Fast Retrieval Method of Forestry Information Features Based on Symmetry Function in Communication Network

Hui Wang [1] and Jie Song [2],*

[1]   School of Automatic Control and Mechanical Engineering, Kunming University, Kunming 650214, China; wh8393@126.com
[2]   School of Information Technology, Kunming University, Kunming 650214, China
*   Correspondence: whsj139@126.com

Abstract: Aiming at the problem of insufficient integration and sharing of forestry information resources under the current communication network and the lack of the concept set of forestry information attributes, which leads to poor information retrieval performance, a fast retrieval method of forestry information features based on symmetry function is studied in depth, and the method is implemented by PDA (Personal Digital Assistant)-BA (Buliding Automation). Using the SED (Stream Editor) forestry information acquisition method under a communication network to collect forestry information, a forestry signal noise cancellation method based on symmetric function method is obtained. In order to improve the accuracy of forestry information acquisition, denoising of the signal in the information was carried out. Constructing forestry information data ontology, integrating forestry resources, establishing a conceptual set of forestry information attributes, distinguishing forestry information attributes, establishing a fast retrieval model of forestry information features based on the synonym library, and completing the fast retrieval of forestry information features. The experimental results show that the recall and precision of this method are 99.25% and 99.24%, respectively, and the retrieval performance is superior, which has a certain application value.

Keywords: communication network; symmetry function; forestry information; fast feature retrieval; PDA; denoising

1. Introduction

In the development of modern forestry planning, the collection of forest information resources is an important basic condition, which can be divided into three categories: The first category is the information of the forest itself. Including the types of forests and the spatial configuration of the various regions, the distribution, growth, and dry loss of the forest and diameter grades, and the like; and the second category is the external information of the forest. It includes forest right, social economy, natural condition, engineering equipment, etc. The third category is the information of forestry production, including the road, engineering project and seedling raising, afforestation, forest, cutting and processing plan, report, cost, and effect, etc., related to forestry. Forestry informationization under the general communication network can be understood as applying various informationization technologies to the whole process of production, management, and service in the whole forestry field, making forestry production highly informationized and intellectualized, thus greatly saving labor costs, improving forestry efficiency, and forestry productivity levels. China's forestry is the basic industry of the national economy, undertaking the important mission of ecological environment construction and promoting social sustainable development. Human beings are the main beneficiary groups in

the construction of the forestry industry. When the forestry industry structure is formed, people will play their respective roles in it [1–3]. In recent years, the advancement of forestry informationization has promoted the sharing of forestry information resources, provided convenience for the public, and promoted the development of the forestry industry. However, in order to make better use of forestry information resources and to serve the vast number of scientific researchers, teaching workers and forestry farmers in the field of Forestry Science in China, it is urgent to implement fast feature retrieval for forestry information.

Relevant experts and scholars have done a lot of research on feature extraction of forestry information. Reference [4] presents the Word Mover's Distance (WMD), a novel distance function between text documents. Their work is based on recent results in word embeddings that learn semantically meaningful representations for words from local co-occurrences in sentences. The WMD distance measures the dissimilarity between two text documents as the minimum amount of distance that the embedded words of one document need to "travel" to reach the embedded words of another document. It shows that this distance metric can be cast as an instance of the Earth Mover's Distance, a well studied transportation problem for which several highly efficient solvers have been developed. Reference [5] extends the regular BoF model by (a) incorporating a weighting mask that allows for altering the importance of each learned codeword and (b) by optimizing the model end-to-end (from the word embeddings to the weighting mask). Furthermore, the BoEW model also provides a fast way to fine-tune the learned representation towards the information need of the user using relevant feedback techniques. Finally, a novel spherical entropy objective function is proposed to optimize the learned representation for retrieval using the cosine similarity metric. Reference [6] introduces a new language representation model called BERT, which stands for Bidirectional Encoder Representations from Transformers. Unlike recent language representation models, BERT is designed to pre-train deep bidirectional representations by jointly conditioning on both the left and right context in all layers. As a result, the pre-trained BERT representations can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and language inference, without substantial task-specific architecture modifications. BERT is conceptually simple and empirically powerful. However, these methods have the problem of low retrieval accuracy in information retrieval.

In order to solve the above-mentioned problems, this paper presents a fast retrieval method of forestry information features based on symmetry function under communication network to realize the fast retrieval of forestry information features and provide a weak force for forestry information construction.

2. Algorithm Definitions

In this part, the realization process of fast retrieval method of forestry information features based on the communication network function will be described in detail. Before fast retrieval of forestry information features, it is necessary to analyze the development significance of forestry information under the general communication network. This is the premise and goal to complete the retrieval of forestry information features, and is the standard to verify the effectiveness of the retrieval methods studied. According to the analysis results, forestry information was collected by PDA equipment. The symmetry function method of communication network is used to denoise the collected forestry information to ensure the accuracy of the collected information and the accuracy of the subsequent information feature extraction results. The forestry information after denoising is integrated, and the information ontology of forestry information is constructed. According to the attributes of information, it is divided into the concept set of forestry information attributes, which improves the speed of forestry information feature retrieval and achieves the purpose of fast retrieval. By introducing a synonym thesaurus, information features are retrieved in the concepts of forestry information attributes, and fast retrieval of forestry information features are completed.

2.1. Development Significance of Forestry Informatization under Communication Network

(1) To promote the development of forestry economy [5]. In the context of implementing the scientific concept of development, information construction is one of the effective ways to solve the bottleneck problem of economic and social development, and is also an inevitable choice to ensure the long-term stable development of economic and social. Forestry is the basic industry of our national economy. We must adapt to the development trend of national informatization construction, actively carry out forestry informatization construction, provide more detailed and abundant forestry information for society, attract social groups to participate in forestry construction, integrate social resources, and provide strong support for promoting forestry economic development [6].

(2) It is conducive to promoting forestry technological innovation. Innovation is the inexhaustible driving force of development, and informatization is the result of scientific and technological innovation, and also a strong support for scientific and technological innovation. Forestry information construction can effectively integrate forestry practical technology, scientific research results and information resources, and build a platform for forestry technological innovation and provide resources [7], so as to realize the rapid popularization and application of forestry technology. Forestry information construction can provide a convenient information platform for modern forestry construction, as well as advanced and scientific management means, and promote forestry to the direction of modernization [8].

(3) It is conducive to promoting the scientific management of forestry. One of the main functions of forestry informatization is to provide timely and comprehensive information support to users of forestry information and provide reliable decision-making basis for forestry management departments. Under the computer network environment, forestry informationization can realize digitalization, informationization, and intellectualization of ecological engineering construction through multi-level information exchange and transmission channels, and quickly transmit the situation of ecological engineering construction to various management departments, so that each management department can fully understand the progress of the project, improve the transparency of forestry management, facilitate timely discovery and solution of problems, and expand communication channels. At the same time, forestry information construction can also publish forestry information uniformly, provide diversified information services for forestry management departments and production units, realize forestry information sharing, and improve forestry information utilization efficiency [9].

2.2. Forestry Information Collection Method Based on PDA in Communication Network

According to the development significance of forestry informatization under the general communication network, the retrieval of forestry information features is carried out. Firstly, PDA is used to collect forestry information.

Personal Digital Assistant (PDA) is a handheld computer which integrates computing, telephone, network and GPS. It has the functions of installing, using, and uninstalling software like PC. For this reason, PDA can be used to collect field data comprehensively in one kind of inventory, two kinds of regulation, desertification popularization and forest quality inspection. By using the palm computer data acquisition and recording system of forest resource regulation, the field data acquisition of forestry can be digitalized and electronically realized. All kinds of calculation and storage functions can be automatically completed by one machine at hand. A large amount of information can be carried around, queried quickly, safely and reliably, and transmitted conveniently [10].

The general information of forestry information collection based on PDA under the general communication network includes "topographic map size", "prefecture", "county", "town", "nature reserve", "forest park", "state-owned forest farm", "collective forest farm", "investigator", "guide", "prosecutor", and "work unit". On the system menu, click "Survey Overall Information", click the name of the item that needs to input data, drop down the data and select input; click the date or time item, the system automatically gives the date or time of the current palm computer, click "OK", and the date or time is recorded [11].

(1) Location and measurement. Location and measurement includes "time and description", "sample plot point" and "sample position". Items can be contracted and displayed when input. Up to three "primer locators (trees)" or "sample plot center point locators (trees)" can be input and primer or sample position information can be plotted according to need.

(2) Lead and perimeter measurements include "sample plot lead measurements" and "perimeter measurements". The horizontal distance and cumulative result can be calculated automatically by inputting "azimuth angle", "inclination angle", and "oblique distance".

(3) Factor investigation. One type of survey includes 75 factors such as land type, dominant tree species, forest species, average diameter, etc. It uses drop-down arrows or "..." to select the input. According to the category factor system, the data will be checked and the system with obvious logic errors will be prompted. It can automatically collect GPS coordinates and manually write coordinate data when the "GPS status" item is displayed as "unpositioned".

(4) In the attributes table of each tree ruler, the items such as "tree species", "later diameter", "azimuth angle", "horizontal distance", and "forest layer" are input, respectively. If there are preliminary data, the items of "standing tree type", "checking ruler type" and "preliminary diameter" will be automatically transferred into the preliminary data. Sample wood survey station is the center of the sample plot, azimuth angle is zero north, clockwise calculation, such as input azimuth angle, horizontal distance beyond the sample plot scope or data has obvious logical errors, the system will give prompts to edit and delete the sample trees; in each wood inspection record, the sample trees can be edited and deleted. The average DBH (Diameter at Breast Height) and height of dominant tree species were calculated.

(5) The sketch map of the sample position can show all the sample positions of the sample plot. Sample position sketch map has many functions, such as full or partial enlargement, selection of sample, display of sample profile, deletion of sample, new sample, and so on.

2.3. A Noise Cancellation Method for Forestry Signals Based on Symmetric Function Method in Communication Network

As the signal received by PDA is susceptible to noise in the course of transmission when collecting forestry information, which affects the integrity and accuracy of information, this paper uses the method of forestry signal noise cancellation based on the symmetric function method under the communication network to filter forestry signal and improve the accuracy of information [12].

The symmetric function method is a mathematical transformation method. Among them, "symmetry" is not symmetry in a geometric sense, but a kind of invariant obtained after many mathematical invariant transformations. The symmetric function method is to transform an object by borrowing another object. Two objects are called symmetric relations. The result of the symmetric function method is equivalent in relations. In this paper, the method of invariant transformation is used to eliminate the noise of the collected information, and the denoising information is equivalent to the collected information.

Definition 1. *Let $x(t)$ be the real signal of forestry information collected by PDA under communication network, these forestry information have fixed characteristics, such as tree species, row numbers, tree numbers, etc. And the signal energy f is limited, then the symmetric correlation function $Wx(t)$ of the real signal of forestry information $x(t)$ at t time is:*

$$Wx(t) = \int_{-\infty}^{\infty} x(t + f/2)x(t - f/2)df \tag{1}$$

Definition 2. *Let $x(t)$ be the real periodic signal of forestry information collected by PDA under the general communication network, and the period is T. For example, forestry economic change information, tree growth information, etc. Then the symmetric correlation function $Wx(t)$ of $x(t)$ at t time is defined as:*

$$Wx(t) = \frac{1}{2T}\int_{-T}^{T} x(t+f/2)x(t-f/2)df \tag{2}$$

$$x(t) = s(t) + n(t) \tag{3}$$

Among them, $s(t)$ is a useful signal for forestry information collected under communication network, $n(t)$ is a random noise with mean of Q and variance of e^2, $s(t)$ and $n(t)$ are independent of each other, and the observation time of $x(t)$ is T. The signal denoising expression of forestry information is obtained as follows:

$$
\begin{aligned}
Wx(t) &= \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x(t+f/2)x(t-f/2)df = \\
&\frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} [s(t+f/2)+n(t+f/2)] \times [s(t-f/2)+n(t-f/2)]df = \\
&\frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} [s(t+f/2)s(t-f/2)+n(t+f/2)s(t-f/2)+s(t+f/2)n(t-f/2)+ \\
&n(t+f/2)n(t-f/2)df = \\
&\frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} [s(t+f/2)s(t-f/2)+n(t+f/2)n(t-f/2)]df = \\
&\frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} s(t+f/2)s(t-f/2)df + \frac{1}{T}n^2
\end{aligned}
\tag{4}
$$

In Formula (4), the longer T is taken, the less noise the $Wx(t)$ is affected by at the t point, and the stronger the ability to suppress noise. Thus, the noise elimination of forestry information signal has been completed.

Based on the above analysis, the elimination of forestry signal noise is completed, which lays a foundation for the fast retrieval of forestry information features.

2.4. A Fast Retrieval Method of Forestry Information Features under Communication Network

With the development of forestry scientific research, the number of forestry information data is increasing. How to make users retrieve accurate and comprehensive information efficiently has become one of the urgent problems in forestry information data sharing. However, traditional information retrieval technology cannot meet the needs of users [13–15]. Based on the forestry signal based on Section 2.3 after noise elimination, this paper uses the fast retrieval method of forestry information features under the communication network to quickly retrieve forestry information features.

2.4.1. Construction of Forestry Information Ontology under Communication Network

At present, forestry information resources in China are heterogeneous and decentralized, so how to integrate these resources and make them have really effective use is a problem that needs to be solved. Information retrieval based on forestry ontology in communication network is a relatively intelligent information retrieval strategy. This method can integrate forestry resources and is of great significance to improve the utilization level of information resources and the quality of information services. In the field of forestry, the application of forestry ontology to develop fast and accurate information retrieval methods is an effective way to improve the level of forestry information, which has strong practical significance. This paper uses the fast retrieval method of forestry information features to construct forestry information ontology and integrate forestry resources by using denoised forestry information [16].

The domain of ontology in this paper is all the knowledge contained in forestry information data. The existing forestry information data mainly includes eight categories: Forest resources, forestry ecological environment, forest protection, timber science, forest cultivation, forestry science and technology foundation, forestry science research topics, and industry development. The amount of data is nearly 800 GB. According to the principle of ontology construction under the communication network, it should be the first step. The domain ontology constructed in other research is selected

for reuse, but because of the different purposes, starting points and application scope of traditional methods, the supporting languages and construction methods are also different, it is impossible to reuse domain ontology [17].

After defining the scope of forestry scientific data ontology, key terms and concepts are enumerated, and core concepts, core concepts, and attributes are constructed. After verification by experts in the field, those who did not pass the audit returned to the higher level, redefined key terms and relationships, generated concepts, attributes, and examples through the audit, and finally formalized coding to generate forestry science data ontology database. The construction flow chart of forestry information data ontology under the communication network is shown in Figure 1.
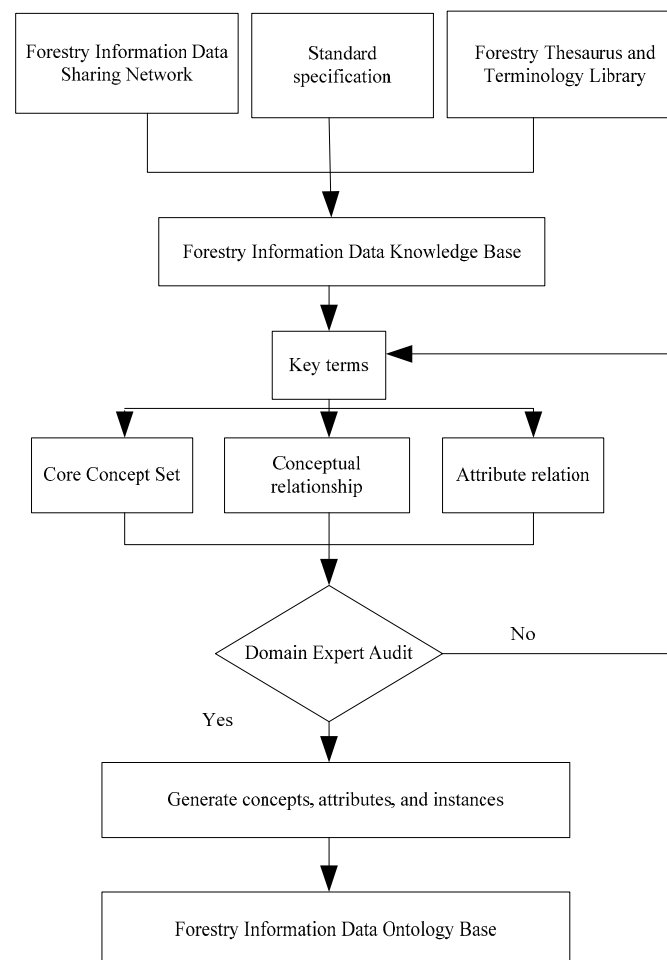

Figure 1. Construction flow chart of forestry information data Ontology.

2.4.2. Establishment of Concept Set of Forestry Information Attribute under Communication Network

According to the above-mentioned forestry information ontology and the attributes of forestry information, the concept set of forestry information attributes is constructed for forestry information ontology. The construction of the concept set of forestry information attributes can greatly improve the retrieval speed of forestry information features.

Concepts are the forms of thinking that reflect the essential attributes of objects, that is, the reflection of objective things in the human brain, and the abstraction and generalization of the common essential characteristics of perceived concepts. Concepts can be either concrete or abstract. Any knowledge is a conceptualization of something or behavior, and ontology is a conceptual normative description system. Ontology corresponds to different application domains, involving different

domains, and different professional terms. Part of the conceptual set of forestry information under the general communication network studied in this paper is described in Table 1.

Table 1. Examples of concepts selected.

Concept Set	Give an Example
Forest types	Tree forest, coniferous forest, broad-leaved forest, evergreen broad-leaved mixed forest, bamboo forest, etc.
Forest tree species	Larix olgensis, Pinus sylvestris var. mongolica, Pinus massoniana, Populus davidiana, Quercus mongolica, Betula platyphylla, Betula platyphylla, Ulmus pumila, etc.
Administrative division	Beijing, Hebei, Inner Mongolia Autonomous Region, Heilongjiang Province, etc.
Geographical distribution	Northeast, North, East, Central, South, Southwest and Northwest China
Terrain type	Plateau, Basin, Plain, Hill, Mountain
Professional terms	Sample plots, forest classes, forest areas, analytical trees, etc.
Forest attribute	Forest age, area, distribution, stock, density, coverage, biomass, growth, etc.
Data type	Forest facies map, distribution map, regionalization map, model, number table

2.4.3. Construction of Fast Retrieval Model of Forestry Information Features Based on Thesaurus

(1) Model framework

This model includes four parts: Standardization of forest information feature terms, query expansion of forest information features, web crawling of forest information features, and weighted ranking of forest information features. Firstly, the forestry Chinese-English-La thesaurus is used to standardize the forestry information feature retrieval words input by users, and the forestry information feature retrieval word K is obtained; secondly, the web page information related to K is captured; then, the related words set and corresponding weights used to query the forestry information feature expansion are obtained by using the algorithm of calculating the semantic similarity between the thesaurus; finally, according to the query expansion words of forestry information features and their corresponding weights, quantitative analysis, and ranking of web page information about forestry information features are carried out [18].

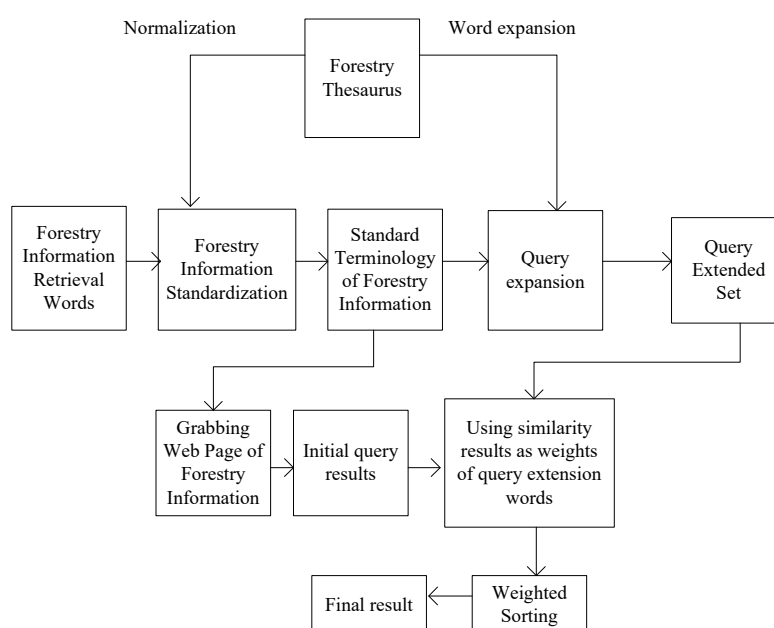The model framework is shown in Figure 2.

Figure 2. A Thesaurus-based fast retrieval model framework for forestry information features.

(2) Standardization of terminology of forestry information features

Firstly, the retrieval terms of forestry information features input by users are extracted, and the standardization of them is judged according to the thesaurus of forestry information features [19–21]. Due to the different user's retrieval needs and input habits, there may be four different situations: If the forestry information feature retrieval words are thesauri, they need not be standardized and can be used directly; if the forestry information feature retrieval words are non-thesauri in thesaurus, they can be transformed into corresponding thesauri through the equivalent relationship of thesaurus; if the forestry information feature retrieval words can match the part of the thesaurus, all the matchable thesauri will be returned for users to select new forestry information feature retrieval words from them; in other cases, the original feature search words are retained without query expansion.

(3) Grabbing web pages of forestry information features

Through standardization of forestry information feature terms, the retrieval feature words are K, and the general search engine uses K as the retrieval feature words of forestry information [22–24]. Take the URL of s results. Htmlparser, an open source web page analysis tool, was used to analyze the forestry information pages corresponding to the s Web site and extract the title, abstract, body, and other information from the pages.

(4) Forestry information feature query expansion

The similarity calculation method is used to calculate the similarity of all K-related forestry information feature words in the forestry information feature thesaurus. By setting threshold, eligible related words are selected and added to the forestry information feature query expansion set N.

The calculation method of similarity between all words in the forestry information feature thesaurus and K-related forestry information feature words is as follows:

Assuming that all the similarities of the two are within $[0, 1]$, if the weight is 0, there is no relationship between the two forestry information descriptions; if the weight is 1, the two forestry descriptions are equivalent. At the same time, it is stipulated that if the two thesauri are located in different conceptual trees, the similarity is considered to be 0.

Considering the similarity of forestry information words C_1 and C_2, the similarity formulas are divided into three categories according to the relationship type between C_1 and C_2: $SimD(C_1, C_2)$ for equivalent similarity, $SimF(C_1, C_2)$ for generic similarity, and $SimW(C_1, C_2)$ for correlation similarity.

a. Equivalent similarity $SimD(C_1, C_2)$. In the forestry information thesaurus, the equivalence relation is equivalent, that is, the two words can be used interchangeably.

$$SimD(C_1, C_2) = 1 \tag{5}$$

b. generic similarity $SimF(C_1, C_2)$:

$$SimF(C_1, C_2) = f_1 \times f_2 \times f_3 \tag{6}$$

Among them, f_1 is based on the similarity of the shortest path, $f_1 = e^{-ad}$ (d is the shortest path length from C_1 to C_2 in the thesaurus, a is the adjustment factor), f_2 is based on the similarity of the nearest root depth, and f_3 is based on the similarity of the semantic vector.

c. Relevant similarity $SimW(C_1, C_2)$:

$$SimW(C_1, C_2) = g_1 \times g_2 \tag{7}$$

Among them, C_1 is the related forestry information descriptor of C_2; g_1 is the similarity of forestry information features based on the depth of the related descriptor; and g_2 is the similarity based on the density of the related descriptor.

(5) Weighted sorting of forestry information features

When calculating the weighting of forestry information features, the results of similarity between the related words of forestry information features in N and K are taken as the weights of the related words. The specific steps of the weighted ranking method of forestry information features are as follows:

Step 1: Statistical forestry information feature query expansion set of each forestry information feature related words in the title of the page and the frequency of occurrence in the body of the page P.

Step 2: Calculate the weights of each web page. The formula is as follows:

$$TW_n = \frac{\sum\limits_{i=1}^{m} W_i \times (\omega \times T_i + P_i)}{WN_n} \tag{8}$$

Among them, TW_n is the total weight of the nth web page; WN_n is the word number of the n-th web page; m is the number of words related to forestry information features in the extended set n for querying forestry information features; W_i is the similarity between the *i*-th forestry information feature related words in n and the forestry information retrieval word K; T_i and P_i are the frequency of the words appearing in the title and text of the first web page, respectively; ω is the ratio of heading to text, which is used to adjust the importance of heading to the final result. The bigger the ω, the bigger the impact of the title on the weight of web pages.

Step 3: Sort the web pages according to the weights of forestry information features from large to small and return them to users [25]. Thus, the fast retrieval of forestry information features is completed.

3. Results

3.1. Experiment Setup

(1) Selection of evaluation index of forestry information feature retrieval effect based on the symmetry function under the communication network.

The retrieval effect of forestry information feature based on symmetry function in communication network refers to the effective result of information retrieval using the retrieval method, which directly reflects the performance of the retrieval method. Generally, recall and precision of retrieval results based on the relevance of forestry information features are the main indicators of traditional search engine evaluation. This paper chooses two indexes of relevance and search length to evaluate the validity of this method.

Considering that most users can only look at the results of the first page returned when retrieving forestry information, this paper chooses the correlation of the first 10 results when evaluating, which is expressed by *P@10*. The calculation method is as follows:

$$P@10 = \frac{a}{a+b} \tag{9}$$

Among them, *a* represents the number of results related to user's forestry information features in the first 10 results of retrieval, while *b* represents the number of results unrelated to user's retrieval terms in the first 10 results. Thus, the average correlation formulas of forestry information features of the first 10 items can be obtained:

$$\overline{P@10} = \frac{1}{n}\sum_{i=1}^{n} P_i \tag{10}$$

P_1 to P_n is obtained from *n* independent experiments.

The search length of forestry information features is set to the number of irrelevant results that need to be viewed to find the first five relevant results, and the search length is expressed in *L*. Similarly, the average search length formula of forestry information features can be obtained.

$$\overline{L} = \frac{1}{n}\sum_{i=1}^{n} L_i \tag{11}$$

Among them, L_1 to L_n are obtained from n independent experiments.

(2) Measurement of weights of relevant parameters for feature retrieval of forestry information based on symmetry function in the communication network.

Two important parameters were determined by experiments: Threshold Q used in forestry information feature query expansion module and heading text ratio ω in weighted ranking. In order to determine the characteristic weights of forestry information as accurately as possible, 10 forestry terms were randomly selected from the experimental data for testing. In the experiment, the first 100 search results of Baidu search engine are selected as the results of the general search engine to grab the forestry information web pages, and the title text of the forestry information feature retrieval is set to 1 first. Relevant forestry personnel confirm whether the returned results are related to the search terms. Use the final result to draw a broken line, as shown in Figure 3. In Figure 3, $\overline{P@10}$ represents the average correlation of forestry information features of the first 10 items, and $\overline{L}$ represents the average search length of forestry information features.
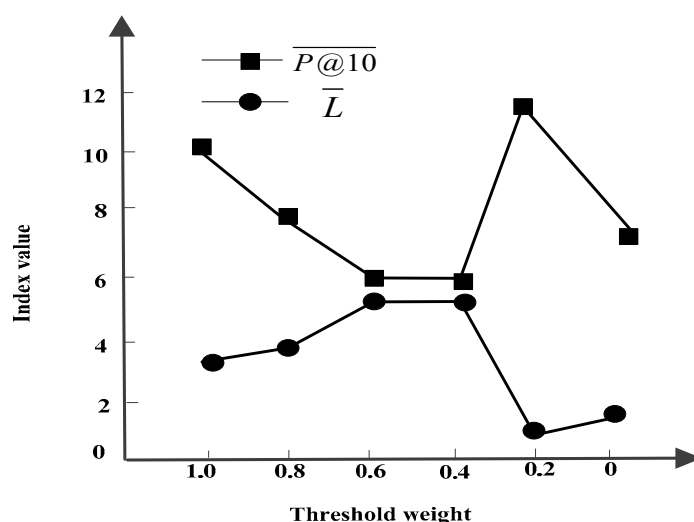

Figure 3. Determination data of threshold value.

Figure 3 shows that when the threshold value is 0.2, the data value of $\overline{P@10}$ is the highest, that is, the correlation degree of the first 10 results is the highest; $\overline{L}$ is the lowest, that is, the least irrelevant results need to be browsed to find the first 10 related results. Therefore, the threshold determination is 0.2.

By using the determined threshold, we can select the words closest to the search terms from the thesaurus for query expansion. Taking summer green forest as an example, the relevant terms can be obtained by defining the threshold value as follows: deciduous broad-leaved forest (0.8179), oak forest (0.6073), alder forest (0.6703), broad-leaved forest (0.6683), evergreen broad-leaved forest (0.5479), illuminated forest (0.5479), evergreen forest (0.5477), and hard-leaved evergreen forest (0.4889), in which the number in brackets is the similarity with the search terms.

After getting the threshold result, the threshold value is adjusted to 0.2, and the text-to-text ratio of titles is continued with these 10 descriptions. Similarly, the final experimental results are used to plot the broken lines, as shown in Figure 4.
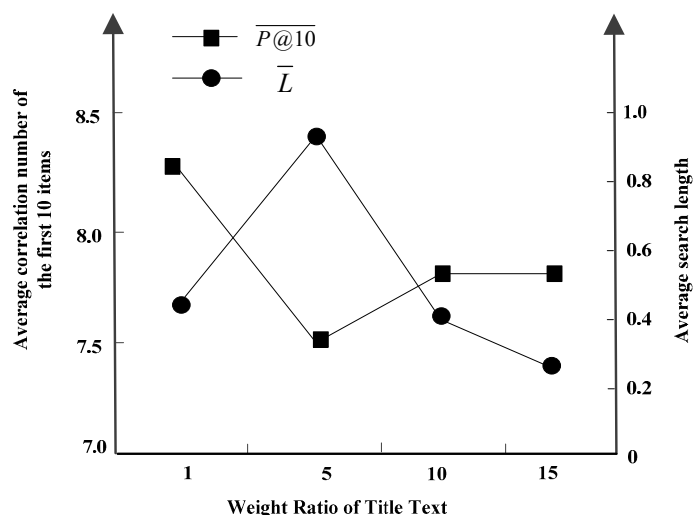
Figure 4. Determination data of title-text rate.

As can be seen from Figure 4, when the heading text ratio ω is 5, the $\overline{P@10}$ value is the highest, while $\overline{L}$ is the smallest. Therefore, by synthesizing two data, the heading text ratio ω is determined to be 5.

3.2. Analysis of Experimental Results

According to the optimal weights determined by 3.1, 15 words with forestry information features were randomly selected from the experimental data and retrieved by incremental crawling retrieval method, Lucid-based multi-channel retrieval method and this method. The $\overline{P@10}$ and $\overline{L}$ indices of the returned results were measured under different conditions, and the experimental results were drawn as Table 2:

Table 2. Comparison of three methods.

Serial Number	Search Terms	$\overline{P@10}$			$\overline{L}$		
		Incremental Crawling Retrieval Method	Lucid-Based Multi-Channel Retrieval Method	Article Method	Incremental Crawling Retrieval Method	Lucid-Based Multi-Channel Retrieval Method	Article Method
1	Summer green forest	5	7	8	5	0	0
2	Rainforest	5	5	7	3	5	1
3	Sparse forest	5	5	5	5	4	5
4	Redwood	2	4	9	14	8	1
5	Red spruce	6	4	9	3	8	0
6	Black spruce	4	4	8	6	7	0
7	White fir	6	5	5	3	2	0
8	Japanese hemlock	5	6	5	3	0	0
9	Fir forest	8	7	8	1	1	0
10	Ye Ye Lin	7	8	9	1	1	0
11	Larix gmelinii forest	6	6	7	4	1	3
12	Evergreen deciduous broad-leaved mixed forest	9	10	10	0	0	0
13	Seed forest	6	6	9	3	2	0
14	General Fazhenglin	6	8	9	2	0	0
15	Pond cypress	8	9	9	0	0	0

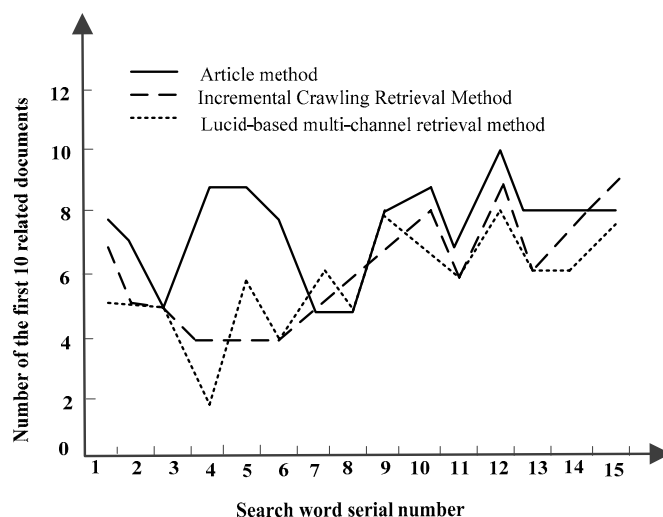According to the results of Table 2, make a broken line chart, as shown in Figures 5 and 6.

Figure 5. Comparison of retrieval results of three systems based on the number of the first 10 related documents.
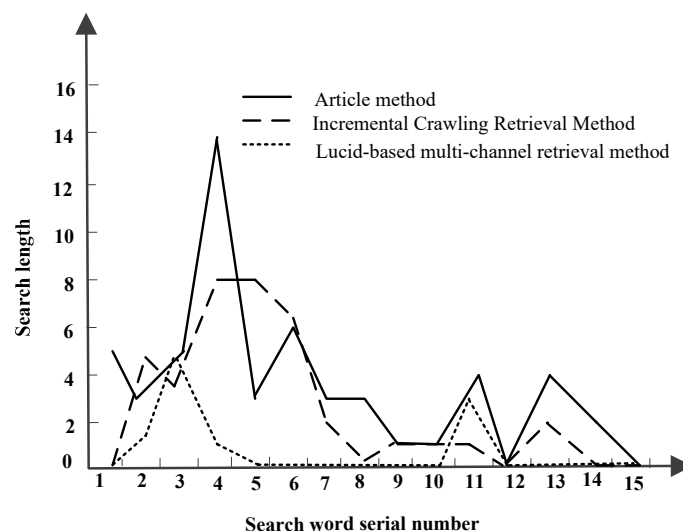

Figure 6. Comparison of retrieval results of three systems based on search length.

Analysis of Figure 5 shows that, under the premise of the same word number in forestry information retrieval, the number of documents related to forestry information retrieved by this method is always more than that of other two methods, and the maximum number of documents retrieved by this method is 10, while the maximum number of documents retrieved by the incremental crawling method and Lucid-based multi-way retrieval method are 9 and 8, respectively. It can be seen that this method can retrieve more comprehensive information according to the features of forestry information and has superior retrieval performance.

Analysis of Figure 6 shows that, under the premise of the same word number in forestry information retrieval, the search length of this method is always greater than that of the other two methods, and the maximum search length is 14, while the maximum number of documents retrieved by incremental crawling retrieval method and Lucid multi-way retrieval method are 8 and 5, respectively. Thus, the retrieval length of forestry information retrieval words in this method is higher.

Ten retrieval terms with different forestry information features are set up, and three methods are used to identify and retrieve them. The recall and precision of the three methods are analyzed. Table 3 describes the related concepts of 10 retrieval terms, and Table 4 compares the results of three methods.

Table 3. Relevant concepts of 10 search terms.

Serial Number	Search Terms	Relevant Concepts (Part)
1	Wood strength	Compressive strength along grain, tensile strength along grain, shear strength along grain, compressive strength along grain, tensile strength along grain, shear strength along grain, wood material properties
2	Wood drying	Log drying, fiber drying, debris drying, wood chips drying, finished wood drying, board drying, embryo drying
3	Wood heating	Microwave heating, high frequency heating, radiation heating, dielectric heating, heater, heating device, contact high frequency mixed heating
4	Early wood	Spring wood and timber
5	Modification treatment	Wood treatment, spraying treatment, oxidation treatment, volume stabilization treatment, plasticizing treatment, strengthening treatment, acid resistance treatment, moisture resistance treatment
6	Flame retardant treatment	Fire retardant treatment, fire resistant treatment, brushing treatment, wood treatment, spraying treatment
7	Wood defects	Shrinkage, central hardening, cracking, abnormal structure, cracking, decay of heartwood, discoloration of heartwood, twill, cracking, natural defects
8	Wood processing	Wood Processing Technology, Wood Processing Machinery, Wood Defects, Wood Processing Plant, Wood Processing Industry, Planer, Reprocessing
9	Wood texture	Spiral, straight, twill, interlaced, waveform
10	Wood preservation	Wood Protective Agents, Wood Preservation Treatment, Wood Enterprises, Gun Injection, Pressure Injection, Wood Moth Prevention

Table 4. Comparison of retrieval effects of three methods (Company/%).

Serial Number	Search Terms	Article Method		Incremental Crawling Retrieval Method		Lucid-Based Multi-Channel Retrieval Method	
		Recall Rate/%	Precision Rate/%	Recall Rate/%	Precision Rate/%	Recall Rate/%	Precision Rate/%
1	Wood strength	98.99	98.54	65.43	63.21	54.24	53.63
2	Wood drying	98.87	99.43	65.76	64.23	54.24	53.63
3	Wood heating	98.99	99.45	67.54	65.34	56.43	56.43
4	Early wood	98.67	99.53	64.56	65.33	53.63	53.63
5	Modification treatment	99.34	99.23	65.35	65.42	54.24	53.63
6	Flame retardant treatment	99.56	99.54	65.32	65.43	54.24	56.43
7	Wood defects	99.54	99.56	65.21	63.23	54.24	56.43
8	Wood processing	99.56	98.99	65.68	66.43	54.24	56.43
9	Wood texture	99.43	98.57	65.67	66.34	54.24	56.43
10	Wood preservation	99.54	99.56	65.24	64.21	54.24	56.43
Mean value	-	99.25	99.24	65.58	64.92	54.41	55.31

The above two tables show that the average recall rate of this method is 99.25%, 33.67% higher than that of incremental crawling retrieval method, 44.84% higher than that of Lucid multi-way retrieval method; the average recall rate of this method is 99.24%, which is 34.32% higher than that of incremental crawling retrieval method and 43.93% higher than that of Lucid multi-channel retrieval

method. Thus, the recall and precision of this method are the highest, and the retrieval performance is the best.

In order to further verify the performance of the proposed method, the retrieval time of the three methods is compared and analyzed. The results are shown in Figure 7.
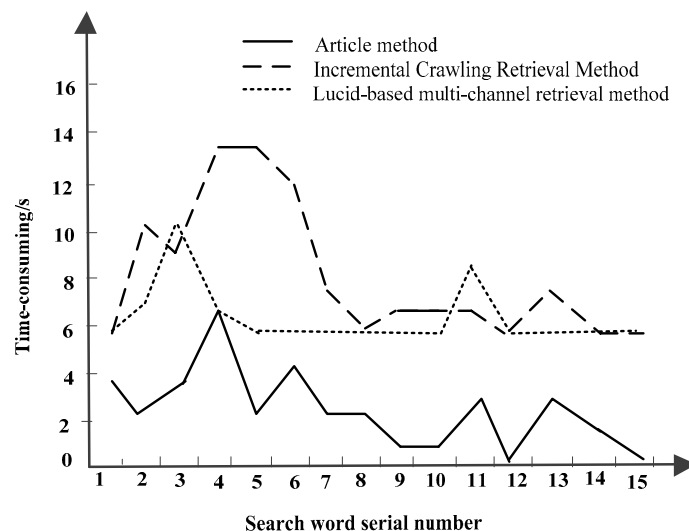

Figure 7. Comparison of retrieval time between different methods.

As can be seen from Figure 7, the retrieval time of the method is less than 6 s, and the time-consuming average of the other two methods is about 10 s, which is much higher than the method in this paper, thus the low time-consuming performance of the method is proved, which shows that the method can realize the quick retrieval of the characteristics of the forestry information.

4. Discussions

Based on the research contents of this paper, the following measures are put forward to speed up the process of forestry informatization construction in China:

(1) To raise awareness and unify ideas. The application of information technology can be said to be a technological revolution, but in the process of forestry informationization construction, we need to have a deep understanding of forestry informationization development from the ideological point of view. Especially as a leading cadre of forestry department, we need to accelerate the transformation of ideas, emancipate our minds, and make full use of forestry informationization to accelerate the realization of forestry modernization. As the leading cadres of forestry departments at all levels, we need to make solid progress in forestry informatization construction, conduct investigation and research, deploy and coordinate in person, supervise the implementation of forestry informatization construction, and check and accept the work in person. As the leader of forestry information construction, we need to speed up the renewal of our knowledge structure, take the lead in learning and propaganda, and speed up the smooth development of forestry information construction.

(2) Establish the forestry information network system covering the whole country as soon as possible. The forestry information system needs to take the information center of the State Forestry Administration as the hub, and connect effectively with the forestry information system of provinces, cities, and counties, so as to realize the integration of forestry information collection, and the use the market-oriented way to carry out the forestry information system. The operation will accelerate the overall coverage of the national forestry information network system and ensure the healthy development of the forestry information network.

(3) Give full play to the information platform to serve the coordinated development of forestry. Through the platform of forestry information network system, we can publicize and popularize

forestry policies, regulations and forestry legal knowledge, better display forestry style, and improve forestry awareness, which is of great significance to the whole society's in-depth understanding of forestry. We can publish the latest forestry policy through the network platform, open up a platform for communication and interaction on the network, and conduct competitive questions and answers on forestry technology information to accumulate more abundant knowledge for forestry workers, so that they can better provide high-quality services for forestry development.

(4) Improving forestry information database. Establish various forestry information databases for collecting, summarizing and storing relevant forestry information. For example, the establishment of databases of forest, sandy land, wetland, and biodiversity resources, or the establishment of public basic databases of basic geographic information and remote sensing image data. In this process, forestry departments at all levels should clearly classify various information resources and clarify the responsibilities and specific authority of each department in the construction and maintenance of databases.

5. Conclusions

This paper presents a fast retrieval method of forestry information features based on symmetry function in communication network. Firstly, forestry information collection method based on PDA is adopted. This method can not only navigate and locate, but also collect data at any time, and automatically complete various calculation and storage functions, which greatly reduces the labor intensity of personnel. As the signal received by PDA is susceptible to the influence of noise when collecting forestry information, which affects the integrity and accuracy of information, this paper adopts the method of forestry signal noise cancellation based on symmetric function method to filter forestry signal and improve the accuracy of information. It can cancel the noise and retain the signal items, use the fast retrieval method of forestry information features to construct the forestry information data ontology after denoising to integrate the forestry resources, and establish the concept set of forestry information attributes to distinguish the forestry information attributes. Finally, a fast retrieval model of forestry information features based on thesaurus is established to complete the fast retrieval of forestry information features.

Compared with incremental crawling retrieval method and Lucid-based multi-channel retrieval method, the recall and precision of this method are 99.25% and 99.24%, which have high recall and precision.

In future studies, we can try to learn embeddings that capture the information that you are currently modelling with traditional NLP approaches, and in-depth study on related issues such as information retrieval.

Author Contributions: Conceptualization, H.W. and J.S.; methodology, H.W.; software, H.W.; validation, H.W. and J.S.; formal analysis, J.S.; investigation, J.S.; resources, H.W.; data curation, J.S.; writing—original draft preparation, H.W.; writing—review and editing, J.S.; visualization, J.S.; supervision, J.S.; project administration, H.W.; funding acquisition, H.W.

References

1. Xu, M.; Tian, S.Q.; Chen, Z. Research on Incremental Crawler Based on the Web Information of Forest Products Trade. *Agric. Netw. Inf.* **2016**, *2*, 18–21.
2. Li, J.; Zhang, X.B.; Xue, J. Building Expert Diagnosis System for Bamboo Pests Based on Lucid Multi-Way Identification. *J. Zhejiang A F Univ.* **2016**, *33*, 122–129.
3. Shaikh, M.; Nava, B.; Kashcheyev, A. A model-assisted radio occultation data inversion method based on data ingestion into NeQuick. *Adv. Space Res.* **2017**, *59*, 326–336. [CrossRef]

4. Kusner, M.J.; Sun, Y.; Kolkin, N.I.; Weinberger, K. From word embeddings to document distances. In Proceedings of the International Conference on International Conference on Machine Learning, Lille, France, 6–11 July 2015.

5. Passalis, N.; Tefas, A. Learning Bag-of-Embedded-Words Representations for Textual Information Retrieval. *Pattern Recognit.* **2018**, *81*, 254–267. [CrossRef]

6. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805.

7. Deschenaux, R.; Martin, S.A.; Vilches, M.T. Switchable Mesomorphic Materials Based on the Ferrocene−Ferrocenium Redox System: Electron-Transfer-Generated Columnar Liquid-Crystalline Phases. *Organometallics* **2017**, *18*, 5553–5559. [CrossRef]

8. Zhou, K.; Martin, A.; Pan, Q.; Liu, Z.-G. Median evidential c-means algorithm and its application to community detection. *Knowl.-Based Syst.* **2015**, *74*, 69–88. [CrossRef]

9. Fu, C.; Song, J.Q. Design and Realization of Web Military Intelligence Mining System Based on Document Clustering. *J. China Acad. Electron. Inf. Technol.* **2015**, *10*, 541–545.

10. Brousseau, L.C.; Williams, D.J.; Kouvetakism, A. Synthetic Routes to Ga(CN)3 and MGa(CN)4 (M = Li, Cu) Framework Structures. *J. Am. Chem. Soc.* **2015**, *119*, 6292–6296. [CrossRef]

11. Zhang, Y.; Lv, P.; Huo, Y.J. An Alternative Certification Method for Conducted Emissions of SMPS. *J. Power Supply* **2016**, *14*, 166–170.

12. Ma, A.; Li, J.; Yuen, P.C. Cross-Domain Person Re-Identification Using Domain Adaptation Ranking SVMs. *IEEE Trans. Image Process.* **2015**, *24*, 1599–1613. [CrossRef] [PubMed]

13. Yang, J.; Jiang, B.; Li, B.; Tian, K.; Lv, Z.H. A Fast Image Retrieval Method Designed for Network Big Data. *IEEE Trans. Ind. Inform.* **2017**, *13*, 2350–2359. [CrossRef]

14. Jie, Z. Fast retrieval algorithm of feature information in database. *J. Discret. Math. Sci. Cryptogr.* **2017**, *20*, 1507–1511.

15. Shen, X.J.; Ye, M.M.; Gan, T.; Han, D.J. Information retrieval based on concept lattice and its tree visualization. *Comput. Eng. Appl.* **2017**, *53*, 95–99.

16. Gupta, T.; Altman, M.; Shukla, A.D. Covalent Assembled Osmium-Chromophore-Based Monolayers: Chemically Induced Modulation of Optical Properties in the Visible Region. *Chem. Mater.* **2015**, *18*, 142–156. [CrossRef]

17. Zhang, J.H. Simulation of Image Retrieval Model by Considering Overlapping Feature Classification. *Comput. Simul.* **2016**, *33*, 431–434. [CrossRef]

18. Chaudhuri, R.; Fiete, I. Computational Principles of Memory. *Nat. Neurosci.* **2016**, *19*, 394–403. [CrossRef] [PubMed]

19. Drémeau, A.; Liutkus, A.; Martina, D.; Katz, O.; Schülke, C.; Krzakala, F.; Gigan, S.; Daudet, L. Reference-less measurement of the transmission matrix of a highly scattering material using a DMD and phase retrieval techniques. *Opt. Express* **2015**, *23*, 11898–11911. [CrossRef] [PubMed]

20. Agliari, E.; Barra, A.; Ferraro, G.D. Anergy in Self-Directed B Lymphocytes: A Statistical Mechanics Perspective. *J. Theor. Biol.* **2015**, *375*, 21–31. [CrossRef]

21. Belloul, M.B.; Hauchecorne, A. Effect of Periodic Horizontal Gradients on the Retrieval of Atmospheric Profiles from Occultation Measurements. *Radio Sci.* **2016**, *32*, 469–478. [CrossRef]

22. Schultz, D.; Spiegel, S.; Marwan, N. Approximation of Diagonal Line Based Measures in Recurrence Quantification Analysis. *Phys. Lett. A* **2015**, *379*, 997–1011. [CrossRef]

23. Kravchenko, A.N.; Robertson, G.P. Statistical Challenges in Analyses of Chamber-Based Soil CO_2 and N_2O Emissions Data. *Soil Sci. Soc. Am. J.* **2015**, *79*, 200–201. [CrossRef]

24. Vandaele, A.C.; Chamberlain, S.; Mahieux, A.; Ristic, B.; Robert, S.; Thomas, I.; Trompet, L.; Wilquet, V.; Belyaev, D.; Fedorova, A.; et al. Contribution from SOIR/VEX to the updated Venus International Reference Atmosphere (VIRA). *Adv. Space Res.* **2016**, *57*, 443–458. [CrossRef]

25. Buchbinder, A.M.; Gibbsdavis, J.M.; Stokes, G.Y. Method for Evaluating Vibrational Mode Assignments in Surface-Bound Cyclic Hydrocarbons Using Sum-Frequency Generation. *J. Phys. Chem. C* **2016**, *115*, 18284–18294. [CrossRef]