ADVANCED FUNCTIONAL MATERIALS
www.afm-journal.de

Automating Materials Exploration with a Semantic Knowledge Graph for Li-Ion Battery Cathodes

Zhiwei Nie, Shisheng Zheng, Yuanji Liu, Zhefeng Chen, Shunning Li, Kai Lei,* and Feng Pan**

The recent marriage of materials science and artificial intelligence has created the need to extract and collate materials information from the tremendous backlog of academic publications. However, this is notoriously hard to achieve in sophisticated application domains, such as Li-ion battery (LIB) cathodes, which require multiple variables for materials selection, making it challenging to automatically identify the critical terms in the text. Herein, a semantics representation framework, featuring a dual-attention module that refines word embeddings through multi-source information fusion, is proposed for literature mining of LIB cathodes. The word embeddings thus produced are biased toward domain-specific knowledge and can enable the detection of deep-seated associations among materials for targeted applications. Based on this framework, we establish a semantic knowledge graph dedicated to LIB cathodes, which allows us to unravel the latent materials relationships from scientific literature and even to discover candidate materials not yet exploited as cathodes before. This work provides a long-sought path to the realization of text-mining-based knowledge management for complicated materials systems with little dependence on domain expertise.

1. Introduction

The discovery of novel materials is mostly born of our innate ability to perceive the correlation of different substances according to their compositions, structures, and properties. For example, we can expect KCl to exhibit physical and chemical properties close to NaCl since they have identical structure (rock salt) and similar compositions,[1] and we can also envisage the replacement of Si field-effect transistors by InP because the bandgap of InP is nearly equal to that of Si.[2] This human intuition requires state-of-the-art knowledge in a specific area, which is generally only known by experts in the corresponding subfield and therefore constitutes the major obstacle in multi- and interdisciplinary research. Although there exist several standardized databases[3] containing the structural information and some basic properties of the known compounds, a comprehensive platform that integrates information on materials characteristics and applications is still lacking. This platform should encompass the scientific knowledge embedded in the text of scholarly literature and transform it into digital information flows, such that all the research results can be seamlessly interlinked with each other and better suited for data mining and knowledge discovery in materials science. Currently, the increasing proliferation of scientific literature has sparked a growing need for such a platform.

Knowledge graph,[4] an effective knowledge management tool, emerges as one of the most suitable techniques for fulfilling the above goal. In a knowledge graph, textual information is represented in a structured manner, which, when combined with association, fusion, and reasoning techniques, can realize the conversion from information to scientific knowledge. This can help researchers to obtain and sort out previous research findings accurately and efficiently, and even make qualitative predictions on materials.[5] The construction of a knowledge graph can be facilitated by natural language processing (NLP) technology.[6] NLP has been successfully applied in the fields of biology and medicine,[7] but its application in the field of materials is still in its infancy.[8] The main reasons lie in that the textual information in materials science literature usually differs with regard to application domains and adopts unstructured or highly heterogeneous formats, which severely hinder the extraction and analysis of the critical terms. Recently, Tshitoyan et al. proposed to encode the textual information as unsupervised information-dense word embeddings and demonstrated that this NLP technique is able to identify potential thermoelectric materials.[9] Despite its appealing and powerful features, their approach in its original form could hardly be extended to other materials systems whose application relies on multiple properties. For example, materials for Li-ion battery (LIB) cathodes[10] must be electrochemically active but highly stable during ion (de)intercalation,[11] with voltage,[12] capacity,[13] and rate capability[14] as essential contributing factors for materials design. Such a sophisticated issue

Z. Nie, S. Zheng, Y. Liu, Z. Chen, S. Li, F. Pan
School of Advanced Materials
Peking University Shenzhen Graduate School
Shenzhen 518055, P. R. China
E-mail: lisn@pku.edu.cn; panfeng@pkusz.edu.cn
K. Lei
Shenzhen Key Lab for Information Centric Networking & Blockchain Technology (ICNLAB)
Peking University Shenzhen Graduate School
Shenzhen 518055, P. R. China
E-mail: leik@pkusz.edu.cn

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adfm.202201437.

DOI: 10.1002/adfm.202201437

necessitates delicate optimization of the word embeddings so as to permit text mining in a sufficiently rich corpus.

In this work, we develop a framework named dual-attention-based word embeddings for materials (DATWEM) to generate representations rich in semantic information of given application domains. We demonstrate its capability to deal with materials for LIB cathodes, a complicated material system that is hard to handle in traditional NLP tasks. In this framework, the word embeddings trained from the inorganic materials corpus are adjusted by two attention modules,[15] one leveraging the word embeddings trained from the cathode materials corpus and the other utilizing the keywords of articles. In so doing, relationships between various kinds of compounds are unearthed in the sense that they possess similar textual information with a bias toward LIB cathode application. A knowledge graph dedicated to this subfield is established based on these relationships, showing transferability and strong robustness in the face of a large corpus. This protocol could enable the discovery of novel materials for complicated applications from the wealth of scientific literature, which can accelerate the design process and inspire innovative ideas for future studies of multifunctional materials.

2. Results and Discussion

The architecture of DATWEM is shown in **Figure 1**. The framework contains two independent word embedding modules that encode the corpus of inorganic materials and the corpus of cathode materials, respectively. The word embeddings obtained from the inorganic material corpus are then processed by a bi-directional long short-term memory (BiLSTM)[16] layer, after which the initial representations of the words are fed into an attention module. At this stage, the domain knowledge obtained from the cathode materials corpus is incorporated into these word embeddings. Afterward, they are subjected to another

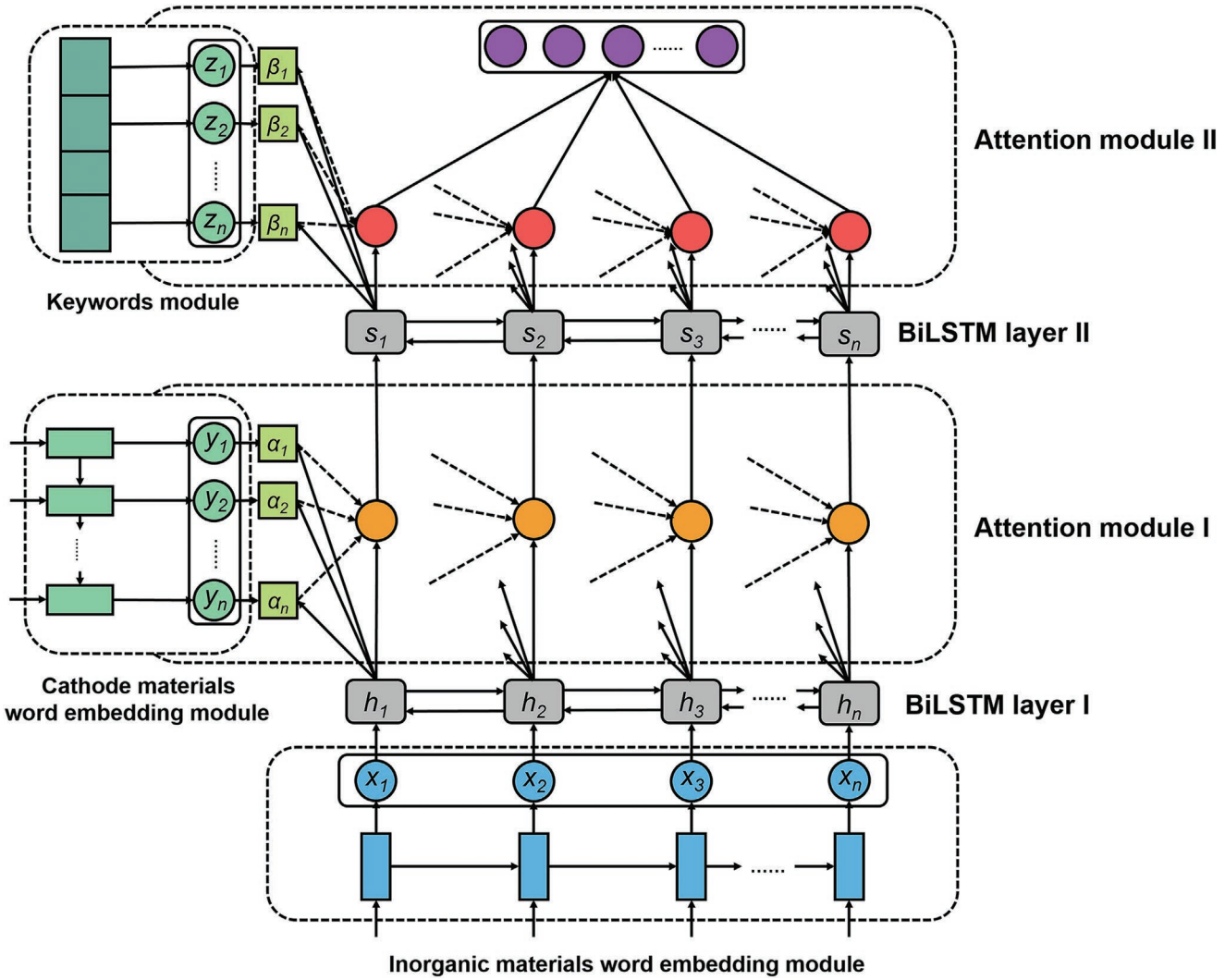

Figure 1. The architecture of DATWEM. It consists of two word embedding modules, a keywords module, two BiLSTM layers and two attention modules. *x* and *y* represent two different sets of word vectors, *z* represents keywords transformed vectors, *h* and *s* represent hidden states, α and β represent the corresponding weight.

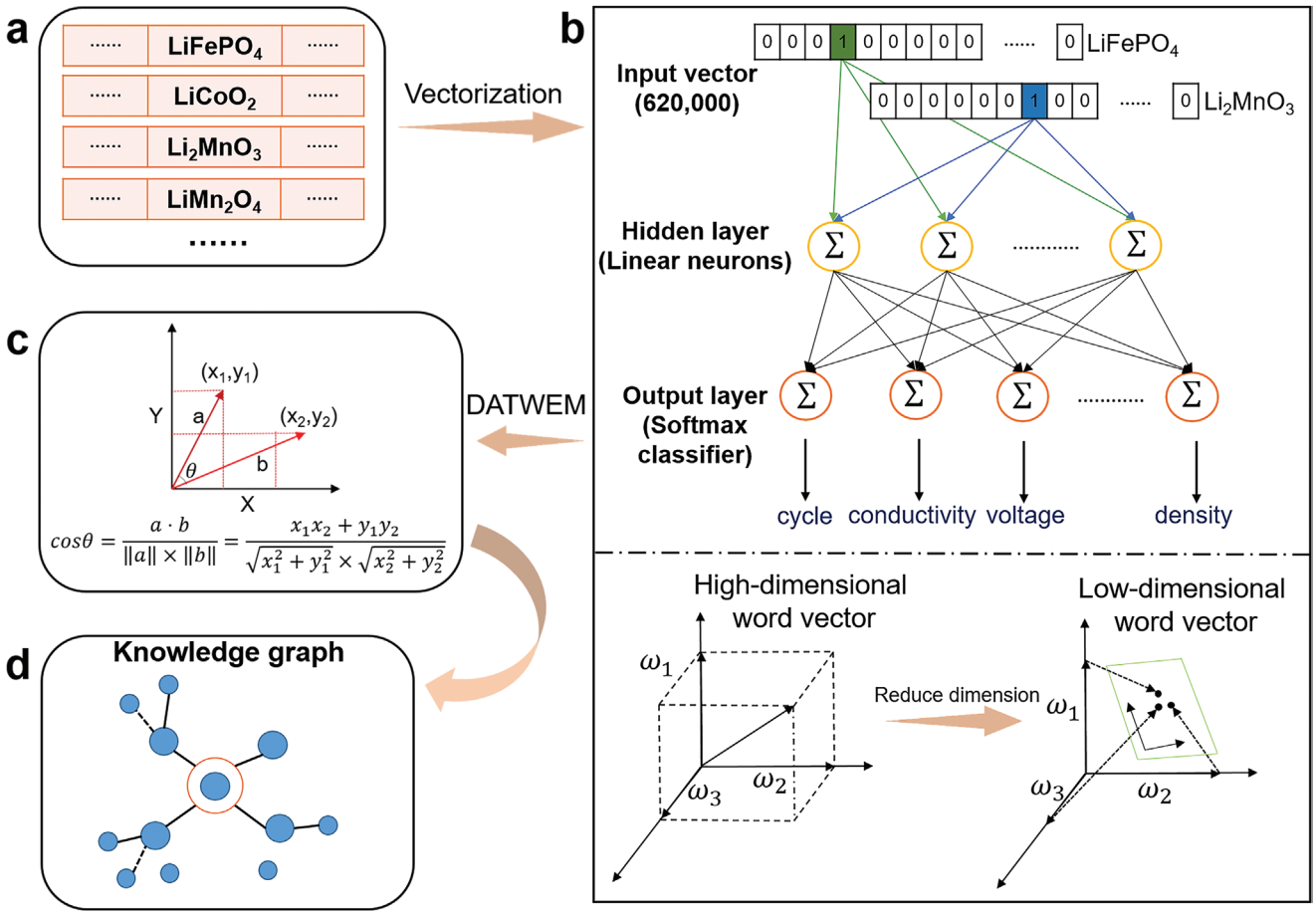

Figure 2. Construction flowchart of a knowledge graph. a) Corpus preprocessing. b) Word embedding training. c) Quantification of similarity between word embeddings. d) Establishment of the knowledge graph.

BiLSTM layer, and the second attention operation with respect to keywords is executed. In this manner, the information of inorganic materials, LIB cathodes, and the main descriptors of the articles are integrated by multi-source information fusion (see more details in Experimental Section and Supporting Information), through which the domain-specific information of the text can be effectively captured and transformed into semantics representations.

The word embeddings for materials intimately related with a targeted application have offered a practical means to quantify their relationships under the same context, which is a prerequisite for the construction of a knowledge graph. As shown in **Figure 2**, the flowchart for building a knowledge graph of cathode materials includes four steps. First, the material words are vectorized using one-hot encoding.[17] Next, the high-dimensional vectors are compressed into low-dimensional ones in the word embedding process. After the separate training of word embeddings in different corpus, they are delivered as the input attributes to DATWEM, producing the final word embeddings. Under the distributional hypothesis,[18] the cosine similarity between the word embeddings can be used as a measure of the correlation between the semantics of two subjects.[19] Accordingly, we construct the knowledge graph of LIB cathode materials, in which the nodes represent the data points corresponding to the relevant materials, and the edges represent the correlation between them using the metric of cosine similarity. It is worth mentioning that most of the conventional databases only consider the direct associations between data, while the knowledge graph can mine deeper data connections and provide a portfolio of expandable networks in the subdivision fields, thus offering a quick understanding of the correlation between materials from a data-driven perspective.

The dual-attention mechanism in DATWEM can offer high interpretability to the word embeddings due to the incorporated domain knowledge. In **Figure 3**a, the capability of the DATWEM framework in capturing the correlation between LIB cathode materials is evaluated by comparing its quality with the traditional word embedding scheme (without attention module) employed in the previous works.[9] Six indicators are taken into consideration, including accuracy, precision, recall, F1-score (F1), area under PR curve (AUPR), and area under ROC curve (AUROC), which can quantify different capabilities of the models. In order to comprehensively verify the ability to identify the correlation of materials, these indicators should be examined simultaneously. The outcomes reveal that the dual-attention module can significantly improve each of the six indicators, thus enabling a much more reliable representation of the contextual characteristics of cathode materials.

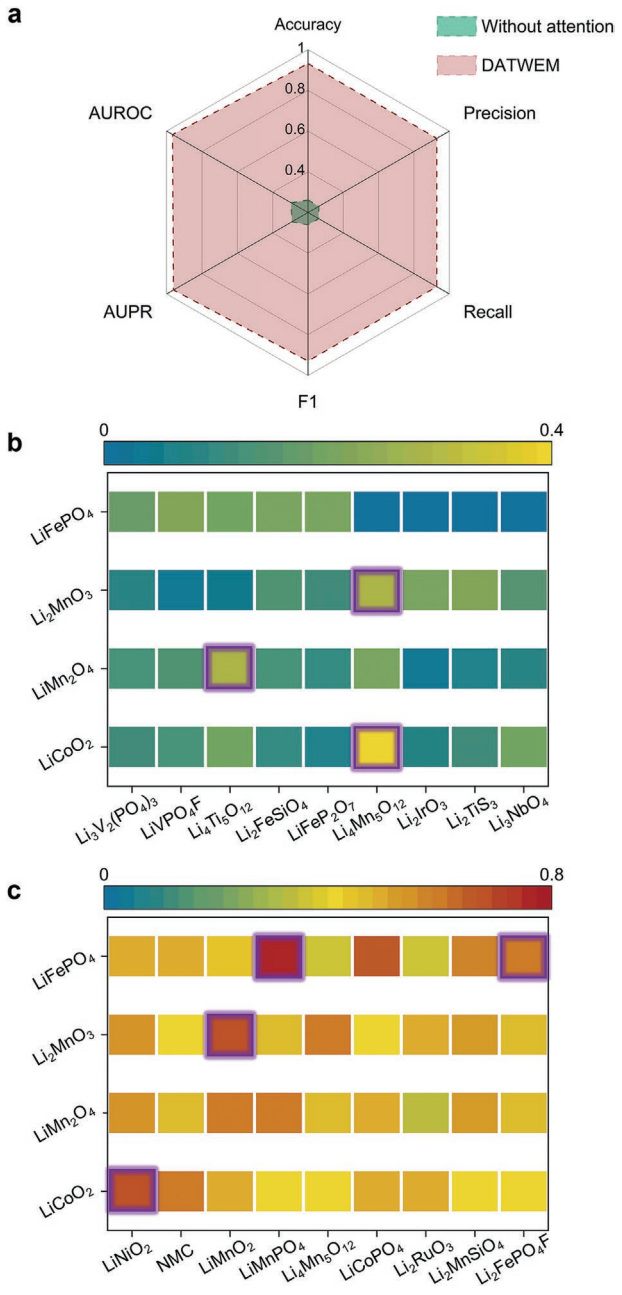

Figure 3. Comparison between different word embedding models. a) Comparison of performance in cathode materials derivation framework between traditional word embedding model without attention and the modified model with dual-attention module (DATWEM). A manual test set of 50 samples including cathode materials and non-cathode materials is used for model evaluation. Heatmaps of the cathode materials relationships obtained from b) the traditional word embedding model and c) the DATWEM model. The colors scale with the values of the cosine similarity between word embeddings.

To permit a more explicit comparison, we analyze the association network of cathode materials outputted by both frameworks. Figure 3b,c displays the degree of similarity between the word embeddings for two groups of cathode materials: the representative ones ($LiCoO_2$, $LiMn_2O_4$, Li_2MnO_3, and $LiFePO_4$) and other materials that show relatively high similarity with the keyword "cathode." The word embeddings of these typical cathode materials contain rich distributed information related to cathode application, which guarantees efficient and high-quality associations and therefore greater probability of discovering potential cathode materials. Under the traditional scheme (Figure 3b), a high similarity between Li_2MnO_3 and $Li_4Mn_5O_{12}$ (highlighted by a purple square) is derived, because they have some identical features, such as element (Mn, O) and valence state (+4 for Mn) that are closely related with their application in cathodes. This conforms to our expectations. However, relationships against our expectation are also prevalent, such as $LiCoO_2$-$Li_4Mn_5O_{12}$, which shows much higher similarity than the Li_2MnO_3-$Li_4Mn_5O_{12}$ pair. More importantly, the similar structure (spinel) between $LiMn_2O_4$ and $Li_4Ti_5O_{12}$ is overemphasized in the word embeddings, leading to the false classification of $Li_4Ti_5O_{12}$ as a cathode, while it is generally used as an anode. These contradictions with domain knowledge most likely stem from the ineffective word embeddings that fail to reproduce the key information relevant to cathode applications. By contrast, cathode materials relationships obtained from DATWEM (Figure 3c) are more consistent with existing knowledge typically recognized by researchers in this field. For example, $LiCoO_2$-$LiNiO_2$ and $LiFePO_4$-$LiMnPO_4$ pairs are extracted due to structural similarity, while other pairs bear resemblance in terms of either composition or electrochemical performance (e.g., Li_2MnO_3-$LiMnO_2$, $LiFePO_4$-Li_2FePO_4F). As compared to the traditional scheme, the calculated degree of similarity from the DATWEM framework is overall considerably higher, indicative of an association network more unambiguously described. The accuracy and robustness of the established knowledge graph allow us to efficiently query and retrieve information of materials for targeted application from the literature.

On the basis of this knowledge graph, we can navigate the materials that have made their appearance in the literature but have not yet been recognized as potential cathode materials. We perform unsupervised clustering (**Figure 4**a) to visualize the semantic similarity between different materials that are identified as related to one of the representative cathode materials in the corresponding clustered group. The parameter settings can be found in the Supporting Information. Notably, nearly all the materials in the vicinity of $LiCoO_2$ are either layered transition-metal oxides (similar structure to $LiCoO_2$) or comprised of Co ions (similar composition to $LiCoO_2$)—both features are correlated with their use as LIB cathodes. By filtering those that are already included in the cathode materials corpus, we discover a material with the formula $Li_2TiMn_3O_8$, which is a potential cathode material according to the knowledge graph but has not been explored as a cathode in the literature. $Li_2TiMn_3O_8$ and $LiCoO_2$ form a direct connection (Figure 4b) through their common features, such as the layered structure, and form an indirect connection through the latent identities, such as elements with variable valence that are appropriate for cathode application. With the aid of this knowledge graph, the connections between cathode materials are unveiled in a logical way, thus enabling the prediction of new material compositions under the guidance of existing known cathode materials.

We would like to note that the DATWEM framework is versatile and can be adopted to a variety of application domains,

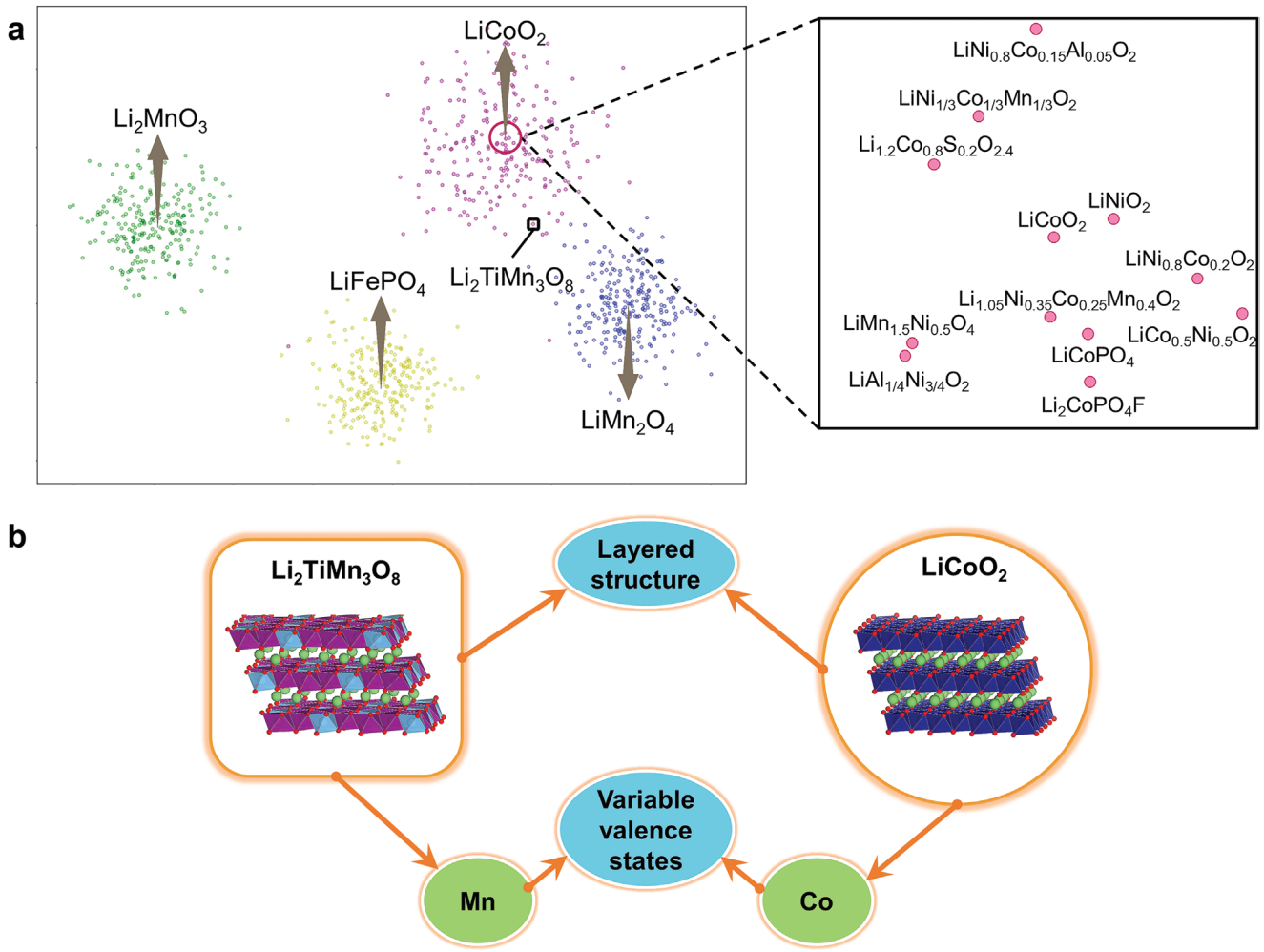

Figure 4. The application of knowledge graph for materials discovery. a) Cluster atlas of cathode materials via unsupervised learning after dimensionality reduction of the word vectors. The total input embeddings are 620 000 and large clusters of four typical cathode materials are retained to visualize the semantic similarity between different materials. b) Automatic identification of similar features between $Li_2TiMn_3O_8$ and $LiCoO_2$. Through the mining of direct or indirect paths between the nodes corresponding to different materials, potential cathode materials with high similarity to typical cathodes can be automatically identified from the corpus of scientific literature.

enabling the exploration of interpretable relationships between materials and avoiding the establishment of knowledge graphs within a black box. This is essential when a substantial amount of materials and properties are related to the investigated domain. For example, most of the reported anodes, electrolytes, and coating materials would bear a strong relationship to the cathodes in the scientific literature, making it a necessity to constrain their participation in the constructed knowledge graph for cathodes only. The attention mechanism can automate the extraction of expert knowledge from the text and therefore give rise to significant improvements in terms of materials classification and prediction. This automation also means that the knowledge graph construction process is largely independent of expertise in the corresponding application domain, thus helping to break down the disciplinary boundaries and offer opportunities for multifunctional materials design. A potential limitation of the present work is that the abstracts of the articles provide very limited information on the detailed electrochemical data, such as the voltage profile and structural evolution during operation. Future incorporation of image data and full-text corpus is therefore warranted to confer predictive power on the electrochemical performance.

3. Conclusion

In this study, we construct a semantic knowledge graph of LIB cathode materials based on a novel materials science knowledge embedding framework, DATWEM, which is especially apt to handle complicated materials systems. This framework utilizes the attention mechanism to refine word embeddings such that semantics representations rich in prior knowledge of the targeted field are generated. High fidelity is verified in the establishment of relationships between materials for cathode application, which ensures the superior quality of the constructed knowledge graph. We demonstrate the feasibility of automatic prediction of LIB cathode materials by leveraging this knowledge graph. The model proposed is transferable in a variety of subdivisions

in materials science as it can guide algorithms to learn specific information by which the interpretability is greatly enhanced. We believe that this work will pave the way for the cross integration of materials science and artificial intelligence so as to realize materials innovation from a data-driven perspective.

4. Experimental Section

Data Collection and Processing: 4.1 million abstracts related to materials, physics, and chemistry were collected through application programming interfaces (APIs)[20] of Web of Science (https://clarivate.com/webofsciencegroup/solutions/xml-and-apis/), Elsevier's Scopus and Science Direct (https://dev.elsevier.com/), and the Springer Nature (https://dev.springernature.com/). The obtained corpus was segmented by ChemDataExtractor[21] to generate tokens. Pymatgen,[22] regular expression, and rule-based techniques were used in combination to normalize chemical formulas. More specifically, the chemical formulas were screened to abandon uncertain variables, with the elements arranged in alphabetical order and the common multipliers normalized to the smallest integers.

Abstract Classification: 2000 abstracts were randomly chosen as training data to train two linear classifiers based on logistic regression[23] to obtain corpora of inorganic materials and cathode materials separately. Randomly selected abstracts were annotated as "relevant" or "not relevant" and each abstract was described using a term frequency-inverse document frequency (TF-IDF)[24] vector. F1-score[25] was selected as the evaluation index of the two classifiers. After training, the F1-score of the inorganic materials classifier reached 92%, and that of the cathode materials classifier reached 94%. Through the classification of the literature, we could remove articles outside the targeted research field.

Word Embedding: The Word2Vec[26] toolkit in genism (https://radimrehurek.com/gensim/) was used to implement word embedding. In order to achieve a better embedding effect, a series of comparative tests were carried out to explore the appropriate model and the combination of parameters. After comparison, the Skip-gram model was selected and the hyperparameters were optimized as follows: the initial learning rate was 0.01, the initial learning rate dropped to 0.0001 within 100 epochs, the embedding was 250-dimensional, the context window was 9, the threshold of subsampling was 10^{-4}, and the number of negative samples was 17.

Attention Module I: Word embeddings trained from inorganic materials corpus will form initial representations of the words after being processed by the BiLSTM layer. Word embeddings trained from cathode materials corpus and the initial representation were operated based on the attention mechanism, so as to obtain new representations rich in the characteristic information of cathode materials. The procedure is as follows.

$$h_i = \mathrm{BiLSTM}(v_i; W_1) \tag{1}$$

$$u_i = f_{\mathrm{lookup}}(u_i, W_{\mathrm{emb}}) \tag{2}$$

$$\alpha_{u_i} = \frac{\exp(h_i^T u_i)}{\sum_l \exp(h_i^T u_l)} \tag{3}$$

$$M_i = \sum_i \alpha_{u_i} u_i \tag{4}$$

where W_1 is the parameter of the first BiLSTM layer, h_i represents its hidden layer state, v_i is the word vector of inorganic materials corpus, f_{lookup} is the table lookup function, W_{emb} is the word embedding matrix of cathode materials corpus, u_i is the word embedding representation of the cathode material corpus, l is the length of input sequence of the cathode materials corpus, α_{u_i} is the corresponding weight, and M_i is the new representation of the words.

Attention Module II: We first extracted keywords from cathode materials corpus based on the TF-IDF algorithm. The top k words with the largest TF-IDF value were selected as keywords. The TF-IDF of word t_i in document d_j was then calculated as follows.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{5}$$

$$idf_i = \log\left(\frac{|D|}{1+\left|\{j : t_i \in d_j\}\right|}\right) \tag{6}$$

$$tfidf_{i,j} = tf_{i,j} \times idf_i \tag{7}$$

where $n_{i,j}$ is the number of times the word appears in the document d_j, $\sum_k n_{k,j}$ is the sum of the number of occurrences of all words in the document d_j, $|D|$ is the total number of documents in the corpus, and $|\{j: t_i \in d_j\}|$ is the number of documents containing the word t_i.

After the extraction of keywords, a keyword set was formed, and then the expression of keywords was obtained through a word embedding matrix W_{emb}^{w} and a fully connected neural network. This representation and the output of the first layer of BiLSTM underwent keywords attention operations to obtain a representation $N_{j,t}$ of each word based on the keywords attention distribution. The procedure is as follows.

$$s_i = \mathrm{BiLSTM}(M_i; W_2) \tag{8}$$

$$e_i = f_{\mathrm{lookup}}\left(e_i, W_{\mathrm{emb}}^{w}\right) \tag{9}$$

$$u_{e_i} = \tanh(W_k e_i + b_k) \tag{10}$$

$$\beta_{u_i} = \frac{\exp\left(s_i^T u_{e_i}\right)}{\sum_l \exp\left(s_i^T u_{e_l}\right)} \tag{11}$$

$$N_i = \sum_i \beta_{u_i} u_{e_i} \tag{12}$$

where W_2 is the parameter of the second BiLSTM layer, s_i represents its hidden layer state, W_k and b_k are the parameters of the fully connected layer, u_{e_i} is the representation of the keyword e_i, l is the length of input sequence of keywords, and N_i is the new expression based on keywords attention distribution.

Statistical Analysis: The corpus was processed by text search and regular expression matching[27] to reduce the statistical noise (extraneous abstracts with titles containing "Foreword," "Comment," etc.) of the corpus. For data presentation, evaluation indicators were rounded to one decimal place and the similarities between word embeddings were rounded to four decimal places. For statistical analysis, a sample size of 620 000 word embeddings was collected, forming the pre-processed vocabulary. Accuracy, Precision, Recall, F1-score, AUROC, and AUPR were adopted as statistical indices to evaluate the performance. In this work, programming languages of Python and R were employed. The model and related code have been released in an open-source format and are available in the Github repository: https://github.com/AI-for-Materials/DATWEM.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work was financially supported by Soft Science Research Project of Guangdong Province (No. 2017B030301013), The Shenzhen Science and Technology Research Grant (No. JCYJ20200109140416788), the Chemistry and Chemical Engineering Guangdong Laboratory (Grant No.

1922018), the National Natural Science Foundation of China (62072012 and 22109003), and Key-Area Research and Development Program of Guangdong Province (2020B0101090003).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

data-driven knowledge discovery, knowledge graph, Li-ion battery cathodes, semantics representation, text mining

Received: February 5, 2022
Revised: March 3, 2022
Published online: March 28, 2022

[1] a) S. Froyen, M. L. Cohen, *J. Phys. C: Solid State Phys.* **1986**, *19*, 2623; b) D. Roessler, W. Walker, *Phys. Rev.* **1968**, *166*, 599.
[2] D. Liang, J. Bowers, *Electron. Lett.* **2009**, *45*, 578.
[3] a) A. Jain, G. Hautier, C. J. Moore, S. P. Ong, C. C. Fischer, T. Mueller, K. A. Persson, G. Ceder, *Comp. Mater. Sci.* **2011**, *50*, 2295; b) S. Curtarolo, W. Setyawan, G. L. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. J. Mehl, H. T. Stokes, D. O. Demchenko, D. Morgan, *Comp. Mater. Sci.* **2012**, *58*, 218; c) J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, *JOM* **2013**, *65*, 1501; d) J. Jie, M. Weng, S. Li, D. Chen, S. Li, W. Xiao, J. Zheng, F. Pan, L. Wang, *Sci. China: Technol. Sci.* **2019**, *62*, 1423; e) S. Li, Y. Liu, D. Chen, Y. Jiang, Z. Nie, F. Pan, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2021**, *12*, e1558; f) L. Zhang, B. He, Q. Zhao, Z. Zou, S. Chi, P. Mi, A. Ye, Y. Li, D. Wang, M. Avdeev, S. Adams, S. Shi, *Adv. Funct. Mater.* **2020**, *30*, 2003087; g) B. He, S. Chi, A. Ye, P. Mi, L. Zhang, B. Pu, Z. Zou, Y. Ran, Q. Zhao, D. Wang, W. Zhang, J. Zhao, S. Admas, M. Avdeev, S. Shi, *Sci. Data* **2020**, *7*, 151.
[4] a) Y. Lin, Z. Liu, M. Sun, Y. Liu, X. Zhu, presented at *Twenty-Ninth AAAI Conf. Artificial Intelligence*, January **2015**; b) J. Pujara, H. Miao, L. Getoor, W. Cohen, *in Int. Semantic Web Conf*, (Eds: H. Alani, L. Kagal, A. Fokoue, P. Groth, C. Biemann, J. X. Parreira, L. Aroyo, N. Noy, C. Welty, K. Janowicz), Springer, Berlin, Heidelberg **2013**, pp. 542–557; c) Q. Wang, Z. Mao, B. Wang, L. Guo, *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 2724; d) Z. Nie, Y. Liu, L. Yang, S. Li, F. Pan, *Adv. Energy Mater.* **2021**, *11*, 2003580.
[5] a) M. R. Karim, M. Cochez, J. B. Jares, M. Uddin, O. Beyan, S. Decker, presented at *Proc. 10th ACM Int. Conf. Bioinformatics, Computational Biology and Health Informatics*, Niagara Falls, NY USA, September **2019**; b) X. Lin, Z. Quan, Z.-J. Wang, T. Ma, X. Zeng, presented at *IJCAI*, May **2020**; c) J. Long, Z. Chen, W. He, T. Wu, J. Ren, *Appl. Soft Comput.* **2020**, *91*, 106205; d) D. M. Bean, H. Wu, E. Iqbal, O. Dzahini, Z. M. Ibrahim, M. Broadbent, R. Stewart, R. J. Dobson, *Sci. Rep.* **2017**, *7*, 16416; e) D. Mrdjenovich, M. K. Horton, J. H. Montoya, C. M. Legaspi, S. Dwaraknath, V. Tshitoyan, A. Jain, K. A. Persson, *Matter* **2020**, *2*, 464.
[6] a) G. G. Chowdhury, *Annu. Rev. Inf. Sci. Technol.* **2003**, *37*, 51; b) P. M. Nadkarni, L. Ohno-Machado, W. W. Chapman, *J. Am. Med. Inform Assoc.* **2011**, *18*, 544.
[7] a) T. C. Rindflesch, M. Fiszman, *J. Biomed. Inf.* **2003**, *36*, 462; b) T. C. Rindflesch, H. Kilicoglu, M. Fiszman, G. Rosemblat, D. Shin, *Inf. Serv. Use* **2011**, *31*, 15; c) Y. Gu, R. Tinn, H. Cheng, M. Lucas, N. Usuyama, X. Liu, T. Naumann, J. Gao, H. Poon, *ACM Trans. Comput. Healthcare* **2021**, *3*, 1; d) L. Hong, J. Lin, S. Li, F. Wan, H. Yang, T. Jiang, D. Zhao, J. Zeng, *Nat. Mach. Intell.* **2020**, *2*, 347; e) M. Manica, R. Mathis, J. Cadow, M. R. Martínez, *Nat. Mach. Intell.* **2019**, *1*, 181.
[8] a) E. A. Olivetti, J. M. Cole, E. Kim, O. Kononova, G. Ceder, T. Y.-J. Han, A. M. Hiszpanski, *Appl. Phys. Rev.* **2020**, *7*, 041317; b) Z. Jensen, S. Kwon, D. Schwalbe-Koda, C. Paris, R. Gómez-Bombarelli, Y. Román-Leshkov, A. Corma, M. Moliner, E. A. Olivetti, *ACS Cent. Sci.* **2021**, *7*, 858; c) O. Kononova, T. He, H. Huo, A. Trewartha, E. A. Olivetti, G. Ceder, *iScience* **2021**, *24*, 102155; d) C. J. Court, J. M. Cole, *npj Comput. Mater.* **2020**, *6*, 18; e) C. B. Cooper, E. J. Beard, Á. Vázquez-Mayagoitia, L. Stan, G. B. Stenning, D. W. Nye, J. A. Vigil, T. Tomar, J. Jia, G. B. Bodedla, S. Chen, L. Gallego, S. Franco, A. Carella, K. R. Justin Thomas, S. Xue, X. Zhu, J. M. Cole, *Adv. Energy Mater.* **2019**, *9*, 1802820.
[9] V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K. A. Persson, G. Ceder, A. Jain, *Nature* **2019**, *571*, 95.
[10] a) E. M. Erickson, F. Schipper, T. R. Penki, J.-Y. Shin, C. Erk, F.-F. Chesneau, B. Markovsky, D. Aurbach, *J. Electrochem. Soc.* **2017**, *164*, A6341; b) F. Wu, J. Maier, Y. Yu, *Chem. Soc. Rev.* **2020**, *49*, 1569.
[11] a) M. Noel, V. Suryanarayanan, *J. Power Sources* **2002**, *111*, 193; b) H.-J. Liang, B.-H. Hou, W.-H. Li, Q.-L. Ning, X. Yang, Z.-Y. Gu, X.-J. Nie, G. Wang, X.-L. Wu, *Energy Environ. Sci.* **2019**, *12*, 3575.
[12] G. Cherkashinin, R. Hausbrand, W. Jaegermann, *J. Electrochem. Soc.* **2019**, *166*, A5308.
[13] B. Saha, K. Goebel, presented at *Annual Conference of the PHM Society*, San Diego CA **2009**.
[14] X. Xiao, P. Liu, J. S. Wang, M. Verbrugge, M. P. Balogh, *Electrochem. Commun.* **2011**, *13*, 209.
[15] a) Z. Niu, G. Zhong, H. Yu, *Neurocomputing* **2021**, *452*, 48; b) C.-W. Huang, S. S. Narayanan, presented at *2017 IEEE Int. Conf. Multimedia and Expo (ICME)*, Hong Kong, July **2017**.
[16] a) S. Hochreiter, J. Schmidhuber, *Neural Comput.* **1997**, *9*, 1735; b) F. A. Gers, J. Schmidhuber, F. Cummins, *Neural Comput.* **2000**, *12*, 2451; c) F. A. Gers, N. N. Schraudolph, J. Schmidhuber, *J. Mach. Learn. Res.* **2002**, *3*, 115.
[17] a) P. Rodríguez, M. A. Bautista, J. Gonzalez, S. Escalera, *Image Vis. Comput.* **2018**, *75*, 21; b) S. Okada, M. Ohzeki, S. Taguchi, *Sci. Rep.* **2019**, *9*, 2098.
[18] a) M. Sahlgren, *Ital. J. Disabil. Stud.* **2008**, *20*, 33; b) L. B. Soares, N. FitzGerald, J. Ling, T. Kwiatkowski, arXiv:1906.03158, **2019**.
[19] a) G. Boleda, *Annu. Rev. Linguist.* **2020**, *6*, 213; b) A. Lenci, *Ital. J. Linguist.* **2008**, *20*, 1.
[20] C. R. de Souza, D. Redmiles, L.-T. Cheng, D. Millen, J. Patterson, presented at *Proc. 2004 ACM Conf. Computer Supported Cooperative Work*, 00A0, Chicago, IL November **2004**.
[21] M. C. Swain, J. M. Cole, *J. Chem Inf. Model.* **2016**, *56*, 1894.
[22] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, G. Ceder, *Comp. Mater. Sci.* **2013**, *68*, 314.
[23] A. Ng, M. Jordan, *Adv. Neural Inf. Process. Syst.* **2001**, *14*, 841.
[24] A. A. Hakim, A. Erwin, K. I. Eng, M. Galinium, W. Muliady, presented at *2014 6th ICITEE*, Yogyakarta Indonesia, October **2014**.
[25] D. Chicco, G. Jurman, *BMC Genomics* **2020**, *21*, 6.
[26] X. Rong, arXiv:1411.2738, **2014**.
[27] P. Bille, M. Farach-Colton, *Theor. Comput. Sci.* **2008**, *409*, 486.