category. In this work, we unified them into the `Property` category because they are sometimes difficult to be distinguished.

[1] https://www.elsevier.com/__data/promis_misc/SMM%20Keywords.pdf

Category	Explanation	Example
Characterization	characterization methods	X-ray diffraction, SEM
Process	synthesis and process	sol-gel, calcination, sputtering
Property	materials properties	electrical, cryogenic, magnetic fields
Material	structural entities, sample descriptors	tetragonal, P4/nmm, bulk, film, grain
Element	elements, compounds	Ti, oxygen, $YBa_2Cu_3O_7$
Doping	doping operation	doping, addition, doped
Value	quantitative information with units	100K, 5-10μm

Table 1: Term categories


1 The effect of Ca substitution in Ba site of Y(Ba1− x Ca x)2Cu3O7−δ, (x =0.00, 0.04, 0.08, 0.1 and 0.125), ceramics prepared by thermal treatment method was investigated.

2 Surface morphology, structural and superconducting were studied using field emission electron microscope (FESEM), X-ray Diffraction (XRD) and four-probe method.

3 FESEM analysis showed an increasing of samples' grain size, homogeneity and compactness with increasing of Ca substitution.

4 From XRD, the samples had orthorhombic crystal structure of space group Pmmm besides small amount of unknown peaks.

5 The critical temperature (Tc R=zero) decreased from 87K for the pure sample to 80K for sample with x =0.08, and it remained the same for samples with x ⩾0.08.

6 Sample with x =0.04 showed the sharpest superconducting transition (ΔT c), which could be due to good microstructure morphology and better crystallinity.

Figure 1: Annotation sample

Material: The Material category lists the terms of structural entities and sample descriptors such as tetragonal crystal symmetry, bulk/film sample, and grain boundary. The information augments the details of the sample whose composition is defined by Element.

Element: The Element category lists names and symbols of elements and compounds such as Ti, oxygen, and $YBa_2Cu_3O_7$ or YBCO. This is strictly distinguishable from the Material category as entities in the Element category can be expressed by chemical composition formulae.

Doping: Since the doping often changes materials properties significantly, we added the Doping category to represent doping operations such as doping, substitute, and addition.

Value: The Value category targets numerical values with units, such as 160K. This enables to extract transition temperatures and process conditions, for example.

3. Corpus Construction Workflow

This section explains the corpus creation workflow from data collection to annotation steps.

We conducted two-round annotation processes. The first round was conducted as a trial annotation since there was no annotations that we can referred to when we started this annotation in 2018. We collected 200 abstracts and manually annotated them. In the second round, we collected 800 abstracts and then conducted pattern matching-based automatic annotation using a term list for each category. We finally manually annotated terms in the 800 abstracts.

In the remainder of this section, we explain document collection, pattern matching, and manual annotation.

3.1. Document Collection

We collected abstracts through ScienceDirect Web API from February to June in 2019.

In the first round, we first selected about 4,000 abstracts, issued from 1983 to 2018, with the query "supercond*" and finally selected 200 abstracts that include:

- at least one superconductivity term,
- at least one term which means doping or transition temperature,
- more than or equal to 100 words.

In the second round, we were more confident about abstract selection based on the experience in the first round and selected about 11,000 abstracts, issued from 2010 to 2019, by the query "superconductivity". Then we finally selected 800 abstracts that are not included in the 200 first round abstracts and include:

- at least one term for each of superconductivity, doping, element/compound,
- more than or equal 100 words.

3.2. Automatic Labeling by Pattern Matching

We then conducted pattern matching-based automatic annotation using a term list for each category complied from the annotated corpus and the keyword list enriched by experts in addition to the original list of *scripta materialia*. This pattern matching-based annotation step was included to alleviate the manual annotation burden.

In the pattern matching of term lists against abstracts, we basically performed matching within generated variations of term list entries. Specifically, we applied lemmatization process and regular expression based pattern matching. In the lemmatization process, we removed entry inflections

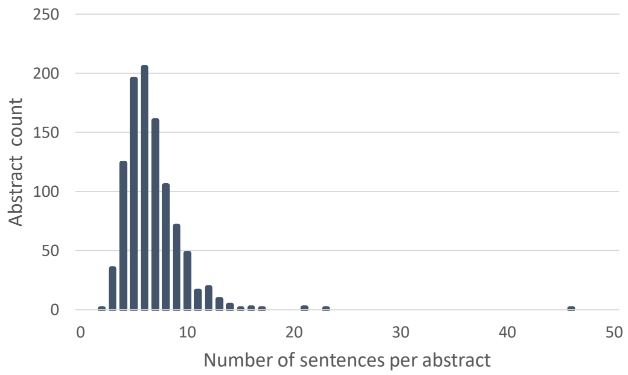


Figure 2: Histogram of #sentences per abstract

such as “ing” and “ed”. By combining a lemma word and suffixes such as “ing” or “ed” with OR operator by using regular expression, we extend the coverage of their word forms. Note that this process targets for term labels other than `Element` and `Value`. This is because `Element` can take various kinds of compound names and `Value` can have a different value each time.

3.3. Manual Annotation

The first round was performed by a well-experienced domain expert and the second round was conducted by four annotators, who are undergraduate and master students in material science or relevant areas according to the annotation guideline created during the first round annotation. We used BRAT (Stenetorp et al., 2012) annotation tool, which is a browser-based annotation environment for collaborative text annotation. After the first and second rounds, we corrected minor errors while checking the data formats, such as dual labeling of the same terms. This is because when a term has two category labels, training and testing of NER are interfered. At the same time, when a term had different categories in a different appearance, we left this dual labels as contextual difference. Figure 1 shows an example of completed annotations that is visualized with BRAT.

4. Corpus Analysis

This section explains corpus statistics and annotator agreement results.

4.1. Corpus Statistics

Table 2 shows statistics of the SC-CoMIcs corpus. Terms in the `Material`, `Property` and `Element` categories appear a lot in the corpus. The `Doping` and `Value` categories also appear sufficient number times; the latter is beneficial to extract quantitative information from text in relation to superconductivity terms. The number of `Characterization` terms is relatively small; however, the frequency of `Characterization` terms *per se* is not a problem as described in Section 5.3. Figures 2 and 3 show the histograms of the numbers of sentences per abstract and tokens per sentence. Most abstracts contain round 4-8 sentences and most sentences have around 20-35 words.

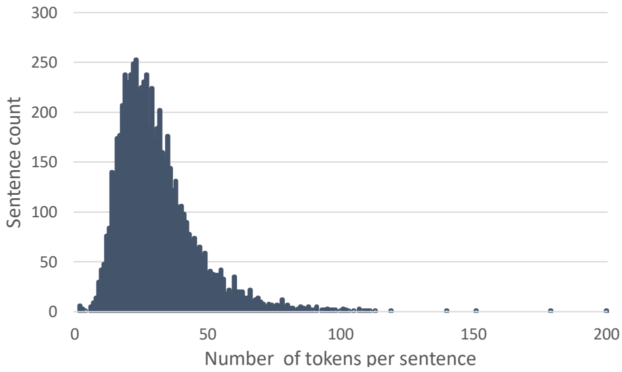


Figure 3: Histogram of #tokens per sentence

#abstracts	1,000
#sentences	6,639
#tokens	204,884
#sentences/abstract	6.64
#tokens/sentence	30.9

Category	Frequency
Characterization	1,789
Material	6,953
Property	15,129
Element	9,526
Doping	2,565
Process	2,173
Value	4,202

Table 2: Corpus statistics

4.2. Annotator Agreements

For the second round, we evaluated annotator agreement. We asked the four annotators to annotate extra 10 abstracts independently, without telling that they were annotating the same 10 abstracts. The domain expert, who annotated in the first round, also created reference annotations for the 10 abstracts.

The annotation scores were calculated based on the both exact-span and label match. Table 3 shows the agreement between each annotator and the domain expert. Annotator A performed slightly better than others, although there is no significant difference in the total scores. On the other hand, there are some variations in the agreement rates by categories. `Characterization` and `Doping` have high agreement rates. `Material` and `Property` have relatively low agreement rates of around 70% because these are difficult or ambiguous for annotation, e.g., “anisotropic” and “order”; the annotator needs to consider contexts based on the domain knowledge to judge the annotation.

We also calculated the Fleiss’ kappa to measure the agreement between the four annotators, which is independent of expert annotation. This is a token level evaluation based on the BIO labels and it takes into account all pairs of four annotators. As a result, we obtained the Fleiss’ kappa score of 83.1%.

After the independent annotation by the four annotators in the second round, all the annotations of 800 abstracts were reviewed by four annotators. This implies that the resulting

Annotator	A			B		
Category	Precision	Recall	F1	Precision	Recall	F1
Characterization	0.947	0.900	0.923	0.905	0.950	0.927
Material	0.867	0.697	0.773	0.771	0.607	0.679
Property	0.817	0.744	0.779	0.774	0.774	0.774
Element	0.964	0.951	0.957	0.837	0.796	0.816
Doping	1.000	1.000	1.000	0.974	0.884	0.927
Process	0.933	0.824	0.875	0.906	0.853	0.879
Value	0.730	0.836	0.780	0.750	0.927	0.829
Total(micro-ave.)	0.879	0.822	0.850	0.812	0.777	0.794

Annotator	C			D		
Category	Precision	Recall	F1	Precision	Recall	F1
Characterization	0.727	0.800	0.762	0.900	0.900	0.900
Material	0.779	0.607	0.682	0.911	0.754	0.825
Property	0.703	0.732	0.717	0.757	0.649	0.699
Element	0.818	0.761	0.788	0.815	0.838	0.826
Doping	0.966	0.651	0.778	0.969	0.721	0.827
Process	0.735	0.735	0.735	0.788	0.765	0.776
Value	0.776	0.945	0.852	0.688	0.800	0.739
Total (micro-ave.)	0.769	0.729	0.749	0.813	0.752	0.781

Table 3: Agreement scores of four annotator's (A, B, C , and D) annotations with respect to the reference annotations

corpus would have better annotation quality than the agreement rates in Table 3.

5. Experiments

5.1. Experimental Settings

Recently, *Bidirectional Encoder Representations from Transformers (BERT)* (Devlin et al., 2018) is used in various NLP applications and contributed to improve performance. We used the pre-trained embedding model SciBERT (Beltagy et al., 2019), which is a BERT model that is trained on a large-scale scientific literature. BERT is a contextual embedding model, which means that an embedding vector of the same word changes according to the surrounding words. We used the SciBERT NER model with the default parameter settings, which is provided along with the SciBERT model. [2]

The input for SciBERT NER was the SciBERT (uncased) pre-trained model. SciBERT NER is a stack of Bi-LSTM→CRF layers and determines category labels in the form of IOB2 tagging. In the preprocessing, we removed inner NE annotations in the nested annotations.

5.2. NER Performance

We evaluated the NER performance using the 10-fold cross validation. In the training, SciBERT embedding models were fixed, *i.e.*, no fine tuning applied. The evaluation results are shown in Table 4. The scores in the table are the averages of ten times of training. Compared to the human annotator agreement rates, the NER achieved quite comparable performance to most of the human annotators. Table 5 shows results for each fold of the cross validation.

5.3. Learning Curves

To see whether the number of training data is sufficient or not, we draw learning curves by changing the number of training data from 100 to 900 by the 100 unit. We trained each model ten times and computed the average of the scores. The results are displayed in Figure 4. The scores grow as the number of data increases. When we used 900 abstracts, the recognition performance of `Doping`, `Characterization`, and `Property` seems to be saturated but that of `Process` and the other categories are still slightly increasing. The number of occurrences of `Characterization` terms in the corpus is relatively small; however, since its NER performance seems to be saturated using 900 training abstracts, the number of `Characterization` terms in the corpus is not a serious problem; while its NER performance has room to improve as its human annotator performance is mostly much higher. We believe that we need to use domain knowledge to improve the recognition performance of `Characterization` terms. In general, the size of the SC-CoMIcs corpus is proved to be mostly sufficient while the larger data is the better for some categories, same as the cases in any machine learning-based approaches.

5.4. Applications with Word Similarity

To see the usefulness of our corpus, we created a term retrieval tool (Figure 5) based on the word similarity using word2vec (Mikolov et al., 2013). We used the word2vec module in the gensim (Řehůřek and Sojka, 2010) Python library. The used algorithm is the skip-gram with negative sampling. The details of the parameter settings are shown in Table 6. While BERT and its family are contextual embedding models and need a large amount of corpus, word2vec is more suitable for generating word vectors from annotated 10,000 abstracts.

[2] `https://github.com/allenai/scibert`

Category	Precision	Recall	F1
Characterization	0.780	0.771	0.774
Material	0.752	0.738	0.743
Property	0.734	0.689	0.711
Element	0.842	0.869	0.855
Doping	**0.960**	**0.964**	**0.961**
Process	0.739	0.729	0.733
Value	0.807	0.635	0.709
Total (micro-ave.)	0.784	0.754	0.768

Table 4: Scores of SciBERT NER

Fold	Precision	Recall	F1
1*	0.793	0.752	0.772
2*	0.706	0.714	0.710
3	0.721	0.754	0.737
4	0.788	0.756	0.772
5	0.809	**0.778**	0.793
6	0.806	0.751	0.778
7	0.786	0.734	0.759
8	0.809	0.757	0.782
9	**0.817**	**0.778**	**0.797**
10	0.809	0.761	0.784

Table 5: Scores for each fold: * indicates the fold in which the first annotated 200 abstracts are used as a test data for SciBERT NER tool.

Every multi-word NE term is converted into a single token except for the `Element` term that consists of a list of compound names, which represents a main target superconducting material entity of the abstract. In the present study, we separated `Element` terms, which include more than one element, into pieces. For example, an `Element` term “$YBa_2Cu_3O_7$” is separated into “Y”, “Ba”, “Cu”, “O”. This simple conversion of compound names is useful to capture distances among elements based on existing compounds and the related entities through embedding.

We automatically annotated 9,000 abstracts using the NER model trained over our corpus. Using the corpus and automatically annotated abstracts, we created word vectors with the 300 dimensionality by the gensim word2vec tool.

The term retrieval tool finds terms relevant to the query term using the trained word2vec word vectors. The special feature of this tool is that we can specify categories to show as relevant terms. Table 7 shows the ranked retrieval results of queries “film” and “bulk” in the `Process` category. Note that, abbreviations are removed from the ranked terms. The terms with * are those which are judged by experts reasonably to be categorized in `Process`. The most of the terms are correctly categorized, legitimating the present NER model; except a few errors such as field-cooled and liquid nitrogen temperature which seem to be in `Process` but should be conditions of `Characterization`. The entities selected with ”film” query include overall typical processing for film samples; those selected with ”bulk” query include reasonably the bulk processing. These `Process` entities can be further selected by combining another query, for example, a compound name. In contrast, without annotated documents, we cannot focus on a specific category of terms as shown in Table 8, where we find few terms categorized in `Process`.

Parameter	Value
dimentionality	300
window size	8
minimum count	5
negative	1

Table 6: Hyper parameters of word2vec model

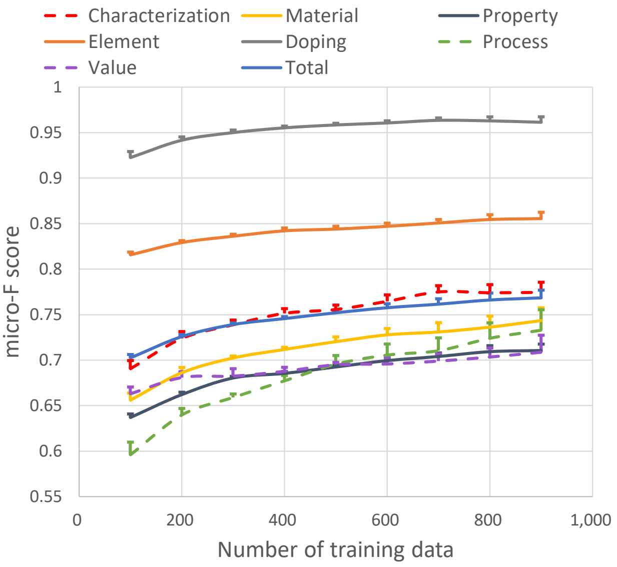
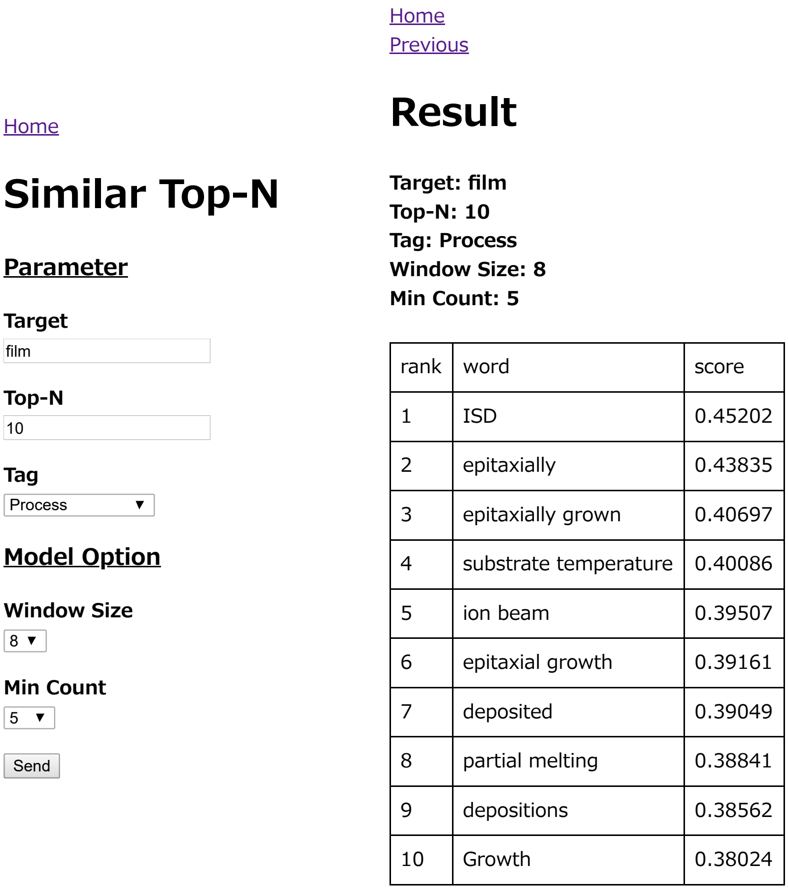


Figure 4: Learning curves by categories

6. Related Work

There are plenty of NER datasets which target general English documents, mostly news articles. The CoNLL 2003 dataset (Sang and De Meulder, 2003) have four named entity tags: `LOC` (location), `ORG` (organization), `PER` (person) and `MISC` (miscellaneous). Ontonotes 5.0 (Pradhan et al., 2013) handles wider range of documents including phone conversations and Web pages. In the biomedical area, there are a lot of annotated corpus, such as the GENIA corpus (Kim et al., 2003). In the chemistry area, CHEMDNER (Krallinger et al., 2015) targets extraction of chemical materials with seven fine grained categories, such as abbreviations and identifiers. i2b2 challenge 2010 data (Uzuner et al., 2011) targets the extraction of medical concepts from radiology reports.

On the other hand, there are few manually annotated corpora in Materials Science and Engineering. Exceptions are matscholar (Weston et al., 2019), the materials synthesis corpus (Mysore et al., 2019), and (Foppiano et al., 2019). The matscholar has annotations of seven entities, such as material names, synthesis methods, and applications. However, it targets the general material science domain while we focus on superconducting domain. In addition, we include numerical expressions, such as temperatures, which are essential for finding new superconductors. For reference, inter-annotator agreement of this corpus was 87.4%. The materials synthesis corpus is annotated information re-



Similar Top-N

Parameter

Target: film

Top-N: 10

Tag: Process

Model Option

Window Size: 8

Min Count: 5

Result

Target: film
Top-N: 10
Tag: Process
Window Size: 8
Min Count: 5

rank	word	score
1	ISD	0.45202
2	epitaxially	0.43835
3	epitaxially grown	0.40697
4	substrate temperature	0.40086
5	ion beam	0.39507
6	epitaxial growth	0.39161
7	deposited	0.39049
8	partial melting	0.38841
9	depositions	0.38562
10	Growth	0.38024

Figure 5: demo page

lated to the synthesis process, such as operations and their typed arguments. In this work, quantitative information is handled as part of typed arguments. However, because it targets only the paragraphs of the synthesis process, the resultant corpus is different from ours and not suitable for the purpose of this study, where we deal with a wide range of corpus related to superconductivity. For reference, its inter-annotator agreements were 20.5-97.1%, which highly depend on the categories. (Foppiano et al., 2019) studies an information extraction task on superconducting materials domain, whose direction is close to our work. They annotated material names and transition temperatures on five full papers, and obtained pair sets of these entities. However, our experimental results showed that we need the amount of around 1,000 annotated abstracts to obtain mostly stable results.

There are a few studies that worked on information extraction using the material science corpora. (Tamari et al., 2019) proposed a method for generating action graphs from material process texts in the framework of reinforcement learning by using the material synthesis corpus. (Onishi et al., 2018) proposed distant supervised learning framework to extract relationships with a small number of annotated texts. Based on the extracted relationships, a process-structure-property-performance (PSPP) design chart (Olson, 2013) was constructed.

Our sizable annotated corpus focuses on comprehensive knowledge of superconductivity described in 1,000 abstracts.

7. Conclusions and Future Work

In this paper, we constructed a novel manually-annotated corpus, SC-CoMIcs, tailored for extracting information related to superconductivity. This corpus is essential for extracting not only material names and process information but also doping and quantitative information. Around 75-85% of annotator agreements is similar to those of other corpora (Weston et al., 2019; Mysore et al., 2019) in MI. We also performed experiments of SciBERT NER on the corpus, and achieved the F1-scores of approximately 77%, which is comparable to the human annotator agreements. The learning curves also support that the size of the corpus is mostly sufficient enough. Our future work includes using the NER results in downstream applications such as relation extraction and summarization, and constructing term as well as document retrieval systems, which would support discovery of new superconducting materials.

Query	film		bulk	
Rank	Term	Similarity	Term	Similarity
1	*epitaxially	0.4384	*melt-processed	0.3608
2	*epitaxially grown	0.4070	*sintered	0.3018
3	*substrate temperature	0.4009	*milled	0.2905
4	*ion beam	0.3951	*welded	0.2762
5	*epitaxial growth	0.3916	*Growth	0.2760
6	*deposited	0.3905	*pressed	0.2729
7	*partial melting	0.3884	*infiltration	0.2678
8	*depositions	0.3856	*ex-situ	0.2659
9	*Growth	0.3802	*heat treatment temperature	0.2599
10	*reactive sputtering	0.3785	*sintering temperatures	0.2598
11	*growth parameters	0.3779	*joined	0.2586
12	*deposition	0.3757	*ball-milled	0.2580
13	*seeding	0.3747	*powder-in-tube method	0.2576
14	*pyrolysis	0.3692	*solidified	0.2570
15	*heat treatment temperature	0.3684	*solid-state reactions	0.2555
16	*melt-processed	0.3682	field-cooled	0.2531
17	*sputtering pressure	0.3678	liquid nitrogen temperature	0.2517
18	*dc sputtering	0.3656	*partial melting	0.2512
19	*growth temperature	0.3647	field-cooling	0.2415
20	*as-deposited	0.3635	*melt growth	0.2385

Table 7: Term retrieval results with `Process` category restriction.

Query	film		bulk	
Rank	Term	Similarity	Term	Similarity
1	films	0.5081	magnetized	0.3870
2	crack-free	0.4472	single-domain	0.3738
3	*epitaxially	0.4384	iron-arsenide	0.3683
4	magnetic shield	0.4339	Bulk	0.3644
5	buffered	0.4270	*melt-processed	0.3608
6	film surface	0.4264	single-grain	0.3591
7	c-axis orientation	0.4230	single grains	0.3552
8	buffer layers	0.4170	levitation force	0.3537
9	substrate	0.4150	*levitation	0.3513
10	sapphire	0.4122	guidance force	0.3497
11	Flat	0.4112	permanent magnet guideway	0.3463
12	domain structure	0.4108	single domain	0.3424
13	substrate surface	0.4104	high-T c superconductor	0.3318
14	buffer layer	0.4072	bulk magnet	0.3241
15	*epitaxially grown	0.4070	pellet	0.3200
16	nanodots	0.4066	magnetic stiffness	0.3198
17	faceted	0.4021	critical current anisotropy	0.3196
18	*substrate temperature	0.4009	bulks	0.3195
19	surface roughness	0.3992	type II superconductor	0.3144
20	*ion beam	0.3951	joint resistance	0.3100

Table 8: Term retrieval results without `Process` category restriction.

Bibliographical References

Bednorz, J. and Müller, K. (1986). Possible hightc superconductivity in the ba-la-cu-o system. *Zeitschrift für Physik B Condensed Matter*, 64(2):189–193.

Beltagy, I., Cohan, A., and Lo, K. (2019). Scibert: Pretrained contextualized embeddings for scientific text. *arXiv preprint arXiv:1903.10676*.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Foltyn, S., Civale, L., and MacManus-Driscoll, J. e. a. (2007). Materials science challenges for high-temperature superconducting wire. *Nature Mater.*, 6:631–642.

Foppiano, L., Thaer, M. D., Suzuki, A., and Ishii, M. (2019). Proposal for automatic extraction framework of superconductors related information from scientific literature. *Letters and Technology News*, 119(66):1–5.

Kim, J.-D., Ohta, T., Tateisi, Y., and Tsujii, J. (2003). Genia corpus – a semantically annotated corpus for bio-textmining. *Bioinformatics*, 19(suppl_1):i180–i182.

Krallinger, M., Leitner, F., Rabal, O., Vazquez, M., Oyarzabal, J., and Valencia, A. (2015). Chemdner: The drugs and chemical names extraction challenge. *Journal of cheminformatics*, 7(1):S1.

Malozemoff, A. P., Mannhart, J., and Scalapino, D. (2005). High-temperature cuprate superconductors get to work. *Physics Today*, April:41–47.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

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. *arXiv preprint arXiv:1905.06939*.

Olson, G. (2013). Genomic materials design: The ferrous frontier. *Acta Materialia*, 61(3):771 – 781. The Diamond Jubilee Issue.

Onishi, T., Kadohira, T., and Watanabe, I. (2018). Relation extraction with weakly supervised learning based on process-structure-property-performance reciprocity. *Science and technology of advanced materials*, 19(1):649–659.

Pradhan, S., Moschitti, A., Xue, N., Ng, H. T., Björkelund, A., Uryupina, O., Zhang, Y., and Zhong, Z. (2013). Towards robust linguistic analysis using ontonotes. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, pages 143–152.

Ramprasad, R., Batra, R., and Pilania, G. e. a. (2017). Machine learning in materials informatics: recent applications and prospects. *npj Comput. Mater.*, 3:54.

Řehůřek, R. and Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May. ELRA. `http://is.muni.cz/publication/884893/en`.

Sang, E. F. and De Meulder, F. (2003). Introduction to the conll-2003 shared task: Language-independent named entity recognition. *arXiv preprint cs/0306050*.

Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., and Tsujii, J. (2012). brat: a web-based tool for nlp-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107. Association for Computational Linguistics.

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 *Proceedings of the Workshop on Extracting Structured Knowledge from Scientific Publications*, pages 62–71, Minneapolis, Minnesota, June. Association for Computational Linguistics.

Uzuner, Ö., South, B. R., Shen, S., and DuVall, S. L. (2011). 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association*, 18(5):552–556.

Weston, L., Tshitoyan, V., Dagdelen, J., Kononova, O., Trewartha, A., Persson, K. A., Ceder, G., and Jain, A. (2019). Named entity recognition and normalization applied to large-scale information extraction from the materials science literature. *Journal of chemical information and modeling*, 59(9):3692–3702.