International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-4, April 2019

Streamline Of Cross Media Retrieval Using Term Frequency-Inverse Document Frequency And Color Histogram

Monelli Ayyavaraiah

Abstract- ***Cross media retrieval provides the different media results such as text, video and audio for the single query. Many researches has been carried out for the cross-media retrieval due to its benefits. In this research, the features selection method such as Term Frequency and Inverse document frequency with color histogram (TFIDF-CH) is proposed for the cross-media retrieval system. Wikipedia dataset is popular dataset for the cross-media retrieval method and this is used to test the function of the proposed method. The text and image from the database are represent in the Bag-of-Words (BoW) and Visual BoW respectively. The TF-IDF feature is extracted from the text and color histogram is selected from the images. The graph is drawn based on these features stored in the dictionary learning. Then applied Minkowski distance to calculate the similarities between the different media. The TFIDF-CH achieves the average Mean Average Precision (MAP of 59.025 % compared with existing method having average MAP of 41.32%.***

Keywords: Color histogram, Cross media retrieval, Dictionary learning, Minkowski distance, and Term Frequency and Inverse document frequency.

I. INTRODUCTION

Traditional image retrieval system involves in the images with manual text annotation and based on the keyword search. Content Based Image Retrieval (CBIR) method involves in multimedia data of single modality, like image, audio and video retrieval [1-3]. Recently, new technique is developed to retrieve the multimedia data based on the single formatted query, as known as Cross-media retrieval system. This method involves in creating the sematic relationship across various types of modality and find the similarity between the data, then provide the retrieved information. This in-turn increase the difficulty in identifying the relationship between different types of modality. Last few years, a number of studies has been proposed and various method are analyzed in the cross-media retrieval. Some of the exiting method involves in the learning the optimal common representation of the different modality [4]. Cross-media retrieval is useful in key-word based multimedia search, when the content-based multimedia retrieval is insufficient [5]. This method faces the more difficulties compared to the traditional method in terms of identifying the sematic from low-level features and for the performance optimization [6].

Manuscript published on 30 April 2019.
***** Correspondence Author (s)
Monelli Ayyavaraiah*, Assistant Professor, Dept.of CSE, Sai Rajeswari Institute of Technology- Proddatur,

Early method estimates the sematic similarity based on the feature similarity and the performance is not efficient due to the sematic gap. Then linear and non-linear feature analysis method to dealt with the underlying data correlation. Another technique is relevance feedback, which shows the effective performance in the retrieval [7]. This technique is based on the feedback provided by the user in terms of positive and negative and used in the long-term and short-term learning [8]. Retrieving from the different modality is the challenging task and this in-turn requires the new method to perform the function. Still, retrieval method facing the difficulties in produce the semantic relation between the text and image. Some method involves in solving the problem by proposing several methods in past five years [9-10]. This study aims to improve the performance of the Cross-Media Retrieval Method (CMRM) using proposed method. The text has been representing with the TF-IDF and the image is representing with color histogram. The dictionary learning stores the data of the features and the graph is plotted from the features. The Minkowski distances is the common similarity measure processed in the graph and find the similar results. The performance of the TFIDF-CHis evaluated in the Wikipedia dataset and compared with different methods.

The organization of the paper, (i) Literature Survey in the section II, (ii) Proposed methodology in the section III, and (iii) Experimental result in the section IV.

II. LITERATURE SURVEY

There are large number of data present in the internet, which makes it difficult to retrieve the data relevant queries. This creates the demand for increase the performance of the cross-media retrieval. Many researches were carried out to enhance the retrieval and notable research are surveyed in this section.

Meijia Zhang, et al. [11] proposed the cross-media retrieval approach dependon the two-learning method. First method involves in the developing the common subspace to represent and find similarity between the heterogeneous data in the isomorphic subspace. Second method involves in the plotting isomorphic and heterogeneous adjacent graph to maintain the correlation of the training and testing data. These two methods are combined together to learn the common space.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

This method is evaluated on the three types of database and various metrics are measured in the experiment. The isomorphic and heterogeneous adjacent graph were built to mine the data structure. This method provides the effective performance in the cross-media retrieval and the clustering process can be applied to automate the process.

T. Yao, et al., [12] presented the similarity method to effectively find the sharing space between the heterogeneous data based on a supervised coarse-to-fine semantic hashing. This in turn also learns a better Hamming space in building the sematic similarity between samples of coarse-to-fine similarity matrix. To derive the local optimal solutions, an iterative updating scheme was used. An orthogonal rotation matrix was constructed from the data by reducing the quantization loss to improve the hashing codes discrimination without affecting the optimality of the relaxed solution. This method is tested on the widely used dataset such as Wikipedia and NUS-WIDE, which shows this method has high performancethe existing method.The performance of the sophisticated learning schemes such as kernel learning was affected by the training cost.

T. Yao, et al., [13] establish a Semantic Consistency Hashing (SCH) to retrieve the heterogeneous data, which considers both inter-modal and intra-modal semantic correlation in the sematic space for learning. A non-negative matrix factorization is constructed from the training data for finding the similarity without affecting the inter-model sematic consistency in each modality. An optimal algorithm is applied to remove the time complexity and extended to the multi-label learning, annotation, image classification and other domains. The training cost of Collective Matrix Factorization Hashing (CMFH) was much higher than the SCH method in large training set. Therefore, the SCH method could not efficiently utilized in small scale applications.

P. Xu, et al., [14] made an attempt to improve the cross-media retrieval using Sketch-based image retrieval and obtained the successful relation on image to text matching. This study considered the different method on the subspace learning and tested with the proposed method in the two fine-grade SBIR database. This method with the sub-space learning having the higher performance. The computational complexity of the analysis method was very high in the SBIR process.

X. Zhai, et al., [15] is another research propose the similarity measure, which considers both inter-media and intra-media graph. This study proposes the Heterogeneous Similarity with Nearest Neighbor (HSNN) to analyze the correlation between the data. The similarity is measured by calculating the probability for the two modalities in the same category. Negative correlation is important in finding the dissimilarity. The correlation was used for the relevant and irrelevant constraint to queries. This method has the capacity to process the correlation among the different types of media.The experimental result shows the HSNN were flexible and can be easily added with any traditional similarity measure. The efficiency of the model can be further increased by fusion of similarity measures through AdaRank method.

To overcome the current defects in the cross-media retrieval, this research proposes architecture and also to improve the performance. The proposed method involves in the dictionary learning based on the graph and Minkowski distance to calculate the similarity. The brief explanation about the proposed method are provided in the next section.

III. PROPOSED METHODOLOGY

The cross-media retrieval receives the more attention due to the availability of vast multimedia data. Many researcheswere carried out to increase the precision of the retrieval system. The main objective of this research is to increase the efficiency of retrieval in various modality. First, the features are extracted from the text and image and these features are used by the dictionary learning. The TF-IDF and the Color Histogram (TFIDF-CH) is proposed for the feature selection technique. Dictionary learning have the data related to the train the data and also for testing process, and the graph is drawn based on the data. Then, the similarity is measured based on the Minkowski distance and retrieve the data, that has high similar to the query data. This method is tested with the Wikipedia dataset and different metrics are evaluated.

A. Benchmark

Each article consists of a pair of image and text description, and it is categorized into 10 semantic classes. The 1,500 pairs are randomly chosen for the training and 500 pairs are used for validation and semantic classes. The database description is provided in the Table (1).

TABLE 1. DATABASE DESCRIPTION

Datasets	Total Class	Total Number	Training	Testing	Image Feature	Text Feature
Wikipedia	10	2866 pairs	2173 pairs	693 pairs	4096-d CNN	100-d LDA

B. Feature Selection

Selection of the features is the important task, because it plays the major role in identifying the data and this helps to extract the relevant information. Different kinds of feature selection are used to describe the text and image. In this research, TF-IDF based on BOW are selected for the text and, Bag of Visual Words and color histogram are selected for the images. The BoW is the common representation for the text and this is constructed based on the frequent words. Image features such as color histogram and Bag of Visual Words are applied to represent the image in the database.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

1) Bag-of-Words

The BoW is the common method for representing text from the document and this highly used in the text classification. BoW method removes the case information and punctuation from the sentences and each feature represent the single word in the document in the training corpus. This method is mainly depending on the frequent word in the document. The word in the document often repeated then the correlation between the words and document is increases. The correlation is not measured in the BoW, this will represent the document in the features. The infrequent words are often removed, which is least related to the queries [16]. This method has the advantages to provide the way to easily remove the irrelevant features due to stop-words and frequency word removal. This tends to remove the dataset size and this is useful especially in the big data. These approach helps to decrease the irrelevant retrieval in the method and this method is combined with TF-IDF.

2) Term Frequency- Inverse Document Frequency

One of the common methods to represent the document and often used in the document classification is TF-IDF[17], [18]. This method considers that the word is important in the document that is often repeated.

$$TF - IDF_{ij} = tf_{ij} \times \log\left(\frac{N}{df_i+1}\right), \quad (1)$$

From the Eq.(1), the tf_{ij} represent the word i present in the document$j$, this will increase the important of word based on the number of times it repeated.

3) Bag of Visual Words

These are constructed around the specific points like scale-space extreme (e.g. color Histogram key points [19]), or in the simple term the windows are extracted from the image at various scales and regular positions. The features are extracted from the image patch like color histogram and this can be easily affected by the noise and denoted in the high dimension spaces, not used as words.

4) Color Histogram

The color distribution probability is represented as the color histogram. In the color space, the color bins are denoted as n. Now, the number of pixel N in the image I then the color histogram is represented as $H(I) = [h_1, h_2, \dots, h_n]$, where $h_i = N_i/N$ is the probability of a pixel in the image present in the ith color bin [20], and N_i is the total number of pixels in the ith color bin. From the probability theorem, h_i can be defined in the Eq. (2).

$$h_i = \sum_{j=1}^{N} P_{(i|j)} P_j = \frac{1}{N}\sum_{j=1}^{N} P_{(i|j)}$$

(2)

Where pixel probability P_j is selected from the image I being the jth pixel, which is$1/N$, and the conditional probability is $P_{(i|j)}$ of the jth pixel present in the ith color bin.

C. Dictionary Learning

The data modeling with the linear combinations of a few elements based on the dictionary has been utilized in the much of recent researches and useful in the cross-media retrieval. The basic issue is to develop the efficient method with which samples can be rebuild from a 'best dictionary' with a 'sparse coefficients'.

Let data matrix is $X \in R^{p\times n}$ to be reconstructed, the learned dictionary is $D \in R^{p\times k}$and the sparse rebuild coefficients is denoted as $\alpha \in R^{k\times n}$ (sparse codes), where p, n and k are the feature space dimensions of the number of data samples and size of the dictionary respectively. The formulation of sparse coding can be represented in the Eq. (3).

$$\min_{D,\alpha} \frac{1}{2}\left|\left|X - D\alpha\right|\right|_F^2 + \lambda\Psi(\alpha) \text{ s.t.} \left|\left|d_i\right|\right| \le 1, \forall_i, \quad (3)$$

Where $\Psi(\alpha)$ denotes the imposed penalty over sparse codes α and d_i is one of the dictionary atoms of D. Typically, the norm l_1 is computed as a penalty to apply sparsity on each sparse codes$\alpha_j(\alpha_j \in \alpha(j = 1, \dots, N))$ is denoted in the Eq. (4).

$$\Psi(\alpha) = \sum_{j=1}^{N} \left|\left|\alpha_j\right|\right|_1 \quad (4)$$

The traditional data-driven method in the dictionary learning is highly function to rebuild the task like noisy signal restoration. The sparse coding is developed into supervised sparse coding to learn the discriminative sparse model instead of usual rebuild one. In the real-world computation, the different data can be assigned to the different groups,a mixed-norm regularization $(l_1/l_2 - norm)$ was carried out in the sparse coding to attain the sparsity and also to increase the samples rebuid from the same group and by the atoms of same dictionary in the Eq. (5).

$$M = \min_{D,\alpha} \frac{1}{2}\left|\left|X - D\alpha\right|\right|_F^2 + \lambda \sum_{l=1}^{J} \sum_{i=1}^{k} \left|\left|\alpha_i, \Omega_l\right|\right|_2$$

$$\text{s.t.} \left|\left|d_i\right|\right| \le 1, \forall_i, \quad (5)$$

Where the number of classes is J, Ω_l is the indices of the examples that belong to the l-th group, and $a_{:,\Omega_l}$ is the coefficient matrix related to the examples in the l-th group.

D. Graph Representation

The intra-modality and intermodality correlation of the M datasets are plotted in the unified graph $G(V, E\ w)$ similar to [10]. The vertex set V denotes the data from all the data sets and the edge set E models the pairwise intra-modality similarity or the inter-modality correlation between the data points. The weight of an edge is measured by some similarity functions correlation between data points.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

The weight is used by the Minkowski distance to find the similarity between the edges and to model the intra-modality similarity within same modality, local similarity metric with a Gaussian kernel. The intra-modality similarity $\omega_{i,j}^{m}$of two data points $x_i^m$and $x_j^m$from modality $m$is defined in Eq. (6).

$$\omega_{i,j}^{m} = \begin{cases} e^{-\frac{|x_i^m - x_j^m|^2}{2\alpha^2}}, & if\ x_i^m \in N_k(x_j^m) or x_i^m \in N_k(x_j^m) \\ 0 & otherwise \end{cases} \quad (6)$$

Where $N_k(x)$ represents the set of k −nearest neighbors of x and $\sigma = \frac{1}{N}\sum_{i,j}|x_i^m - x_j^m|^2$ is the expectation over all the pairwise distance in X^m.

This require $O(N^2p^m)$ time to process an intra-modality similarity matrix. When N is large, some approximate methods like anchor graph structure to construct this similarity matrix effectively. The k-NN graph is used in this method.

The inter-modality correlation of the data from the two modalities as modality a and b and $a \neq b$ and $a, b \in \{1,2, \dots, M\}$, the similarity function $\omega_{i,j}^{a,b}$ for the two data x_i^a and x_j^b is defined in Eq. (7).

$$\omega_{i,j}^{a,b} = \begin{cases} 1, & if\ x_i^a\ has\ known\ correlation\ with\ x_j^b \\ 0, & otherwise \end{cases} \quad (7)$$

Moreover, to better understand the semantics of data, we additionally exploit the category information. Let Cbe the category-labels set indicating the category label of each data point (i.e., each vertex in G), the final category-labeled unified graph is denoted as $G(V, E, \omega, C)$.

E. Minkowski Similarity

Minkowski similarity is measured from the graph and the similar values are extracted, based on these values data is retrieved. Minkowski distance is the popular method in measuring the similarities, the data established near to each other having higher similarity. The Minkowski distance can be measured using the Eq. (8). $Minkowski_p(x_i, x_j) = \left(\sum_{k=1}^{m}|x_{ik} - x_{jk}|^p\right)^{\frac{1}{p}}$ (8)

Where x is property vector, m is the dimension of the data.

In the Minkowski metric, the Manhattan metric has $p = 1$, while special case is a Euclidean distance where $p = 2$. However, there are no rules present for selecting a measure for given application.Once the similarity has been measure, then the point related to the data are retrieved. The proposed method is evaluated in the Wikipedia database and several metrics are measured. The experimental result demonstratesthat the effectiveness of the TFIDF-CH and compared with state of art CMRM.

IV. EXPERIMENTAL RESULT

The CMRM has the increasing importance and having more benefits that the single-media retrieval. The outcome of the CMRM contains multimodal data related to the sematic query. In this research, designed a method based on the feature selection for the increase the reliability of the cross-media retrieval. The text are represent in the Bag-of-Words and the images are represent in the Visual Bag-of-words. The TF-IDF is created on the words and the color histogram feature is extracted from the images. The dictionary learning is used to increase the learning of the text and image. The Minkowski distance identify the similarity between the features. Once the similarity is measured between the query and different media in the dictionary learning, the higher similarity images and text are retrieved from the database. The Wikipedia database is used for the testing the function of the TFIDF-CH and different metrics are evaluated from the proposed method. The proposed method is evaluated in the environment of python 3.7, opencv 3.0, JupyterLaband Anaconda 1.9.6.

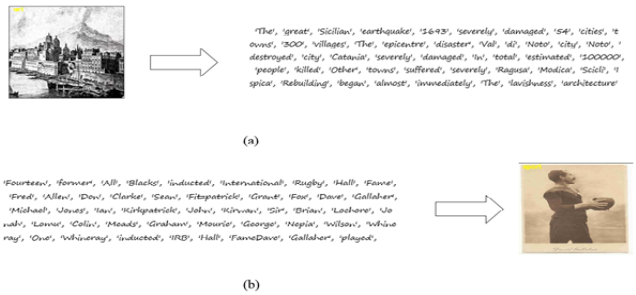

Figure 1. Retrieved results by the proposed method (a) I2T, (b) T2I

The proposed method retrieved image of the I2T and T2I from the Wikipedia datasets are shown in the Fig. (1) (a-b). In the Fig. (1) (a), Image is taken as the query and text is retrieved as the results. In the Fig. (1) (b), text is given as query and image is retrieved as the result based on the features of the image. Similarly, the proposed method is processed for the 10 categories for all the data in the dataset. The different metrics are measured for the proposed method and compared with existing methods.

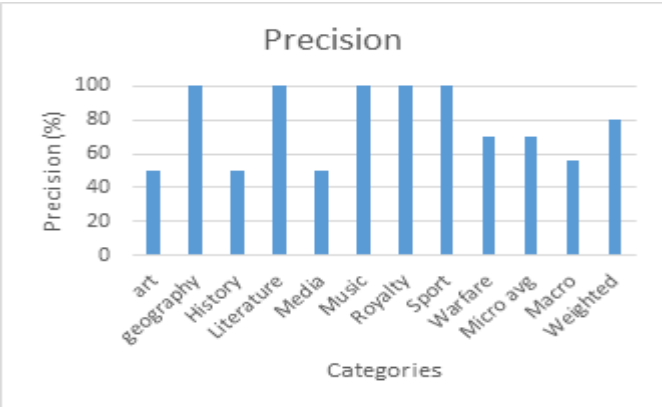

Figure 2. Proposed method precision value for different categories

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

The precision value is calculated for the different categories and this shows the effectiveness of the TFIDF-CH. The precision value is high for some categories like geography, literatureand sports etc., and low for few categories like art, and history. The history categories having low color value and has the less precision. The precision value is high for sport and this have rich color value. The precision value for different categories using proposed method is shown in the Fig. (2).

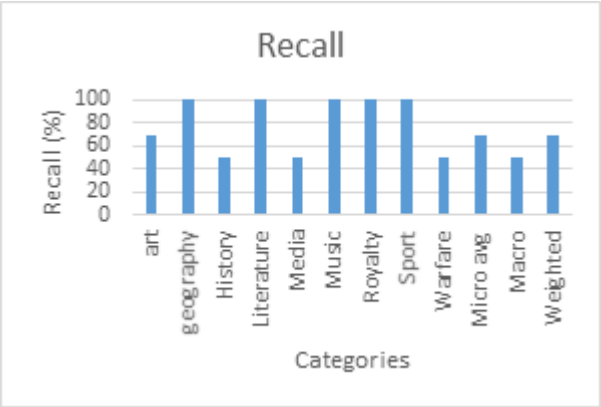

Figure 3. Proposed methods recall value for all categories

The recall value is higher for some categories like royalty, sports etc., and lower for some categories like History and media. The history categories contain less features in the image and this is having low features to compare with other data in the dataset, due to this history having low performance. The recall value is measured for the different categories using proposed method and shown in the Fig. (3).

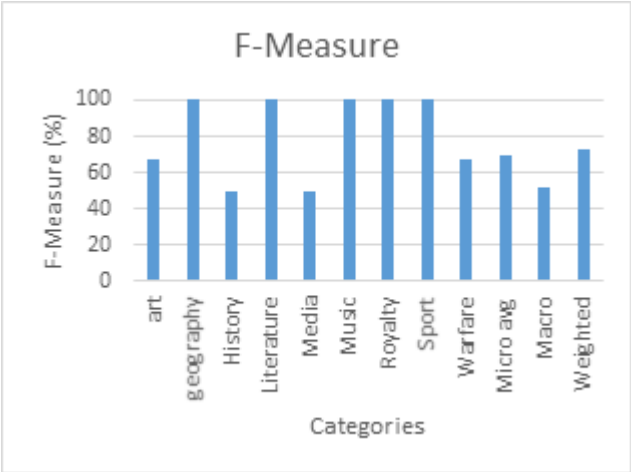

Figure 4. F-Measure metrics

The F-measure is calculated for the different categories in the dataset and shown in the Fig. (4). This shows that the geographic, music and royalty etc., having the higher value compared to other categories. The function of the TFIDF-CHis high for the images with higher color values in the database. The proposed method has the higher F-measure value for the most of the categories.

TABLE 2. COMPARATIVE ANALYSIS OF MAP FOR DIFFERENT METHODS

Methods	I2T	T2I	Average
PLS [21]	35.95%	35.10%	35.53
CCA [21]	33.16%	31.66%	32.41
SM [21]	36.85%	38.67%	37.76
SCM [21]	37.48%	39.26%	38.37
GMMFA [21]	28.41%	24.87%	26.64
GMLDA [21]	30.03%	28.06%	29.05
MDCR [21]	41.07%	37.75%	39.41
JLSLR [21]	39.38%	36.91%	38.15
MJSL[21]	44.32%	38.32%	41.32
TFIDF-CH	**60.73%**	**57.32%**	**59.025**

The features are selected from the query images or text and data in the database to identify the similarities. The more similar related to the query are retrieved and different parameters are measured. The proposed methods are evaluated in the two scenario such as Image to Text (I2T) and Text to Image (T2I). The metrics such as Mean Average Precision (MAP) is measured for the different methods and shown in the table. (2). This shows that the TFIDF-CHhas the higher MAP value compared to other existing methods in cross-media retrieval. The proposed method is compared with existing technique such as Partial Least Squares (PLS), Canonical Correlation Analysis (CCA), Semantic Matching (SM), Semantic Correlation Matching (SCM), Generalized Multiview Marginal Fisher Analysis (GMMFA), Generalized Multiview Linear Discriminant Analysis (GMLDA), Modality dependent Cross-media Retrieval (MDCR) model and Joint Latent Subspace Learning and Regression for Cross-Modal Retrieval (JLSLR) and Multi-class Joint Subspace Learning algorithm (MJSL) [21].

This shows that the TFIDF-CHhas the higher performance than the state-of-art method. The proposed method can be applicable for the practical use in the cross-media retrieval process.

V. CONCLUSION

The major objective of the method is to improve the performance of the cross-media retrieval using the proposed TFIDF-CH.The Bow and the Visual Bow are used to represent the text and image from the database of theCMRM. The features are measured from the query and the data in the dictionary learning. The graph is plotted based on the features of text and image from the dictionary learning method. TheMinkowski distance measuresthe similarity between the features and shorter distance have higher similarity. The more similar data are retrieved and different metrics are measured from the output of the proposed method. The TFIDF-CH method has the higher MAP when compared to the existing methods. The TFIDF-CH has the 60.73 % MAP in the I2T and 57.32 % in the T2I. The experimental result shows the effectiveness of the proposed method.

REFERENCES

1. B. Jiang, J. Yang, Z. Lv, K. Tian, Q. Meng, and Y. Yan, "Internet cross-media retrieval based on deep learning," *Journal of Visual Communication and Image Representation*, vol. 48, pp. 356-366, 2017.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

2. X. Zhai, Yuxin Peng, and Jianguo Xiao, "Learning cross-media joint representation with sparse and semisupervised regularization," *IEEE Transactions on Circuits and Systems for Video Technology* vol. 24, no. 6, pp. 965-978, 2014.
3. Y. Peng, X. Zhai, Y. Zhao, and X. Huang, "Semi-supervised cross-media feature learning with unified patch graph regularization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 3, pp. 583-596, 2016.
4. Y. Wei, Y. Zhao, C. Lu, S. Wei, L. Liu, Z. Zhu, and S. Yan, "Cross-modal retrieval with cnn visual features: A new baseline," *IEEE transactions on cybernetics*, vol. 47, no. 2, pp. 449-460, 2017.
5. Y. Wei, Y. Zhao, Z. Zhu, S. Wei, Y. Xiao, J. Feng, and S. Yan, "Modality-dependent cross-media retrieval," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 7, no. 4, pp. 57, 2016.
6. Y. X. Peng, W. W. Zhu, Y. Zhao, C. S. Xu, Q. M. Huang, H. Q. Lu, and W. Gao, "Cross-media analysis and reasoning: advances and directions," *Frontiers of Information Technology & Electronic Engineering*, vol. 18, no. 1, pp. 44-57, 2017.
7. H. Zhang, Yun Liu, and Zhigang Ma, "Fusing inherent and external knowledge with nonlinear learning for cross-media retrieval," *Neurocomputing* vol. 119, pp. 10-16, 2013.
8. S. Vigneshwari, S., and M. Aramudhan, "Personalized cross ontological framework for secured document retrieval in the cloud," *National Academy Science Letters* vol. 38, no. 5, pp. 421-424, 2015.
9. L. Huang, and Yuxin Peng, "Cross-media retrieval by exploiting fine-grained correlation at entity level," *Neurocomputing* vol. 236, pp. 123-133, 2017.
10. J. He, B. Ma, S. Wang, Y. Liu, and Q. Huang, "Multi-label double-layer learning for cross-modal retrieval," *Neurocomputing*, vol. 275, pp. 1893-1902, 2018.
11. Zhang, M., Zhang, H., Li, J., Wang, L., Fang, Y. and Sun, J., 2019. Supervised graph regularization based cross media retrieval with intra and inter-class correlation. Journal of Visual Communication and Image Representation, 58, pp.1-11.
12. T. Yao, X. Kong, H. Fu, and Q. Tian, "Supervised Coarse-to-Fine Semantic Hashing for cross-media retrieval," *Digital Signal Processing*, vol. 63, pp. 135-144, 2017.
13. T. Yao, X. Kong, H. Fu, and Q. Tian, "Semantic consistency hashing for cross-modal retrieval," *Neurocomputing*, vol. 193, pp. 250-259, 2016.
14. P. Xu, Q. Yin, Y. Huang, Y. Z. Song, Z. Ma, L. Wang, and J. Guo, "Cross-modal subspace learning for fine-grained sketch-based image retrieval," *Neurocomputing*, 2017.
15. X. Zhai, Yuxin Peng, and Jianguo Xiao, "Cross-media retrieval by intra-media and inter-media correlation mining," *Multimedia systems* vol. 19, no. 5, pp. 395-406, 2013.
16. Passalis, N. and Tefas, A., 2018. Learning bag-of-embedded-words representations for textual information retrieval. Pattern Recognition, 81, pp.254-267.
17. Kim, D., Seo, D., Cho, S. and Kang, P., 2019. Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec. Information Sciences, 477, pp.15-29.
18. Erra, U., Senatore, S., Minnella, F. and Caggianese, G., 2015. Approximate TF–IDF based on topic extraction from massive message stream using the GPU. Information Sciences, 292, pp.143-161.
19. Ali, H., Lali, M.I., Nawaz, M.Z., Sharif, M. and Saleem, B.A., 2017. Symptom based automated detection of citrus diseases using color histogram and textural descriptors. Computers and Electronics in agriculture, 138, pp.92-104.
20. Yan, J., Zhang, H., Sun, J., Wang, Q., Guo, P., Meng, L., Wan, W. and Dong, X., 2018. Joint graph regularization based modality-dependent cross-media retrieval. Multimedia Tools and Applications, 77(3), pp.3009-3027.
21. Yu, E., Li, J., Wang, L., Zhang, J., Wan, W. and Sun, J., 2018. Multi-class joint subspace learning for cross-modal retrieval. Pattern Recognition Letters.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)