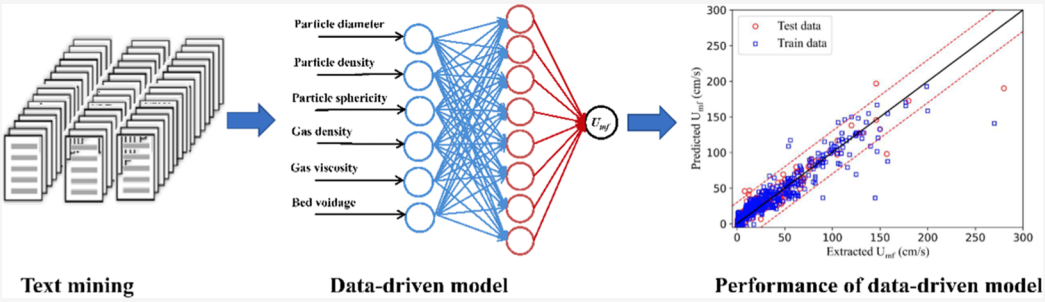
I&EC research
Industrial & Engineering Chemistry Research
pubs.acs.org/IECR
Article

Data-Driven Prediction of Minimum Fluidization Velocity in Gas-Fluidized Beds Using Data Extracted by Text Mining

Jibin Zhou, Duiping Liu, Mao Ye,* and Zhongmin Liu

Cite This: *Ind. Eng. Chem. Res.* 2021, 60, 13727−13739
Read Online



ABSTRACT: Minimum fluidization velocity (U_{mf}) is of fundamental importance in gas fluidization. Lots of empirical correlations have so far been reported in the literature to calculate U_{mf}. However, U_{mf} is affected by numerous factors, including, among others, the operation conditions and physical properties of both solids and gases. The applicability of empirical correlations relies essentially on the experiments upon which they were developed, and in practice, the choice of U_{mf} is a matter of the knowledge and experience of chemical engineers. In this work, we proposed to establish a database by extracting experimental data of U_{mf} from open literature using the text mining technique. We first presented a pipeline of natural language processing to identify and extract the functional parameters related to U_{mf} with 83% accuracy from ∼40 000 papers. A database of U_{mf} containing eight impacting factors, i.e., particle diameter, particle density, particle sphericity, bed voidage at minimum fluidization, gas density, gas viscosity, operating temperature, and pressure, was created. We then used a data-driven machine learning method with the extracting data to predict U_{mf}, which is shown superior over the empirical correlations by achieving higher accuracy for a much wider range of gas−solid systems. We expect this work illustrates a potential and promising approach to make use of the huge amount of experimental data in the literature and replace the empirical correlations in practical chemical engineering design and operations.

1. INTRODUCTION

Due to the excellent performance of mass and heat transfer, fluidized beds have been widely employed in industrial processes.[1,2] The design and operation of industrial fluidized beds rely heavily on the understanding of particulate two-phase flows and demand well-established methods to calculate key parameters describing critical hydrodynamic characteristics.[3] The superficial velocity at incipient fluidization, namely, the minimum fluidization velocity (U_{mf}), is one of the most fundamental parameters required in fluidized bed design and operation.[4,5] It defines the moment when the drag force acting on the particles balances the total gravitational force and hence constitutes a reference for the evaluation of fluidization intensity at higher superficial velocities.[6] U_{mf} is normally determined by plotting pressure drop across the bed against the superficial velocity via experiments in a fluidized bed for a particular solid−fluid system.[7] This is, however, rather time-consuming and costly, especially considering the fact that real industrial fluidized beds always operate under elevated temperatures and pressures. Chemical engineers always fitted the measured U_{mf} with the properties of particles and gases and established corresponding empirical correlations for convenience.[5,8−10] However, the accuracy and consistency of empirical correlations are impaired since most of these correlations were developed based on the data with limited experiments, which are only applicable for the situations with the same solid−fluid system under similar experimental conditions.[5,8,9,11,12] For example, Gupta et al. investigated the suitability of 79 correlations for fine tailings materials and found that most correlations underestimate the values of U_{mf}.[13] Anantharaman et al. also found that most correlations are highly empirical and system-specified, which hinders their applicability if tested on systems outside their scope.[5] Actually, it is hard to reach the consensus on which correlation of U_{mf} exhibits the best performance owing to the widespread of

Received: June 16, 2021
Published: September 8, 2021

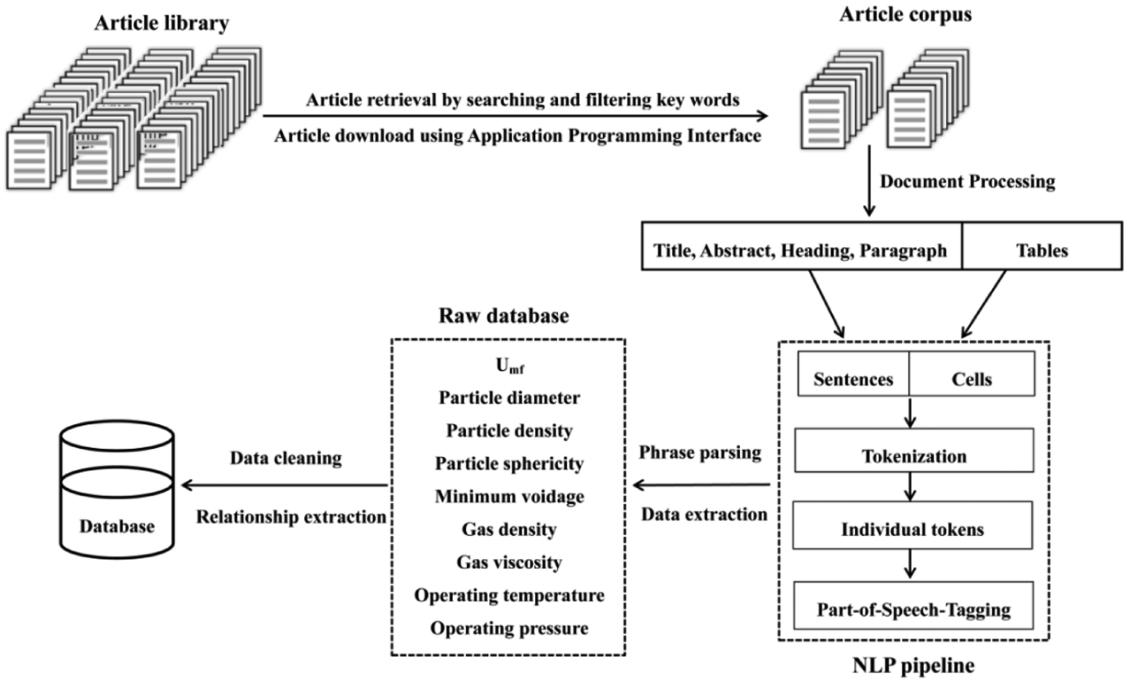


Figure 1. Schematic diagram of the automatic extraction framework based on the modified ChemDataExtractor used in this work.

solid−fluid systems encountered in practices. The performance and scope of a given empirical correlation are considerably constrained by the limited experimental data employed in developing the correlation. To address this issue, enlarging the population of the dataset is highly desired. Recently, data-driven models based on machine learning can accelerate the development of new materials and aid in learning new complex relationships.[14,15] Therefore, it may be feasible to develop new correlations to predict the value of U_{mf} with machine learning, the bottleneck of which, however, is also the lack of a reliable database.

Fortunately, with decades of development of fluidization technology, abundant experimental data on U_{mf} have been recorded and published in open literature. But it is impractical to manually collate and extract the fragmented experimental data from the large volume of literature.[16,17] It is shown in recent years that text mining based on natural language processing (NLP) offers a quite efficient way of extracting valuable data from scientific articles.[18] In fact, NLP, as efficient computer technology, has been widely used to understand human natural language and transform the unstructured text into normalized and structured data.[19] For example, Jensen et al. developed a new structural descriptor for the topology of zeolite with machine learning based on a zeolite synthesis database extracted by the NLP technique.[15] So far, several well-established NLP open-source tools have been developed for automatically extracting the information from scientific publications.[20−22] Among them, ChemDataExtractor,[22] a toolkit for mapping the unstructured text of scientific articles onto a structured format using the NLP and machine learning techniques, has drawn wide attention. ChemDataExtractor provides an effective means to extract information with high precision from scientific articles in physics, chemistry, and materials science.[23−25] Court et al. used ChemDataExtractor to successfully autogenerate the material database of Curie and Néel temperatures, which provides a basis for the discovery of magnetic materials.[23] Huang et al. presented a database of battery material using the ChemDataExtractor, which can be used for the design and prediction of battery materials.[25] Herein, we proposed an automatic extraction framework (as shown in Figure 1) based on the modified ChemDataExtractor, which consists of five steps: (1) article retrieval and download to build a relevant article corpus; (2) document processing to convert raw articles into a document structure; (3) NLP pipeline to convert the document structure into individual tokens; (4) phrase parsing to extract information from the text and table; and (5) data cleaning to remove the invalid data.

In this work, we aim at extracting U_{mf} and functional properties related to U_{mf} from vast scientific publications using the automatic extraction framework based on the modified ChemDataExtractor. We first built a database consisting of ∼1400 effective data records from a total of ∼40 000 articles by automatically extracting U_{mf} and eight parameters (i.e., particle diameter, particle density, particle sphericity, voidage at minimum fluidization, gas density, gas viscosity, operating temperature, and pressure) that significantly affect U_{mf}.[26,27] With this database, we then evaluate the performances of Ergun and Wen−Yu correlations. Moreover, machine learning models based on the artificial neural network (ANN) are also implemented to predict the U_{mf}. It shows that this data-driven method based on machine learning is superior over the empirical correlations in terms of higher prediction accuracy.

2. METHODS

In this section, methods used to identify and automatically extract the relevant information from the text and table embedded in the scientific literature will be discussed. In particular, modifications made to the original version of ChemDataExtractor[22] will be emphasized.

2.1. Article Retrieval and Download. Articles considered in this work are published in Elsevier (https://dev.elsevier.com/) and Wiley (https://onlinelibrary.wiley.com/) as they open their full paper in Xtensible Markup Language (XML) and Hypertext Markup Language (HTML) formats for text

https://doi.org/10.1021/acs.iecr.1c02307
Ind. Eng. Chem. Res. 2021, 60, 13727−13739

Table 1. Parse Expressions for Nine Properties

property	specifier (above) and units (below) of each property
U_{mf}	R("[UV]mf, re.I") | (R("[Mm]in")+ R("fluid")+ R("vel"))
	R("[cm]?m") + (W("/") + W("s")) | (R("[cm]?m") + R("[-−]") + W("1"))
particle diameter	(R("particle | solid, re.I") + I("diameter") | I("size")) | R("d[ps], re.I")
	(R("μ") + W("m")) | (R([μcm]?m))
particle density	(R("particle | solid | actual | skelet, re.I") + I("density")) | R("ρ[ps], re.I")
	R("[Kk]?g") + (W("/") + W("c?m3")) | (R("c?m") + R("[-−]") + W("3"))
particle sphericity	R("spheric | roundness") | (I("shape") + I("coefficient")) | R("φ[ps], re.I")
gas density	(I("gas") | I("air") + I("density")) | R("ρ[gf], re.I")
	R("[Kk]?g") + (W("/") + W("c?m3")) | (R("c?m") + R("[-−]") + W("3"))
gas viscosity	R("viscosity | viscidity," re.I) | R([μη]g, re.I)
	R("Pa·s") | (I("kg") + W("/") + W("m") + W("/") + W("s"))
bed voidage	(I("minimum") | I("bed") + I("voidage," re.I)) | (R("ε") + I("mf"))
temperature	I("particle") | I("bed") + I("temperature")
	(W("°") + R("[CFK]")) | R("^K$")
pressure	I("operating") + I("pressure")
	R("[km]?Pa, re.I") | R("bar") | R("atm")

mining purposes.[28,29] Article retrieval and download can be completed by the following three steps:

(1) Building the relevant Digital Object Identifiers (DOIs) by searching with the keyword "minimum fluidization velocity": To construct the article corpus, DOIs, serving as the unique article identifier, are obtained using the CrossRef search Application Programming Interface (API).[29] It yields the DOIs of almost ∼100 000 publications.
(2) Filtering the irrelevant DOIs: With DOIs obtained from the first step, we download the titles and abstracts of articles using API with the click-through service.[29] Since this work focuses on the gas−solid system, we then refine the DOIs by discarding the gas−liquid−solid, liquid−solid systems when the keywords ("liquid," "water", "oil", etc.) appear in the titles or abstracts. As a result, we create a library of ∼60 000 DOIs.
(3) Downloading the corresponding full-text papers: Most articles published before the year 2000 are in the PDF format, which is not conducive to parsing. Hence, only articles published after the year 2000 in the format of XML/HTML are selected. After this operation, the number of DOIs further drops to ∼40 000, and the full-text articles are also programmatically downloaded using API with permissions.[29]

2.2. Document Processing. As typical markup languages, XML and HTML provide explicit tags that can be used to identify section and subsection headers, which makes them easy to transform the document to title, abstract, heading, paragraph, figure, and table elements. In this way, any article with the XML/HTML format can be converted into a standard structure using the "reader" package in ChemDataExtractor.[22]

2.3. NLP Pipeline. After document processing, an NLP pipeline is run to convert the obtained elements (i.e., title, abstract, heading, paragraph, and table) into single tokens via the following steps. The first step is tokenization, where the text is split into multiple sentences and the table is split into cells to create the sentence-level tokens.[22] In the subsequent step, each sentence and cell are further split into words and/or punctuation to obtain the individual word-level tokens. In the final step, the part-of-speech (POS) tagging is applied to identify the syntactic function (e.g., noun or verb) of the individual tokens. For example, the sentence "The mean size of FCC particles is found to be about 120 μm." can be transformed to "The (DT) mean (JJ) size (NN) of (IN) FCC (NNP) particles (NNS) is (VBZ) found (VBN) to (TO) be (VB) about (RB) 120 (CD) μm (NN). (.)" after the NLP pipeline. Note that DT, JJ, NN, IN, NNP, NNS, VBZ, VBN, VB, and CD represent the determiner, adjective, noun, preposition, proper noun, plural nouns, verb (third-person singular present), verb (past participle), adverb, and cardinal number, respectively.[30]

2.4. Phrase Parsing. To make it suitable for this project to extract information related to U_{mf}, some modifications have been made to the original version of ChemDataExtractor.[22] It is anticipated that U_{mf} is closely related to the properties of both particle and fluidization gas, as well as the operating conditions (e.g., temperature and pressure).[26,27] Thus, we have tailored nine property parsers using the specific parse expressions of ChemDataExtractor for data extraction. The parse expressions consist of parser elements connected by operators (e.g., "+" or "|"). In general, each property parser contains the specifier (i.e., keywords represent the extracted properties), value, and unit (if exists), as listed in Table 1. Normally, the parser element "R" stands for regex, which can match the text patterns using regular expressions. For example, the expression of "R("[UV]mf, re.I")" can match the tokens of "Umf", "umf", "Vmf", "vmf", even "Umf,exp", and so on. "W" matches a case-sensitive individual token, while "I" matches a case-insensitive individual token. "+" operator can connect the individual tokens into a sequence, and "|" operator is used when only one of the multiple alternatives is needed.[22,25] What needs to be emphasized is that the names of property parser for the same property that appears in text and table are different, for example, when U_{mf} appears in the text, it is represented by "text_Umf", and when it appears in the table, it is represented by "table_Umf". For a property parser, the specifier is important as it is used to distinguish the attribute of the property. Given the diversity of property definitions, a perfect specifier requires a deal of effort. For example, particle diameter in different articles can be represented as the surface diameter, equivalent diameter, Sauter diameter, and even the weighted mean diameter. Similarly, the particle density may be written as the skeleton density, real density, apparent density, and bulk density.

https://doi.org/10.1021/acs.iecr.1c02307
Ind. Eng. Chem. Res. 2021, 60, 13727−13739

(a) "Powders were Geldart's group B **alumina particles** (Al_2O_3) with a **mean Sauter diameter** of **329 µm**. The alumina particles revealed to be non-porous, therefore their bulk density was equal to the **density of alumina** (i.e. **3900 kg/m^3**). From shape factor measurements by laser granulometry, we could estimate a value of about **0.55** for the **sphericity** of the particles The minimum fluidization velocity of these powders U_{mf} was found equal to **12.5 cm/s** at **20 °C** and to **8 cm/s** at **600 °C**."

Red: Particle name
Blue: Specifier
black: Value
Green: Units

(b):

```
[
  {'cem': [{'alumina particles'}]},
  {'text-particle-diameter': [{'value': '329', 'units': 'µm'}]},
  {'text-particle-density': [{'value': '3900', 'units': 'kg/m3'}]},
  {'text-particle-sphericity': [{'value': '0.55'}]},
  {'text-Umf': [{'value': '12.5 and 8', 'units': 'cm/s'}]},
  {'text-temperature': [{'value': '20', 'units': '°C'}, {'value': '600', 'units': '°C'}]}
]
```

Figure 2. Example paragraph (a) for text parsing from the article[31] and corresponding extraction result (b) based on our text parsing.

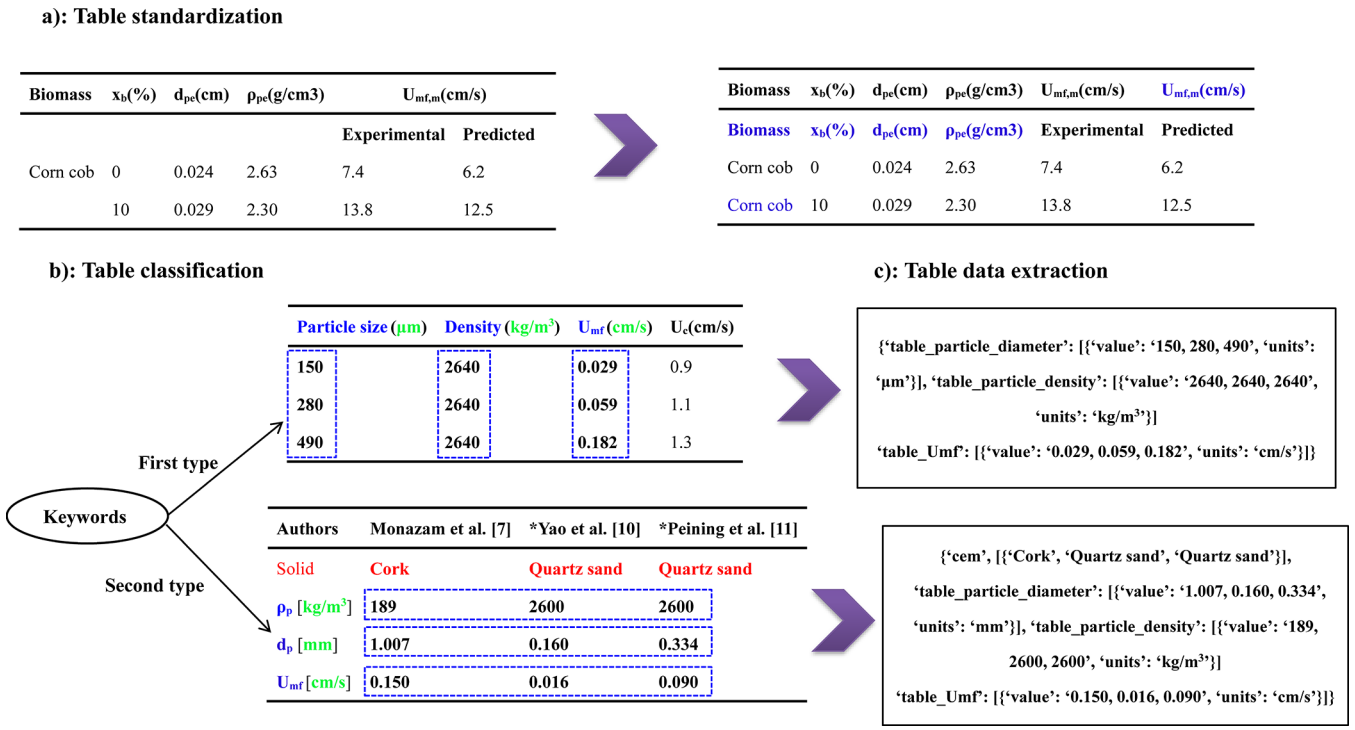


Figure 3. Table standardization (a), classification (b), and table data extraction (c). The tables are taken from the published literature.[33−35]

2.4.1. Text Parsing. To accurately extract the information embedded in the text, several property-related grammars are defined. Following the previous study,[25] we also divided the parsing grammar into five cases according to the order of occurrence: prefix-value-cem, prefix-cem-value, value-prefix-cem, cem-value-prefix, and cem-prefix-value. Therein, "cem" represents the particle name (e.g., FCC, Al_2O_3), "value" contains the numerical value with units (if exists), and "prefix" contains the specifier and the corresponding definition text around it.[25] Text parsing works at the sentence level and stores the output data of a sentence in the form of Python dictionary format.[23,25] If none of these five cases is found within a sentence, the parsing moves to process the next sentence. Considering that it is almost impossible that the information of all properties appears in a single sentence, the total data records are restored in the form of the Python list format after traversing through a paragraph.

Figure 2 demonstrates the example of a paragraph containing the parsing grammar of cem-prefix-value, prefix-cem-value, and value-prefix-cem. In Figure 2a, based on the

https://doi.org/10.1021/acs.iecr.1c02307
Ind. Eng. Chem. Res. 2021, 60, 13727−13739

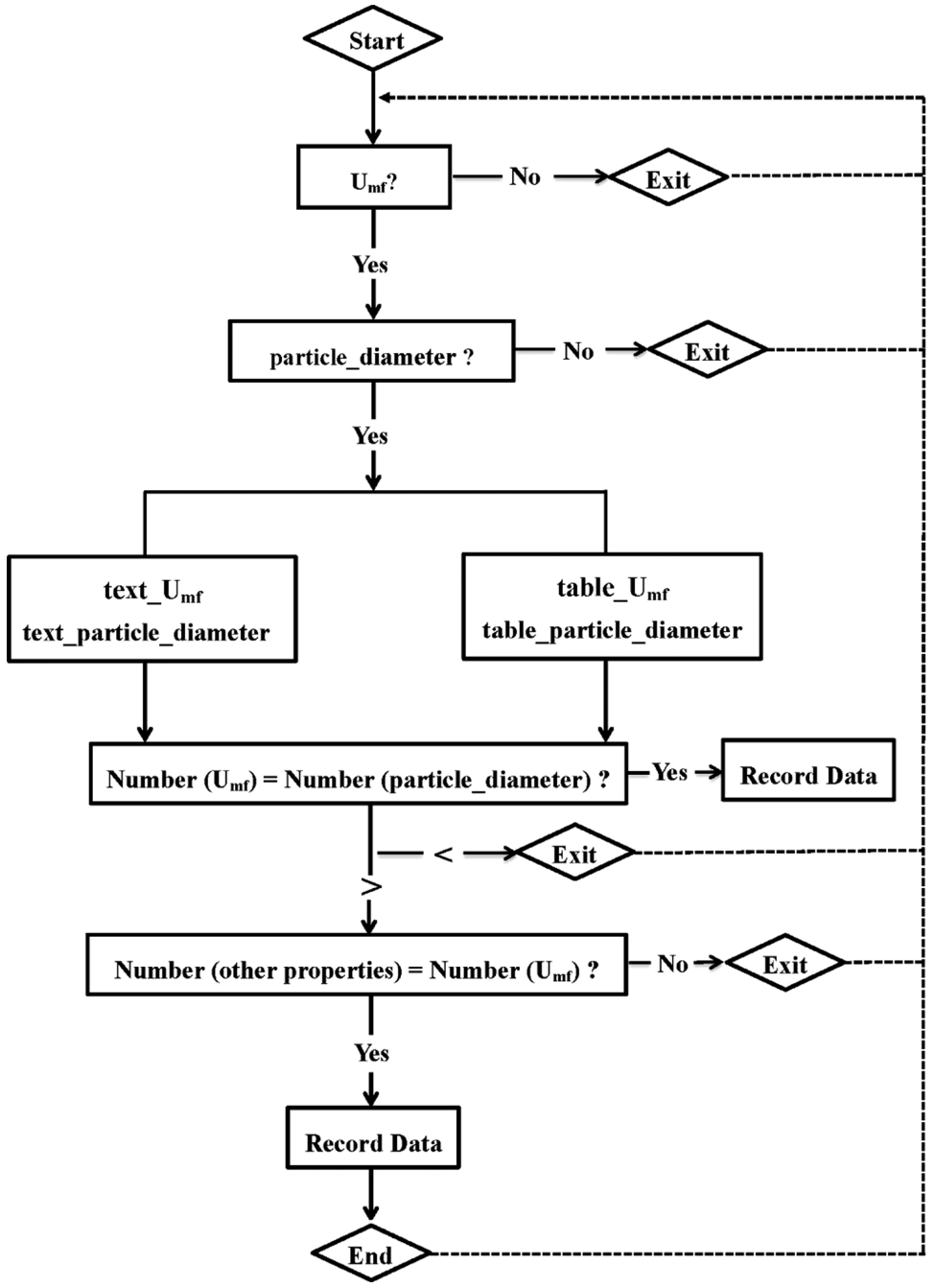


Figure 4. Basic logic for combining the data extracted from text and table.

format of cem-prefix-value, the "cem" represents the "alumina particles" and the "prefix" represents "a mean Sauter diameter of" where the specifier is "Sauter diameter." "density and U_{mf}" are also the specifiers, and "329 μm, 3900 kg/m^3, 12.5 cm/s, 20 °C, 8 cm/s, 600 °C" represent the "value". Based on the format of value-prefix-cem, we can identify that "for the sphericity" represents the "prefix" and "0.55" represents the "value". Figure 2b shows the corresponding extracted results based on our text parsing grammar. Obviously, the parsing grammar works very well in extracting the data and deriving the relations between U_{mf} and the corresponding functional parameters if all of the information appears in the same paragraph. However, in some cases, the information may not appear in a single paragraph alone. Hence, to obtain the contextual information, data records of each paragraph are stored. Then, the records are merged if there is a similarity between the extracted information, e.g., records have the same "cem". An inherent drawback of the parser is that it is too strict, and even a minor mismatch can cause its failure.[25] This undoubtedly leads to high accuracy and low recall (here, recall quantifies the fraction of relevant data found by the mining process). To improve the recall, it needs to manually collate the information related to the specifier and prefix of each functional property in the article and update the property parser.[25]

2.4.2. Table Parsing. Compared to the information embedded in the unstructured text, tables are another type of highly attractive resource for information extraction due to high data density and structured diagrams.[32] The challenge is that the complexity and diversity of table formats make table

https://doi.org/10.1021/acs.iecr.1c02307
Ind. Eng. Chem. Res. 2021, 60, 13727−13739

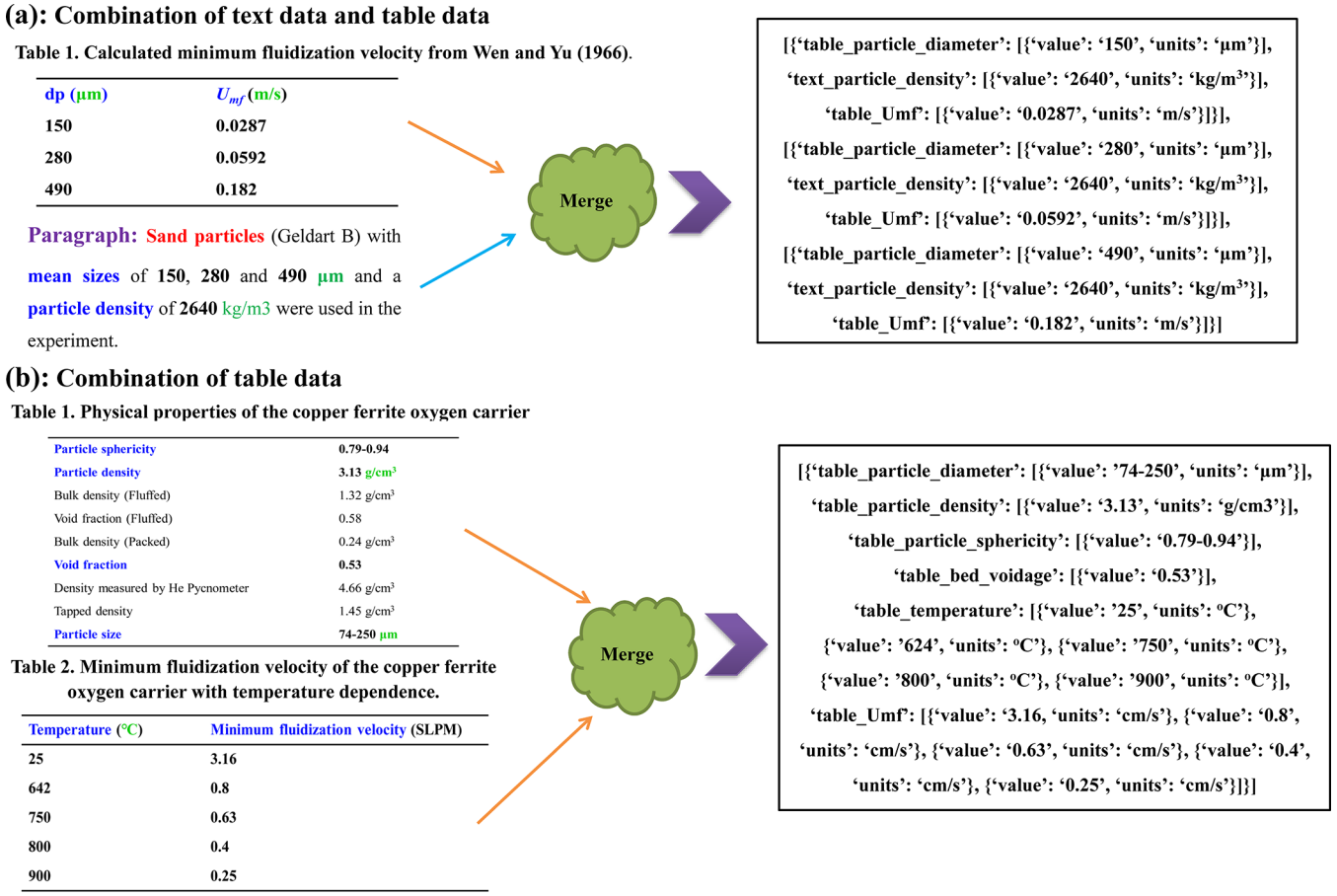


Figure 5. Examples of data combinations. Data were selected from open literature.[36,37]

parsing relatively cumbersome. Thus, we use the following steps for table parsing:

(1) Splitting the cell spanning into individual cells: Given that cell spanning (horizontal, vertical, or both) is universal in tables, it needs to be standardized before the table parsing. In the original version of ChemDataExtractor,[22] certain blank cells were simply added at the end of the row or column according to the number of cell spans, which will inevitably affect the subsequent data extraction and reduce the accuracy. In this regard, we have made some modifications based on the specific formats of the table in XML/HTML. Taking the XML format as the example, for a horizontal span, "namest" in the "entry" element indicates the starting column name and "namend" indicates the ending column name. For example, in Figure 3a, the "$U_{mf,m}$ (cm/s)" appears in a horizontal span that starts in column 5 and ends in column 6. Correspondingly, "namest" is "col5" and "namend" is "col6". For a vertical span, "morerows" in the format of integer indicates the number of rows that need to be added. In Figure 3a, "Biomass" occurs in a vertical span and "morerows" is 1, which means that it should span downward by one additional row. With these modifications, the cell spans are well split into several individual cells with the same tokens and further standardized into the format shown in Figure 3a.

(2) Identifying the category of the table: In general, a table can be divided into two categories: (1) specifiers locate in the first row, and their values are arranged in the same column below; (2) specifiers locate in the first column and their values are arranged in the next column right. The original version of ChemDataExtractor can only work for the first type of table.[22] However, the second type of table is also quite frequently used in scientific literature. Therefore, in this work, we developed a simple method to handle this type of table. As a single cell is regarded as a separate text domain, cells of the first row (heading) are first scanned. If both the specifiers of U_{mf} and other functional properties are identified, the table is classified into the first type. Then, it extracts the data from the cells of the column below. Otherwise, it belongs to the second type, and the data extraction begins from the cells on the right (Figure 3b).

(3) Extracting the data: After the type of table is identified, the cells of the heading or first column are first parsed to determine the type of properties, followed by further processing the subsequent row or column to derive the values. All of the data extracted from a table is stored in a Python dictionary, where the data for each functional property is stored in a Python list, as shown in Figure 3c. Besides, the information stored in the table captions and footnotes is also extracted through text parsing.

2.5. Data Combination. After text and table parsing, it is necessary to resolve the data interdependency and combine the data accurately into a single intact data record. The data combination processing is schematically shown in Figure 4.

It should be emphasized that since the purpose of this work is to establish a database of U_{mf}, the data record that does not

https://doi.org/10.1021/acs.iecr.1c02307
Ind. Eng. Chem. Res. 2021, 60, 13727−13739

contain the information of U_{mf} or only contain U_{mf} but with no other functional properties will be discarded. Therefore, at the beginning of the combination, we need to determine if U_{mf} is extracted from the text or table; if not, the combination will be exited and a new combination starts. If U_{mf} exits, the program then starts to search the other properties (i.e., particle property, gas property, and operating conditions). For example, if particle diameter is found, the data record will be temporarily combined and recorded. Otherwise, the program will be exited again. Then, we need to judge whether the combined data records are reasonable. Here, if one of the following four conditions is satisfied, the data record will be retained: (a) the number of text_Umf is equal to the number of text_particle_diameter; (b) the number of text_Umf is equal to that of table_particle_diameter; (c) the number of table_Umf is equal to that of table_particle_diameter; and (d) the number of table_Umf is equal to that of text_particle_diameter. If the number of U_{mf} is smaller than that of particle diameter, the data record will be discarded. However, if the number of U_{mf} is larger than the number of particle diameter, it needs to further count the numbers of other properties (e.g., particle density, gas property, operating temperature, or pressure) and then compare them with the number of U_{mf}. If they are equal, the data record is also recorded. If not, the combination ends and the data record is discarded, and a new program of combination starts again.

Figure 5 shows an example of data combination. In Figure 5a, the information of particle diameter and U_{mf} are embedded in the table but the information of particle density and particle diameter appears in the text. Therefore, according to the combination method, the particle density from text is directly assigned to the data extracted from the table. In Figure 5b, the number of U_{mf} extracted from the table is 5, which is larger than the number of particle diameter (only 1) extracted from another table but equals to the number of operating temperatures extracted from the same table. Therefore, all of the data extracted can also be well merged. In some cases where the "cem" appears in both table captions and tables, we need to make an additional comparison to check whether the "cem" is consistent.

2.6. Data Cleaning. The combined database may contain a range of invalid data and duplicated data, which makes it impossible to directly available for large-scale analysis. Therefore, a data cleaning process is applied to remove the invalid and duplicated data.

2.6.1. Removing the Obviously Abnormal Data. Although table parsing can work well for most tables in scientific articles, there are still some atypical tables whose data cannot be accurately extracted. In this case, the invalid data are removed by limiting the values of properties to a reasonable range, i.e., particle diameter in the range of 10 μm to 15 mm, U_{mf} in the range of 0−5 m/s, and particle sphericity in the range of 0.2−1. Besides, some semiempirical correlations stored in the tables may also be incorrectly extracted as the values.[38] These data are also removed by applying some predefined regular rules.

2.6.2. Removing the Non-numeric Data. In some articles, the values are problematic, such as a range or order of magnitude for the property rather than a specific value (e.g., $d_p < 0.05$, $U_{mf} \gg 1$[39]). The data extracted from this uncertain expression are also removed. However, it needs to be pointed that the data extracted from such expression, e.g., "$420 < d_p < 800$",[40] is retained selectively. In subsequent data analysis, the value of d_p (610) is taken as the average of the two values.

There is no doubt that the accuracy of the extracted datasets can be improved through the data combination and cleaning. Given that there may be inconsistencies between units of each property in different articles, a data standardization process is performed on the cleaned data to convert them to the same unit. The standardized units of diameter, density, viscosity, temperature, pressure, and U_{mf} are μm, kg/m^3, Pa·s, K, kPa, and cm/s, respectively.

In total, the automatic extraction processes yield a final set of ∼1425 effective data records, including U_{mf}, particle diameter, and particle density, from a small set of only 765 articles, showing that in the majority of the articles, either information is not available or our code could not identify them. Table 2 depicts the total number of extracted data for each property. The data size of each property is different. A smaller number of gas properties and operating conditions may be due to the fact that most experiments are carried out at room temperature and pressure, and air or nitrogen is used as the fluidizing gas, so no specific values are available. At the same time, due to the difficulty in measuring particle sphericity and bed voidage experimentally, their data sizes are also smaller.

Table 2. Total Number of Extracted Data for Each Property

property	number	property	number	property	number
particle diameter	1781	particle density	1425	gas density	146
particle sphericity	163	bed voidage	239	gas viscosity	139
temperature	190	pressure	121	U_{mf}	1781

3. RESULTS AND DISCUSSION

3.1. Technical Validation. Manually check the extracted data versus the embedded data and evaluate the performance by calculating the precision, recall, and *F*-score, which are defined as follows:[22]

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$F - \text{score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{3}$$

where TP represents true positive, FP represents false positive, and FN represents false negative. More specifically, precision indicates the ability to extract the correct data out of the database. Recall quantifies the fraction of relevant data found by the mining process. F-score provides a score that balances both the concerns of precision and recall. The validated results are depicted in Table 3 and the overall scores, as well as the scores for each property, are included. Overall, we achieved a high precision (83%), with precision on each property ranging from 77 to 87%. The lower precisions of particle and gas density may be due to the fact that the specifiers of gas and particle density are the same in some tables, resulting in confusion between gas and particle density. The overall recall of 74.7% may indicate that a small amount of data is cleaned due to the strict criteria applied to the data combination and cleaning processes.

In the end, the database of U_{mf} is obtained, and Figure 6 shows the data distribution for the nine properties. As seen in

https://doi.org/10.1021/acs.iecr.1c02307
Ind. Eng. Chem. Res. 2021, 60, 13727−13739

Table 3. Performance for Data Extraction in View of Precision, Recall, and *F*-Score

property	precision (%)	recall (%)	*F*-score (%)
particle diameter	86.2	72.6	78.8
particle density	80.2	69.6	74.5
particle sphericity	90.5	76.0	82.6
bed voidage	84.1	74.5	79.0
gas density	81.4	72.3	76.6
gas viscosity	85.7	73.4	79.1
temperature	76.8	76.9	76.8
pressure	77.4	79.2	78.3
U_{mf}	87.3	77.8	82.3
overall	83.3	74.7	78.8

Figure 6a, the particle diameter can be up to 10 mm, although mainly distributed below 2 mm. There is also a wide distribution of particle density, shown in Figure 6b. Correspondingly, the coverage of U_{mf} is wide, with a maximum of nearly 3 m/s (Figure 6i). To the best of our knowledge, we have created the most extensive collection of particle and gas properties and U_{mf}, which allows us to conduct more in-depth analyses.

3.2. Performances of Empirical Correlations.

In this section, we will evaluate the correlations of Ergun[41] and Wen–Yu[26] with the extracted database. Ergun correlation, which is present as a function of the Archimedes number (Ar) and Reynolds number (Re_{mf}) at the minimum fluidization, provides a conventional approach to determine U_{mf} based on the pressure drop in a packed bed[41]

$$\frac{1.75}{\varepsilon_{mf}^3\varphi}Re_{mf}^2 + \frac{150(1-\varepsilon_{mf})}{\varepsilon_{mf}^3\varphi^2}Re_{mf} = Ar \tag{4}$$

where

$$Ar = \frac{\rho_g d_p^3(\rho_p - \rho_g)g}{\mu^2} \tag{5}$$

$$Re_{mf} = \frac{\rho_g U_{mf} d_p}{\mu} \tag{6}$$

Wen et al. simplified the expressions by assigning values to the two groups in eqs 7 and 8[26]

$$\frac{1}{\varepsilon_{mf}^3\varphi} \approx 14 \tag{7}$$

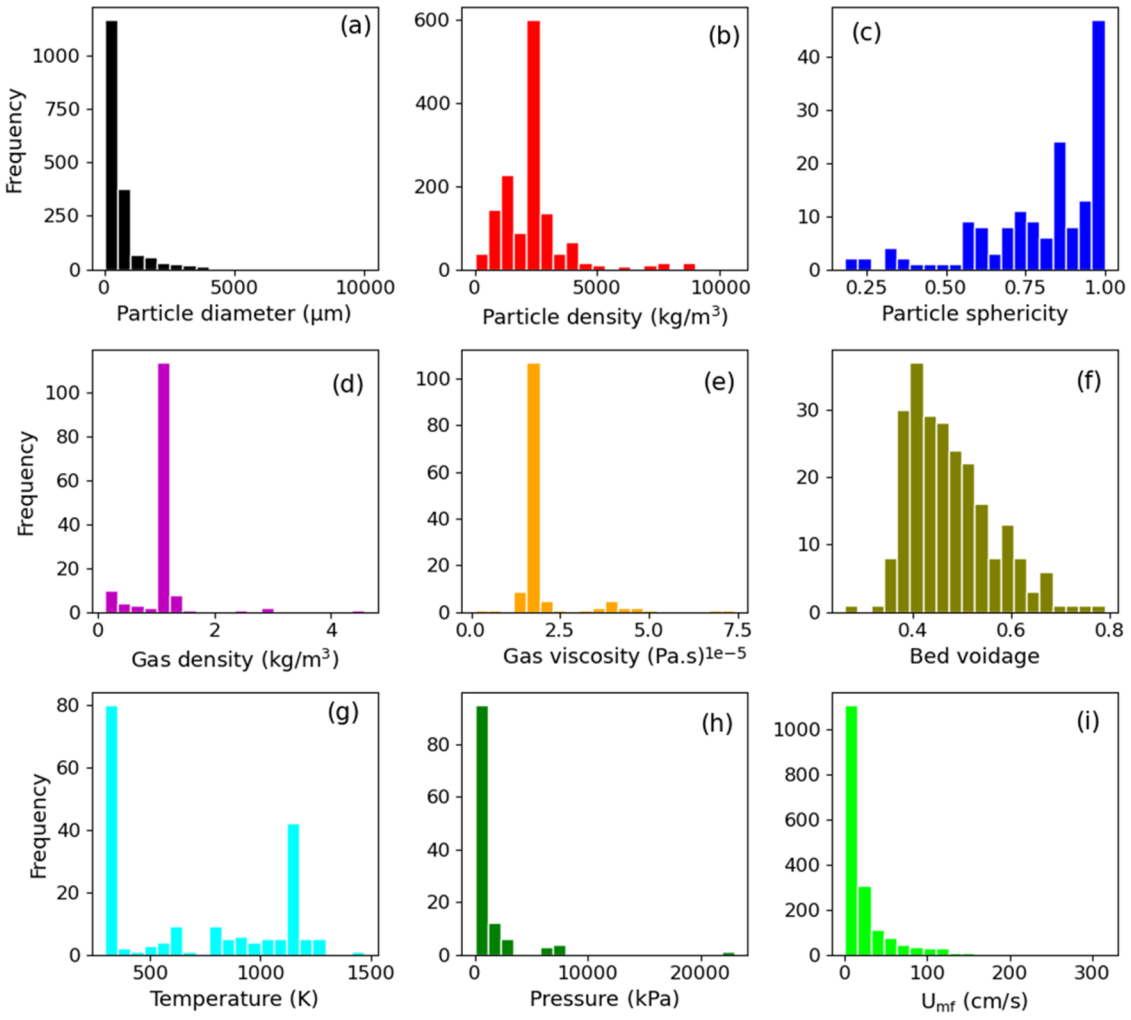


Figure 6. Autoextracted data distribution of the nine properties. (a) Particle diameter, (b) particle density, (c) particle sphericity, (d) gas density, (e) gas viscosity, (f) bed voidage, (g) operating temperature, (h) operating pressure, and (i) U_{mf}.

https://doi.org/10.1021/acs.iecr.1c02307
Ind. Eng. Chem. Res. 2021, 60, 13727−13739

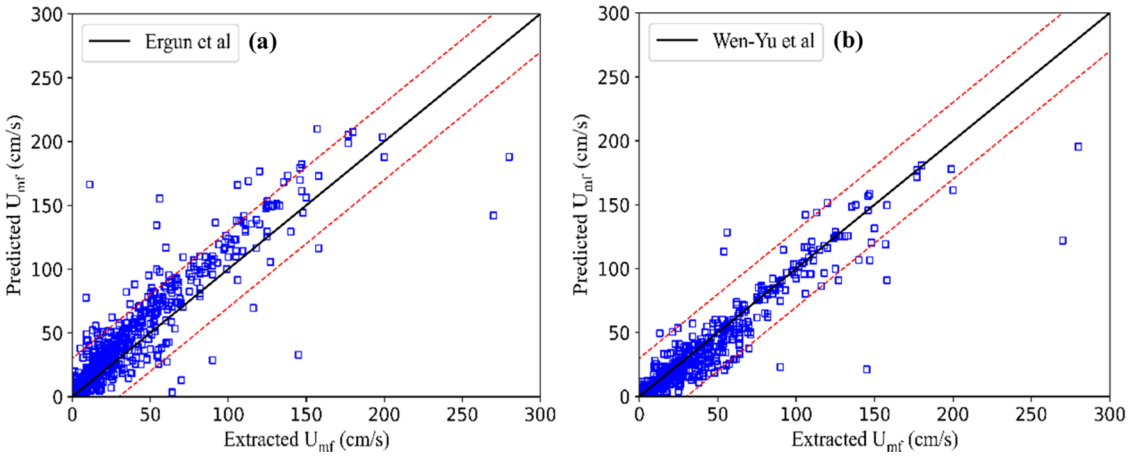


Figure 7. Comparison between the predicted and extracted values of U_{mf} for the Ergun correlation (a) and the Wen−Yu correlation (b).

$$\frac{(1-\varepsilon_{mf})}{\varepsilon_{mf}^3\varphi^2} \approx 11 \tag{8}$$

Then, Wen−Yu correlation is obtained:

$$Re_{mf} = (33.7^2 + 0.0408Ar)^{0.5} - 33.7 \tag{9}$$

Considering that the operating temperature and pressure mainly affect the gas density and viscosity, in this work, the effects of temperature and pressure are implicitly reflected on the gas density and viscosity by the following formula:[8]

$$\rho_g = 1.2 \times \frac{293}{T} \times \frac{P}{0.1} \tag{10}$$

$$\mu_g = 1.46 \times 10^{-6} \times \frac{T^{1.504}}{T + 120} \tag{11}$$

After checking the extracted database, we found that there were 21 sets of data in which both the gas properties (density and viscosity) and operation temperature or pressure are included, as depicted in Table S1. Then, we first calculated the values of gas density and viscosity according to eqs 10 and 11. As shown in Table S1, the differences between the calculated and the extracted values of gas density and viscosity are very small. In other words, it is reasonable to reflect the influences of temperature and pressure on gas density and viscosity according to eqs 10 and 11. In addition, for those data records in which the gas properties, particle sphericity, and bed voidage cannot be extracted through analysis of the experimental part in the corresponding articles, we found that most of them contain the keywords "air", "nitrogen", and "atmospheric". Therefore, for simplicity, the gas density and viscosity are considered to be 1.2 kg/m^3 and 0.000018 Pa·s, respectively, and the particle sphericity is designated as 1, while the bed voidage is the average of the available values.

To compare the performances of different correlations, three commonly used evaluation metrics are considered, namely, root-mean-squared error (RMSE), mean absolute error (MAE), and determination coefficient (R^2). Thereinto, MAE uses the absolute operator to explain how the model fared among the median values. RMSE uses the square root to assess the model's ability to predict the larger values. For RMSE and MAE, the lower the better. R^2 is utilized to judge the fitting effect of the model. The larger the value of R^2, the better the fitting effect.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y^i - \check{y}^i)^2} \tag{12}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y^i - \check{y}^i| \tag{13}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y^i - \check{y}^i)^2}{\sum_{i=1}^{N}(y^i - \bar{y})^2} \tag{14}$$

where y^i and $\check{y}^i$ are the extracted value and predicted value, respectively. $\bar{y}$ is the mean value of the extracted values. N is the total number of samples.

Figure 7 shows the performances of the Ergun correlation and the Wen−Yu correlation between the calculated and the extracted U_{mf} with the boundaries corresponding to ±10% error. Table 4 depicts the quantitative evaluation of the performance of these two correlations. For the Ergun correlation, the RMSE, MAE, and R^2 are 15.110, 8.721, and 0.797, respectively. For the Wen−Yu correlation, the RMSE and MAE decrease to 10.406 and 4.445, respectively, and the R^2 increases to 0.904. This result manifests that the Wen−Yu correlation shows a slightly better performance. To further illustrate the difference between these two correlations in more detail, a comparison result is depicted in Figure S1 and S2 and Table S2 in view of Geldart group A, B, and D particles according to the Geldart classification.[12] As shown in Table S2, for the Geldart group A particle, the Ergun correlation has the lower RMSE, MAE, and the higher R^2, indicating that when used to predict the U_{mf} of the Geldart group A particle, the Ergun correlation has a better performance than the Wen−Yu correlation. However, for Geldart group B and D particles, the Wen−Yu correlation shows a better performance than the Ergun correlation due to the lower RMSE, MAE, and the

Table 4. Performances of Different Empirical Correlations and ANN Models

	RMSE	MAE	R^2
Ergun correlation	15.110	8.721	0.797
Wen−Yu correlation	10.406	4.445	0.904
ANN mode with six parameters	9.144	2.357	0.918
ANN mode with four parameters	7.989	2.181	0.937

https://doi.org/10.1021/acs.iecr.1c02307
Ind. Eng. Chem. Res. 2021, 60, 13727−13739

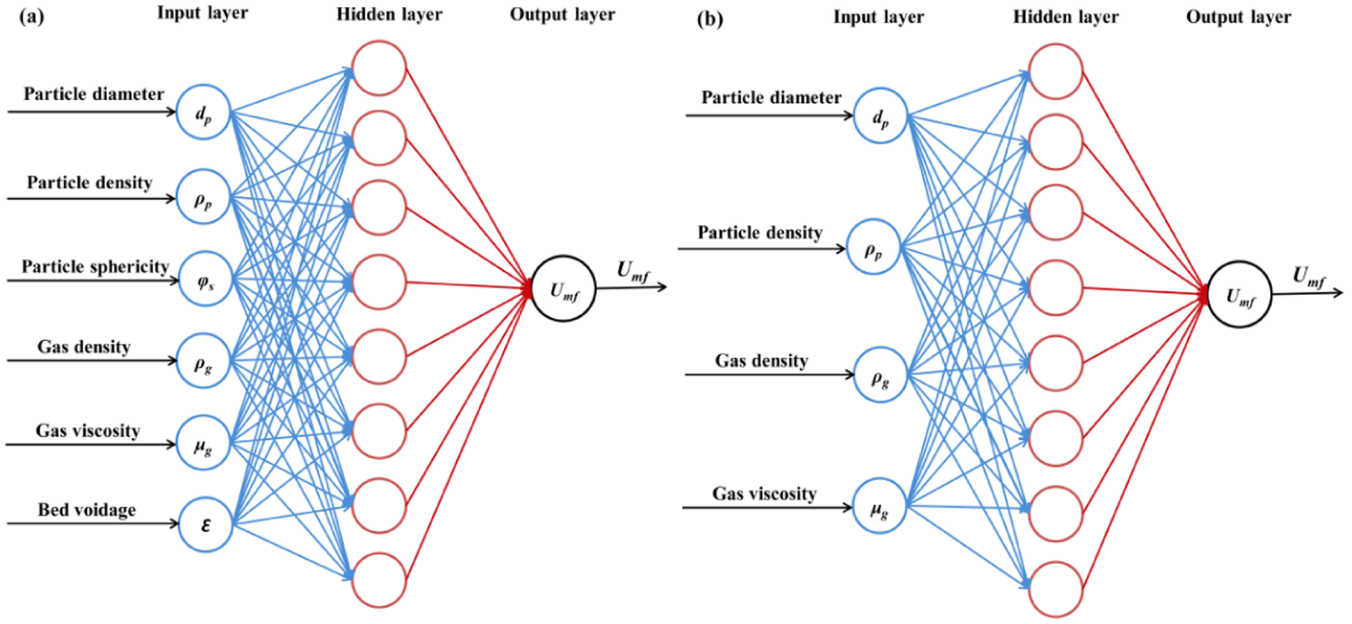


Figure 8. Schematic of the ANN models with inputs of six parameters (a) and four parameters (b).

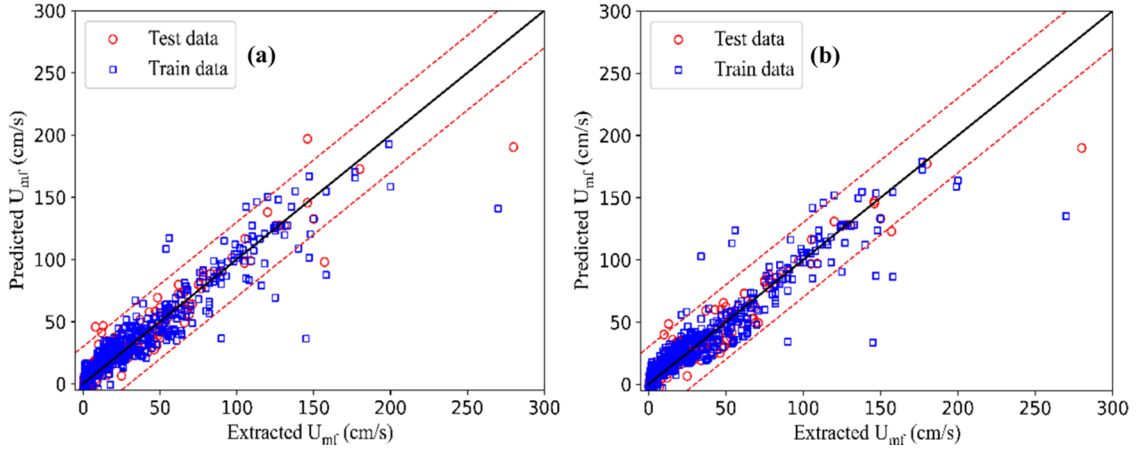


Figure 9. Comparison between the predicted and extracted values of U_{mf} for the ANN models with inputs of six parameters (a) and four parameters (b).

higher R^2. This result once again illustrates that empirical correlations always are system-specified and have their own scope of applicability.

3.3. Performances of ANN Models. Artificial neural network (ANN), as a universal approximator, is capable of modeling complex problems.[42] Especially, it can map a set of inputs to the correct output by training a large amount of data. A typical ANN model consists of an input layer, one or more hidden layers, and an output layer. The input layer is used to receive the data, the hidden layer performs nonlinear operations, and the output layer produces the results. The units in hidden layers represent nonlinear functions, which contain the parameters (i.e., weights, biases) and activation function. In this work, the inputs of the ANN model have six parameters (e.g., particle property (diameter, density, and sphericity), gas property (density and viscosity), and bed voidage, shown in Figure 8a) or four parameters (particle property (diameter and density) and gas property (density and viscosity), shown in Figure 8b); the U_{mf} is determined as the output. MAE, RMSE, and R^2 are also used to evaluate the performances of the ANN models. Rectified linear units (ReLU)[43] is selected as the activation function. The back-propagation algorithm is used to train the network due to its self-organizing, adaptive, and self-learning function.[44] To train the ANN model, data are always divided into the training set and testing set, in which the training set is used to train and optimize the model, and the test set is used to evaluate the model. By studying the influence of the number of hidden layers and neurons, as well as the proportion of test set, the numbers of the hidden layer, neurons, and proportion of test set are determined to be 1, 16, and 0.3 (Table S3), respectively. That is, in this work, 70% of the data are randomly selected as the training set and the rest as the test set.

Figure 9 shows the performances of the two ANN models by comparing the predicted values with the extracted values of U_{mf}. It can be seen that most of the data fall within the satisfactory confidence intervals (±90%). As shown in Table 4, the RMSE, MAE, and R^2 are 9.144, 2.357, and 0.918 for the ANN model with six parameters, respectively. However, for the ANN model with four parameters, the RMSE and MAE

https://doi.org/10.1021/acs.iecr.1c02307
Ind. Eng. Chem. Res. 2021, 60, 13727−13739

decrease to 7.989 and 2.181, respectively, and the R^2 increases to 0.937. This result indicates that the ANN model is superior to the empirical correlations; especially, the ANN model with four parameters is better than the ANN model with six parameters. That is, to calculate the U_{mf} with the trained ANN model, we only need to know the properties of particles and gases. Actually, since it is difficult to measure the particle sphericity and bed voidage experimentally, especially when dealing with the irregular beds and coarse particles,[11] most correlations have been proposed without considering the particle sphericity and bed voidage. However, it needs to be emphasized that reducing the number of input parameters of the ANN model does not necessarily improve the accuracy. Actually, for an ANN model, the performance depends not only on the quantity of input data but also on the quality of input data. In this work, for the ANN model with six parameters, the numbers of particle sphericity and bed voidage are only 163 and 239, respectively. However, to train the ANN model, we have artificially assumed that the remaining particle sphericity is 1 and the remaining bed voidage is the average of the available values. To some extent, although this assumption increases the amount of input data, it inevitably reduces the quality of input data. Therefore, the accuracy of the ANN model with six parameters is slightly lower than that of the ANN model with four parameters.

With the extracted database and the trained ANN model, an in-depth analysis is also carried out to illustrate the importance of different variables to U_{mf} from both the qualitative and quantitative aspects. First, the relationships between the U_{mf} and different variables are depicted in Figure S3, wherein an approximately linear trend between the particle diameter and U_{mf} is observed in Figure S3a. However, the relationships between other variables and U_{mf} are not clear or even irregular. These results qualitatively explicate that particle diameter has the greatest effect on U_{mf} among all variables. Second, when the original database is used to train the ANN model, the permutation importance (PI)[45] value of each feature for prediction can be obtained, and the result is shown in Figure S4. Its principle is to make predictions by shuffling the order of each feature value in turn and then compare the performances to that of the original database. The attenuation of performance represents the importance of the feature being shuffled. According to the normalized PI value shown in Figure S4, the particle diameter is the most important to U_{mf}, followed by particle density, particle sphericity, gas viscosity, gas density, and bed voidage. It needs to be emphasized that the obtained PI value largely depends on the amount of data of this feature. That is, the results shown in Figure S4 need to be used with caution because the amount of data for each feature in this extracted database is different, as listed in Table 2.

Considering that the minimum fluidization has been frequently studied with dimensionless parameters (see eq 4), we also attempt to use the dimensionless Re_{mf} and Ar as the inputs of the ANN model and then evaluate the performance of the model. The results can be seen in Figure 10. The RMSE, MAE, and R^2 are 11.309, 2.674, and 0.874, respectively. Obviously, the performance is worse than the ANN models with inputs of dimensional variables. This may be because when calculating the Re_{mf} and Ar, some parameters such as the particle sphericity and bed voidage are difficult to measure. For simplicity, some presumed values are allocated to these variables, which may affect the data quality and thus reduces the accuracy of the model.

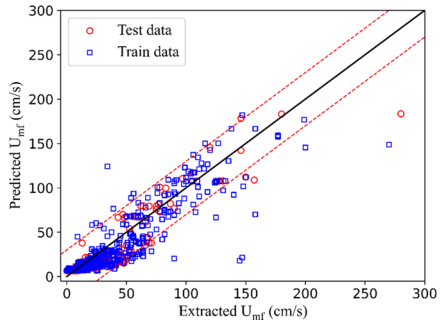


Figure 10. Comparison between the predicted and extracted values of U_{mf} for the ANN models with inputs of dimensionless Ar and Re_{mf}.

4. CONCLUSIONS

In this work, we established a general database for U_{mf} by extracting experimental information from the open literature using the text mining technique. We developed a pipeline to identify and extract the functional parameters related to U_{mf} from ~40 000 papers using a modified ChemDataExtractor toolkit. A database covering a wide range of eight factors that significantly influence U_{mf}, i.e., particle diameter, particle density, particle sphericity, voidage at minimum fluidization, gas density, gas viscosity, operating temperature, and pressure, was created. The most commonly used Ergun correlation and Wen−Yu correlation have been evaluated against the extracted database and show their limitations. We further showed that this database together with ANN machine learning models offers the better capability to predict U_{mf} for the given properties of particles and gases, and a more accurate prediction of U_{mf} could be achieved for a wide range of gas−solid systems with a large range of temperatures and pressures, compared to the empirical correlations. Note that most of the correlations that are critical for chemical reactor design and operation optimization are empirical and largely dependent upon the experimental data; this work shows the possibility of establishing a general database using the large quantity of data embedded in the literature of chemical engineering via the text mining technique. The machine learning model driven by the data from this database is expected to have a wider applicable range and better accuracy than that of empirical correlations in predicting the parameter of interest.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.iecr.1c02307.

Extracted and calculated values gas density, gas viscosity, pressure, and temperature; results of Ergun and Wen−Yu correlations for different Geldart group particles; comparisons of the predicted and extracted values of U_{mf} for the Ergun correlation in view of Geldart group A, B, and D particles; comparisons of the predicted and extracted values of U_{mf} for the Wen−Yu correlation in view of Geldart group A, B, and D particles; the performances of ANN models with different parameters; relationships between the U_{mf} and particle diameter (a), particle density (b), particle sphericity (c), gas density (d), gas viscosity (e), and bed voidage (f); and the normalized PI value of the features (PDF)

https://doi.org/10.1021/acs.iecr.1c02307
Ind. Eng. Chem. Res. 2021, 60, 13727−13739

AUTHOR INFORMATION

Corresponding Author

Mao Ye − *National Engineering Laboratory for Methanol to Olefins, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China;* orcid.org/0000-0002-7078-2402; Email: maoye@dicp.ac.cn

Authors

Jibin Zhou − *National Engineering Laboratory for Methanol to Olefins, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China*

Duiping Liu − *National Engineering Laboratory for Methanol to Olefins, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China*

Zhongmin Liu − *National Engineering Laboratory for Methanol to Olefins, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China; University of Chinese Academy of Sciences, Beijing 100083, China;* orcid.org/0000-0002-7999-2940

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.iecr.1c02307

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank the financial support from the National Natural Science Foundation of China (Grant no. 91834302) and the Natural Science Foundation of Liaoning Province of China (2021-BS-011).

NOTATION

d	particle diameter, μm
U	velocity, m/s
T	temperature, K
P	pressure, KPa
Re	Reynolds number
Ar	Archimedes number

SUBSCRIPTS

p	particle
g	gas
mf	minimum fluidization

GREEK LETTERS

ρ	density, kg/m^3
μ	viscosity, Pa·s
φ	particle sphericity
ε	bed voidage

REFERENCES

(1) Sadeghbeigi, R. *Fluid Catalytic Cracking Handbook: An Expert Guide to the Practical Operation, Design, and Optimization of FCC Units*, 3rd ed.; Elsevier Science, 2012.

(2) Tian, P.; Wei, Y.; Ye, M.; Liu, Z. M. Methanol to Olefins (MTO): From Fundamentals to Commercialization. *ACS Catal* **2015**, *5*, 1922−1938.

(3) Tan, K. K.; Tan, Y. W.; Tey, B. T.; Look, K. Y. On the onset of incipient fluidization. *Powder Technol.* **2008**, *187*, 175−180.

(4) Okhotskii, V. B. Hydrodynamic Processes in Fluidization. *Theor. Found. Chem. Eng.* **2005**, *39*, 542−547.

(5) Anantharaman, A.; Cocco, R. A.; Chew, J. W. Evaluation of correlations for minimum fluidization velocity (Umf) in gas-solid fluidization. *Powder Technol.* **2018**, *323*, 454−485.

(6) Li, H.; Yan, W.; Wu, W.; Wang, C. Characteristics of fluidisation behaviour in a pressurised bubbling fluidised bed. *Can. J. Chem. Eng.* **2013**, *91*, 760−769.

(7) Cocco, R.; Karri, S.; Knowlton, T. Introduction to Fluidization. *Chem. Eng. Prog.* **2014**, *110*, 21−29.

(8) Shao, Y.; Li, Z.; Zhong, W.; Bian, Z.; Yu, A. Minimum fluidization velocity of particles with different size distributions at elevated pressures and temperatures. *Chem. Eng. Sci.* **2020**, *216*, No. 115555.

(9) Sangeetha, V.; Swathy, R.; Narayanamurthy, N.; Lakshmanan, C. M.; Miranda, L. R. Minimum Fluidization Velocity At High Temperatures Based on Geldart Powder Classification. *Chem. Eng. Technol.* **2000**, *23*, 713−719.

(10) Subramani, H. J.; Mothivel Balaiyya, M. B.; Miranda, L. R. Minimum fluidization velocity at elevated temperatures for Geldart's group-B powders. *Exp. Therm. Fluid Sci.* **2007**, *32*, 166−173.

(11) Coltters, R.; Rivas, A. L. Minimum fluidation velocity correlations in particulate systems. *Powder Technol.* **2004**, *147*, 34−48.

(12) Geldart, D. Types of gas fluidization. *Powder Technol.* **1973**, *7*, 285−292.

(13) Gupta, S.; Agarwal, V.; Singh, S. N.; Seshadri, V.; Mills, D.; Singh, J.; Prakash, C. Prediction of minimum fluidization velocity for fine tailings materials. *Powder Technol.* **2009**, *196*, 263−271.

(14) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547−555.

(15) Jensen, Z.; Kim, E.; Kwon, S.; Gani, T. Z.; Román-Leshkov, Y.; Moliner, M.; Corma, A.; Olivetti, E. A machine learning approach to zeolite synthesis enabled by automatic literature data extraction. *ACS Cent. Sci.* **2019**, *5*, 892−899.

(16) Raccuglia, P.; Elbert, K. C.; Adler, P. D. F.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J. Machine-learning-assisted materials discovery using failed experiments. *Nature* **2016**, *533*, 73−76.

(17) Kim, E.; Huang, K.; Saunders, A.; Mccallum, A.; Ceder, G.; Olivetti, E. Materials Synthesis Insights from Scientific Literature via Text Extraction and Machine Learning. *Chem. Mater.* **2017**, *29*, 9436−9444.

(18) Rajman, M.; Besançon, R. *Text Mining: Natural Language Techniques and Text Mining Applications*; Springer, 1998.

(19) Anne, K.; Stephen, R. P. *Natural Language Processing and Text Mining*; Springer, 2007.

(20) Jessop, D. M.; Adams, S. E.; Willighagen, E. L.; Hawizy, L.; Murray-Rust, P. OSCAR4: a flexible architecture for chemical text-mining. *J. Cheminf.* **2011**, *3*, No. 41.

(21) Hawizy, L.; Jessop, D. M.; Adams, N.; Murray-Rust, P. ChemicalTagger: A tool for semantic text-mining in chemistry. *J. Cheminf.* **2011**, *3*, No. 17.

(22) Swain, M. C.; Cole, J. M. Modeling, ChemDataExtractor: A toolkit for automated extraction of chemical information from the scientific literature. *J. Chem. Inf. Model.* **2016**, *56*, 1894−1904.

(23) Court, C. J.; Cole, J. M. Auto-generated materials database of Curie and Néel temperatures via semi-supervised relationship extraction. *Sci. Data* **2018**, *5*, No. 180111.

(24) Hiszpanski, A. M.; Gallagher, B.; Chellappan, K.; Li, P.; Liu, S.; et al. Nanomaterial Synthesis Insights from Machine Learning of Scientific Articles by Extracting, Structuring, and Visualizing Knowledge. *J. Chem. Inf. Model.* **2020**, *60*, 2876−2887.

(25) Huang, S.; Cole, J. M. A database of battery materials auto-generated using ChemDataExtractor. *Sci. Data* **2020**, *7*, No. 260.

(26) Wen, C. Y.; Yu, Y. H. A generalized method for predicting the minimum fluidization velocity. *AIChE J.* **1966**, *12*, 610−612.

(27) Olowson, P.; Almstedt, A.-E. Influence of pressure on the minimum fluidization velocity. *Chem. Eng. Sci.* **1991**, *46*, 637−640.

(28) Van Noorden, R. Elsevier opens its papers to text-mining. *Nature* **2014**, *506*, No. 17.

(29) Lammey, R. CrossRef text and data mining services. *Sci. Ed.* **2015**, *2*, 22−27.
(30) Marcus, M. P.; Marcinkiewicz, M. A.; Santorini, B. Building a large annotated corpus of English: the penn treebank. *Comput. Linguist.* **1993**, *19*, 313−330.
(31) Reuge, N.; Cadoret, L.; Caussat, B. Multifluid Eulerian modelling of a silicon Fluidized Bed Chemical Vapor Deposition process: Analysis of various kinetic models. *Chem. Eng. J.* **2009**, *148*, 506−516.
(32) Gurulingappa, H.; Mudi, A.; Toldo, L.; Hofmann-Apitius, M.; Bhate, J. Challenges in mining the literature for chemical information. *RSC Adv.* **2013**, *3*, 16194−16211.
(33) Paudel, B.; Feng, Z.-G. Prediction of minimum fluidization velocity for binary mixtures of biomass and inert particles. *Powder Technol.* **2013**, *237*, 134−140.
(34) Ziaei-Halimejani, H.; Zarghami, R.; Mostoufi, N. Investigation of hydrodynamics of gas-solid fluidized beds using cross recurrence quantification analysis. *Adv. Powder Technol.* **2017**, *28*, 1237−1248.
(35) Kim, D.; Lee, G. W.; Won, Y. S.; Choi, J.-H.; Joo, J. B.; Ryu, H.-J.; Jo, S.-H. Effect of level of overflow solid outlet on pressure drop of a bubbling fluidized-bed. *Adv. Powder Technol.* **2019**, *30*, 2564−2573.
(36) Karimi, M.; Mostoufi, N.; Zarghami, R.; Sotudeh-Gharebagh, R. Nonlinear dynamics of a gas−solid fluidized bed by the state space analysis. *Chem. Eng. Sci.* **2011**, *66*, 4645−4653.
(37) Siriwardane, R.; Benincosa, W.; Riley, J.; Tian, H.; Richards, G. Investigation of reactions in a fluidized bed reactor during chemical looping combustion of coal/steam with copper oxide-iron oxide-alumina oxygen carrier. *Appl. Energy* **2016**, *183*, 1550−1564.
(38) Li, X.; Shen, Y.; Wei, L.; He, C.; Lapkin, A. A.; Lipiński, W.; Dai, Y.; Wang, C.-H. Hydrogen production of solar-driven steam gasification of sewage sludge in an indirectly irradiated fluidized-bed reactor. *Appl. Energy* **2020**, *261*, No. 114229.
(39) Almendros-Ibáñez, J. A.; Fernández-Torrijos, M.; Díaz-Heras, M.; Belmonte, J. F.; Sobrino, C. A review of solar thermal energy storage in beds of particles: Packed and fluidized beds. *Sol. Energy* **2019**, *192*, 193−237.
(40) van Putten, I. C.; van Sint Annaland, M.; Weickert, G. Fluidization behavior in a circulating slugging fluidized bed reactor. Part I: Residence time and residence time distribution of polyethylene solids. *Chem. Eng. Sci.* **2007**, *62*, 2522−2534.
(41) Ergun, S. Fluid Flow Through Packed Columns. *J. Chem. Eng. Prog.* **1952**, *48*, 89−95.
(42) Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Networks* **1989**, *2*, 359−366.
(43) Nair, V.; Hinton, G. E. *Rectified Linear Units Improve Restricted Boltzmann Machines*; Omnipress: Haifa, Israel, 2010; pp 807−814.
(44) Gu, J.; Yin, G.; Huang, P.; Guo, J.; Chen, L. An improved back propagation neural network prediction model for subsurface drip irrigation system. *Comput. Electr. Eng.* **2017**, *60*, 58−65.
(45) André, A.; Laura, T.; Oliver, S.; Thomas, L. Permutation importance: a corrected feature importance measure. *Bioinformatics* **2010**, *26*, 1340−1347.

https://doi.org/10.1021/acs.iecr.1c02307
Ind. Eng. Chem. Res. 2021, 60, 13727−13739