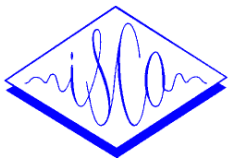
ISCA Archive
http://www.isca-speech.org/archive



First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

Design Principle of Language Model for Speech Recognition

Toshiya SAKANO Tsuyoshi MORIMOTO

ATR Interpreting Telephony Research Laboratories
Seika-cho, Souraku-gun, Kyoto 619-02, Japan

ABSTRACT

In speech recognition, the recognition rate can be improved by using statistical information about linguistic texts. However, there is no criterion for selecting sample linguistic texts from those available. If unbalanced linguistic text samples are selected, the information extracted would not be suitable for a linguistic model. To solve this problem, we need to quantitatively analyze linguistic texts.

In this paper, we introduce a method to quantitatively analyze linguistic texts, and propose a method for selecting linguistic texts. Moreover, we describe the possibility of developing a criterion for selecting test texts needed for system evaluation, and show the relationship between recognition system performance and the features of the sample text group used for the linguistic model.

1 Introduction

Modeling is a significant element of speech recognition. Language models are necessary in order to achieve effective recognition, for example, a stochastic grammar which can parse sentences effectively. Of course, a language model is established from linguistic data such as conversation texts. Because it is necessary that linguistic data be parsed, and their grammatical structures determined, it is very difficult to gather sufficient linguistic data to reflect all natural language features. Accordingly, until now, as much linguistic data as possible has been gathered to establish language models with regard to whether the linguistic data are necessary or not. Such linguistic data are very contrived, that is, not natural. When the gathered linguistic data are unbalanced with respect to unnecessary features, the language model made from them may also be unbalanced with respect to the unnecessary features. If a language model is to be made with respect to grammatical features, the unbalance with respect to unnecessary features must be removed. Therefore, sample data to make a language model with respect to grammatical features should be selected uniformly with respect to unnecessary features. Similarly, in order to test the language model uniformly, the linguistic test data should be selected uniformly with respect to grammatical features.

This paper propose a metric between linguistic texts, and describes the characteristics of a text space derived from the metric by using quantification theory IV.

2 Quantification Theory IV

The quantification theory is one of the quantification methods proposed by Chikio Hayashi. There are four quantification theories, and quantification theory IV is the fourth one. Quantification theory IV can quantify data which have no numerical value. For example, because linguistic texts are not numerical, we adopted the quantification theory to quantify the linguistic texts.

For this procedure, similarity among data must be defined. By this method, data whose similarity values are approximately the same are quantified closely, and data whose similarity values are very different are quantified distantly. Mathematically, quantification theory IV is used to calculate the eigenvalues of a certain symmetric matrix and their eigenvectors. The matrix is derived from data similarity.

3 Metric and Similarity of Conversation Texts

3.1 Eample 1: *Topical Metric and Similarity*

Conversation texts include nouns, and the patterns of noun co-occurrence are all different. We think that the pattern of noun co-occurrence in a conversation text reflects its topics. The distribution of noun occurrence in a conversation text can be regarded as its pattern of noun co-occurrence.

Let t_i and t_j be conversation texts and let P_i and P_j be the probability distribution functions for the nouns in t_j and t_j respectively. For example, $P_i(ns)$ denotes the probability that noun ns exists in the conversation text t_i. Then, the metric d_w between t_i and t_j is defined as follows:

$$d_w(t_i, t_j) = \sqrt{\sum_{w \in t_i \cup t_j} \{P_i(w) - P_j(w)\}^2}$$

If t_i is the same conversation text as t_j, $d_w(t_i, t_j)$ is 0. However, the converse is not true. The value range of $d_w(t_i, t_j)$ is from 0 to $\sqrt{2}$(this is proved very easily).

Now we can define the noun pattern similarity e_w as follows:

$$e_w(t_i, t_j) = -d_w(t_i, t_j)$$

3.2 Example 2: *Grammatical Metric and Similarity*

Each sentence in a linguistic text is able to be parsed using a linguistic grammar. Usually, the parsed result is a sequence of rewriting grammar rules. That is, a parsed result of a sentence reflects its pure grammatical structure. Therefore, a conversation text is regarded as a set consisting of sequences of rewriting grammar rules by a linguistic grammar.

In the same manner as noun pattern similarity, a grammatical pattern similarity can be defined. Let t_i and t_j be conversation texts and let P_i and P_j be the probability distribution functions for the parsed results in t_j and t_j respectively. Then, the metric d_g between t_i and t_j is defined as follows:

$$d_g(t_i, t_j) = \sqrt{\sum_{g \in G} \{P_i(g) - P_j(g)\}^2}$$

(G is a treated grammar and g means a grammar pattern defined by the grammar G.)

Now we can define the grammatical pattern similarity e_g as follows:

$$e_g(t_i, t_j) = -d_g(t_i, t_j)$$

4 Sample Data for a Language Model

4.1 Filtering Out Unnecessary Elements from Sample Data

When statistical features are introduced into a linguistic grammar such as a stochastic grammar, they must be established by a linguistic corpus. Then, as many natural linguistic data as possible are necessary. The words "natural linguistic data" mean that the data is not unbalanced with respect to unnecessary elements other than grammatical properties. For example, generally, topics have nothing to do with such a grammatical elements. If the linguistic data is unbalanced with respect to properties which are not grammatical properties, the established stochastic grammar is effective to the superior elements which have nothing to do with grammatical properties of liguistic texts, but not effective to the inferior elements. Because we should make a stochastic grammar reflecting the pure grammatical properties, it is needed to reduce as many unnecessary properties as possible for the stochastic grammar.

For this purpose, we define a metric and a similarity between the texts with respect to unnecessary elements other than grammatical properties, and quantify the texts by using quantification theory IV with the metric and the similarity to represent the texts in the n-dimensional text space. In order to filter out the unnecessary elements, the mesh method can be used to select the sample texts from the n-dimensional text space with respect to the unnecessary elements. By this way, sample texts uniform with respect to unnecessary elements other than grammatical properties can be selected from linguistic text database. The details are explained in the following sections.

4.2 Mesh Method

Generally, conversation texts quantificated by using quantification theory IV are represented as points in an n-dimensional Euclidean space. We call this space n-dimensional text space. In order to uniformly select the sample texts from the text space, a mesh method is usefull. In this method, first, the text space is represented by cubes(called meshes) defined as a direct product of each axis separated into equally long segments, and texts of the same number are selected from each mesh. The number of selected texts depends on the size of the mesh and the number of texts selected from one mesh. However, the number of texts selected from one mesh depends on the size of the mesh. If the mesh is very small, there are some meshes with no texts. Therefore, the size of the mesh and the number of texts selected from one mesh should be determined prudently.

4.3 Leaning Degree

Definition 1 *Let $X = \{x_i\}_{0 \le i \le r}$ be a real number sequence such that*

$$x_0 \le x_1 \le \cdots \le x_r$$

Let $D = \{d_i\}_{1 \le i \le r}$ be the intervals set such that

$$d_1 = x_1 - x_0,\ d_2 = x_2 - x_1, \cdots,\ d_r = x_r - x_{r-1}$$

If $x_0 \neq x_r$, we define the leaning degree L of X as follows,

$$L = \frac{r\sqrt{V[D]}}{x_r - x_0} \quad (V[D] \text{ means the variance value of } D)$$

In this definition, if all numbers in X line up at regular intervals, the leaning degree of X is 0. The leaning degree L of X means how unbalanced the data in X are, that is, not uniform.

Definition 2 *Let $X = \{x_i\}_i$ be a sequence of points in a n-dimensional Euclidean space, and let $\{a_i\}_i$ be a family of*

orthonormal axes. Then, the leaning degree L of X is defined as the mean of the leaning degrees for $\{a_i\}_i$. *That is,*

$$L = \frac{\sum_{i=0}^{r} L_i}{n} \quad (L_i \textit{ is the leaning degree of } a_i.)$$

	Topics					
Texts	A	B	C	D	E	F
data.K010	⊗				⊗	
data.K011	⊗		⊗			
data.K012	⊗	⊗		⊗	⊗	
data.K013		○		○	○	○
data.K014		○		○	○	○
data.K015	○		○			
data.K016	⊗			⊗	⊗	
data.K017	○		○			
data.K018	⊗		⊗			
⋮						
data.T010	○		○		○	
data.T011	○		○			○
data.T012	⊗		⊗	⊗		⊗
data.T013	○		○			
data.T014	○		○			

A:conference B:transportation C:procedure
D:hotel E:information F:sight seeing

Table 1: Contents of Texts Files

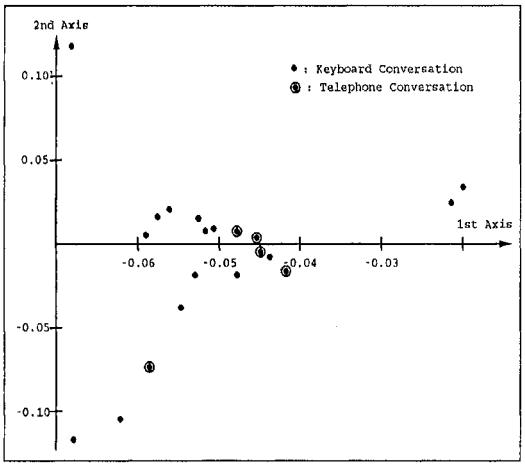


Figure 1: First and Second Axes of the Text Space

4.4 Example

We experimented with 20 linguistic texts. These linguistic texts are composed of simulated keyboard conversations

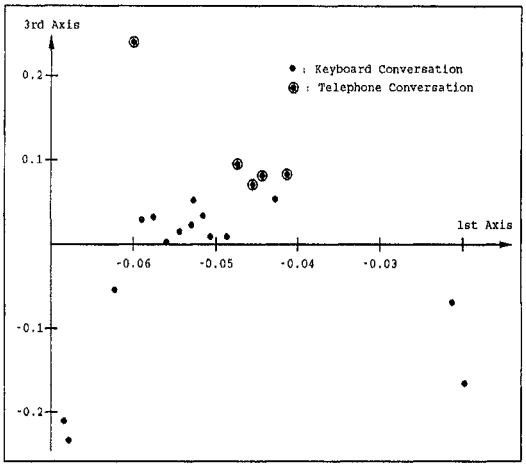


Figure 2: First and Third Axes of the Text Space

and telephone conversations . The keyboard conversation texts are stored in 15 files, and the telephone conversations texts in 5. Table 1 shows the texts and their contents as topics. The topics in table 1 are determined by reading them. We analyzed these texts using quantification theory IV with respect to topics of texts by using distributions of nouns in the texts. As a result, the texts can be represented in three-dimensional Euclidean space. Figures 1 and 2 show the space. Judging from the results, the first axis tends to be about sight-seeing or business management, the second axis tends to be about documents or payment, and the third axis tends to be about keyboard conversation or telephone conversation. Last, we selected six files, which are marked in the table 1, from all files by using the mesh method. The leaning degree of all 20 files is 3.032, and the leaning degree of the selected files is 1.462. From table 1, it can be easily understood that these selected files can cover any topic. This shows the effectiveness of the similarity e_w of the conversation text and the mesh method.

5 Density of Text Space

Density of text space is a significant feature. Leaning degree of text space is a statistical feature about the whole of points in the text space. The other side, density of text space is statistical feature about each point in the text space. The relationship between the density of text space with respect to grammar properties effects the performance of a speech recognition system. The definition of density of text space and the relationship between it and the performance of a speech recognition system are described in the following.

5.1 Leaning Density

Definition 3 *Let* $X = \{x_i\}_i$ *be a sequence of points in a*

n-dimensional Euclidean space. Then, the leaning density δ_i of x_i is defined as follows:

$$\delta_i' = \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{d_g(x_i, x_j)^2}$$

$$\delta_i = \frac{\delta_i'}{\sum_{j=1}^{n} \delta_j'}$$

The leaning density δ_i means a measure how many data crowd around x_i.

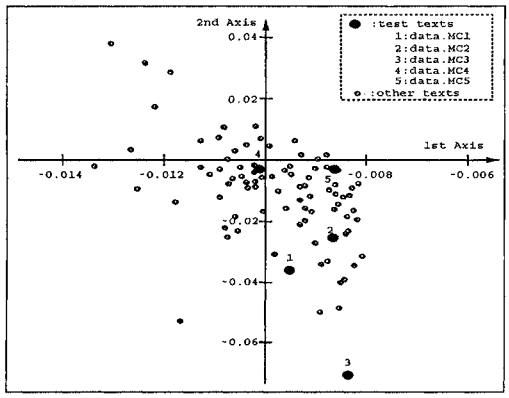


Figure 3: Grammar Text Space

	Texts	Leaning Density	Recognition Rate
1	data.MC1	0.0035	84 [%]
2	data.MC2	0.0036	81 [%]
3	data.MC3	0.0031	65 [%]
4	data.MC4	0.0043	74 [%]
5	data.MC5	0.0046	90 [%]

Table 2: Leaning Density and Recognition Rate of Test Texts

5.2 Relationship between Leaning Density and System Performance

First, we investigated the text space of 100 sample linguistic texts including the test texts by using quantification theory IV with respect to grammar features. Figure 3 shows the result, and the emphasized points in this figure indicate the test texts. The test texts exist at rather distant points.

Next, we obtained the relationship between leaning density and the performance of a speech recognition system using the test data. Table 2 shows the leaning densities of the five test texts in this text space and the recognition rate. There is a relationship between the leaning densities and the recognition rate of our system(except for 'data.MC4'). That is, the higher the leaning density, the worse the recognition rate of the system becomes.

Judging from these results, leaning density can be used as a measure to presume the recognition rate of a text. Then,the following methods can be possible:

- In the case that some target samples are given for speech recognition, it can improve the performance of the speech recognition system to learn the stochastic grammar of this system with the texts which increase the leaning density of the target samples.
- In the case that a stochastic grammar of a speech recognition system is learned with given sample texts, in order to precisely evaluate the system, the most suitable data(texts) can be selected from text space by mesh method.

6 Conclusion

In this paper, we proposed the metric of linguistic texts, and described that a text space can be derived from a metric, and leaning degree and leaning density can be defined for a measure to presume the recognition rate of a text. Moreover, suitable sample or test texts for modeling or evaluating a speech recognition system can be selected by using mesh method. These will be needed in the future when speech recognition systems become practical. While the experiments we tried are preliminary, in the future, we would like to experiment in detail with more linguistic texts.

ACKNOWLEDGMENTS

The authors are grateful to Dr.Kurematsu, the president of ATR Interpreting Telephony Research Laboratories, and all the members of the Knowledge and Data Base Department for their constant help and encouragement.

REFERENCES

[1] Biber, Douglas (1986). *Spoken and wriiten textual dimensions in English: resolving the contradictory findings.* Language 62,384-414

[2] Biber, Douglas (1989). *A typology of English texts.* Linguistic 27,3-43

[3] K.Kita,T.Kawabata,T.Hanazawa (1990). *HMM Continuous Speech Recognition Using Stochastic Language Models.* ICASSP90