ISSN 2347-1921

Application of Search Model to Detect Urhobo Names in Niger Delta Region of Nigeria: A Preliminary Study

F. Z. Okwonu
Department of Mathematics and Computer Science,
Delta State University, Abraka, Nigeria.

ABSTRACT

This paper is designed to determine the accuracy of search words or names from a database using search vector. Generally, database consists of large collection of information. By specifying the search words or names this model is able to retrieve the required information from the database relying strictly on the search vector. In this regards, the target names are assigned one if the names corresponds to the search vector and zero otherwise. This information is applied to develop the target matrix. Based on the target matrix, we can determine precisely the number of names that corresponds to the different name books. To further analyze the performance of this technique, the model is validated by post multiplying the variance by the target mean. The result revealed that the search model was able to detect accurately the required names.

Indexing terms/Keywords

Target name; Search name; Target matrix; Search vector

Academic Discipline And Sub-Disciplines

Mmathematics, Statistics

SUBJECT CLASSIFICATION

Mathematics Subject Classification:03C45, 03C65

Council for Innovative Research

Peer Review Research Publishing System

Journal: JOURNAL OF ADVANCES IN MATHEMATICS

Vol.11, No.3

www.cirjam.com , editorjam@gmail.com

ISSN 2347-1921

1. INTRODUCTION

The advent of digital libraries on the internet has given rise to different procedures to store information and this also led to different techniques of retrieving information on the internet. In recent past, different retrieval procedures have been developed which strictly depends on matrix theory (Leon S. J. 2006). These data are stored in a database, database comprises of collection of different documents. How do we search for the document of interest in the database? We may desire to search the database to obtain information that interests our search desire. However, we may assume that the database comprises of k documents and there are l defined words that can be applied to search for the target words. Generally, common words are not applicable in this consideration hence the target words are well defined. In this consideration, if the defined words are arranged in an orderly manner, then the words can be defined in a matrix form, say $k \times l$ matrix X. In this manner, a document is denoted by a column of the matrix, as such the first entry in the X matrix is found in the nth column which defines the number describing the relative frequency of the first defined target words in the nth document and the process continues for the second, third entries and so forth. To obtain the target words or names, a list of search words are well defined in a vector form, say $u,(k \times 1)$. The nth entry of the vector is one if the nth word or name is in the list of defined words and corresponds to the word in the desired target word or name, else is zero if the target word does not corresponds to the defined word. Based on the above discussion, how do we achieve the above feat?

Intuitively, one may consider the linear system or considering matrix theory or linear algebra. These methods directly relate to the statistical classification and discrimination procedure based on the Fisher's technique in passing. Conventionally, the Fisher linear discriminant technique is applied to discriminate, separate, classify and predict. The search model is a linear system just like the Fisher's approach. On the contrary, this technique does not apply the strictly basic Fisher's assumptions. Fukunaga noted that the Fisher's technique can be implemented based on the pooled covariance matrix (Fukunaga 1990). However, this search approach differs in this consideration in that the homoscedasticity assumption of the Fisher's technique is violated. Although, the Fisher linear discriminant analysis is based on distributional, homoscedasticity assumptions, dimension reduction and separation. Unlike the Fisher's technique, this model can be applied to search for specific words or names provided the database or information exists in well defined storage system. In passing, the equations are similar but differ with computational requirements and implementations. This paper is designed to search for target names from the list of well defined Urhobo names using artificial database developed using seven clans. This can be achieved by using the linear system otherwise called search model. This method is basically linear in nature, it encompasses target matrix and search vector. The search vector is used to search for the specified information.

The remainder of this paper is organized as follows. Section Two describe the search model and Section Three contains the application and analysis. Conclusion is contained in Section Four.

Search model

This section entails the description of how search vector is used to obtain desired information from a database. Although, we are specific in using local names in testing the model. The generalization to search for vital and complex information on the internet or well established database is straight forward. We assumed that keywords, brand names and virtually every existing thing can be transformed into matrix. In this consideration, we are interested in transforming letters or words into numbers, say corresponding search is assigned one and unmatched search is assigned zero. We express the preceding statement as follows;

$$x_{ij} = \begin{cases} 1 \; if \; the \quad desired \quad word \quad is \quad in \quad the \quad list \quad of \quad defined \quad words \\ 0 \; otherwise \end{cases} \qquad (1)$$

Based on the information provided by Equation (1) we can transform the output into the following system, say

$$W = X^T u, \qquad (2)$$

$$\bar{W} = w_i / k, i = 1, 2, .., k. \qquad (3)$$

Where X is $k \times l$ matrix and u is $k \times 1$ vector. Equation (2) is called the target matrix for the corresponding or target information and Equation (3) is the within target mean. To apply this model to search for names or words, we have to define key words or names which in this case correspond to the Urhobo given names. This kind of procedure requires a database consisting of the Urhobo names or simply put, one can pick different books consisting of the Urhobo names, say. The vital aspect of this search model depends on the search vector u which determine the target objectives. Whatever, the $k \times l$ matrix contains the corresponding names with one indicating correspond and zero not correspond. In order to determine the accuracy of this model, the following equation is applied

$$Z = \bar{W} \sqrt{diag(S^2(W))}, \qquad (4)$$

Where $S^2(W)$ is the target covariance matrix, say

$$S^2(W) = \frac{\sum_{i=1}^{k}(w_j - \bar{w}_i)(w_j - \bar{w}_i)'}{k-1}, \ j = 1, 2, \ldots. l.$$

Where w_j is the profile of the target word searched and $\bar{w}_i$ is the within search mean. Equation (4) is applicable if we want to determine the frequency or accuracy of search information obtained based on Equation (2). The search frequency of this technique depends on the value of Z. Like the Fisher linear discriminant analysis, if the coefficient of the profile variable is large it implies that the profile variable contribute more and hence discriminate more (Johnson 2007). In this case, the value of Z determine the frequency of the target word. Equation (2) determines how many of the target word are in each document.

2. Application and analysis

Suppose Table 1 is the database that contains names of seven Urhobo clans (X1, X2, ….,X7). The implication of this database is that the first clan (X1) contains three names; say, Efe, Ochuko and Eloho, same goes to the other clans with different name strength. The idea is that we are keen in searching for the target names in each of the clan and also to determine if our search method is accurate in determining the target names in each clan. We further determine the frequency of target names using Equation (4).

Table 1: Matrix representation for the database of Urhobo names

X1	**EFE**	**OCHUKO**	**ELOHO**	**-**
X2	KESIENA	ELOHO	ELO	OREVA
X3	ONOME	UFOUMA	EFE	-
X4	OCHUKO	ELO	UFOUMA	-
X5	OREVA	ELO	ONOME	-
X6	KESIENA	ELOHO	OCHUKO	-
X7	UFOUMA	ONOME	ELO	EFE

To transform the above information into database matrix, we represent each entry with one if the search name corresponds to the target name and zero if the search name does not correspond to the target name, say Equation (1), respectively. In the following, Table 1 is transformed into Table 2. In this case, the search technique only recognizes Eloho and read Elo as Eloho. In any case, any entry with Elo in the database will be translated by the search technique to Eloho as such, any entry with Elo will appear as non-corresponding target name. This is due to the sensitivity of the search model to the first three letters in both names (E, L, O), respectively.

Table 2. Database matrix and corresponding search vector

Names	**X1**	**X2**	**X3**	**X4**	**X5**	**X6**	**X7**
EFE	1	0	1	0	0	0	1
OCHUKO	1	0	0	1	0	1	0
ELOHO	1	1	0	0	0	1	0
KESIENA	0	1	0	0	0	1	0
OREVA	0	1	0	0	1	0	0
UFOUMA	0	0	1	1	0	0	1
ONOME	0	0	1	0	1	0	1

Using Equation (2), the information in Table 2 translates to the information in Table 3. In Table 3 below, the values in the diagonal correspond to the exact number of target names in each clan and the off diagonal corresponds to common target names that belong to all clans. Suppose we are keen in reporting clan in which all the names correspond to the target names, then the diagonal of Table 3 is reported. Suppose that we are also interested in reporting the corresponding names in each case we can report the off diagonal.

Table 3: Target names

	X1	X2	X3	X4	X5	X6	X7
X1	3	1	1	1	0	2	1
X2	1	3	0	0	1	2	0
X3	1	0	3	1	1	0	3
X4	1	0	1	2	0	1	1
X5	0	1	1	0	2	0	1
X6	2	2	0	1	0	3	0
X7	1	0	3	1	1	0	3

In practice, often common names do occur. Regionally for instance, the name Chukwu is common among the south east region (Ibos) of Nigeria but such names are not common in the south west of Nigeria. In this case, the uncommon names are represented with zero and the common names with one and the diagonal represent the frequency of common names. Based on the above, we apply Equation (4) to determine the frequency of common names. Using the equation, the following values are obtained

$$Z=[1.2229, 1.1547,1.6117, 0.5399,1.3885, 1.6117].$$

The computation revealed that X3 and X7 have the highest number of common names followed by X6 and X1, respectively.

3. Conclusion

The search model applied to search for target names actually detected the exact names and revealed that the model is effective and accurate in performing search. The model was able to reveal common names and uncommon names based on the application of the search vector. We also developed model that help to determine the frequency of the target names and the analysis corresponds to the result obtained. In general, the computational requirements of these models are simple and easy to implement.

Reference

1. Fukunaga, K., Ed. (1990). Introduction to statistical pattern recognition, Academic Press, Boston.
2. Johnson, R. A., and Wichern, D. W., Ed. (2007). Applied multivariate statistical analysis, Prentice Hall, New York.
3. Leon S. J., Ed. (2006). Linear Algebra with application, Pearson education international.