Analyzing 2D Material Synthesis Parameters from Scientific Literature with Natural Language Processing

A Technical Report for CS 4980

Presented to the Faculty of the School of Engineering and Applied Sciences
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Engineering Major

Author

Rohan Taneja
May 11, 2021

Technical Project Team Members

Rohan Taneja

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Signature _____Rohan Taneja_____ Date _____

Approved _____ Date _____
Prasanna Balachandran, Department of Materials Science and Engineering

Approved _____ Date _____
Yangfeng Ji, Department of Computer Science

Analyzing 2D Material Synthesis Parameters from Scientific Literature with Natural Language Processing

Rohan Taneja
rt4uw@virginia.edu
Faculty Advisors: Prasanna Balachandran, Yangfeng Ji
University of Virginia

May 10, 2021

Abstract

Material synthesis involves the fabrication and tailoring of material properties at the micro and even nano-structure level and has the potential to drastically change industries and introduce new functionalities and applications for existing materials. In particular, material monolayers, which are ultrathin, 2D films of molecular compounds used in metals and semiconductors, have been at the forefront of advancements in materials synthesis due to their emerging applications involving the fabrication of microelectronic devices. Our ability to characterize, identify, and synthesize these unique microstructures relies heavily on understanding properties and observed material behaviors from past experiments. Fortunately, the screening of high-performance materials via deep learning methods has offered a completely new approach for expediting materials characterization by allowing scientists to quickly recognize patterns in high-dimensional data [1]. As part of an ongoing research project in the Materials Science department at the University of Virginia, this project focuses on utilizing methods of natural language processing (NLP) and text mining to extract and codify synthesis parameters from scientific literature. By providing an automated method for collecting and codifying materials synthesis parameters from prior literature, we believe this

approach has the potential to significantly expedite the time it takes to discover new material properties and understand the unique behavior of their monolayers.

1 Background

Recent developments in the fields of machine learning and natural language processing have gave way to new applications of artificial intelligence (A.I.) relating to materials synthesis and compound identification. Natural Language Processing (NLP) is one such branch of A.I. that deals with deciphering and deriving meaning from human natural language, and has led to some of the most prominent technologies used commonly today such as speech recognition, entity recognition, predictive typing, and more. The intersection of materials science and NLP-based methods have shown significant potential in accelerating the discovery and knowledge of novel materials, specifically 2D materials [2].

However, NLP developments within the materials science community have been bottlenecked by the limited availability of training data, primarily due to the closed nature of niche scientific communities. New developments and insights into material parameters can often be overlooked or hidden in different scientific journals, effectively requiring researchers to constantly stay updated with the most recent scientific literature, even in other specialized fields of Materials Science. Prior to the era of big data, materials science research was limited by a lack of an extensive data set that would otherwise better help to understand material properties. Thus, a recent trend in materials informatics has involved exploring new ways to consolidate and codify parameters from research papers for different types of materials and microstructures. By congregating new information from both successful and failed experiments in an automated fashion, we believe scientists will soon have the data necessary to be able to make use of NLP methods to find hidden insights on the unique properties of monolayers.

1.1 Related Work

Scientific databases such as Google Scholar, Crossref, and PubMed offer thousands of consolidated research articles across several different scientific

disciplines. However, knowledge extraction from these databases is far too limited to be used solely from the query engines of these tools, which offer only minimal query options such as keywords and titles but are not able to view the individual texts. Furthermore, AtomWork, one of the largest of such databases with records of over 55,000 properties of materials, is able to query material parameters by properties such as basic crystal structure and phase diagram data. However, these queries still are not able to provide key parameters from experimentation such as annealing temperatures or grain size to reach desired properties, and thus are not able to independently provide enough information to adequately inform materials design in the current landscape of materials synthesis research.

One prior study has aimed to begin approaching this dilemma by utilizing weakly supervised learning and NLP-based techniques to identify entity relations within scientific texts from several different scientific literature sources and Web texts [3]. The focus of this study was to extract processing-structure-property-performance (PSPP) relationships, such as how a phenomenon such as annealing at a certain temperature can produce a a specific physical or chemical property within a material. In order to construct these graph relationships, several keywords were used to identify process (i.e. 'annealing'), structure (i.e. 'grain refining'), and property (i.e. 'strength') parameters from a fixed list. Weakly supervised learning allows sentences to be automatically labeled by a relational descriptor, such that these factors can be identified in different lingual contexts and categorized appropriately. A convolutional neural network (CNN) model with multiple hidden layers was trained on a corpus of 5000 sentences to approximate the probability of a desired relation in order to correctly place instances of these factors into a PSPP chart. The outcome of this study was a model for representing extracted knowledge from scientific literature as relationships, although the limited factor collection and single-labeling limited the scope and accuracy of this model.

A similar study focused on building a framework for using NLP techniques to compile metal oxide synthesis parameters across thousands of scholarly publications. This study even demonstrated how machine learning method could use the data extracted to *predict* parameters needed to synthesize titania nanotubes via hydrothermal methods [2]. The system developed was able to automatically retrieve articles and parse and codify materials syn-

thesis parameters in order to build a database containing information about synthesis conditions and properties of the materials they produce. The corpus was constructed from the CrossRef Application Programming Interface and sentences were classified as materials synthesis paragraphs manually to enable supervised learning, which were then tokenized and parsed into dependency trees. This data was transformed into word embeddings that could be fed into neural network models implemented with *tensorflow* and *sci-kit learn* libraries in order to classify metal oxide parameters. The same approach was conducted using a Support Vector Machine (SVM) to further validate the accuracy of binary classification problems used in this model. Although the training data is primarily centric to metal-oxide parameters, we believe the open-source sentence dependency algorithm and other aspects of the classifier will serve useful in our classification model for monolayer synthesis parameters.

1.2 Motivation

We believe the next step in utilizing NLP-based methods to push advancements in material design will be to construct a scalable and machine-readable database for codified synthesis parameters of monolayer materials. Monolayers can take the form of a single-molecule thick film of organic or inorganic compounds that spontaneously form on surfaces by adsorption [4].They have become a highly relevant topic of discussion in the materials science community due to their emerging applications in microelectronic devices and potential for developments in nanotechnology. They are known to be the thinnest material today and have been used in thin computer chips and semiconductor technology. One single molecule layer of a material can serve as a switch or a logic device and allow utilization of about 10^{13} units/cm^2, as opposed to present utilization in these technologies of about 10^8 units/cm^2 [5].

Little is known in the scientific community about what the most reliable way of extracting monolayers of materials is and how they can be synthesized. Most of what is known about monolayer synthesis is through self-assembled monolayers (SAMs), which are molecular assemblies that form from chemisorption between the substrate of a molecular compound and a metal surface [5]. Scientists have made some progress in recent years in understanding how SAMs can be prepared in the laboratory through various methods. One study demonstrated that adding salt to reduce the melting

point can reinforce the growth of monolayer films [6]. Another prior study demonstrated that simply dipping the desired substrate in a millimolar solution for a set time can increase the chances of extracting a monolayer from a material [7]. The compound Molybdenum Disulfide (MoS_2), an inorganic compound of molybdenum and sulfur, has become a material of particular interest after it began being used to build transistors over a decade ago. MoS_2 is composed of closely compact monolayers with weak interlayer bonding that give it unique properties such as charge density waves, although integrating compounds such as this into modern electronic materials can be challenging to achieve the right transport signature [8]. Still, the proliferation of MoS_2 growth studies has led to significant progress in understanding how monolayers of this material are formed.

We believe that applying new advancements in the fields of machine learning and deep learning will allow us to build an automated tool that can build these constructive datasets and provide us more insight into the synthesis of monolayer materials. One of the major bottlenecks in the application of machine learning methods for materials discovery is the database building process. Our project aims to improve our understanding of how monolayers can be used to develop revolutionary technologies by using PDF documents from prior experiments to form a database that is machine-readable. Our preliminary study revealed some substrates of key interest such as graphene and quartz, as well as common procedural methods amongst these studies including Atomic force spectroscopy and Raman spectroscopy. MoS_2 thin films are typically grown using chemical vapor deposition methods and we will also consider specific growth parameters, such as growth temperature, pressure, and time, to use as keywords in our database construction.

The purpose of this project is to utilize methods of machine learning and text mining to extract and codify synthesis parameters from research papers for material monolayers. By providing an automated way to collect and classify information from previous experiments, we will be able to train machine learning models using this data to develop new key insights into the synthesis of monolayers. This project will focus on using machine learning to gain a better understanding of the unique structure and properties of 2D materials or monolayers. To address these research questions, we will be using MoS_2 as a template material because significant research has been done on the structure and synthesis of MoS_2 monolayers.

2 Prior Work

2.1 Manually Building Parameter Database

We began this research project in Spring of 2019, and the following methods detail the work that we accomplished on the project prior to the Spring 2021 semester during which I took my Capstone course.

The first step of our project was to begin manually constructing a database for materials synthesis parameters for training and evaluating our model performance, which involved identifying and downloading over 50 different scientific papers relevant to MoS_2 monolayer growth. Each paper was labeled with a numeric identifier, and the experiments/methods section was extracted for analysis from each paper. We prioritized key parameters that we could find across all articles to include in our database in order to provide the most useful information relating to the growth conditions of monolayer films within each experiment. The "CVD" parameter was used to indicate the form of chemical vapor deposition (double or single vapor). The "Mo_precursor" would indicate whether a Molybdenum film or powder was synthesized, and "Mo_precursor_T" would indicate the growth temperature of the Molybdenum precursor. "S_precursor" was used to tag the Sulfur precursor used (powder or vapor), and "S_precursor_T" to indicate the growth temperature of the Sulfur. The substrate on which the MoS_2 was grown was also tagged under "substrate", and the highest growth temperature and time for which this temperature was used in the experiment were recorded under the "highest_growth_T_degC" and "growth_time_mins" columns respectively. Growth Pressure in Torrs was marked under the parameter "growth_P_torr". The thickness of a monolayer was marked under "thickness_coverage" and labeled as monolayer, bilayer, or multilayer, in order to describe the molecular thickness of the film that was synthesized in a given experiment. Lastly, we created a "distance_btwn_MoS_2_Raman_peaks" parameter to provide additional information about the molecular structure of the synthesized films from Raman Spectroscopy analyses. Furthermore, for all qualitative parameters, we provided an additional column and mapped the possible parameter values to a unique numeric identifier so that all of our data could be repre-

Paper_Num	CVD	CVD_Number	Mo_precursor	Mo_precursor_Number	Mo_precur_T	S_precurs	S_precur_T
1	double vapor	2	MoO3 powder	1	300	S powder	180
1	double vapor	2	MoO3 powder	1	300	S powder	180
2	double vapor	2	MoO3 powder	1	700	S powder	250
3	single vapor	1	MoO3 film	2	900	S powder	180
9	double vapor	2	MoO3 powder	1	830	S powder	180
35	double vapor	2	MoO3 powder	1	650	S powder	145
35	double vapor	2	MoO3 powder	1	650	S powder	145
16	double vapor	2	MoO3 powder	1	800	S powder	120
28	double vapor	2	MoO3 powder	1	650	S powder	180
28	double vapor	2	MoO3 powder	1	650	S powder	180
28	double vapor	2	MoO3 powder	1	650	S powder	180
18	double vapor	2	MoO3 powder	1	540	S powder	130
27	double vapor	2	MoO3 powder	1	700	S powder	150
20	double vapor	2	MoO3 powder	1	700	S powder	150
21	double vapor	2	MoO3 powder	1	400	S powder	200
40	double vapor	2	MoO3 powder	1	650	S powder	120

substrate	substrate_Number	highest_growth_T_degC	growth_time_mins	growth_P_torr	thickness_coverage	thickness_coverage_quan	distance_btwn_MoS2_Ra
h-BN	2	800	55	760	monolayer	1	20.5
SiO2	5	800	55	760	monolayer	1	20.5
sapphire	3	700	200	760	monolayer	1	20.2
sapphire	3	900	60	760	multilayer	3	23
SiO2/Si	1	830	15	760	monolayer	1	18
graphene	4	650	10	0.75	bilayer	2	22.4
SiO2/Si	1	650	10	1	bilayer	2	22
graphene	4	800	15	40	monolayer	1	20.6
Au_100_nm	5	650	3	760	monolayer	1	20
graphene	4	650	3	760	monolayer	1	25
h-BN	2	650	3	760	monolayer	1	21
SiO2/Si	1	750	50	0.67	monolayer	1	19.8
SiO2/Si	1	700	30	760	monolayer	1	20.3
SiO2/Si	1	700	8	760	monolayer	1	20
SiO2/Si	1	800	30	760	monolayer	1	20.9
SiO2/Si	1	650	60	0.02	monolayer	1	19.5

Figure 1: Manually built Database for MoS_2 parameters.

sented quantitatively. Although this manual database could provide a useful and condensed source of information for experts in the field of MoS_2 synthesis, the primary purpose of this database was to create a baseline to compare the parameters we expect our model to extract. The result of this manual building was a database of over 50 articles with no null values. A sample of codified parameters in our database is shown in Figure 1.

2.2 Article Retrieval and Data Preprocessing

The scientific papers were retrieved from both CrossRef and Web sources and compiled into a single directory in order for us to begin an automated procedure for extracting parameters from literature. For each paper we iden-

tified previously, we created a new directory with the original file renamed with its unique identifier appended, as well as the XML version of the PDF file if it existed for more accurate parsing. For scientific papers with an XML format, the metadata was used to directly access different parts of the paper, particularly the sections containing variations of keywords "Methods", "Experiment", "Recipe" or "Procedure". For those with only a PDF provided, we compared several different Python libraries for extracting text from PDF files and found the most success with the *tika* PDF converter library. By combining a python script implementing tika with the *natural language tool kit* (NLTK), we could convert our PDFs to text and remove stopwords (words like "the" or "in" that are not needed for our database) as well as punctuation and non-scientific special characters. The *NLTK* library suite also allowed us to perform other preprocessing factors like word tokenization so that our text could be represented as a vector of words and prepared for sentence embeddings.

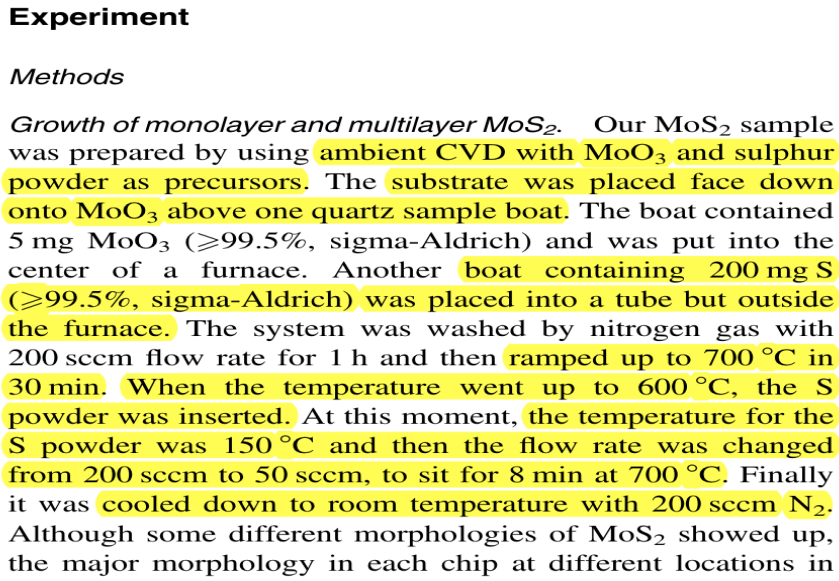
2.3 Data Labeling

Before we could begin tagging parameters dynamically with our model, we needed to extract only the relevant experimentation sections from the scientific papers to create our training data. We selected 20 PDFs of scientific papers and manually extracted the synthesis paragraphs from each of the papers. The paragraphs were then tagged for the same material parameters used in the MoS_2 database using the format $[parameter]\_[LABELCATEGORY]\_[LABEL]$. Figure 2 shows a sample of the *Methods* section for the PDF from the paper "Strongly enhanced photoluminescence in nanostructured monolayer MoS_2 by chemical vapor deposition" by Zhu et. al. Figure 3 shows the corresponding extracted text tagged with our parameters.

3 Methods

3.1 Identifying Synthesis Paragraphs with Binary Classification

Using this training data, we began building a binary classification model to classify experimental synthesis paragraphs. To train our model, the full scientific papers corresponding to the extracted text paragraphs were first



Experiment

Methods

Growth of monolayer and multilayer MoS_2. Our MoS_2 sample was prepared by using ambient CVD with MoO_3 and sulphur powder as precursors. The substrate was placed face down onto MoO_3 above one quartz sample boat. The boat contained 5 mg MoO_3 ($\geqslant$99.5%, sigma-Aldrich) and was put into the center of a furnace. Another boat containing 200 mg S ($\geqslant$99.5%, sigma-Aldrich) was placed into a tube but outside the furnace. The system was washed by nitrogen gas with 200 sccm flow rate for 1 h and then ramped up to 700 °C in 30 min. When the temperature went up to 600 °C, the S powder was inserted. At this moment, the temperature for the S powder was 150 °C and then the flow rate was changed from 200 sccm to 50 sccm, to sit for 8 min at 700 °C. Finally it was cooled down to room temperature with 200 sccm N_2. Although some different morphologies of MoS_2 showed up, the major morphology in each chip at different locations in

Figure 2: Original PDF for Reference No. 7 [9]

```
Growth of monolayer and multilayer MoS2. Our MoS2 sample was prepared by using ambient CVD with
[MoO3]_PRECURSOR_[Mo Precursor] and  [sulphur powder]_PRECURSOR_[S Precursor] as precursors. The
substrate was placed face down onto MoO3 above one quartz sample boat. The boat contained 5 mg
MoO3 (99.5%, sigma-Aldrich) and was put into the center of a furnace. Another boat containing 200
mg S (99.5%, sigma-Aldrich) was placed into a tube but outside the furnace. The system was washed
by nitrogen gas with 200 sccm flow rate for 1 h and then ramped up to 700 °C in 30 min. When the
temperature went up to 600 °C, the S powder was inserted. At this moment, the temperature for the
S powder was  150 °C and then the flow rate was changed from 200 sccm to 50 sccm, to sit for [8
min]_Growth_[Growth Time] at [700 °C]_GROWTH_[Growth T Highest]. Finally it was cooled down to
room temperature with 200 sccm N2.
```

Figure 3: Codified, tagged text for Reference No. 7.

split into 600 paragraphs and preprocessed to remove stop words, punctuation, and special characters. The paragraphs were tagged with a 1 if they were a part of a synthesis paragraph, and a 0 if they were not, as shown in Figure 4. The paragraphs were then converted to numeric features using the CountVectorizer from the sklearn package to construct a Bag of Words (BoW) representation. The paragraph data was split so that 2/3 of the sentences would be training data and the remaining 1/3 would be used as test data. A Logistic Regression model was then fitted to the training set to classify synthesis paragraphs based off common keywords. A chart of the word frequency and feature weights was graphed to determine how the frequency of keywords affected the model's weighting scheme. As seen in Figure 5, the highest weighted words had the highest word frequencies, and the 5 highest words were identified as 'tube', 'min', 'temperature', 'grown', and 'powder', which verified that the model was correctly weighting words that are likely to appear in synthesis paragraphs. The resulting confusion matrix revealed that the model correctly identified 164 paragraphs as non-synthesis, 12 paragraphs correctly identified as synthesis, 1 non-synthesis paragraph incorrectly identified as synthesis, and no false negatives. The model achieved an accuracy of 0.9944, precision of 0.9231, and recall of 1.0 on this testing set, indicating that a Logistic Regression model could successfully identify experimental paragraphs in scientific papers and was sufficient to solve this binary classification problem.

3.2 Parameter Tagging with a Conditional Random Field (CRF) Model

Using the labeled paragraphs as training data, we constructed a Conditional Random Field (CRF) model to identify and predict parameters in synthesis paragraphs based on contextual information and hand-crafted features. A CRF model allowed us to build features based off the word's characteristics as well as the location of a word in a sentence and the context words around it to make our predictions. The paragraphs were split into sentences and once again preprocessed to remove stop words and special characters. We compiled a list of features based off context words before and after a labeled parameter, as well as word-specific features such as its part of speech and the word's stem. Using the CRF's highest weighted features for each label, we were able to adjust and refine these features to improve our model's accuracy

1	mos2 ribbons grown cvd method sio2 si substrates optimized experimental parameters synthesis executed two individually controlled quartz tube furnaces using argon 97 hy
1	results presented paper reproduced times phenomenon domain shape change place chip along ow direction always existed used cvd grow mos2 two furnaces control tempe
1	si wafers 300 nm thermally grown sio2 cleaned o2 plasma used growth substrates alumina boats containing solid moo3 99 sigmaaldrich 9995 sigmaaldrich powders used mo
1	graphitized sic wafer placed 1 downstream si face alumina boat containing 10 mg molybdenum trioxide powder moo3 9998 trace metal sigmaaldrich located center heating z
1	substrates si sio2 sapphire ultrasonically degreased immersion acetone methanol isopropyl alcohol ipa solution deionized di water prior mos2 layers deposition cleaned subst

0	interestingly one additional mode 2893 cm1 e observed sfmos2 single layered 1hmos2 1t 1tmos2 2d crystal without inversion symmetry
0	comparison raman measurements carried normal ofmos2 figure 4c shows raman spectra excited 532 nm laser normal black solid curve c
0	figure 4 raman spectra excited 532 nm laser sfmos2 red solid curve sio2 si substrate blue dotted curve c normal black solid curve ofmos

Figure 4: Labeled Sentences for Supervised Learning

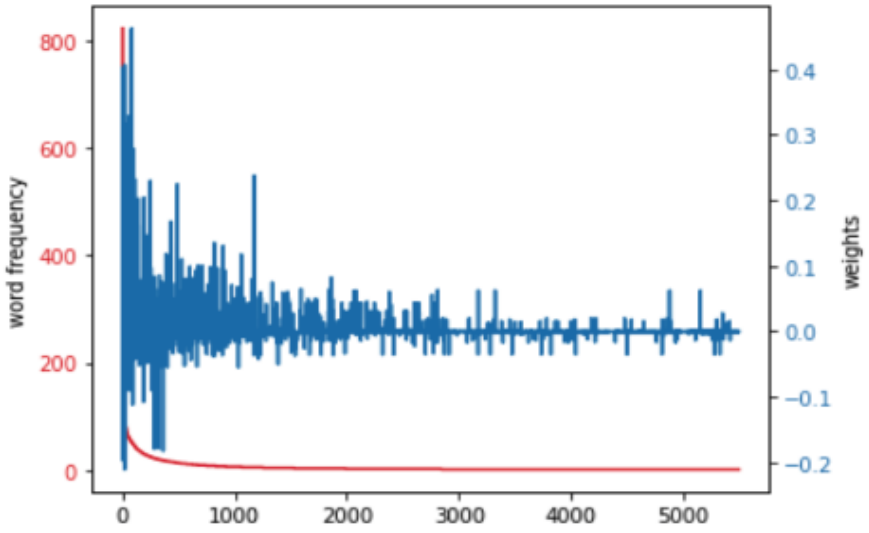


Figure 5: Word Frequency vs Weights in Experimental Paragraphs

```
Input X:
bias:   1.0
word.lower():   heat
word.stem():    heat
word[-2:]:      at
word.isupper(): False
word.isdigit(): False
word.containsDigit():   True
postag: NN
postag[:2]:     NN
-1:word.lower():        placing
-1:word.isupper():      False
-1:postag:      VBG
-1:postag[:2]:  VB
-2:word.lower():        by
-2:word.isupper():      False
-2:postag:      IN
-2:postag[:2]:  IN
+1:word.lower():        insulator
+1:word.istitle():      False
+1:word.isupper():      False
+1:postag:      NN
+1:postag[:2]:  NN

Target y:
O
```

Figure 6: Hand-Crafted Features for CRF Entity Recognition

and recall on the test set. A full list of features can be found in Figure 6 for the example word 'insulator'. A 5-cross fold validation was used to optimize the L1 and L2 regularization coefficients and fit our training paragraphs to a CRF model over 100 iterations. The model was tested on 36 validation sentences to locate and predict parameter labels, and achieved a maximum weighted F1 score of 0.65 and an average recall and precision of 0.653 and 0.526, respectively.

3.3 Future Work

The primary bottleneck for our CRF model was a limited amount of supervised training data on our test set, and future work on this project will entail various methods we can use to improve the model's accuracy. Because the model was only trained on the sentences of synthesis paragraphs and not the entirety of the scientific paper, we will need $30 - 50$ more hand-labeled

paragraphs to see significant improvements in our model's accuracy. Data augmentation can also be used to provide more training data, such as flipping digits in numeric parameters like 'highest growth temperature' and swapping context words around parameters. Random deletion and insertion of context words may also help to augment our training set. Furthermore, marginal improvements can be made by experimenting with other hand-crafted features and changing the normalization coefficients.

After further refining our model, we will be able to extract parameters from scientific articles and automatically assemble a database of monolayer synthesis parameters for the MoS2 compound. We are starting to look into other compounds relevant to monolayer synthesis such as Tungsten diselenide (WSe_2), which is also starting to be used more in material design for monolayer synthesis. Furthermore, we would like to add more automation to our model by being able to web-scrape for relevant articles to feed our model automatically, rather than us having to find the set of articles to extract information from. We will be able to utilize query engines in scientific databases like Google Scholar to pick the articles we need, and will thus be able to add more training data in a much more efficient manner. Lastly, we see potential to build off this project in the future by using unsupervised learning methods to detect patterns in synthesis conditions for certain monolayers in our database. For example, a K-Means clustering model could be used for partitional clustering so that a set of optimal growth conditions for a WSe_2 monolayer or bilayer could be identified and singled out.

4 Conclusion

The data preparation, pre-processing, and initial classification model construction are currently being used in the final step of this project - having our system identify parameters in various contexts so that we can dynamically take in new research papers and grow our database. With our paragraph classifier and CRF model for tagging parameters, we are able to automate the entirety of the process of converting a PDF to text, identifying the experimental section, and extracting relevant information about the successful syntheses of monolayers. We hope that we can continue refining these tools to more accurately codify parameters from various contexts, and future work will allow us to build on existing machine learning pipelines to predict optimal properties for monolayer growth. This would help provide some direction

for materials scientists working on synthesizing these new materials, and thus the long term ramifications of this tool might be that new materials can be discovered much quicker than they ever have before.

Acknowledgements

I would like to thank my faculty advisor, Professor Prasanna Balachandran of the Materials Science Department, for his overall guidance and supervision on this project since its beginning. I would also like to thank Wanyu Du for providing her help in constructing the CRF model and providing her NLP expertise throughout the project. Lastly, I would like to thank my faculty advisor Professor Yangfeng Ji from the CS Department for his guidance in using NLP techniques for this project and for helping with the direction of our current efforts to design and fine-tune our model for synthesis parameter extraction.

References

[1] Liu, Y., Zhao, T., Ju, W., & Shi, S. (2017). Materials discovery and design using machine learning. *Materials Discovery and Design Using Machine Learning,*3(3), 159-177. doi:https://doi.org/10.1016/j.jmat.2017.08.002

[2] Kim, E., Huang, K., Saunders, A., Mccallum, A., Ceder, G., & Olivetti, E. (2017). Materials Synthesis Insights from Scientific Literature via Text Extraction and Machine Learning. *Chemistry of Materials*,29(21), 9436-9444. doi:10.1021/acs.chemmater.7b03500

[3] Takeshi Onishi, Takuya Kadohira & Ikumu Watanabe (2018)Relation extraction with weakly supervised learning based on process-structure-property-performance reciprocity, *Science and Technology of Advanced Materials*, 19:1, 649-659, DOI:10.1080/14686996.2018.1500852

[4] Love; et al. (2005). "Self-Assembled Monolayers of Thiolates on Metals as a Form of Nanotechnology". *Chem. Rev.* 105 (4): 1103–1170. doi:10.1021/cr0300789. PMID 15826011.

[5] Chaki, N. K., Aslam, M., Sharma, J., & Vijayamohanan, K. (2001). Applications of self-assembled monolayers in materials chemistry. *Journal of Chemical Sciences*,113(5-6), 659-670. doi:10.1007/bf02708798

[6] Zhou, J., Lin, J., Huang, X., Zhou, Y., Chen, Y., Xia, J., . . . Liu, Z. (2018). A library of atomically thin metal chalcogenides. *Nature*,556(7701), 355-359. doi:10.1038/s41586-018-0008-3

[7] Bishop, A. R., & Nuzzo, R. G. (1996). Self-assembled monolayers: Recent developments and applications. *Current Opinion in Colloid & Interface Science*,1(1), 127-136. doi:10.1016/s1359-0294(96)80053-7

[8] Hu, Jia-Qi, et al. "Dependence of Electronic and Optical Properties of MoS2 Multilayers on the Interlayer Coupling and Van Hove Singularity." *Nanoscale Research Letters*, vol. 14, no. 1, 19 Aug. 2019, doi:10.1186/s11671-019-3105-9.

[9] Zhu, Yi, et al. "Strongly Enhanced Photoluminescence in Nanostructured Monolayer MoS_2 by Chemical Vapor Deposition." *Nanotechnology*, vol. 27, no. 13, 22 Feb. 2016, doi:10.1088/0957-4484/27/13/135706.