Part 1 Methodology

Combining both quantitative and qualitative analyses, engineering front study applied deep integration of data analysis and expert advice to select a total of 94 engineering research fronts and 96 engineering development fronts in 2018. The number distribution of engineering fronts of nine fields corresponding to the nine academic divisions of the Chinese Academy of Engineering (CAE) is shown in Table 1.

Clarivate Analytics conducted quantitative analysis and obtained 702 candidate engineering research fronts based on the co-citation clustering method and 53 ThemeScape maps of 53 disciplines of nine fields.

In addition, experts of nine fields proposed candidate engineering fronts based on their insights before the original dataset was determined and after data mining was conducted. This was done to supplement potential gaps in the content coverage for each field in quantitative analysis. After several rounds of expert group meetings and questionnaires, approximately 10 engineering research fronts and 10 engineering development fronts were determined. Among them, three key engineering research fronts and three key engineering development fronts were selected for detailed interpretation.

1 Identification of engineering research fronts

In this report, the basic materials of engineering research fronts were extracted in the following two methods. In the first method, *Clarivate Analytics* clustered the top 10% highly cited papers of journal literatures and conference proceedings indexed in the *Web of Science Core Collection* based on the co-citation clustering method to obtain the topics. The second method was based on expert proposal. These two methods yielded 482 candidate engineering research fronts. After expert group meetings and questionnaires, approximately 10 engineering research fronts were selected from each field.

1.1 Production of topics

Clarivate Analytics matched the *Web of Science Core Collection* journal/conference proceedings to the nine CAE academic divisions, including a total of 12 882 journals and 28 626 conference proceedings. Experts manually reviewed the journal/conference proceeding mapping table for each field and finalized the list by making appropriate adjustments with their domain knowledge. Additionally, *Clarivate Analytics* reassigned publications in 67 Multidisciplinary Sciences journals such as *Nature* and *Science* to their most relevant subject area. By using the information found in the cited references of each publication, it was possible, in most cases, to algorithmically reassign them to a subject area. The final publication dataset comprised of articles, reviews, and proceeding papers that were published in matched journals and conference proceedings from the above methods. These

Table 1 Number distribution of engineering fronts of nine fields

Field	Number of engineering research fronts	Number of engineering development fronts
Mechanical & Vehicle Engineering	10	10
Information & Electronic Engineering	10	10
Chemical, Metallurgy & Materials Engineering	12	12
Energy & Mining Engineering	13	14
Civil, Hydraulic & Architecture Engineering	10	10
Environmental & Light Textile Engineering	10	10
Agriculture	10	10
Medicine & Health	9	10
Engineering Management	10	10
Total	94	96

publications were published between January 2012 and 2017, with citation counts to the end of February 2018.

In each field, *Clarivate Analytics* identified publications that were in the world's top 10% by number of citations received, considering the publication year and Journal Subject Category (JSC). Because citation rates differ for different modes of publication, conference proceeding papers and journal papers were treated separately. For each field, *Clarivate Analytics* clustered these top 10% highly cited publications based on co-citation clustering analysis. The first step of quantitative analysis was to obtain clustered topics.

The second step was to select 50 topics for each of the nine fields. To reflect the emerging feature of the fronts, the topics with the average publication year later than 2016 were paid more attention. For topics with the average publication year between 2016 and 2017, 20 clustered topics with no significant overlapping were selected by the order of indicators of the number of core publications, total citations, and the percentage of consistently cited publications. For topics with the average publication year before 2016, 30 clustered topics with no significant overlapping were selected by the order of indicators of the number of core publications, total citations, average publication year of core publications, and the percentage of consistently cited publications. Following this two-step process, 702 topics for nine fields were identified.

1.2 Expert review

Expert insights about engineering fronts are necessary supplements for bibliometric analysis.

Before data mining was conducted, experts from nine fields proposed key engineering research topics based on their expertise and then librarians for intelligence analysis of different subjects transferred the topics to search queries. This was an important component of the original dataset.

After data mining, to supplement potential gaps in the content coverage for each field, experts proposed key words, typical papers, or typical journals of specific subject to support *Clarivate Analytics* to perform custom search.

When reviewing the 702 topics of bibliometric analysis, experts proposed key engineering research topics once again, and librarians for intelligence analysis of different subjects provided data support.

The topics from bibliometric analysis and expert proposal were merged and condensed to produce a total of 482 candidate engineering research fronts.

Finally, after several rounds of expert group meetings and questionnaire, 94 engineering research fronts were selected for nine fields.

2 Identification of engineering development fronts

The basic materials of engineering development fronts were also obtained via two methods. In the first method, *Clarivate Analytics* clustered the top 5000 highly cited patent families from *Derwent Innovation* to obtain a total of 53 ThemeScape maps of nine fields. From these maps, experts obtained potential engineering development fronts. The second method was based on expert proposal. These two methods yielded 415 candidate engineering development fronts. Then, approximately 10 engineering development fronts of every field were selected after several rounds of expert group meetings and questionnaires.

2.1 Production of ThemeScape map

Clarivate Analytics established the mapping relationship between Derwent Manual Code and the specialty division criteria system of the CAE's academic divisions to determine the original patent search scope. Following the experts' manual review, the patent search strategy was constructed for 53 subjects in nine fields. Original data for patent analysis were obtained by searching on *Derwent Innovation*. These patents were published between January 2012 and 2017, with citation counts to the end of February 2018.

With comprehensive consideration to average citation of patent family and technical coverage width, we obtained the top 5000 highly cited patent families of each subject. Based on the semantic similarities of patent texts, 53 ThemeScape maps, which can visually present the distribution of engineering technologies were obtained.

With the assistance of librarians for intelligence analysis of different subjects, experts extracted technology information from ThemeScape maps to obtain candidate engineering development fronts of each subject.

2.2 Expert review

Experts played a crucial role in this study. Before data mining was conducted, experts from nine fields proposed key engineering development topics based on their expertise and then librarians for intelligence analysis of different subjects transferred the topics to the patent search queries. This was an important component of the original dataset.

After data mining, experts proposed key issues once again to supplement potential gaps in the content coverage for each field.

The results from ThemeScape map interpretation and expert proposal were used to produce 415 candidate engineering development fronts. Ninety-six engineering development fronts were obtained through expert group meetings and questionnaires.

Furthermore, from the perspective of future development prospects, 27 key engineering research fronts and 27 key engineering development fronts were selected for detailed interpretation.

3 Terminologies

Citations: The citation count is the number of times that a citation has been recorded for a given publication since it was published.

Publications/Papers: Publications/Papers include substantive peer-reviewed research journal articles, review articles, and conference proceedings.

Highly cited papers: Papers in the top 10% in terms of citation frequency are considered to be highly cited, taking into account the year of publication and Journal Subject Category (JSC).

Topics: Topics are clusters of the top 10% highly cited papers based on co-citation clustering analysis.

Core papers: Core papers are a group of highly cited papers forming a clustered topic.

Citing papers: Citing papers refer to the papers that cite core papers.

Mean publication year: The mean publication year refers to the average number of publication years of all the core papers of a topic.

Consistently cited publications: Publications in the top 10% in terms of citation velocity are considered to be consistently cited, taking into account the year of publication and JSC.

Citation velocity: Citation velocity is a measure of the rate of citation accumulation based on a certain frequency within a set period of time. The citation velocity for each paper is a measure of the rate of citation accumulation per month, i.e., from the month it is published until February 2018. Note that publications that are only cited in the same month when published will have citation velocity of 0.

ThemeScape map: ThemeScape map is a themed panoramic view that can visually reflect the overall outlook of an industry or a technical field. It is a visual presentation in map form obtained by analyzing semantic similarities in patent literature, adopting scientific statistical analysis methods for precision analysis, and sorting of patent text contents.

Technical coverage width: Technical coverage width is measured by the number of Derwent Manual Codes that each patent family belongs to.

Percentage of published papers/patents: When counting the output of core papers/patents of countries/regions or institutions, if one paper/patent is of co-authors from different countries/regions or institutions, each country/region or institution will add one, so the total percentage of all countries/regions or institutions is more than 100%.

Specialty division criteria system of the CAE's academic divisions: It includes the specialized fields covered by the various academic divisions of engineering science and technology. It is determined in accordance with the *Academic Divisions and Specialty Division Criteria of the Chinese Academy of Engineering for Election of Academicians (for Trial Implementation).*