IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 27, 2020, accepted December 12, 2020, date of publication December 21, 2020, date of current version December 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3046294

A Novel Semantic Cohesion Approach for Chinese Airworthiness Regulations: Theory and Application

HAOTIAN NIU[1], CUNBAO MA[1], (Member, IEEE), PEI HAN[2], SIYUAN LI[1], AND QI MA[1]
[1]School of Civil Aviation, Northwestern Polytechnical University, Xi'an 710072, China
[2]Communication Division, 20th Institute, China Electronics Technology Group Corporation, Xi'an 710072, China

Corresponding author: Cunbao Ma (caecmacb@nwpu.edu.cn)

ABSTRACT Airworthiness regulation documents are critical, which contain massive safe constraints for system design and airworthiness certification. The textual information of airworthiness regulations is organized and stored in a hierarchical and scattered form due to the unique provision structure. The linguistic form of legal provision brings the problems of incompleteness semantic information and syntactic structure to each legal clause, which limits the application of Natural Language Processing (NLP) technology in the field of airworthiness safety. In this article, a novel theory of semantic cohesion is proposed for Chinese airworthiness regulations, in which four critical elements are contained, including definition, model, theorem and rules. The definition of attributive directed graph model is proposed to achieve the graphical representation of airworthiness regulation texts. Moreover, the graphical issue of regulation texts is proved in accordance with the characteristic of hierarchical provision structure based on graph theory, and the rules of edge structure construction are provided. Based on the theory, a new node content link algorithm is proposed to achieve the tasks of semantic cohesion and structure conversion for Chinese airworthiness regulation texts. According to the hierarchical structure characteristic of regulation texts, the algorithm can convert the provision structure to the general narrative linguistic form by constructing edges with semantic cohesion relation. As a result, provision sentences with complete semantic information and syntactic structure can be generated. The algorithm has been deployed on the current 119 airworthiness regulation texts to verify the validity and feasibility. The experimental results show that the algorithm achieves efficient performance in the tasks of structure conversion and semantic cohesion, and the algorithm performs reliably with an accuracy rate of 100%. The problem of application limitation for NLP technology to Chinese airworthiness regulation texts has been resolved, which promotes the intelligent and automatic development of airworthiness safety.

INDEX TERMS Airworthiness safety, attributive directed graph, Chinese airworthiness regulations, general linguistic form, hierarchical provision structure, NLP, node content link algorithm, semantic cohesion theory.

I. INTRODUCTION

Airworthiness regulations are essential guidance documents in the field of airworthiness safety, which comprise all necessary safe constraints for Air Transport System (ATS). System safety engineers and other airworthiness engineers should conform to the regulations when designing and maintaining the ATS. Satisfying all the requirements of airworthiness regulations is an essential prerequisite for civil aircraft to the market of civil aviation [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Sun-Yuan Hsieh.

Airworthiness regulation documents, as the most crucial standard of aviation industry, have the characteristic of universality. For all types of aircraft, safe design constraints are contained in the documents, which makes airworthiness regulations comprise massive legal provisions with varied pertinence. In order to achieve airworthiness safety design and airworthiness compliance verification, safety engineers are supposed to select appropriate provisions according to different design tasks and integrate the specific rules into system design as safe constraints [2], [3]. The effective processes of airworthiness safety analysis and design can help eliminate the negative effects brought by the repetition revision of design scheme for successful airworthiness certification [4].

Meanwhile, the problem of increasing cost due to scheme modifications can be avoided [5]. Airworthiness safety analysis and design is critical for ensuring flight safety and reducing aviation accidents [6]–[8].

Airworthiness regulations are organized and stored in texts with the form of legal provision structure. The safe information of system is scattered in the texts because legal provisions are massive, which makes it hard for system safety engineers to acquire useful information quickly and accurately. Therefore, the negative influence is brought to the tasks of airworthiness certification and safety design. With the rapid development of artificial intelligence theory in recent years, NLP technology is proposed and adopted to achieve text analysis and processing in an intelligent and automatic way. The NLP technology combines human linguistics and computer science, which helps researchers acquire useful information from texts and create relevant knowledge architecture conveniently. The technology is of great significance because it can help to improve the effectiveness of communication and control process between humans and machines [9], [10].

For NLP tasks, textual data is the primary research object. Semantic and syntax analysis are carried out to text corpus with different levels, such as word, sentence, document. Multiple tasks can be realized through NLP technology, including machine translation [11], public opinion analysis [12], intelligent question-answering [13], information retrieval [14], sentiment analysis [15] and knowledge graph [16], etc. Airworthiness regulation documents are essential data in the field of airworthiness safety. Moreover, the development of NLP technology with the object of airworthiness regulation texts can promote the automatic and intelligent development of airworthiness certification and safety analysis.

However, in author's knowledge, NLP technology fails to achieve desired results in engineering applications on the foundation of airworthiness regulation texts. Due to the unique textual structure (as well as the hierarchical provision structure of law and regulation texts), the applications are limited. Airworthiness safety is critical with a great deal of documents that can guide the safety design and maintenance of ATS. Moreover, it is urgent to improve the intelligence and automation level in analyzing and processing airworthiness texts, which also promotes the research of airworthiness safety [17]–[19]. However, the unique hierarchical structure of legal provisions limits the application of NLP technology to airworthiness regulation texts. Therefore, it is of significant meaning to develop the discourse analysis approach for the special texts with hierarchical provision structure. Eliminating negative influence caused by deficient sentence components and complicated hierarchical relationships in legal provisions structure, is critical to promote the application of NLP technology for airworthiness safety.

The relevant research work of NLP is mainly based on the texts in general linguistic form, such as news, social dialogues, film reviews, and so on [20], [21]. It is more suitable to implement NLP tasks on these text data because their textual structures are relatively simple. Meanwhile, the semantics expression and sentence structure are complete. The problem of lacking sentence components is almost not existing in the sentences with general linguistic form. However, airworthiness provisions are stored and organized according to hierarchical provision structure, which makes them different from the texts with general linguistic form. For each elementary provision unit, complete semantic structure is not existing with the influence of hierarchical provision structure. Provision structure is the unique text structure that limits the application of NLP technology in the field of airworthiness safety [22], [23].

Linguistics structure expresses the mode of information organization in texts, which affects the implementation effectiveness of NLP tasks. For the text data in general linguistic form, sentence-level data units conform to the text structure in which complete semantic and syntactic structure are ensured. This kind of text structure can help to improve the effectiveness of relevant NLP tasks as prior knowledge [24], [25].

Both syntactic parsing and discourse analysis are important components of text analysis [26], [27]. For syntactic parsing, the sentences with complete semantic components are usually viewed as basic units to analyze dependence relations between words, and the dependence relations can act as linguistic characteristics to improve the effectiveness of NLP tasks. Therefore, it is critical for text structure modeling and analysis to NLP tasks [28]. Many scholars have conducted research on syntactic and discourse structure analysis. Li *et al.* [29] combined dependency parsing with deep learning architecture to achieve Named Entity Recognition (NER) and Relation Extraction (RE) for medical texts, the similar work refers to [30]–[32]. Xue *et al.* [33] designed an intelligent question-answering system based on syntactic structure analysis. The syntactic structure theory was adopted by Dashtipour *et al.* [34] to achieve the sentiment analysis task for Persian social media texts. Zulkarnain and Meziane [35] constructed the text structure model for ultrasonic diagnosis reports based on discourse structure analysis theory, and realized the task of standardized structure processing for this kind of text. The above research work shows that the general linguistic form is helpful for extracting the linguistics characteristic of texts, which is of benefit to NLP tasks. The general linguistic form ensures that each sentence-level unit contains complete semantic information and syntactic structure. Both textual structure processing and syntactic parsing are essential for NLP tasks under the architecture of artificial intelligence, including machine learning and deep learning. However, syntactic parsing can only be used for sentence-level text unit with complete semantic and syntactic information, but not for airworthiness regulations in which semantic information and syntactic structure are incomplete.

For airworthiness regulations, the unique provision structure makes the texts represent hierarchical structure characteristics. Thus, the regulation texts are obviously different from the texts in general linguistic form. As a basic knowledge unit, every single elementary provision is stored in an independent textual row space. Moreover, multiple elementary provisions are organized to form an article-level provision content of airworthiness regulation, which makes it hard to distinguish the boundary of legal sentences [36]. The problem of incomplete semantic and syntactic structure is brought to each elementary provision unit due to the hierarchical structure of airworthiness regulation texts. So, it is difficult to realize syntactic and discourse analysis for the texts. In addition, document-level hierarchical relations are existing among elementary provision units due to the provision structure. Namely, all legal provisions belong to one specific regulation document simultaneously in a paratactic way. The complex hierarchical structure limits the application of relevant NLP tasks to airworthiness regulation texts. For airworthiness regulation texts, both document-level and sentence-level structural characteristics should be considered in NLP tasks, which increases the complexity of textual semantic analysis and structure processing.

In this article, a novel semantic cohesion theory is proposed to guide discourse representation and text modeling for Chinese airworthiness regulations. The complex hierarchical provision structure can be converted to the general linguistic form with complete semantic information and syntactic structure in narrative expression according to provision structure characteristics and hierarchical relations among elementary provision units of airworthiness regulations. The standardized processing method for text structure is achieved. Moreover, the problem of complex discourse structure and incomplete semantic information brought by the hierarchical provision structure of airworthiness regulation texts can be resolved through the proposed approach.

Significant results are acquired by combining graph-theory-based modeling approaches and NLP technology for different textual analysis and processing tasks. Graph theory is often adopted in constructing textual models to achieve syntactic and discourse analysis [37]. In graph theory, node and edge are used to represent the practical issues in a mathematical way, moreover, the theoretical base and mathematics description can be provided for complex process modeling. Graph theory has been widely adopted for model construction in various fields [38], such as social network [39], transportation [40], and semantic network [41], etc. Therefore, it is suitable to achieve the model representation of texts with graph structure in which entity and relation can both be expressed appropriately [42]. Moreover, graph theory can be used to construct the textual structure model not only on sentence-level, but also on document-level.

Semantic analysis and structure processing are essential for understanding textual information so as to achieve NLP tasks [43]. Kim [44] analyzed the importance of graph theory for the construction of textual model, and he pointed out that textual structure models can basically keep consistent for the texts in different languages. For news reports, Uçkan and Karci [45] constructed document-level textual model based on graph theory, and achieved the task of automatic summarization based on the model. Bijari *et al.* [46] used graph theory to construct textual structure model, and realized the task of sentiment analysis for social network texts. The above research work shows that it is suitable to use graph theory to construct the model of texts in general linguistic form. However, for Chinese airworthiness regulation texts with unique hierarchical provision structure, the corresponding text modeling and structure processing approaches are still lacking at present.

In this article, a novel semantic cohesion theory for the hierarchical provision structure of Chinese airworthiness regulations is presented. Based on the theory, a new algorithm of structure conversion and semantic cohesion is provided. The problem of incomplete semantic and syntactic structure caused by complex hierarchical provision structure is abstractly expressed as a graph-theoretic issue (namely, the construction of attribute directed graph model and the association of node contents). Moreover, the tasks of semantic cohesion and structure conversion can be achieved through the node content link algorithm according to the inherent hierarchical relation of airworthiness regulations. Consequently, the special hierarchical provision structure of Chinese airworthiness regulation texts can be converted to the general linguistic form with complete semantic information and syntactic structure, which is convenient for the subsequent textual analysis and processing. The contributions of our study are generalized as follows:

- A novel theory of semantic cohesion for Chinese airworthiness regulations is proposed in which four critical elements are contained, including definition, model, theorem, and rules. The semantic cohesion theory can help to achieve the graphical model representation and textual analysis for Chinese airworthiness regulations.
- A new node content link algorithm is proposed based on the semantic cohesion theory, which helps to achieve the tasks of semantic cohesion and structure conversion for Chinese airworthiness regulations. Through the proposed algorithm, the special hierarchical provision structure of airworthiness regulation texts can be converted to the general linguistic form with complete semantic information and syntactic structure.
- The problem of application limitation for NLP technology to Chinese airworthiness regulation texts caused by the unique hierarchical provision structure is solved through the proposed approach, which can promote the intelligent and automatic development of airworthiness safety.

The rest of paper is organized as follows. The upcoming section presents the problem of structure conversion and semantic cohesion for Chinese airworthiness regulation texts. The following section introduces the theory of semantic cohesion in detail. In a further section, the link algorithm of node content for the hierarchical provision structure of airworthiness regulation texts is introduced. In the penultimate section, experiments followed by discussions are illustrated to verify the validity and effectiveness of the proposed approach. The final section concludes and outlines for future research.

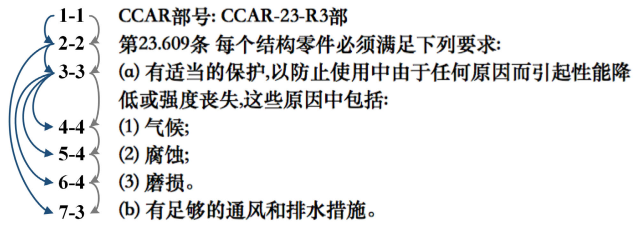
II. PROBLEM MODELING AND DESCRIPTION

A. PRACTICAL PROBLEM DESCRIPTION

The complex hierarchical provision structure of airworthiness regulation texts is the primary problem that limits the application of NLP tasks in the field of airworthiness safety. The unique provision structure is obviously different from the general linguistic form, and this special structure is mainly used for information organization in law and regulation texts. The particular provision structure makes airworthiness regulation texts have the characteristic of hierarchical structure. Furthermore, the characteristic may result in incomplete semantic information and syntactic structure, which affect the application of NLP tasks to law and regulation texts. The main problems brought by the unique hierarchical provision structure of Chinese airworthiness regulations are generalized as follows:

- Provision structure leads to complex hierarchical relations among multiple elementary provision units, while every single elementary provision unit has the problems of incomplete syntactic and semantic structure, and components deficiency. Meanwhile, elementary provision units with the relation of semantic cohesion are separated at different row spaces which are not adjacent.
- Provision structure results in the problem of sentence boundary ambiguity, which makes it difficult to achieve sentence segmentation and syntactic analysis correctly for airworthiness regulation texts.
- Provision structure results in the problem of incoherent text information, which affects the effectiveness of NLP tasks, such as entity relation extraction and knowledge graph construction, etc.

In order to clarify the unique hierarchical provision structure of Chinese airworthiness regulation texts, part of regulations (CCAR-23-R3, Article 23.609) is excerpted and presented in Fig. 1. As shown in Fig. 1, seven different titles are contained in the provision structure, including CCAR, Article, (a), (1), (2), (3), and (b). Each title corresponds to an elementary provision unit stored in an independent row space, respectively. In Fig. 1, the combined number before each elementary provision unit denotes the number of rows (as well as the order of nodes) and the corresponding title level, respectively. The number on the left side of symbol '-' denotes the order of nodes under hierarchical provision structure, and the number on the right side of symbol '-' is adopted to indicate the corresponding title level. Under hierarchical provision structure, text information is organized and represented according to the order of rows, and the arranged sequence is shown in grey arrows in Fig. 1. Because of the hierarchical characteristic brought by multiple title levels, complex structural relations exist in elementary provision units. To ensure complete semantic and syntactic structure, elementary provision units with semantic cohesion relations are supposed to be combined. As shown in Fig. 1, the semantic cohesion relations among elementary provision units are represented in blue arrows. Consequently,



(a)

Serial Number: CCAR-23-R3

Article 23.609 Each structure component must satisfy the following requirements:

(a) appropriate protective measures are set to prevent performance degradation or strength loss, including:

(1) climate

(2) corrosion

(3) abrasion

(b) adequate ventilation and drainage

(b)

FIGURE 1. Hierarchical provision structure of Chinese airworthiness regulations.

in order to eliminate the negative influence caused by the special hierarchical provision structure to the relevant NLP tasks. It is necessary to develop document-level semantic cohesion approaches for the texts with hierarchical provision structure. The problem of structure conversion and semantic cohesion for Chinese airworthiness regulations is defined as follow, namely, connecting elementary provision units with semantic cohesion relation according to hierarchical structure characteristics, and converting the special structure to the general linguistic form with complete semantic information and syntactic structure.

B. PROBLEM MODELING

In this article, the problem of structure conversion and semantic cohesion for elementary provision units is abstractly expressed as a graph-theory issue. The purpose of structure conversion is to transform the unique hierarchical provision structure of law and regulation texts to the general linguistic form. The purpose of semantic cohesion is to solve the problem of incomplete semantic and syntactic structure by connecting elementary provision units with semantic cohesion relations. Therefore, legal provision sentences with complete semantic information and syntactic structure can be achieved.

From the view of graph theory, the problem of semantic cohesion and textual structure conversion for Chinese airworthiness regulations is described as follow, namely, constructing attributive directed graph model for Chinese airworthiness regulation texts according to hierarchical provision structure, and linking node contents based on the graph model. The task is realized by acquiring subgraphs and directed paths with complete semantic information and syntactic structure. Problem modeling and technology framework are represented in Fig. 2. As shown in Fig. 2, four critical steps are needed to form the technology framework of our approach, including the construction of node set for Chinese

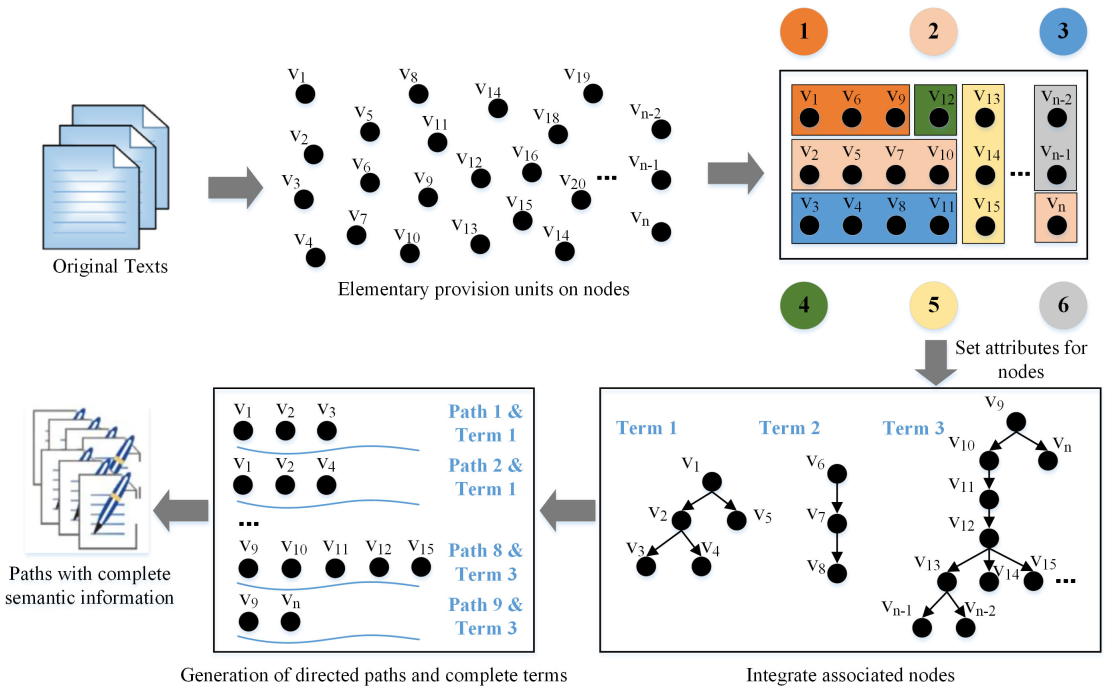


FIGURE 2. Problem description and technology framework.

airworthiness regulations, the construction and assignment of node attribution set, the judgment of edge structure, and the acquisition of directed paths.

In this article, on the foundation of graph theory, the problem of semantic cohesion and structure conversion has been converted to node content link and the construction of attributive directed graph model for Chinese airworthiness regulation texts. With the consideration of hierarchical provision structure, elementary provision units are selected to be the nodes of graph model when constructing attributive directed graph model for airworthiness regulation texts. Because of the unique structure characteristic of law and regulation texts, nodes with semantic cohesion relation are usually separated and located at two different rows which are not adjoining. Consequently, for Chinese airworthiness regulations, the node structure of textual graph model is deterministic while the edge structure is uncertain. Meanwhile, in the special provision structure, attributions are brought to the nodes because of hierarchical multi-level titles.

The problem of structure conversion and semantic cohesion for airworthiness regulation texts is expressed as a graph theory issue. Moreover, the problem can be further expressed as setting judgment mechanism of node contents with the consideration of node ($v, v \in V$) and corresponding attribute characteristic ($vl, vl \in VL$). In order to construct directed edge e and edge set E for the attributive directed graph model of airworthiness regulation texts, it is necessary to determine whether there is an edge structure for different pairs of nodes.

Moreover, the set of directed paths with complete semantic information can be acquired, which is expressed as: $Path = \{Path_1, Path_2, \ldots, Path_n\} = \{(e_1, \ldots, e_a), (e_1, \ldots, e_b), \ldots, (e_m, \ldots, e_n)\}$.

Then, the problem is defined as follow:

$$V(VL(L)) \to E \to Path \tag{1}$$

In order to solve the problem mentioned above, a new theory of semantic cohesion is proposed. And, the concept of attributive directed graph model is defined. Based on the hierarchical characteristic of provision structure, the graphical issue of Chinese airworthiness regulation texts is proved when elementary provision units are taken as nodes. Meanwhile, on the foundation of graph theory, a new link algorithm of node contents is proposed. In the proposed algorithm, the node contents with semantic cohesion relation are linked according to the judgment mechanism (as well as the construction of edge structures). The set of directed paths with complete semantic information and syntactic structure can be generated through the proposed algorithm, which solves the problem of structure conversion and semantic cohesion for Chinese airworthiness regulations.

III. THEORY OF SEMANTIC COHESION

A. DEFINITION OF ATTRIBUTIVE DIRECTED GRAPH

Definition 1 (Attributive Directed Graph): Suppose that there are vertex set $V = \{v_1, v_2, \cdots, v_n\}$ and attribution set $VL = \{vl_1, vl_2, \cdots, vl_n\}$. Each node has at least one specific attribution vl in the attribution set VL, the corresponding set of attribution value is expressed as $L = \{l_1, l_2, \ldots, l_n\}$. Then, a vertex v_i can be denoted as $v_i\ (VL_i(L_i))$. Suppose that there is an edge set $E = \{e_1, e_2, \ldots, e_n\}$, where $e_i = (v_i, v_j) \in E$

indicates that one specific edge structure conforms to the constraints of critical attribution. The parameter ψ denotes the judgment function of edge structure, which is used to judge whether two nodes conform to the constraints of critical attribution. If the vertex set and the edge set can satisfy the following four conditions:

(1) Each node has at least one attribute and corresponding attribute value.

(2) Graphical issue of degree sequence.

(3) Unidirectional connectivity of edge structure.

(4) Judgement function $\psi(v_i, v_j)$: the function is designed to judge whether the edge structure exists between the two nodes.

$$\psi(v_i, v_j) = \begin{cases} 1 & \textit{if the two nodes are connceted} \\ 0 & \textit{otherwise} \end{cases} \tag{2}$$

then, $GL = (V(VL(L), E, \psi)$ is defined as the attributive directed graph, and GL for short.

Definition 2 (Critical Attribution): For the attributive directed graph, the vertex attribution set can be expressed as $VL = \{vl_1, vl_2, \ldots, vl_n\}$, and the critical attribution indicates one specific attribution which can completely distinguish whether the edge structure is existing.

Theorem 1: (Graphical of degree sequence) Suppose that d is a nonnegative integer sequence, and $d = (d_1, d_2, \ldots, d_n)$. The sequence is graphical if and only if the sum of all vertex-degrees ($z = \sum_{i=1}^{n} d_i$) is an even number.

B. CONSTRUCTION APPROACH OF TEXTUAL GRAPH MODEL

It is necessary to combine the hierarchical structure characteristic in constructing the attributive directed graph model of Chinese airworthiness regulation texts. Specifically, elementary provision units are supposed to be selected as the nodes of graph model, and the inherent hierarchical relation of provision structure is viewed as the critical attribution. For airworthiness regulation texts, the edge structures of textual graph model are uncertain, which depend on the hierarchical relation of nodes. Therefore, in this section, the proof of graphical issue for Chinese airworthiness regulation texts is provided by combining with the hierarchical characteristic of provision structure.

In graph theory, graph is a mathematical model which is used to express the real things in an abstract way. The graph model comprises of vertex set and edge set, which denote entities and corresponding relations, respectively. Specially, node selection patterns are varied in constructing textual graph model. For NLP tasks, words and sentences are usually adopted as nodes in textual graph model, and edge structures exist between two adjacent nodes. However, with the consideration of the particularity of airworthiness regulation texts, it is necessary to select nodes and edges based on hierarchical provision structure when constructing textual graph model. The construction approach of attributive directed graph for Chinese airworthiness regulation texts is represented in Fig. 3.

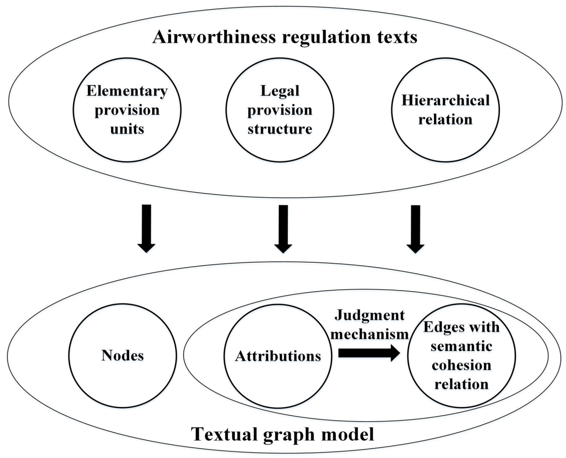


FIGURE 3. Construction approach of attributive directed graph model for airworthiness regulation texts.

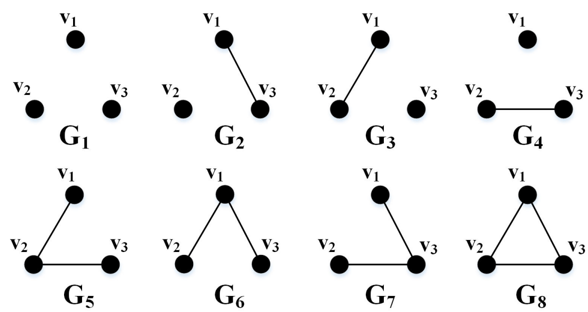


FIGURE 4. The uncertainty of edge structure in constructing attributive directed graph.

With the consideration of provision structure, elementary provision units of airworthiness regulations are selected as nodes of attributive directed graph, and hierarchical relations brought by multi-level titles are adopted as node attributions. Then, judgements are carried out for each two adjacent nodes to find out whether semantic relation is existing between the two nodes. The directed paths with complete semantic information are formed by constructing edge structures between the nodes with semantic cohesion relation.

In the attributive directed graph model of Chinese airworthiness regulation texts, the vertex structure is deterministic, and the edge structure is uncertain. The edge structure depends on the attribution of nodes and the hierarchical relation of provision structure. For multiple nodes, edge structures may exist in one or more than one pairs of nodes. The uncertainty of edge structures in constructing attributive directed graph model for Chinese airworthiness regulation texts is shown in Fig. 4.

C. PROOF OF CRITICAL THEOREM

The graph-based textual model usually builds on the text data in general linguistic form, and takes words as nodes to achieve model representation. However, airworthiness regulation texts have the special linguistic structure. Therefore, in our study, elementary provision units are selected as vertex to construct the attributive directed graph model of airworthiness regulation texts.

Because of the unique hierarchical provision structure, airworthiness regulation texts are different from the text data in general linguistic form. Therefore, based on graph theory, the graphical issue of airworthiness regulation texts is proved with the combination of hierarchical characteristic of provision structure. It is determined that whether the attributive directed graph model of airworthiness regulation texts can be constructed when elementary provision units are taken as vertex and edge structures are constructed between the nodes with semantic relation.

In graph theory, a simple graph G is defined as an ordered set (V, E), and $G = (V, E)$. The parameters V and E denote vertex set and edge set, respectively. The vertex set V is expressed as follow:

$$V = (v_1, v_2, \ldots, v_i, \ldots, v_n) \tag{3}$$

where parameter v_i indicates one certain node in V, and $v_i \in V$.

In graph model, the definition of edge set is expressed as follow:

$$E = \{(v_1, v_2) | \forall v_1, v_2 \in V, v_1 \neq v_2\} \tag{4}$$

where (v_1, v_2) denotes an edge structure of graph G.

In attributive directed graph, edge structure exists in the nodes v_1 and v_2 which have semantic cohesion relation. Because of the unique hierarchical provision structure, the airworthiness regulation texts are different from the texts with general linguistic form. Therefore, it is essential to combine the hierarchical characteristic of provision structure when developing structure conversion and semantic cohesion for this type of text. Namely, elementary provision units are supposed to be selected as vertex in constructing the textual graph model of airworthiness regulations. Meanwhile, provision structure makes each elementary provision unit has the characteristic of hierarchical attribution. The model of attributive directed graph is proposed to solve the problem of structure conversion and semantic cohesion for Chinese airworthiness regulations. The structure characteristic of airworthiness regulation texts is analyzed, and node contents with semantic cohesion relation are connected to form edge structures. Then, the attributive directed graph model of Chinese airworthiness regulation texts can be constructed successfully. The hierarchical provision structure of airworthiness regulation texts makes each elementary provision unit have a specific and deterministic title level. According to different title levels, every single elementary provision unit has been arranged with a specific attribution value corresponding to its title.

Therefore, the vertex set of attributive directed graph can be further expressed as the combination of vertex V, attribution VL, and the corresponding attribution value L (namely the title level), as shown below:

$$V(VL(L)) = \{(v_1, l_1), (v_2, l_2), \ldots, (v_i, l_i), \ldots, (v_n, l_n)\} \tag{5}$$

where parameters v_i and l_i denote the content of node i and the corresponding attribution value under hierarchical provision structure, respectively.

The parameter GL is used to indicate the attributive directed graph model of airworthiness regulation texts with the vertex set of elementary provision units. The attributive directed graph is shown as follow:

$$GL = (V(VL(L), E, \psi) \tag{6}$$

where parameter VL denotes the critical attribution set of attributive directed graph. The attributions are corresponding to the multi-level titles of textual graph model. The parameter L indicates the set of attribution values. In textual graph model of airworthiness regulations, the specific attribution value of node is corresponding to the title level. The parameter E denotes the set of edge structure with semantic cohesion relation, and the parameter ψ is the judgment function of semantic cohesion and node contents link.

Node content link is the crucial step in constructing attributive directed graph model for airworthiness regulation texts. Moreover, it is also the basis to realize structure conversion and semantic cohesion, which helps convert the hierarchical provision structure to the general linguistic form with complete semantic information and syntactic structure. Furthermore, the directed paths which represent complete semantic information are generated by constructing edges between the nodes with semantic cohesion relation. The process is described as follow:

$$T \rightarrow GL(V(VL(L)), E, \psi) \rightarrow Path \tag{7}$$

where T denotes airworthiness regulation texts, and Path indicates directed paths with complete semantic information and syntactic structure.

For the problem of structure conversion and semantic cohesion for airworthiness regulation texts, textual graph model can neither be created on sentence-level nor word-level, and it should be created on the level of elementary provision unit. Namely, for each airworthiness regulation text, vertex set V is comprised of all elementary provision units, and each node v_i indicates an elementary provision unit, $v_i \in V$. The edge set E is composed of node pairs with semantic cohesion relation, which can form the directed paths with complete semantic information.

Note that in this article, the graphical issue for Chinese airworthiness regulation texts is proved while considering the uncertain edge structures of textual graph model. Namely, elementary provision units are stored sequentially in row spaces of text according to hierarchical provision structure, but not organized conforming to semantic relations. Therefore, edge structures with semantic relations are uncertain. For Chinese airworthiness regulation texts, the initial mode of information organization and the desired semantic cohesion relation are both represented in Fig. 1. Hierarchical dependency relations among elementary provision units are brought due to multi-level titles in airworthiness regulation texts. Simultaneously, the particular provision structure of airworthiness regulations makes the vertex-degrees (including in-degrees and out-degrees) on different title levels show different characteristics. Moreover, through analyzing the hierarchical provision structure, a common structural characteristic of law and regulation texts is generalized in this article, which is represented as property 1.

Property 1: The provision structure of law and regulation texts has the characteristic of directed branching.

Explanation: Provision structure is the unique information organization mode for law and regulation texts, which enables every single elementary provision unit can derive several subordinate units downwardly, but can only attach to a certain superior unit upwardly.

In our study, the characteristics of vertex-degrees (including in-degrees and out-degrees) on different title-levels are analyzed according to property 1. The hierarchical structure of airworthiness regulation texts can be divided into three different parts in total, including top-level, middle-level, and bottom-level. The top-level nodes denote the elementary provision units with the highest title level which indicates CCAR serial number. The bottom-level nodes indicate leaf nodes in attributive directed graph. These nodes are all terminals without any child nodes, and all of them are located at the end of directed paths with complete semantic information. Except for top-level and bottom-level nodes, the rest of the nodes are defined as middle-level nodes. Note that, in attributive directed graph model, the nodes with different hierarchical structures have different degree characteristics. Specifically, top-level nodes only have out-degrees, but not have in-degrees. Meanwhile, bottom-level nodes only have in-degrees but not have out-degrees. For the nodes located in middle level, they have both in-degrees and out-degrees.

Through proving the graphical issue of degree sequences, the feasibility for constructing the attributive directed graph of Chinese airworthiness regulation texts while taking elementary provision units as nodes, can be proved. For each airworthiness regulation text, the hierarchical structure is different with others, and the composition of edge structures is uncertain. Therefore, the specific value of degree sequence for one certain airworthiness regulation text, can hardly be acquired.

In this article, for Chinese airworthiness regulation texts, the graphical issue of degree sequence has been proved while considering the characteristic of hierarchical provision structure. The provision structure can be divided into six different levels (including Level I-Level VI) which express different title levels of Chinese airworthiness regulations, respectively. The attributive directed graph model of airworthiness regulation texts is represented in Fig. 5.

The highest title level in the provision structure of airworthiness regulation texts is six. The parameter denotes the set comprised of all title levels (as well as the node attribution value of attributive directed graph model) corresponding to

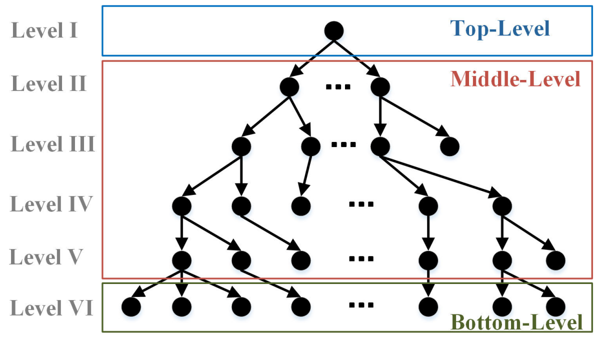


FIGURE 5. The attributive directed graph model of airworthiness regulation text.

each elementary provision unit (as well as the node of attributive directed graph model) under the hierarchical provision structure of airworthiness regulations, as shown below:

$$L = (1, 2, 3, 4, 5, 6) \tag{8}$$

For each airworthiness regulation text, degree sequence can be divided into three different parts according to the characteristic of vertex-degrees, including top-level, middle-level, and bottom-level. The degree sequence of every single airworthiness regulation text is represented as follow:

$$D = \{D_{Top}, D_{Middle}, D_{Bottom}\} \tag{9}$$

where parameter D denotes the degree sequence of airworthiness regulation text, and parameter D_{Top} indicates the degree sequence of top-level nodes. The parameter D_{Middle} means the degree sequence of middle-level nodes, and the parameter D_{Bottom} is used to denote the degree sequence of bottom-level nodes.

For each airworthiness regulation text, the total number of elementary provision units corresponding to each title level i is assumed to be n_i. Namely, the parameter n_i indicates the total number of elementary provision units with title-level i in attributive directed graph model. For airworthiness regulation texts with I-type provision structure, the parameter N denotes the set which is comprised by the total number of elementary provision units with different title-levels, as shown below:

$$N = \{n_1, n_2, n_3, n_4, n_5, n_6\} \tag{10}$$

Note that, for each airworthiness regulation text, the total number of elementary provision units with title-level 1 is 1 under hierarchical provision structure, namely $n_1 = 1$.

The characteristics of vertex-degrees are different corresponding to title levels when taking elementary provision units as nodes to construct textual graph model. As shown in property 1, the nodes with title-level 1 (namely the nodes on top-level) only have out-degrees, but not have in-degrees. Because of the property of directed branching, the out-degrees number of the node with title-level 1 is equal to the total number of in-degrees for the nodes with title-level 2. Moreover, the total number of in-degrees for the nodes with title-level 2 is equal to the total number of nodes with title-level 2. Consequently, the degree sequence for the node with top-level, is shown as below:

$$D_{Top} = d_1^+ = d_2^- = n_2 \tag{11}$$

where parameter d_1^+ denotes the out-degrees number for the node with title-level 1 under hierarchical provision structure. The parameter d_2^- indicates the in-degrees number for the nodes with title-level 2. The parameter n_2 means the total number of nodes with title-level 2.

For middle-level nodes, the total number of in-degrees is equal to the total number of out-degrees for the nodes with superior title level. Meanwhile, it is also equal to the total number of nodes with the current title level. For middle-level nodes, the total number of out-degrees is equal to the total number of in-degrees for the nodes which are subordinate to the current level, and it is also equal to the total number of subordinate nodes. The degree sequence of middle-level nodes can be expressed as follow:

$$\{\exists i, 1 < i \leq 5, d_i \in D_{Middle} :: \begin{matrix} d_i^- = d_{i-1}^+ = n_i \\ d_i^+ = d_{i+1}^- = n_{i+1} \end{matrix}\} \tag{12}$$

where parameter d_i denotes the total number of vertex-degrees, namely the sum of vertex-degrees (including in-degrees and out-degrees) corresponding to the nodes with title-level i under hierarchical provision structure. The parameter d_i^- indicates the total number of in-degrees for the nodes with level i, and the parameter d_i^+ indicates the total number of out-degrees for the nodes with level i. The parameter d_{i-1}^+ indicates the total number of out-degrees for the nodes with level $i-1$, and the parameter d_{i+1}^- indicates the total number of in-degrees for the nodes with level $i+1$. The parameters n_i and n_{i+1} denote the total number of nodes with level i and the total number of nodes with level $i+1$, respectively.

For bottom-level nodes, they only have in-degrees, but not have out-degrees. The total number of in-degrees for bottom-level nodes is equal to the total number of out-degrees corresponding to the nodes with superior level, which also equals to the total number of bottom-level nodes. The degree sequence of bottom-level nodes is shown as below:

$$D_{Bottom} = d_m^- = d_{m-1}^+ = n_m \tag{13}$$

where parameter m denotes the highest title level in current airworthiness regulation text. The parameter d_m^- denotes the total number of in-degrees for the nodes with level m, and the parameter d_{m-1}^+ indicates the total number of out-degrees for the nodes with level $m-1$. The parameter n_m indicates the total number of nodes with level m.

In this article, elementary provision units are selected as nodes to construct the attributive directed graph model for Chinese airworthiness regulation texts. The degree sequence of textual attributive directed graph is represented as follow:

$$\begin{aligned} D &= \{D_{Top}, D_{Middle}, D_{Bottom}\} \\ &= \{D_1, D_2, \ldots, D_m\} \\ &= \{d_1^+, d_2^- + d_2^+, \ldots, d_{m-1}^- + d_{m-1}^+, d_m^-\} \end{aligned} \tag{14}$$

After converting the degree sequence of textual attributive directed graph to the number of nodes on different levels, the degree sequence of textual graph model can be further expressed as follow:

$$D = \{n_2, n_2 + n_3, \ldots, n_i + n_{i+1}, n_m\} \tag{15}$$

where parameter m denotes the highest title level of airworthiness regulation text.

Sum over all elements in the degree sequence (as shown in (15)), the result is represented as follow:

$$\begin{aligned} z &= \sum_{i=1}^{n} d_i = n_2 + n_2 + n_3 + \cdots + n_i + n_{i+1} + \cdots + n_m \\ &= 2(n_2 + n_3 + \cdots + n_m) \end{aligned} \tag{16}$$

where parameter n_i must not be all zero. Therefore, for the degree sequence of attributive directed graph of Chinese airworthiness regulation texts, the sum of the vertex-degrees $z = \sum_{i=1}^{n} d_i$ is determined to be an even number.

With the consideration of hierarchical provision structure for airworthiness regulation texts, the degree sequence of textual graph model which takes elementary provision units as nodes, can satisfy the condition of graphical in theorem 1. In other words, the degree sequence is graphical. Therefore, the attributive directed graph model can be constructed successfully by selecting elementary provision units as nodes and taking pairs of nodes with semantic cohesion relations as edges. Consequently, the graphical issue of airworthiness regulation texts is proved.

Moreover, according to property 1 (namely, the property of directed branching), there is no circle structure between any pairs of nodes but only one directed edge. The graph model of Chinese airworthiness regulation text which takes elementary provision units as nodes, satisfies the definition of unidirectional connectivity. Namely, the graph model of airworthiness regulation text is unidirectional connectivity.

Consequently, it is illustrated that the attributive directed graph model of airworthiness regulation texts can be constructed while taking elementary provision units as nodes and creating edges between the nodes with semantic cohesion relation. Moreover, the proof of graphic issue provides theoretical base for the node content link algorithm.

D. THE RULES OF EDGE STRUCTURE CONSTRUCTION

In order to solve the problem of structure conversion and semantic cohesion for Chinese airworthiness regulations, the concept of attributive directed graph is proposed. Then, the problem of semantic cohesion is transformed to the graph theory issue of linking node contents and constructing attributive directed graph model. As mentioned above, it is deterministic that elementary provision units are taken as nodes, and edge structures are uncertain. Thus, the problem of constructing attributive directed graph model can essentially be interpreted as the reconstruction of edge structures, as well as creating edge structure between nodes with semantic cohesion relation. Moreover, the attributive directed graph of airworthiness regulation texts and the directed paths with complete semantic information and syntactic structure can be successfully achieved.

The parameter L is used to denote the set of attribution value (namely, the set comprises of title levels corresponding to nodes). The hierarchy transfer relation among elementary provision units can be abstractly expressed as the change of node attribution values $l, l \in L$. The elementary provision units are organized and stored sequentially in the row spaces of airworthiness regulation texts. For any two adjacent nodes which are represented sequentially, the judgment mechanism of semantic cohesion is carried out to determine whether edge structure should be constructed between the nodes, as shown below:

$$\{\forall v_i, v_{i-1} \in V, 1 < i < n :: \psi(v_i, v_{i-1}) \in [0, 1]\} \quad (17)$$

where parameters v_i and v_{i-1} denote two adjacent nodes under the provision structure of airworthiness regulation texts. The parameter n indicates the maximum number of nodes under hierarchical provision structure. The Boolean function ψ is used to judge whether the semantic cohesion relation is existing between the two nodes. The function $\psi(v_i, v_{i-1}) = 1$ denotes that the edge structure is existing between the pair of nodes, (v_i, v_{i-1}). On the contrary, there is no semantic cohesion relation between the two nodes.

The judgment function of edge structure in constructing the attributive directed graph model of Chinese airworthiness regulation texts, is shown as follow:

$$F = \Delta l = (l_i - l_{i-1}), \quad \forall l \in L, \ 1 < i < n \quad (18)$$

where parameters v_i and v_{i-1} denote two different nodes stored in adjacent row spaces under hierarchical provision structure. The parameters l_i and l_{i-1} indicate the attribution values corresponding to nodes v_i and v_{i-1}, as well as title levels of the two nodes. The function $F = \Delta l$ denotes the difference between the attribution values of the two adjacent nodes, which is used to control the hierarchy transfer process in the task of semantic cohesion.

The function $F = \Delta l$ has three possible values, which correspond to three different hierarchy transfer processes, respectively. The different value conditions of the function corresponding to the construction mechanism of edge structures are represented and discussed as follows:

$$\begin{aligned} F = (l_i - l_{i-1}) &\rightarrow \begin{cases} > 0, & \psi(v_i, v_{i-1}) = 1 \\ = 0, & \psi(v_i, v_{i-1}) = 0 \\ < 0, & \psi(v_i, v_{i-1}) = 0 \end{cases} \\ &\rightarrow \begin{cases} \psi(v_i, v_{i-1}) = 1 \\ \psi(v_i, v_{i-1}) = 0 \end{cases} \\ &\rightarrow \begin{cases} e(v_i, v_{i-1}), & \mathit{Active} \\ e(v_i, v_{i-1}), & \mathit{Inactive} \end{cases} \end{aligned} \quad (19)$$

Explanations:

Case 1: if $F > 0$, then $\psi(v_i, v_{i-1}) = 1$. It means that semantic cohesion relation exists between node v_i and node v_{i-1}, and v_i is the subordinate node of v_{i-1}. Therefore, edge structure should be created between the pair of nodes. The edge structure is expressed as $e(v_{i-1}, v_i)$. The two nodes are supposed to be connected to form the directed path, and the corresponding elementary provision units have the relation of semantic cohesion.

Case 2: if $F = 0$, then $\psi(v_i, v_{i-1}) = 0$. It means that semantic cohesion relation does not exist between node v_i and node v_{i-1}, namely, v_i and v_{i-1} have the same attribution value. Therefore, the current node v_i is supposed to be attached to another node which is subordinate to v_i (with the attribution value of $l_i - 1$) in the list, the node can be expressed as $(v_f, l_i - 1)$. The node v_f indicates the superior node of v_i, and the attribute value of node v_f is $l_i - 1$. Therefore, the node $(v_f, l_i - 1)$ and the node (v_i, l_i) are supposed to be connected to form the directed path, and the corresponding elementary provision units have the relation of semantic cohesion. The edge structure with semantic cohesion relation should be created between the two nodes v_f and v_i, which can be expressed as $e(v_f, v_i)$.

Case 3: if $F < 0$, then $\psi(v_i, v_{i-1}) = 0$. Similar with case 2, the edge structure with semantic cohesion relation can be expressed as $e(v_f, v_i)$. It means that semantic cohesion relation does not exist between node v_i and node v_{i-1}, and v_i is the superior node of v_{i-1}. Therefore, the current node v_i is supposed to be attached to another node which is subordinate to v_i (with the attribution value of $l_i - 1$) in the list, and the node can be expressed as $(v_f, l_i - 1)$. The node v_f indicates the superior node of v_i, and the attribution value of node v_f is $l_i - 1$. Therefore, the node $(v_f, l_i - 1)$ and the node (v_i, l_i) are supposed to be connected to form the directed path, and the corresponding elementary provision units have the relation of semantic cohesion. The edge structure with semantic cohesion relation should be created between the two nodes v_f and v_i.

Consequently, the construction mechanism of edge structure with complete semantic cohesion relation is generalized and represented as follow:

$$\begin{aligned} &\{\exists(v_i, l_i), \exists(v_{i-1}, l_{i-1}), \exists(v_f, l_i - 1), (v_{i-1}, v_i, v_f) \\ &\quad \in V, (l_{i-1}, l_i, l_i - 1) \in L, v_i \neq v_{i-1}\} :: \\ &\left\{ \begin{array}{l} l_i > l_{i-1}, \psi(v_i, v_{i-1}) = 1 \\ l_i = l_{i-1}, \psi(v_i, v_{i-1}) = 0 \\ l_i < l_{i-1}, \psi(v_i, v_{i-1}) = 0 \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} e(v_{i-1}, v_i) \\ e((v_f, l_i - 1), v_i) \\ e((v_f, l_i - 1), v_i) \end{array} \right\} \rightarrow E \end{aligned} \quad (20)$$

According to different cases, the construction mechanism of edge structure with semantic cohesion relation for the attributive directed graph, is represented in Fig. 6.

The attributive directed graph model of airworthiness regulation texts can be achieved by constructing edge structures with semantic cohesion relation. Moreover, the task of structure conversion for regulation texts is realized, and directed path refers to the sequence of nodes in attributive directed graph model. The directed paths with complete semantic information are acquired by combining nodes with semantic

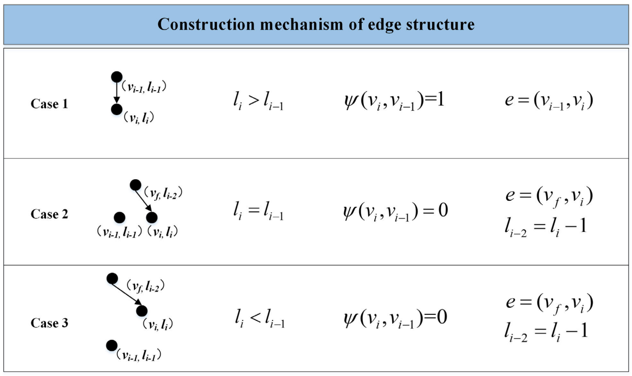


FIGURE 6. Construction mechanism of edge structure with semantic cohesion relation.

cohesion relation, as shown below:

$$\begin{aligned} E &= \{e_1, e_2, \ldots, e_n\} \rightarrow Path \\ &= \{Path_1, Path_2, \ldots, Path_n\} \\ &= \{(e_1, \ldots, e_i), (e_1, \ldots, e_j), \ldots, (e_m, \ldots, e_k)\} \end{aligned} \quad (21)$$

where *Path* denotes the set of directed paths with complete semantic information in the attributive directed graph model of airworthiness regulation texts. The parameters e and E indicate one certain directed edge structure and the set of directed edges in graph model, respectively.

IV. NODE CONTENT LINK ALGORITHM

In this article, a new node content link algorithm is proposed based on the attributive directed graph model of Chinese airworthiness regulation texts. The purpose of the algorithm is to figure out the nodes with semantic cohesion relation by analyzing the hierarchical characteristic of provision structure and create edges between the nodes which are semantic cohesion related. Moreover, the directed paths which can express complete semantic information are achieved. The practical problem in constructing the attribute directed graph model of Chinese airworthiness regulation texts, is to reconstruct directed edges between the nodes with semantic cohesion relation based on hierarchical provision structure. Then, the original textual graph model is converted to a new textual graph model (as well as attributive directed graph). The construction approach of attributive directed graph model and the key steps of node content link algorithm are represented in Fig. 7. Meanwhile, the application object in Fig. 7 is the referenced airworthiness regulations shown in Fig. 1.

In order to construct the attributive directed graph model of airworthiness regulation texts, a new node content link algorithm is proposed. In the algorithm, the hierarchical characteristic of provision structure is analyzed to set corresponding attribution value for each node. For any two adjacent nodes (as well as elementary provision units) organized and stored sequentially in original airworthiness regulation texts, the judgment mechanism of edge structure is carried out. If the dependence relationship exists between the two nodes, edge structure is supposed to be created. On the contrary, if there is no dependence relationship between the two nodes, edge structure should be created between the current node and other one which has semantic cohesion relation with the current node. For Chinese airworthiness regulation texts, both attributive directed graph model and directed path set (namely, the set comprises of directed paths with complete semantic information and syntactic structure) can be achieved through the proposed node content link algorithm.

The flow of the proposed node content link algorithm is represented in Algorithm 1.

Algorithm 1 Link Algorithm for Node Content

Initialization: Chinese airworthiness regulation texts, T
Output: Processed texts with provision sentences, T'
1: Design rule template for hierarchical provision structure through regular expressions
for each line in T **do**
2: Set elementary provision unit as node v_i
3: Identify title level, and arrange corresponding attribute value l_i for node v_i
judgements:
if $F > 0$, $\psi(v_i, v_{i-1}) = 1$ **then**
4: Construct directed edge structure (v_{i-1}, v_i) between nodes v_{i-1} and v_i
elseif $F = 0$, $\psi(v_i, v_{i-1}) = 0$ **then**
5: Output the current directed path $Path_i$ in list, and clear the list
6: Construct list according to node v_i, and create edge structure between node v_i and the node with attribution value of $l_i - 1$
elseif $F < 0$, $\psi(v_i, v_{i-1}) = 0$ **then**
7: Output the current directed path $Path_i$ in list, and clear the list
8: Construct list according to node v_i, and create edge structure between node v_i and the node with the attribution value of $l_i - 1$
end if
end for
9: Create and output directed path according to the last node content in T
10: $T \rightarrow T'$

V. EXPERIMENT AND ANALYSIS

In this article, a novel node content link algorithm is proposed with the consideration of hierarchical provision structure of Chinese airworthiness regulations. The purpose of the algorithm is to transform the unique structure of regulation texts to the general linguistic form combined with the inherent characteristic of multi-level titles. Furthermore, through the proposed algorithm, document-level airworthiness regulation texts are converted to sentence-level text data with complete semantic information and syntactic structure, which helps to promote the application of NLP technology to airworthiness safety. With the consideration of hierarchical structure

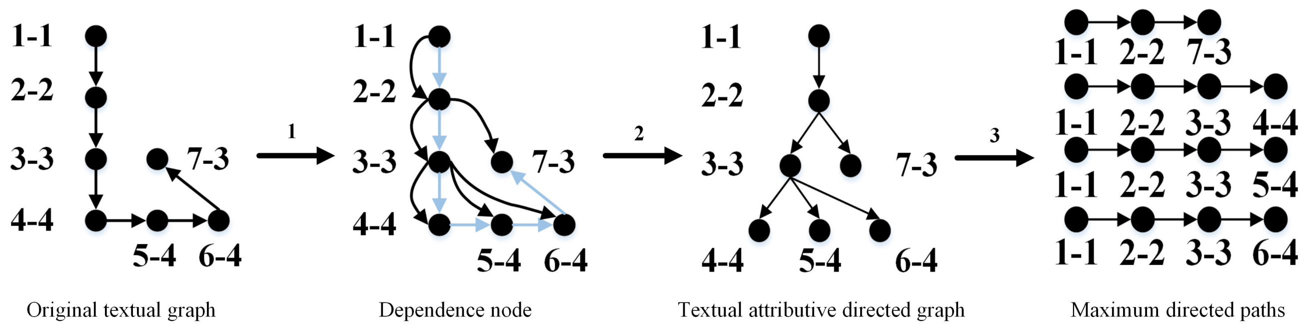


FIGURE 7. The construction approach of attributive directed graph and the critical steps of node content link algorithm (corresponding to the referenced airworthiness regulation in Fig. 1).

characteristics, two sets of templates are generalized and proposed to extract node content (elementary provision unit) and arrange corresponding attribution value (title level) for each node. The current 119 airworthiness regulation texts in force are downloaded as experimental objects to analysis and verify the feasibility and effectiveness of the node content link algorithm.

A. STRUCTURE CHARACTERISTIC ANALYSIS OF AIRWORTHINESS REGULATION TEXTS

In order to verify the feasibility and effectiveness of the node content link algorithm, in total 119 airworthiness regulation texts are obtained from the official website of China Civil Aviation Administration (CCAA). Furthermore, the structure characteristics of airworthiness regulation texts are analyzed and generalized. Analysis results show that there are two different types of hierarchical provision structures in Chinese airworthiness regulation texts. Meanwhile, the two types of structure characteristics are represented with different multi-title forms. Structure analysis and statistics are implemented to Chinese airworthiness regulation texts with different hierarchical structures, statistical results are represented in Table. 1. In this article, airworthiness regulation texts in the first type of format are called as I-type provision structure (32 regulation texts in total), in which six different title levels are contained and the highest title level is 6. The content in textual row is defined as an elementary provision unit. For the airworthiness regulation texts with I-type provision structure, the maximum value and minimum value for the number of elementary provision units in a single airworthiness regulation text are 3215 and 46, respectively. Moreover, the average number of elementary provision units in a single airworthiness regulation text is 897.

In this article, airworthiness regulation texts in the second type of format, are called II-type provision structure (87 regulation texts in total). There are six different title levels contained in the airworthiness regulation texts with II-type provision structure, as well as the highest level is 5. For the airworthiness regulation texts with II-type provision structure, the maximum value and minimum value for the number of elementary provision units in a single airworthiness regulation text are 1826 and 9, respectively. The average number of elementary provision units in a single airworthiness regulation text is 153.

TABLE 1. The statistical results of airworthiness regulation texts.

ARs	Num	HL	Max	Min	Aver
I-type structure	32	6	3215	46	897
II-type structure	87	5	1826	9	153

ARs=Airworthiness regulation texts; Num=Numbers; HL=Highest level in provision structure; Mx=Maximum of elementary provision units in single regulation text; Min=Minimum of elementary provision units in single regulation text; Aver=Average number of elementary provision units in single regulation text.

Therefore, the hierarchical structure of airworthiness regulation texts with I-type provision structure is much more complex than the regulation texts with II-type provision structure. The number of elementary provision units for airworthiness regulation texts with II-type provision structure is much higher than the I-type airworthiness regulation texts. For the airworthiness regulation texts with two different types of hierarchical structures (I-type and II-type), the rule templates are designed to extract elementary provision units as nodes and arrange corresponding attribution value for each node.

According to the two different structure characteristics of Chinese airworthiness regulation texts, the corresponding templates of node extraction are designed. Through the templates, elementary provision units are extracted as nodes, and the corresponding title level is arranged to each node as attribution value.

For airworthiness regulation texts with I-type provision structure, the specific structure characteristics (the highest title level is 6) are shown in Table. 2. For Chinese airworthiness regulations, the hierarchical provision structure is a unique information organization mode which is different from the general linguistic form. Multi-level titles are adopted to represent the hierarchical structure characteristic, including CCAR, Article, Paragraph, Sub-paragraph, Item, and Sub-item. Among them, the extraction rules for node content with title level 1 (CCAR) and the node content with title level 2 (Article) are designed according to character features. The node contents with title levels of 3 to 6 (namely Para-

TABLE 2. Hierarchical characteristics of airworthiness regulations with I-type provision structure.

Hierarchical structure	Title level	Hierarchical structure feature	Example
CCAR	1	character features	CCAR-23-R3
Article	2	character features	第 61.13 条
Paragraph	3	(lowercase letter)	(a)
Sub-paragraph	4	(Arabic number)	(2)
Item	5	(Roman number)	(IV)
Sub-item	6	(Capital letter)	(A)

graph, Sub-paragraph, Item, Sub-item) also have different linguistic features, and the corresponding extraction rules can be designed according to the features. As shown in Table. 2, the structure feature of title level 3 is represented by the linguistic form of lowercase English letter with parentheses (for example (a)). The structure features of title level 4 to title level 6 are represented by linguistic forms of Arabic numerals, Roman numerals, Capital English letter with parentheses, respectively. Meanwhile, there are space symbols existed between titles and elementary provision units. Therefore, based on the structure characteristics, six different extraction rules are designed as the template to extract node and assign corresponding attribution value according to the linguistic features of Chinese airworthiness regulations with I-type provision structure.

For airworthiness regulation texts with II-type provision structure, the specific features of hierarchical provision structure are represented in Table. 3. As shown in Table. 3, the highest title level for this type of text is 5. Among them, similar with I-type provision structure, the extraction rules for the node contents with title level 1 and the node contents with title level 2 are designed according to character features, respectively. The structure feature of title level 3 is represented by the linguistic form of Chinese number with parentheses. The structure feature of title level 4 is represented by the linguistic form of Arabic number with the symbol '.', and the structure feature of title level 5 is represented by the linguistic form of Arabic number with parentheses (for example (1)). Meanwhile, space symbols are existing between titles and elementary provision units. For Chinese airworthiness regulations with II-type provision structure, based on the structure features, five different rules are designed as the rule template to extract node and assign corresponding attribution value for each node.

Consequently, two sets of templates are formed for Chinese airworthiness regulations corresponding to the two different types of provision structures. Based on the rule templates, node content extraction and attribution value allocation can be achieved.

B. DATA PREPROCESSING AND POST-PROCESSING

In this article, a new node content link algorithm is proposed for Chinese airworthiness regulation texts. The purpose of the algorithm is to realize semantic cohesion for

TABLE 3. Hierarchical characteristics of airworthiness regulations with II-type provision structure.

Hierarchical structure	Title level	Hierarchical structure feature	Example
CCAR	1	character features	CCAR-229-R1
Article	2	character features	第一条
Paragraph	3	(Chinese number)	(一)
Sub-paragraph	4	Arabic number.	1.
Item	5	(Arabic number)	(1)

elementary provision units of Chinese airworthiness regulation texts according to the hierarchical characteristics of provision structure. Therefore, the unique text structure can be converted to the general linguistic form with complete semantic information and syntactic structure. The airworthiness regulation texts are stored on the website in (.pdf) format which need to be converted to texts in (.txt) format before the processing of NLP tasks. Noise will inevitably be brought to the process of structure conversion and semantic cohesion, including wrong words, unrecognized punctuations, incorrect formats, etc. Therefore, data preprocessing is necessary for the task of textual structure conversion because errors can be identified and revised by checking repeatedly. The consistent of text structures can be ensured through data preprocessing. The purpose of data post-processing is to clear the redundant contents and revise the improper punctuations which may exist in the processed airworthiness regulation texts to improve the readability. In addition, in our work, both data preprocessing and post-processing are achieved by three postgraduate students (with the knowledge background of airworthiness safety) in the form of alternate checks.

C. ANALYSIS OF EXPERIMENTAL RESULTS

In order to verify the feasibility and validity of the proposed node content link algorithm, the tasks of semantic cohesion and structure conversion are implemented to the two types of airworthiness regulation texts with different structure forms. The experimental results are generalized and analyzed in this section. All experiments are performed on a 2.3 GHz Intel Core (i5) system with 16 GB of memory, and the software program is realized based on Python 3.7 programming language.

1) APPLICATION CASES ANALYSIS

The processes of semantic cohesion and structure conversion are illustrated in detail through an application case. Take airworthiness regulations of CCAR-23-R3, Article 23. 609 as an example (as shown in Fig. 1), the proposed node content link algorithm is implemented to the regulations for the tasks of semantic cohesion and structure conversion, so as to convert the hierarchical provision structure of airworthiness regulation texts to the general linguistic form. In this case, there are six elementary provision units contained in this regulation (CCAR-23-R3, Article 23. 609), and the highest level of hierarchical provision structure is 4 (as shown in Fig. 1). According to the characteristics of hierarchical provision

[CCAR-23-R3], 第23.609条 每个结构零件必须满足下列要求:(a) 有适当的保护,以防止使用中由于任何原因而引起性能降低或强度丧失,这些原因中包括:(1) 气候。
[CCAR-23-R3], 第23.609条 每个结构零件必须满足下列要求:(a) 有适当的保护,以防止使用中由于任何原因而引起性能降低或强度丧失,这些原因中包括:(2) 腐蚀。
[CCAR-23-R3], 第23.609条 每个结构零件必须满足下列要求:(a) 有适当的保护,以防止使用中由于任何原因而引起性能降低或强度丧失,这些原因中包括:(3) 磨损。
[CCAR-23-R3], 第23.609条 每个结构零件必须满足下列要求:(b) 有足够的通风和排水措施。

FIGURE 8. The results of structure conversion and semantic cohesion for Chinese airworthiness regulations.

[CCAR-23-R3], Article 23.609 Each structure component must satisfy the following requirements: (a) appropriate protective measures are set to prevent performance degradation or strength loss, including: (1) climate.
[CCAR-23-R3], Article 23.609 Each structure component must satisfy the following requirements: (a) appropriate protective measures are set to prevent performance degradation or strength loss, including: (2) corrosion.
[CCAR-23-R3], Article 23.609 Each structure component must satisfy the following requirements: (a) appropriate protective measures are set to prevent performance degradation or strength loss, including: (3) abrasion.
[CCAR-23-R3], Article 23.609 Each structure component must satisfy the following requirements: (b) adequate ventilation and drainage.

FIGURE 9. Translated results of the processed provision sentences.

(['CCAR-23-R3'], ['23.609'], ['(a) '], ['(1) '])
(['CCAR-23-R3'], ['23.609'], ['(a) '], ['(2) '])
(['CCAR-23-R3'], ['23.609'], ['(a) '], ['(3) '])
(['CCAR-23-R3'], ['23.609'], ['(b) '])

FIGURE 10. Four directed paths with complete semantic information and syntactic structure achieved through the proposed algorithm.

structure, there are four directed paths with complete semantic information existing in this regulation provision, which are calculated through the proposed algorithm.

The task of structure conversion is achieved by connecting node contents according to semantic cohesion relations, as well as constructing edges between nodes with semantic cohesion relations. The results of structure conversion and semantic cohesion for the referenced airworthiness regulations in this case, are represented in Fig. 8. The processed provision sentences (as shown in Fig. 8) are translated into English in order to ensure the readability. The translated results are represented in Fig. 9.

The four directed paths with complete semantic information and syntactic structure achieved through the proposed algorithm (corresponding to CCAR-23-R3, Article 23. 609) are shown in Fig. 10.

Take CCAR-135-R1 airworthiness regulation text as an example (both original and processed texts) to illustrate the feasibility and validity of the proposed node content link algorithm for the tasks of structure conversion and semantic cohesion on document-level. The original airworthiness regulation text (CCAR-135-R1) and the processed regulation text (after the processing of structure conversion and semantic cohesion) are represented in Fig. 11 and Fig. 12, respectively. The original airworthiness regulation text is represented in Fig. 11, which shows that the regulation text has complex hierarchical provision structure. Then, the proposed algorithm is deployed to the regulation texts, so as to convert the special hierarchical provision structure to the general linguistic form with complete semantic information and syntactic structure. The conversion results are shown in Fig. 12.

For Chinese airworthiness regulation texts with II-type provision structure, an original regulation document is selected to illustrate the validity of the proposed node content link algorithm. The original airworthiness regulation text (CCAR-139CA-R2) and the processed regulation text (after the processing of structure conversion and semantic cohesion) are represented in Fig. 13 and Fig. 14, respectively. The hierarchical provision structure of Chinese airworthiness regulation texts can be converted to the general linguistic form, and provision sentences with complete semantic information and syntactic structure can also be generated through the proposed algorithm, successfully.

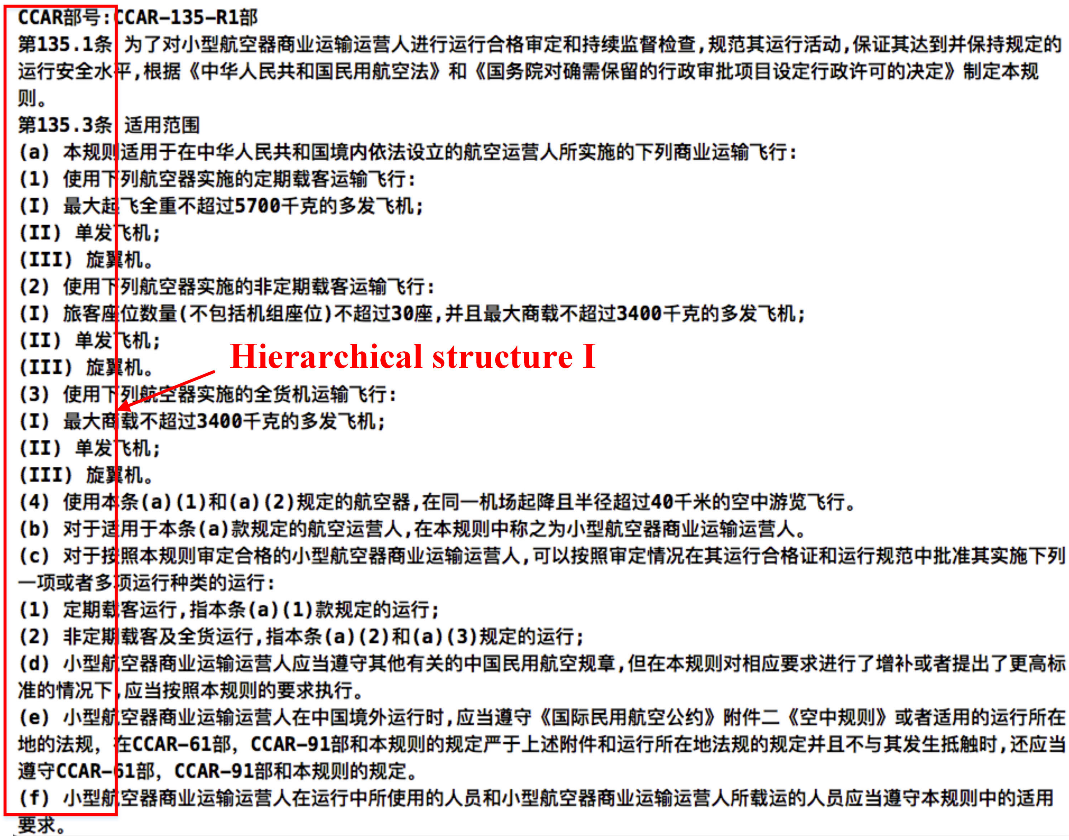
The proposed semantic cohesion theory and approach can solve the application limitation problem for NLP technology to Chinese airworthiness regulation texts caused by the unique hierarchical provision structure. The corresponding explanations are provided as follows:

Based on the proposed semantic cohesion theory and approach, the special hierarchical provision structure of airworthiness regulation texts can be converted to the general linguistic form, which contains the complete semantic information and syntactic components. Moreover, sentence-level textual data is usually taken as elementary unit for NLP tasks because complete syntactic and semantic contents are both contained. However, for law and regulation texts which are affected by the special hierarchical provision structure, linguistic components of sentences are incomplete, which leads to the problem of sentential components deficiency (like subject or object of sentences). The negative phenomenon makes it hard to deploy NLP technology to Chinese airworthiness regulation texts, as well as the problem of application limitation. For law and regulation texts, based on the proposed semantic cohesion theory and approach, the regulation texts can be processed and converted to the sentence-level textual data with general linguistic form. Complete semantic information and syntactic structure are contained in the processed textual data so that NLP tasks can be deployed. Namely, the problem of application limitation for NLP technology to Chinese airworthiness regulation texts can be solved by converting the special hierarchical provision structure to the general linguistic form with complete semantic and syntactic information, so as to acquire the sentence-level textual data.

2) ALGORITHM PERFORMANCE ANALYSIS

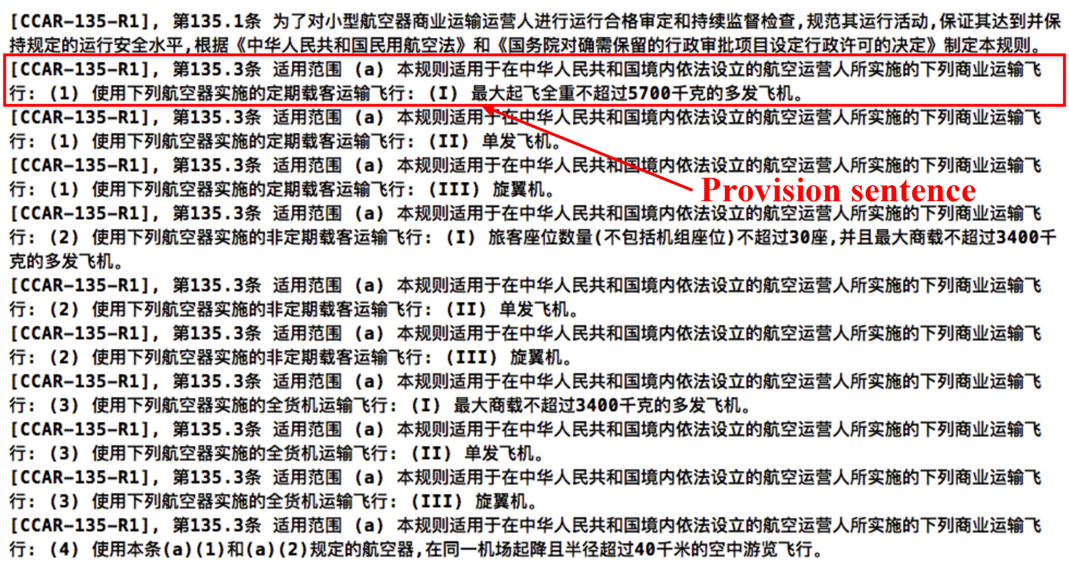
In our study, the algorithm has been deployed to Chinese airworthiness regulation texts with two different provision structures, respectively. The CPU time for the proposed node content link algorithm to achieve the tasks of structure conversion and semantic cohesion is counted to illustrate and analyze the performance of the algorithm.

Firstly, for the airworthiness regulation texts with I-type provision structure, the hierarchical structure is more



CCAR部号：CCAR-135-R1部
第135.1条 为了对小型航空器商业运输运营人进行运行合格审定和持续监督检查，规范其运行活动，保证其达到并保持规定的运行安全水平，根据《中华人民共和国民用航空法》和《国务院对确需保留的行政审批项目设定行政许可的决定》制定本规则。
第135.3条 适用范围
(a) 本规则适用于在中华人民共和国境内依法设立的航空运营人所实施的下列商业运输飞行：
(1) 使用下列航空器实施的定期载客运输飞行：
(I) 最大起飞全重不超过5700千克的多发飞机；
(II) 单发飞机；
(III) 旋翼机。
(2) 使用下列航空器实施的非定期载客运输飞行：
(I) 旅客座位数量(不包括机组座位)不超过30座，并且最大商载不超过3400千克的多发飞机；
(II) 单发飞机；
(III) 旋翼机。
(3) 使用下列航空器实施的全货机运输飞行：
(I) 最大商载不超过3400千克的多发飞机；
(II) 单发飞机；
(III) 旋翼机。
(4) 使用本条(a)(1)和(a)(2)规定的航空器，在同一机场起降且半径超过40千米的空中游览飞行。
(b) 对于适用于本条(a)款规定的航空运营人，在本规则中称之为小型航空器商业运输运营人。
(c) 对于按照本规则审定合格的小型航空器商业运输运营人，可以按照审定情况在其运行合格证和运行规范中批准其实施下列一项或者多项运行种类的运行：
(1) 定期载客运行，指本条(a)(1)款规定的运行；
(2) 非定期载客及全货运行，指本条(a)(2)和(a)(3)规定的运行；
(d) 小型航空器商业运输运营人应当遵守其他有关的中国民用航空规章，但在本规则对相应要求进行了增补或者提出了更高标准的情况下，应当按照本规则的要求执行。
(e) 小型航空器商业运输运营人在中国境外运行时，应当遵守《国际民用航空公约》附件二《空中规则》或者适用的运行所在地的法规，在CCAR-61部，CCAR-91部和本规则的规定严于上述附件和运行所在地法规的规定并且不与其发生抵触时，还应当遵守CCAR-61部，CCAR-91部和本规则的规定。
(f) 小型航空器商业运输运营人在运行中所使用的人员和小型航空器商业运输运营人所载运的人员应当遵守本规则中的适用要求。

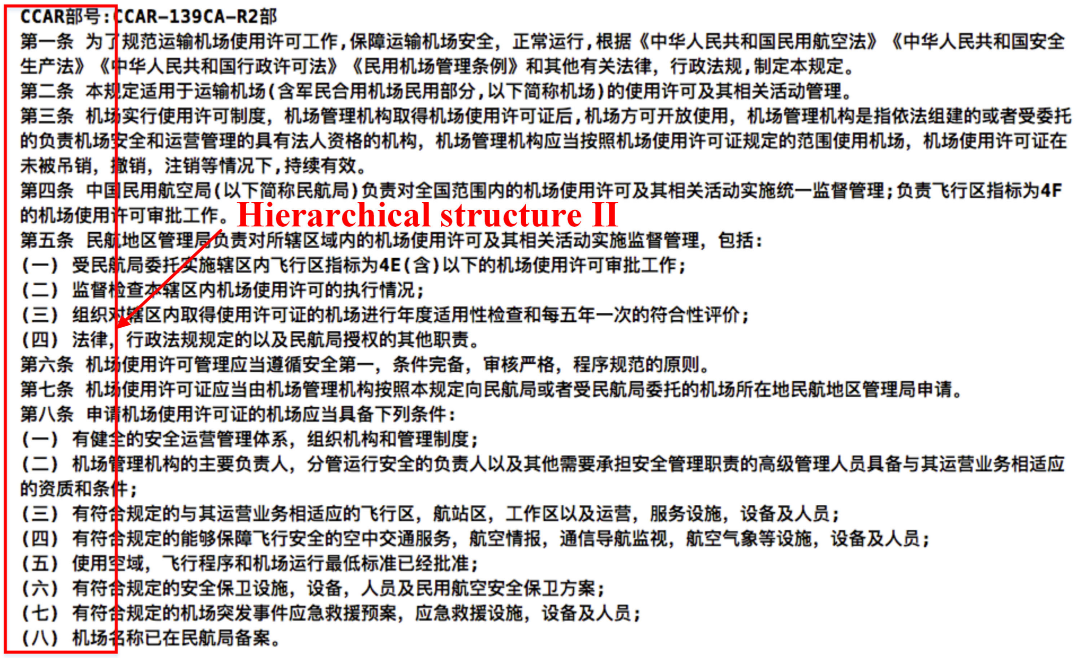
FIGURE 11. The original Chinese airworthiness regulations with I-type provision structure.



[CCAR-135-R1]，第135.1条 为了对小型航空器商业运输运营人进行运行合格审定和持续监督检查，规范其运行活动，保证其达到并保持规定的运行安全水平，根据《中华人民共和国民用航空法》和《国务院对确需保留的行政审批项目设定行政许可的决定》制定本规则。
[CCAR-135-R1]，第135.3条 适用范围 (a) 本规则适用于在中华人民共和国境内依法设立的航空运营人所实施的下列商业运输飞行：(1) 使用下列航空器实施的定期载客运输飞行：(I) 最大起飞全重不超过5700千克的多发飞机。
[CCAR-135-R1]，第135.3条 适用范围 (a) 本规则适用于在中华人民共和国境内依法设立的航空运营人所实施的下列商业运输飞行：(1) 使用下列航空器实施的定期载客运输飞行：(II) 单发飞机。
[CCAR-135-R1]，第135.3条 适用范围 (a) 本规则适用于在中华人民共和国境内依法设立的航空运营人所实施的下列商业运输飞行：(1) 使用下列航空器实施的定期载客运输飞行：(III) 旋翼机。
[CCAR-135-R1]，第135.3条 适用范围 (a) 本规则适用于在中华人民共和国境内依法设立的航空运营人所实施的下列商业运输飞行：(2) 使用下列航空器实施的非定期载客运输飞行：(I) 旅客座位数量(不包括机组座位)不超过30座，并且最大商载不超过3400千克的多发飞机。
[CCAR-135-R1]，第135.3条 适用范围 (a) 本规则适用于在中华人民共和国境内依法设立的航空运营人所实施的下列商业运输飞行：(2) 使用下列航空器实施的非定期载客运输飞行：(II) 单发飞机。
[CCAR-135-R1]，第135.3条 适用范围 (a) 本规则适用于在中华人民共和国境内依法设立的航空运营人所实施的下列商业运输飞行：(2) 使用下列航空器实施的非定期载客运输飞行：(III) 旋翼机。
[CCAR-135-R1]，第135.3条 适用范围 (a) 本规则适用于在中华人民共和国境内依法设立的航空运营人所实施的下列商业运输飞行：(3) 使用下列航空器实施的全货机运输飞行：(I) 最大商载不超过3400千克的多发飞机。
[CCAR-135-R1]，第135.3条 适用范围 (a) 本规则适用于在中华人民共和国境内依法设立的航空运营人所实施的下列商业运输飞行：(3) 使用下列航空器实施的全货机运输飞行：(II) 单发飞机。
[CCAR-135-R1]，第135.3条 适用范围 (a) 本规则适用于在中华人民共和国境内依法设立的航空运营人所实施的下列商业运输飞行：(3) 使用下列航空器实施的全货机运输飞行：(III) 旋翼机。
[CCAR-135-R1]，第135.3条 适用范围 (a) 本规则适用于在中华人民共和国境内依法设立的航空运营人所实施的下列商业运输飞行：(4) 使用本条(a)(1)和(a)(2)规定的航空器，在同一机场起降且半径超过40千米的空中游览飞行。

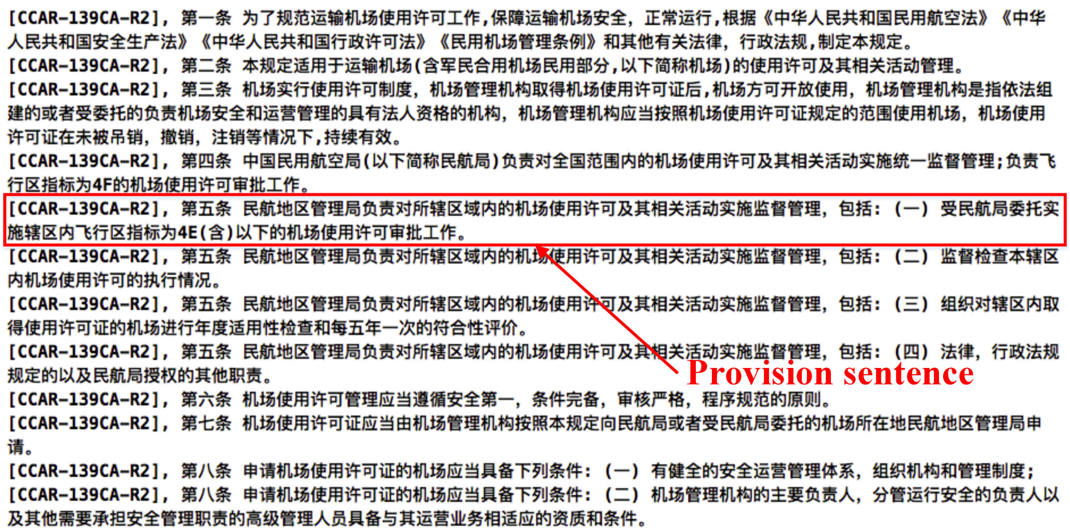
FIGURE 12. The results of structure conversion and semantic cohesion through the proposed algorithm for Chinese airworthiness regulations with I-type provision structure.

complicated than the airworthiness regulation texts with II-type provision structure. More elementary provision units are concluded in the catalog of Article, which means more subordinate provision units are derived from the provision unit of title level 2. The statistical results of elementary provision units for the Chinese airworthiness regulation texts with I-type provision structure are represented in Fig. 15. As shown in Fig. 15, X-axis denotes the serial number of airworthiness regulation documents, and Y-axis indicates the total number of elementary provision units corresponding to the different Chinese airworthiness regulation documents.



CCAR部号：CCAR-139CA-R2部
第一条 为了规范运输机场使用许可工作，保障运输机场安全，正常运行，根据《中华人民共和国民用航空法》《中华人民共和国安全生产法》《中华人民共和国行政许可法》《民用机场管理条例》和其他有关法律，行政法规，制定本规定。
第二条 本规定适用于运输机场（含军民合用机场民用部分，以下简称机场）的使用许可及其相关活动管理。
第三条 机场实行使用许可制度，机场管理机构取得机场使用许可证后，机场方可开放使用，机场管理机构是指依法组建的或者受委托的负责机场安全和运营管理的具有法人资格的机构，机场管理机构应当按照机场使用许可证规定的范围使用机场，机场使用许可证在未被吊销，撤销，注销等情况下，持续有效。
第四条 中国民用航空局（以下简称民航局）负责对全国范围内的机场使用许可及其相关活动实施统一监督管理；负责飞行区指标为4F的机场使用许可审批工作。
第五条 民航地区管理局负责对所辖区域内的机场使用许可及其相关活动实施监督管理，包括：
（一） 受民航局委托实施辖区内飞行区指标为4E（含）以下的机场使用许可审批工作；
（二） 监督检查本辖区内机场使用许可的执行情况；
（三） 组织对辖区内取得使用许可证的机场进行年度适用性检查和每五年一次的符合性评价；
（四） 法律，行政法规规定的以及民航局授权的其他职责。
第六条 机场使用许可管理应当遵循安全第一，条件完备，审核严格，程序规范的原则。
第七条 机场使用许可证应当由机场管理机构按照本规定向民航局或者受民航局委托的机场所在地民航地区管理局申请。
第八条 申请机场使用许可证的机场应当具备下列条件：
（一） 有健全的安全运营管理体系，组织机构和管理制度；
（二） 机场管理机构的主要负责人，分管运行安全的负责人以及其他需要承担安全管理职责的高级管理人员具备与其运营业务相适应的资质和条件；
（三） 有符合规定的与其运营业务相适应的飞行区，航站区，工作区以及运营，服务设施，设备及人员；
（四） 有符合规定的能够保障飞行安全的空中交通服务，航空情报，通信导航监视，航空气象等设施，设备及人员；
（五） 使用空域，飞行程序和机场运行最低标准已经批准；
（六） 有符合规定的安全保卫设施，设备，人员及民用航空安全保卫方案；
（七） 有符合规定的机场突发事件应急救援预案，应急救援设施，设备及人员；
（八） 机场名称已在民航局备案。

FIGURE 13. The original Chinese airworthiness regulations with II-type provision structure.



[CCAR-139CA-R2]，第一条 为了规范运输机场使用许可工作，保障运输机场安全，正常运行，根据《中华人民共和国民用航空法》《中华人民共和国安全生产法》《中华人民共和国行政许可法》《民用机场管理条例》和其他有关法律，行政法规，制定本规定。
[CCAR-139CA-R2]，第二条 本规定适用于运输机场（含军民合用机场民用部分，以下简称机场）的使用许可及其相关活动管理。
[CCAR-139CA-R2]，第三条 机场实行使用许可制度，机场管理机构取得机场使用许可证后，机场方可开放使用，机场管理机构是指依法组建的或者受委托的负责机场安全和运营管理的具有法人资格的机构，机场管理机构应当按照机场使用许可证规定的范围使用机场，机场使用许可证在未被吊销，撤销，注销等情况下，持续有效。
[CCAR-139CA-R2]，第四条 中国民用航空局（以下简称民航局）负责对全国范围内的机场使用许可及其相关活动实施统一监督管理；负责飞行区指标为4F的机场使用许可审批工作。
[CCAR-139CA-R2]，第五条 民航地区管理局负责对所辖区域内的机场使用许可及其相关活动实施监督管理，包括：（一） 受民航局委托实施辖区内飞行区指标为4E（含）以下的机场使用许可审批工作。
[CCAR-139CA-R2]，第五条 民航地区管理局负责对所辖区域内的机场使用许可及其相关活动实施监督管理，包括：（二） 监督检查本辖区内机场使用许可的执行情况。
[CCAR-139CA-R2]，第五条 民航地区管理局负责对所辖区域内的机场使用许可及其相关活动实施监督管理，包括：（三） 组织对辖区内取得使用许可证的机场进行年度适用性检查和每五年一次的符合性评价。
[CCAR-139CA-R2]，第五条 民航地区管理局负责对所辖区域内的机场使用许可及其相关活动实施监督管理，包括：（四） 法律，行政法规规定的以及民航局授权的其他职责。
[CCAR-139CA-R2]，第六条 机场使用许可管理应当遵循安全第一，条件完备，审核严格，程序规范的原则。
[CCAR-139CA-R2]，第七条 机场使用许可证应当由机场管理机构按照本规定向民航局或者受民航局委托的机场所在地民航地区管理局申请。
[CCAR-139CA-R2]，第八条 申请机场使用许可证的机场应当具备下列条件：（一） 有健全的安全运营管理体系，组织机构和管理制度；
[CCAR-139CA-R2]，第八条 申请机场使用许可证的机场应当具备下列条件：（二） 机场管理机构的主要负责人，分管运行安全的负责人以及其他需要承担安全管理职责的高级管理人员具备与其运营业务相适应的资质和条件。

FIGURE 14. The results of structure conversion and semantic cohesion through the proposed algorithm for Chinese airworthiness regulations with II-type provision structure.

In Fig. 15, the dark blue parts of histograms denote the total number of provision units with level 2 in airworthiness regulation texts (namely, the number of elementary provision units with the title of Article). The grey parts express the total number of elementary provision units corresponding to each Chinese airworthiness regulation text (as well as the number of elementary provision units with different title levels), which also express the multi-level hierarchical structure of the text. Airworthiness regulation texts with I-type provision structure are sorted according to the total number of elementary provision units with title level 2 in ascending order (namely, the dark blue parts of histograms in Fig. 15). As is shown in Fig. 15, the structure complexity represents a tendency of increasing gradually.

As is shown in Fig. 15, for any airworthiness regulation texts with I-type provision structure, multiple subordinate elementary provision units (with title levels of 3 to 6), are derived from the corresponding provision unit of title level 2. The more subordinate elementary provision units derived from the corresponding elementary provision unit of title level 2, the more complex the hierarchical provision structure of airworthiness regulation text is. Note that the node content

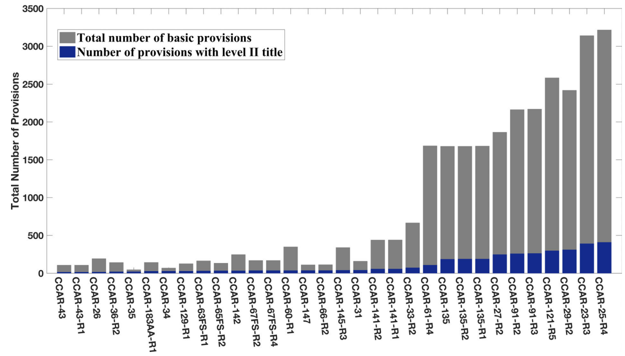


FIGURE 15. The statistical results of elementary provision units for the airworthiness regulation texts with I-type provision structure.

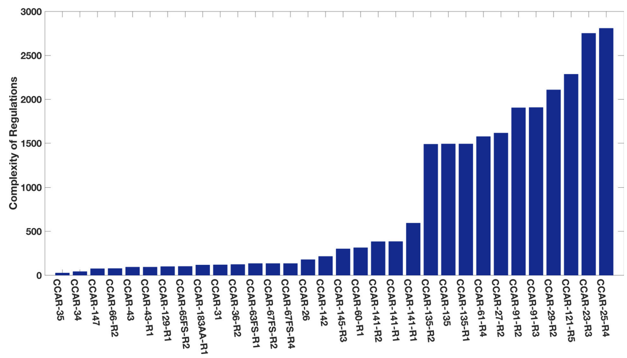


FIGURE 16. The structure complexity of airworthiness regulation texts with I-type provision structure.

with title level 1 denotes the serial number of airworthiness regulation text, and title level 2 indicates the level of Article.

In this article, we use the difference value between the number of all elementary provision units and the total number of provision units with title level 2, to denote the structure complexity of airworthiness regulation texts. For airworthiness regulation texts with I-type provision structure, the structure complexity is represented in Fig. 16. For different airworthiness regulation texts, the structure complexities are also different. The regulation texts are sorted according to the structure complexity in ascending order, and the sorted result is represented in Fig. 16. For airworthiness regulation texts with I-type provision structure, the new generated provision sentences with complete semantic information and syntactic structure are counted and analyzed. The statistical results are represented in Fig. 17. As shown in Fig. 17, the dark blue parts of histograms denote the total number of provision units with title level 2 in airworthiness regulation texts. The grey parts of histograms indicate the total number of subordinate elementary provision units which are derived from the provision units of title level 2.

In Fig. 17, the number of new generated provision sentences with complete semantic and syntactic structure represents the same variation tendency with the structure complexity of airworthiness regulation texts (corresponding to the structure complexity sorted and represented in Fig. 16). It is illustrated that the more complex the text structure is, the

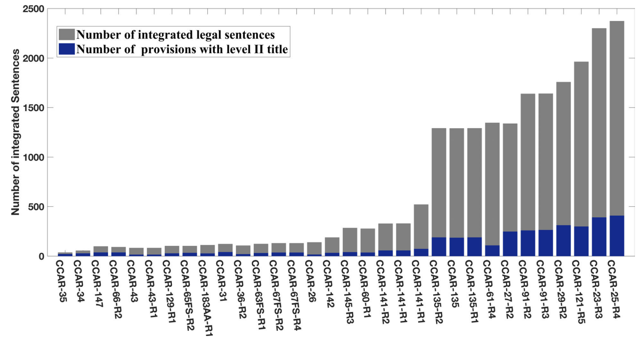


FIGURE 17. Statistical results for provision sentences with complete semantic information and syntactic structure achieved through the proposed algorithm.

more provision sentences with complete linguistic structure and semantic information will be generated. Meanwhile, the total number of new generated provision sentences with complete semantic information and linguistic structure is less than the total number of original elementary provision units, which conforms to the linguistic characteristic of hierarchical provision structure of airworthiness regulations. Many provision sentences with complete semantic information and syntactic structure are generated on the foundation of the provision unit with title level 2 through the proposed algorithm.

For Chinese airworthiness regulation texts with different complexities of provision structure, the required CPU time for structure conversion and semantic cohesion through the proposed algorithm is recorded and compared. The experiments are carried out 50 times for all airworthiness regulation texts. Then, the per CPU time is recorded, and the average time for the 50 times experiments is calculated as the standard CPU time for the tasks of structure conversion and semantic cohesion. The CPU time for structure conversion and semantic cohesion of airworthiness regulations with I-type provision structure through the proposed node content link algorithm, is shown in Fig. 18. In Fig. 18, it is illustrated that the required CPU time keeps the same tendency with the structure complexity of texts. For airworthiness regulations (I-type provision structure) with the highest complexity of structure, the task of structure conversion can be achieved within a short calculation time (less than 0.2 seconds) through the proposed algorithm. Moreover, in total 2379 provision sentences with complete semantic information and syntactic structure are generated through the proposed algorithm. It shows that the algorithm can effectively achieve the tasks of structure conversion and semantic cohesion for Chinese airworthiness regulation texts.

The second part of experiments is carried out for airworthiness regulation texts with II-type provision structure. The provision structure for this kind of text is simpler than the regulation texts with I-type provision structure. Moreover, the airworthiness regulation texts with II-type provision structure have a large proportion in the whole airworthiness regulation texts. For the airworthiness regulation texts with II-type provision structure, the highest title level is 5. Moreover,

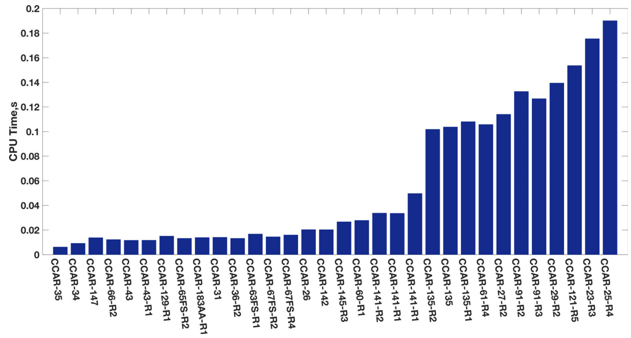


FIGURE 18. CPU time of structure conversion and semantic cohesion for Chinese airworthiness regulation texts with I-type provision structure.

the total number of subordinate elementary provision units derived from the provision units of title level 2, is less than the regulation texts with I-type provision structure.

The statistical results of elementary provision units for Chinese airworthiness regulation texts with II-type provision structure are represented in Fig. 19. In Fig. 19, X-axis denotes the serial number of airworthiness regulation texts with II-type provision structure, and Y-axis indicates the total number of elementary provision units corresponding to the different regulation documents. Each histogram denotes the number of elementary provision units for an airworthiness regulation text with II-type provision structure. Among them, the dark blue parts of histograms denote the total number of provision units with title level 2 in airworthiness regulation texts. The grey parts of histograms express the total number of elementary provision units corresponding to each airworthiness regulation text, namely, the number of subordinate elementary provision units derived from the provision units of title level 2. The airworthiness regulation texts with II-type provision structure (the horizontal ordinate in Fig. 19) are sorted according to the total number of provision units with title level 2 (the dark blue parts of histograms in Fig. 19) in ascending order. Then, the airworthiness regulation texts with II-type provision structure are sorted again according to the structure complexity of texts in ascending order. The sorted result is represented in Fig. 20.

The tasks of structure conversion and semantic cohesion are carried out to the airworthiness regulation texts with II-type provision structure through the proposed node content link algorithm. The total number of new generated provision sentences with complete semantic and syntactic structure is represented in Fig. 21. The X-axis denotes the serial number of airworthiness regulation documents, which are sorted according to structure complexity in ascending order (coherence with Fig. 20).

For the different structure complexity of airworthiness regulations, the required CPU time for structure conversion and semantic cohesion is recorded and analyzed. The statistical results are represented in Fig. 22. As shown in Fig. 22, with the increasing complexity of textual hierarchical structure, the required calculation time is also increased. The CPU time presents the same variation trend with the complexity of text

TABLE 4. Analysis results of textual characteristic and algorithm performance for the two types of airworthiness regulation texts.

ARs	AverNum	Aver CPU time	CPU time	Accurate rate
I-type provision structure	897	0.0576	$6.43*10^{-5}$	100%
II-type provision structure	153	0.0163	$1.06*10^{-4}$	100%
Average value	525	0.0369	$8.515*10^{-5}$	/

ARs=Airworthiness regulation texts; AverNum=Average number of elementary provision units for per document; Aver CPU time=Average CPU time for per document (s); CPU time=CPU time for single elementary provision unit (s).

structure. For airworthiness regulations (II-type provision structure) with the highest degree of structure complexity, the tasks of structure conversion and semantic cohesion can be achieved within a short calculation time (0.11 seconds) through the proposed node content link algorithm, while in total 1568 provision sentences with complete semantic information and syntactic structure are generated successfully. It is illustrated that the proposed algorithm can effectively achieve the tasks of structure conversion and semantic cohesion with a high calculation efficiency. Moreover, the unique provision structure of airworthiness regulation texts can be successfully converted to the general linguistic form with complete semantic information and syntactic structure, which helps to promote the application of NLP technology in the field of airworthiness.

For the two types of airworthiness regulation texts with different provision structures, the comparative result of CPU time required for the tasks of structure conversion and semantic cohesion through the proposed algorithm is represented in Fig. 23. As shown in Fig. 23, the blue line is used to represent the CPU time needed for structure conversion and semantic cohesion through the proposed algorithm for Chinese airworthiness regulation texts with I-type provision structure. The red line indicates the required CPU time for the proposed algorithm to achieve the tasks for airworthiness regulation texts with II-type provision structure. The X-axis denotes the serial numbers of Chinese airworthiness regulation documents (coherence with Fig. 18 and Fig. 22). In Fig. 23, for both types of Chinese airworthiness regulation texts, the calculation time required for the tasks of semantic cohesion and structure conversion, is less than 0.2 seconds. Consequently, the tasks of structure conversion and semantic cohesion for Chinese airworthiness regulation texts can be achieved successfully and effectively through the proposed node content link algorithm.

For the two types of airworthiness regulation texts with different provision structures, the required calculation time for structure conversion and semantic cohesion through the proposed node content link algorithm is counted and analyzed. The statistical results are shown in Table. 4. As shown in Table. 4, for airworthiness regulation texts with I-type provision structure, the average number of elementary provision units for per document is 897, and the average CPU

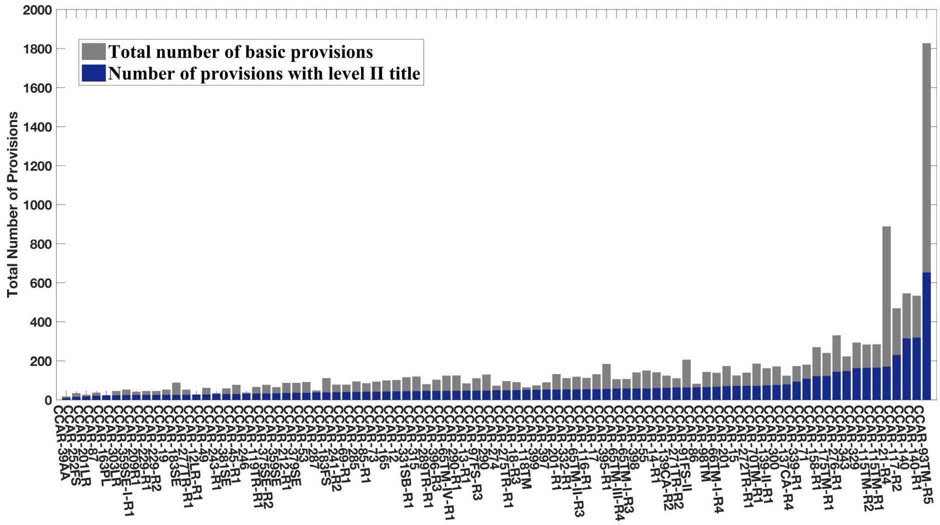


FIGURE 19. The statistical results of elementary provision units for the airworthiness regulation texts with II-type provision structure.

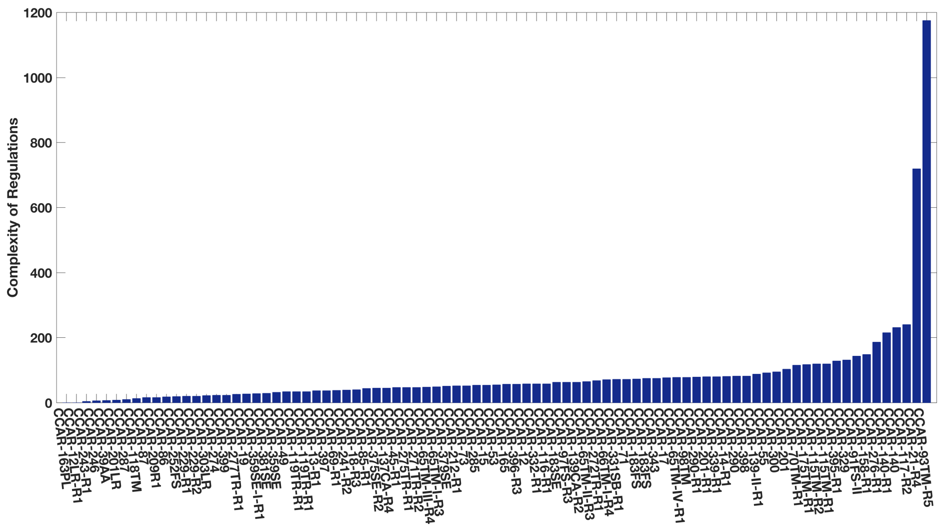


FIGURE 20. The structure complexity of airworthiness regulation texts with II-type provision structure.

time needed for per document is 0.0576 seconds. Moreover, the average CPU time for the proposed algorithm to process every single elementary provision unit is $6.43*10^{-5}$ seconds.

For airworthiness regulation texts with II-type provision structure, the average number of elementary provision units is 153, and the average CPU time for per document is 0.0163 seconds. The average CPU time for the proposed algorithm to process every single elementary provision unit is $1.06*10^{-4}$ seconds. Consequently, for all airworthiness regulation texts (including I-type and II-type), the average number of elementary provision units is 525. The average CPU time for the algorithm to achieve the tasks of structure conversion and semantic cohesion is 0.0369 seconds. Furthermore, the average CPU time for the algorithm to process each elementary provision unit is $8.515*10^{-5}$ seconds. It is illustrated that provision sentences with complete semantic structure can be achieved within a short calculation time through the proposed node content link algorithm. Therefore, the algorithm can realize the tasks of structure conversion and semantic cohesion for Chinese airworthiness regulation texts while considering the special hierarchical provision structure effectively.

D. ACCURATE RATE EVALUATION

The hierarchical provision structure of Chinese airworthiness regulation texts can be converted to the general linguistic form with complete semantic information and syntactic structure through the proposed node content link algorithm. For airworthiness regulation documents provided by the CCAA's official website, the results of structure conversion and semantic cohesion acquired through the proposed algorithm are checked and evaluated with the method of manual proofreading. Three postgraduate students with the background knowledge of airworthiness safety are asked to help us verify the experimental results of the proposed algorithm. The statistical results show that the proposed algorithm achieves the

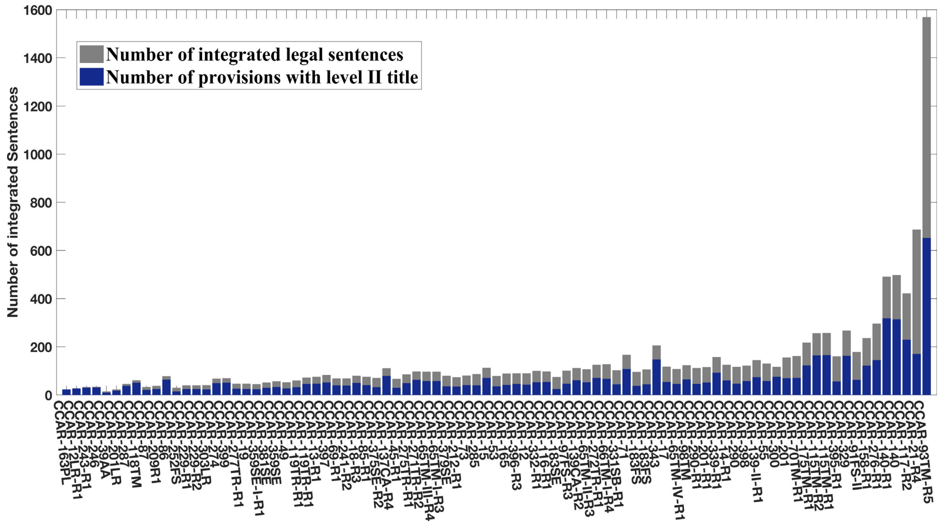


FIGURE 21. Statistical results for the provision sentences with complete semantic information and syntactic structure achieved through the proposed algorithm.

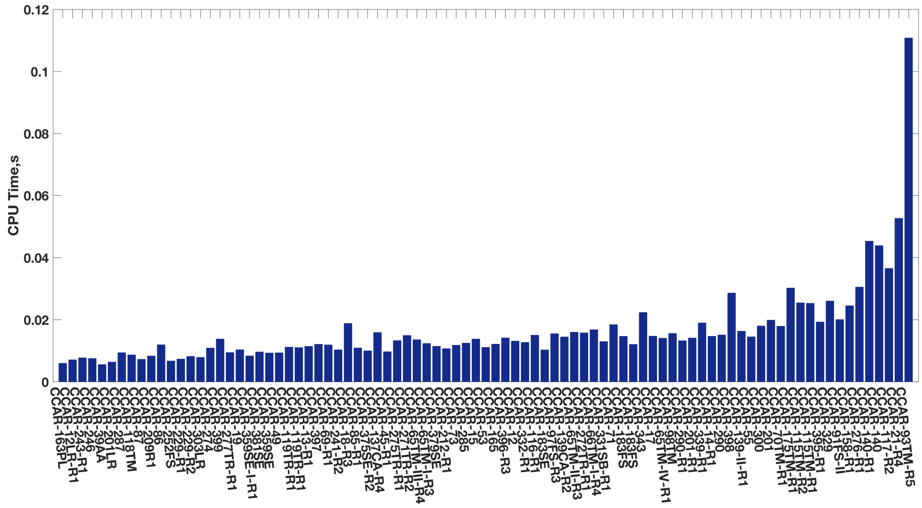


FIGURE 22. CPU time of structure conversion and semantic cohesion for Chinese airworthiness regulation texts with II-type provision structure.

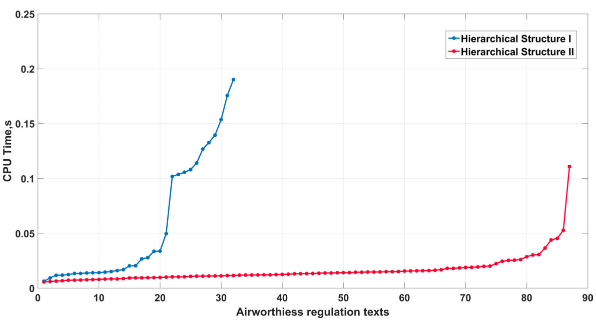


FIGURE 23. The comparative result for the node content link algorithm to achieve structure conversion and semantic cohesion for the two types of Chinese airworthiness regulation texts.

accurate rate of 100% for the tasks of structure conversion and semantic cohesion to Chinese airworthiness regulation texts. Consequently, it is illustrated that the proposed node content link algorithm can achieve the tasks of structure conversion and semantic cohesion successfully and effectively, as well as satisfy the requirements of reliability and validity.

VI. CONCLUSION

The unique hierarchical provision structure of airworthiness regulation texts is different from the general linguistic form, which limits the application of NLP technology to airworthiness safety. In our study, a novel theory of semantic cohesion is proposed to provide mathematical description of discourse representation for Chinese airworthiness regulations. In this theory, the concept of attributive directed graph model is provided, the graphical issue of airworthiness regulation texts is proved by combining with the special provision structure, and the rules of edge structure construction is provided. The tasks of structure conversion and semantic cohesion are abstractly expressed as a graph theory issue, namely, the problem of constructing directed edges and attributive directed graph. Based on the theory, a new node content link algorithm is proposed to achieve structure conversion and semantic cohesion in which the node contents with semantic relation are connected to form provision sentences. The unique hierarchical provision structure can be converted to the general linguistic form according to hierarchy transfer process through the proposed algorithm, and the sentence-level textual data with complete semantic information are generated. Therefore, the problem of application limitation for NLP technology to airworthiness regulation texts has been resolved. Finally, experiments are carried out for the current 119 airworthiness regulation documents, and the performance of the algorithm is analyzed to verify the effectiveness and validity. It is illustrated that the proposed algorithm can achieve the tasks of structure conversion and semantic cohesion for Chinese airworthiness regulation texts effectively.

In future work, the algorithm will be integrated into intellectualized NLP tasks based on Chinese airworthiness regulation texts, such as information acquisition, knowledge graph construction, etc. More experiments about entity relation extraction will be carried out on the foundation of new generated sentence-level textual data and dependency parsing. Furthermore, the application for the node content link algorithm to other legal texts, such as criminal and economic laws, will also be taken into consideration.

REFERENCES

[1] M. B. Şenol, "Evaluation and prioritization of technical and operational airworthiness factors for flight safety," *Aircr. Eng. Aerosp. Technol.*, vol. 92, no. 7, pp. 1049–1061, 2020.

[2] B. Nakgon, "A study on airworthiness for fuel system of unmanned air vehicle," *Int. J. Aeronaut. Space Sci.*, vol. 21, pp. 858–864, Aug. 2020.

[3] L. Purton and K. Kourousis, "Military airworthiness management frameworks: A critical review," *Procedia Eng.*, vol. 80, pp. 545–564, Jan. 2014.

[4] E. Yazgan and A. K. Yilmaz, "Prioritisation of factors contributing to human error for airworthiness management strategy with ANP," *Aircr. Eng. Aerosp. Technol.*, vol. 91, no. 1, pp. 78–93, 2019.

[5] M. A. Blooshi, S. Jafer, and K. Patel, "Review of formal agile methods as cost-effective airworthiness certification processes," *J. Aerosp. Inf. Syst.*, vol. 15, no. 8, pp. 471–484, 2018.

[6] J. Cai, H. Bao, H. F. Zuo, and Y. Huang, "Safety evaluation of airworthiness requirement of bird-strike on aeroplane," *Eng. Failure Anal.*, vol. 102, pp. 407–416, Aug. 2019.

[7] H. T. Niu, C. B. Ma, P. Han, and X. Y. Sun, "Study analysis method to task process of civil aircraft weather radar system," *J. Northwestern Polytechnical Univ.*, vol. 38, no. 1, pp. 84–94, 2020.

[8] H. T. Niu, C. B. Ma, and P. Han, "A decentralized method for collision detection and avoidance applied to civil aircraft," *Proc. Inst. Mech. Eng. G, J. Aerosp. Eng.*, Aug. 2020. Accessed: Aug. 27, 2020, doi: 10.1177/0954410020953045.

[9] E. Cambria and B. White, "Jumping NLP curves: A review of natural language processing research," *IEEE Comput. Intell. Mag.*, vol. 9, no. 2, pp. 48–57, May 2014.

[10] E. Saquete, D. Tomás, P. Moreda, P. Martínez-Barco, and M. Palomar, "Fighting post-truth using natural language processing: A review and open challenges," *Expert Syst. Appl.*, vol. 141, Mar. 2020, Art. no. 112943.

[11] M. Araújo, A. Pereira, and F. Benevenuto, "A comparative study of machine translation for multilingual sentence level sentiment analysis," *Inf. Syst.*, vol. 512, pp. 1078–1102, Feb. 2020.

[12] Y. Z. Li, H. L. Zhou, Z. L. Lin, Y. F. Wang, S. J. Chen, C. Liu, Z. Y. Wang, D. Gifu, and J. B. Xia, "Investigation in the influences of public opinion indicators on vegetable prices by corpora construction and WeChat article analysis," *Future Gener. Comput. Syst.*, vol. 102, pp. 876–888, 2020.

[13] Y. L. Xi, Y. N. Zhang, S. T. Ding, and S. H. Wan, "Visual question answering model based on visual relationship detection," *Signal Process., Image Commun.*, vol. 80, Feb. 2020, Art. no. 115648.

[14] M. M. Mirończuk, "Information extraction system for transforming unstructured text data in fire reports into structured forms: A Polish case study," *Fire Technol.*, vol. 56, no. 2, pp. 545–581, 2020.

[15] K. Zhang, Y. W. Zhu, W. J. Zhang, W. L. Zhang, and Y. H. Zhu, "Transfer correlation between textual content to images for sentiment analysis," *IEEE Access*, vol. 8, pp. 35276–35289, 2020.

[16] X. J. Chen, S. B. Jia, and Y. Xiang, "A review: Knowledge reasoning over knowledge graph," *Expert Syst. Appl.*, vol. 141, Mar. 2020, Art. no. 112948.

[17] M. Tyan, J. Yoon, N. V. Nguyen, J. W. Lee, and S. Kim, "Design airworthiness integration method for general aviation during early development stage," *Aircr. Eng. Aerosp. Technol.*, vol. 91, no. 7, pp. 1067–1076, 2019.

[18] Y. Wang and X. Zhang, "Requirements management applied in airworthiness certification in the civil aircraft," in *Proc. IOP Conf. Mater. Eng.*, Sanya, China, Jan. 2019, Art. no. 012028.

[19] H. Y. Xu and P. Wang, "Real-time reliability verification for UAV flight control system supporting airworthiness certification," *PLoS ONE*, vol. 11, no. 12, 2016, Art. no. e0167168.

[20] P. Gorjan, S. B. Korousic, and E. Tome, "A survey of named entity recognition methods for food information extraction," *IEEE Access*, vol. 8, pp. 31586–31594, 2020.

[21] S. I. Johannes, S. Jurgen, and D. Jens, "Knowledge acquisition from chemical accident databases using an ontology-based method and natural language processing," *Saf. Sci.*, vol. 129, Sep. 2020, Art. no. 104747.

[22] T. D. Bui, S. T. Nguyen, and Q. B. Ho, "Towards a conceptual search for Vietnamese legal text," in *Proc. Int. Conf. Comput. Inf. Syst. Ind. Manage.*, Berlin, Germany, 2014, pp. 175–185.

[23] M. Böschen, R. Bogusch, A. Fraga, and C. Rudat, "Bridging the gap between natural language requirements and formal specifications," in *Proc. Int. Conf. Requirements Eng. Found. Softw. Qual.*, 2016, pp. 1–11.

[24] H. W. Zhou, S. X. Ning, Y. L. Yang, Z. Liu, C. K. Lang, and Y. Y. Lin, "Chemical-induced disease relation extraction with dependency information and prior knowledge," *J. Biomed. Informat.*, vol. 84, pp. 171–178, Aug. 2018.

[25] B. Arfé, L. Mason, and I. Fajardo, "Simplifying informational text structure for struggling readers," *Reading Writing*, vol. 31, no. 9, pp. 2191–2210, 2018.

[26] X. Q. Li, Y. F. Yang, and Y. Lu, "How and when prosodic boundaries influence syntactic parsing under different discourse contexts: An ERP study," *Biol. Psychol.*, vol. 83, no. 3, pp. 250–259, 2010.

[27] B. Webber and A. Joshi, "Discourse structure and computation: Past, present and future," in *Proc. ACL Special Workshop Rediscovering 50 Years Discoveries*, Jeju, South Korea, 2012, pp. 42–54.

[28] C. J. Ji and H. Y. Wu, "Cascade architecture with rhetoric long short-term memory for complex sentence sentiment analysis," *Neurocomputing*, vol. 405, pp. 161–172, May 2020.

[29] Z. H. Li, Z. H. Yang, C. Shen, J. Xu, Y. Y. Zhang, and H. Xu, "Integrating shortest dependency path and sentence sequence into a deep learning framework for relation extraction in clinical text," *BMC Med. Informat. Decis. Making*, vol. 19, no. 1, p. 22, 2019.

[30] V. Suárez-Paniagua, R. M. R. Zavala, I. S. Bedmar, and P. Martínez, "A two-stage deep learning approach for extracting entities and relationships from medical texts," *J. Biomed. Informat.*, vol. 99, Nov. 2019, Art. no. 103285.

[31] A. Iftikhar, S. W. Ul Q. Jaffry, and M. K. Malik, "Information mining from criminal judgments of lahore high court," *IEEE Access*, vol. 7, pp. 59539–59547, 2019.

[32] J. Latif, C. B. Xiao, S. S. Tu, S. U. Rehman, A. Imran, and A. Bilal, "Implementation and use of disease diagnosis systems for electronic medical records based on machine learning: A complete review," *IEEE Access*, vol. 8, pp. 150489–150513, 2020.

[33] M. F. Xue, C. X. Yuan, J. Wang, and W. Q. Liu, "DPAEG: A dependency parse-based adversarial examples generation method for intelligent Q&A robots," *Secur. Commun. Netw.*, vol. 2, pp. 1–15, Jan. 2020.

[34] K. Dashtipour, M. Gogate, J. P. Li, F. L. Jiang, B. Kong, and A. Hussain, "A hybrid Persian sentiment analysis framework: Integrating dependency-grammar-based rules and deep neural networks," *Neurocomputing*, vol. 380, pp. 1–10, Mar. 2020.

[35] N. Z. Zulkarnain and F. Meziane, "Ultrasound report standardization using rhetorical structure theory and domain ontology," *J. Biomed. Inform. X*, vol. 1, Mar. 2019, Art. no. 100003.

[36] G. Sanchez, "Sentence boundary detection in legal text," in *Proc. Natural Legal Lang. Process. Workshop*, Minneapolis, MN, USA, Jun. 2019, pp. 31–38.

[37] C. C. Dourado, R. Galante, M. A. Gonalves, and R. S. Torres, "Bag of textual graphs: A general graph-based text representation model," *J. Assoc. Inf. Sci. Technol.*, vol. 70, no. 8, pp. 817–829, 2019.
[38] O. Batarfi, R. EI Shawi, A. G. Fayoumi, R. Nouri, S. M. R. Beheshti, A. Barnawi, and S. Sakr, "Large scale graph processing systems: Survey and an experimental evaluation," *Cluster Comput.*, vol. 18, no. 3, pp. 1189–1213, 2015.
[39] M. L. Shao, P. Y. Sun, and J. X. Li, "Tree decomposition based anomalous connected subgraph scanning for detection and forecasting events in attributed social media networks," *Neurocomputing*, vol. 407, pp. 83–93, May 2020.
[40] N. Sun, H. Z. Shi, G. J. Han, B. Wang, and L. Shu, "Dynamic path planning algorithms with load balancing based on data prediction for smart transportation systems," *IEEE Access*, vol. 8, pp. 15907–15922, 2020.
[41] D. Antonakaki, P. Fragopoulou, and S. Ioannidis, "A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks," *Expert Syst. Appl.*, vol. 164, Feb. 2021, Art. no. 114006.
[42] T. T. Zhang, B. Z. Lee, Q. H. Zhu, X. Han, and E. M. Ye, "Multi-dimension topic mining based on hierarchical semantic graph model," *IEEE Access*, vol. 8, pp. 64820–64835, 2020.
[43] W. Etaiwi and A. Awajan, "Graph-based Arabic text semantic representation," *Inf. Process. Manage.*, vol. 57, no. 3, 2020, Art. no. 102183.
[44] K. Kim, "Automatic measure of cross-language text structures," *Technol., Knowl. Learn.*, vol. 23, no. 2, pp. 301–314, 2018.
[45] T. Uçkan and A. Karci, "Extractive multi-document text summarization based on graph independent sets," *Egyptian Informat. J.*, vol. 21, no. 3, pp. 145–157, 2020.
[46] K. Bijari, H. Zare, E. Kebriaei, and H. Veisi, "Leveraging deep graph-based text representation for sentiment polarity applications," *Expert Syst. Appl.*, vol. 144, Apr. 2020, Art. no. 113090.

HAOTIAN NIU received the M.S. degree from the School of Electronic Information Engineering, Xi'an Technological University, in 2015. He is currently pursuing the Ph.D. degree with the School of Civil Aviation, Northwestern Polytechnical University (NPU), China. His research interests include autonomous system of collision detection and avoidance, system safety analysis, and natural language processing.

CUNBAO MA (Member, IEEE) graduated from Northwestern Polytechnical University (NPU), China, in 1984. He received the M.S. and Ph.D. degrees from NPU, in 1987 and 2011, respectively. He is currently a Professor with the School of Civil Aviation, NPU. His research interests include system safety analysis, flight simulation, and system modeling.

PEI HAN received the bachelor's and M.S. degrees from School of Aeronautics, Northwestern Polytechnical University (NPU), in 2015 and 2018, respectively. She is currently a System Engineer with the 20th Institute, China Electronics Technology Group Corporation. Her research interests include situation fusion and simulation, and airworthiness safety.

SIYUAN LI received the bachelor's degree from School of Electrical Engineering and Automation, Xi'an University of Architecture and Technology, in 2018. She is currently pursuing the M.S. degree with the School of Civil Aviation, Northwestern Polytechnical University. Her research interests include airworthiness safety and flight simulation.

QI MA received the bachelor's degree from School of Information Technology, Northwest University, in 2019. She is currently pursuing the M.S. degree with the School of Civil Aviation, Northwestern Polytechnical University. Her research interests include situation simulation and airworthiness safety.