A STUDY ON FEATURE SUBSET SELECTION USING IMAGE

MRS.DEEPIKA.S[1], MS.KALAIYARASI.M[2]
MRS.S.DEEPIKA
ASSISTANT PROFESSOR,DEPARTMENT OF B.COM BUSINESS ANALYTICS
PSGR KRISHNAMMAL COLLEGE FOR WOMEN, COIMBATORE
deepika@psgrkcw.ac.in
MS.M. KALAIYARASI
UG SCHOLAR, B.COM (BUSINESS ANALYTICS),
PSGR KRISHNAMMAL COLLEGE FOR WOMEN, COIMBATORE
kalaipsgr2114@gmail.com

ABSTRACT:
The paper mainly aims at retrieve the images based on normal attributes in image retrieval system. An image retrieval system is a computer system for browsing, searching and retrieving images from a large databases of digital images. The image retrieval often needs to optimize the trade of between efficiency and precision. This is analysed using created various vectors with the attributes and plotted seaborn graph and scatter plot.

KEY WORDS: normal attributes, image retrieval, vectors, seaborn graph, scatter plot.

I. INTRODUCTION

The analysis is made with the python the dataset has been imported to jupyter notebook and the python is a object oriented and a high level programming language. The analysis is made through the vectors creating, seaborn graph and scatter plot. Seaborn graph is library for making statistical graphics in python. Seaborn helps explore and understand data easily. A scatter plot is a diagram where the each value in the dataset is represented by a dot.In python matplotlib , the scatter plot can be created using the pyplot.plot().
Image retrieval system to retrieve more semantically relevant images. This target is closely related to the goal of distance metric learning in machine learning field. Distance metric learning is to learn a distance metric for the input space of data from a given collection of labeled points that preserves the distance relation among the training data [1].

II. REVIEW OF LITERATURE

Image retrieval as the search for the best database match to a user-provided query image, it was quickly realized that the design of fully functional retrieval systems would require support for semantic queries [2].
Equivalently, unsupervised learning has been adapted to speedup retrieval process and enhances visualization performance when the images are not labelled or annotated [3].
The typical goal of image retrieval system is to find relevant images to a given query when meta data is missing or unavailable. However, the uploaded digital images on a daily basis to image databases are rarely coupled with relevant labels or keywords. This triggered researches on automatic image annotation approaches [4].
Researches have confirmed the limitations of single similarity measure to yield perceptually meaningful and robust image ranking. Learning based solutions have been proposed as promising an alternative to overcome this weakness. In particular, image categorization/classification has been designed as a preprocessing phase to speed up image retrieval from the large collection [5].

III. METHODOLOGY

A. Creating vector

A vector is a single dimensional array. It occupies the elements in a similar manner as that of a python. Creating vector for comparing the attributes. Vectors are build from components, which are ordinary numbers and attributes.

B. Seaborn graph

Seaborn is a python visualization library and it is used for plotting it provides high level visualising graphs. Seaborn graph is a easily understandable graphical method. It is plotting functions operate on dataframes
and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots.

C. Scatter plot

A scatter plot is a diagram where each value in the dataset is represented by a dot. Draw a graph with the independent variable on the horizontal and the dependent variable on the vertical axis. The plot method gives output in dot graphical format.

IV.   RESULT

INDEXING VECTOR

Fig 1

To create a new vector in jupyter consists of attributes and retrieve the data using normal attributes. Creating new vector with image feature, rank, category, sub category. It shows the most searching category by using ranking method with normal attributes. The above image represent most searching item is food category it was displayed by ranking.

Str1 = data[['image feature, 'sub category', 'rank', 'category']]

Fig 2

Creating vector with image quality, pixel ratio, image type, location. The above image represent the pixel ratio to all the normal attributes. Here the high pixel ratio value is 16:32 with image quality attribute. The searching image type ico quality value is 4k in urban location. So the top most pixel ratio value 16:32 is compared with other normal attribute.

str2 = data[['image feature', 'image quality', 'pixel ratio', 'image type', 'location']]

SEABORN GRAPH

Fig 3
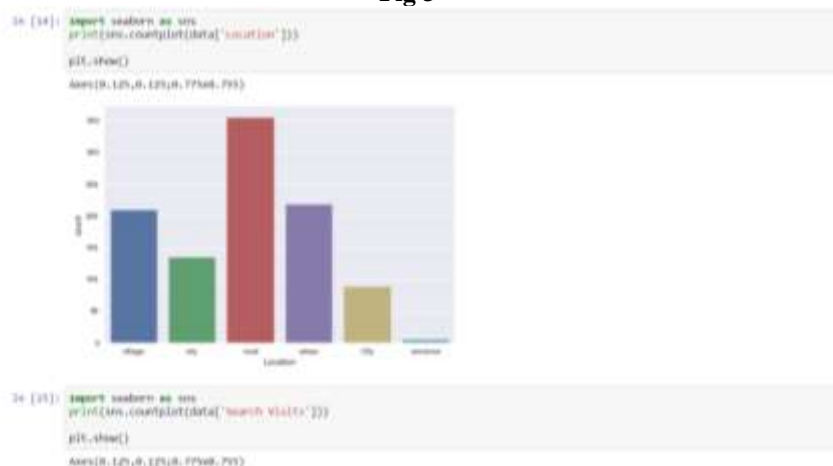
Seaborn is python visualization library. Its used to visualize the data in interactive graph. In this above graph location attributes is taken here for visualization location is like where from people are searching like rural, urban, village, city. So, here more people from rural location searched more than other city, village location people.The above graph shows the highest searching in rural peoples.
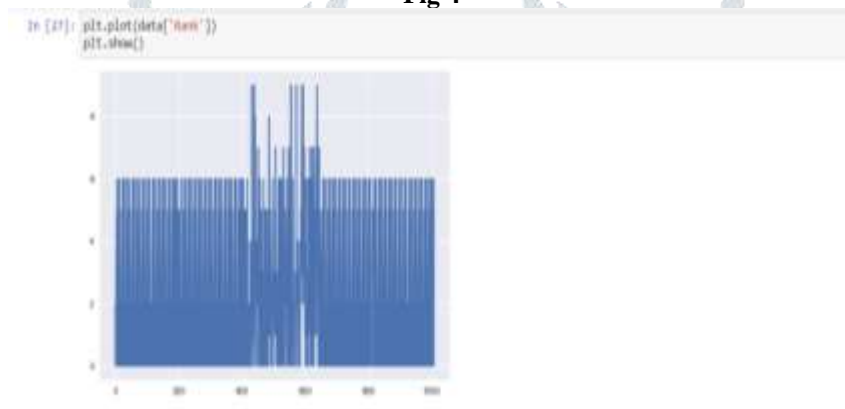
Import seaborn as sns

print(sns.countplot(data['location'])

plot.show()

SCATTER PLOT

Fig 4

A scatter plot suggest various kinds of correlation between the variables with a certain confidence interval. The above scatter plot rank was taken to plot in this graph to find the relationship of the attributes. For which image people gave rank 0 to 9 for the same image feedback range 0 to 5. The above graphs shows rank attributes is plotting using plot(). Rank range from 0 to 10 was ranked for 1000 datasets was displayed.

plt.plot (data['rank'])

plt.show()

FINDINGS:

> This paper incorporates the analysis made on the normal attributes image retrieval by the ranking, feedback, image feature, category, sub category etc.
> Creating vector with normal attributes it shows the all attributes has been compared.
> Seaborn graph shows the more than searching in rural compared with other cities using normal attributes.
> Scatter plot shows rank attribute is plot rank range from 0 to 10 was ranked for 1000 datasets was displayed using normal attributes.

V. CONCLUSION

This paper shows on using normal attributes with image retrieval dataset. Vectors was created for display the normal attributes. Seaborn graph represents the highest searching location compared with the all normal attributes. Scatter plot shows the normal attributes was ranked from 0 to 10.

FURTHER WORK:

It is suggested to use better upcoming method like elbow graph method it represents the number of clusters that have been occurred in the normal attributes. Use k means algorithm is used to find out the similarities between the attributes. . The K-Means nature is to show theattempts to classify data without having first been trained with a labelled data.

REFERENCES:

[1] Yang, L. and Jin, R. (2006). Distance metric learning: A comprehensive survey.

[2] R. Picard, "Digital Libraries: Meeting Place for High-Level and Low-Level Vision," Proc. Asian Conf. Computer Vision, Dec. 1995.

[3] H. Alraqibah, O. Bchir and M. M. Ben Ismail, "X-Ray Image Retrieval System Based on Visual Feature Discrimination", Proceeding SPIE, vol. 9159, International Conference on Digital Image Processing (ICDIP), Athens, April 2014.

[4] H. Feng, T.-S. Chua, A boostrapping approach to annotating large image collection, Workshop on Multimedia Information Retrieval in ACM Multimedia, 2003, pp. 5562.

[5] Shereena V.B. and Julie M. David, content based image retrieval: classification using neural network, The International Journal of Multimedia & Its Applications (IJMA) 6(5)(2014), 1-14.