IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Supervised Social Image Understanding Using Deep Matrix Factorization

Ms.Tambe Sonali[1]
Department of Computer Engg.
TCOER, Pune sona.

Asst.Prof. Madane Supriya[1]
Department of Computer Engg.
TCOER, Pune

***Abstract*—In recent years amount of images assigned with tagging information has increased, this tagging information is insufficient to describe the contents of an image, sometimes irrelevant and noisy, which makes image retrieval less efficient. The system will be responsible handle these issues. Images understanding tasks such as image tag refinement, image tag assignment, and various types of image retrieval techniques such as tag-based image retrieval, content-based image retrieval, and image retrieval based on Semantic are performed simultaneously. The system creates latent representations of images and tags using a deep matrix factorization algorithm. latent representa- tions are deployed in the subspace of latent by simultaneously exploring the visual structure, the semantic structure, and the tagging information. The semantic and visual structures are combined to learn a semantic subspace without over-fitting the irrelevant, incomplete, or subjective tags. Wordnet dictionary is used to optimize, semantic-based image retrieval module for better results. Broadly examining on social image databases led to the successful image understanding.**

***Index Terms*—Tag Refinement, Tag Assignment, Tag-Based Image Retrieval, Content-Based Image Retrieval, Social Image Understanding, Deep Matrix Factorization**

I. INTRODUCTION

Social image understanding refers to the various operations needed to be performed for effective search on social media images. The proposed system focus on performing tasks like image tagging, tag refinement, tag assignment, cross-modal search(TBIR). Due to tremendous growth in web 2.0 applica- tions and social media, users are sharing a no.of multimedia objects on popular social media websites such as Flickr, Facebook, Tumblr. In recent years is users uploading images with rich contextual information attached to them such as user-provided tags. These user-defined tags presents the semantic content of images to some extent, which can be used for many tasks related to social image understanding, such as Content-Based Image Retrieval (CBIR)[2] and Tag-Based Image Re- trieval (TBIR)[1].It is beneficial to commonly explore the rich information of community-contributed images which is fre- quently available along with social media images. Since users tag images arbitrarily, information is weakly supervised which means there are images without information of tagging.User-provided tags are inadequate, subjective, and noisy, to use them for retrieval we need to refine them. Input to this system is images with context tags which is then represented in their latent image and tag representation using weakly supervised deep matrix factorization algorithm[2]. Latent representation is deployed in the latent subspace by simultaneously uncovering the weakly supervised tagging information, that is the visual and semantic structure of the images along with tags.

II. LITERATURE SURVEY

1) Deep Collaborative Embedding for Social Image Understanding[1], Zechao Li, Jinhui Tang, and Tao Mei ,2018

In this work author focused on the Challenge of acquiring knowledge from the community- contributed images with rich weakly-supervised tagging information, which can benefit multiple image understanding tasks simultaneously, such as social image tag refinement and assignment, content-based image retrieval, tag-based image retrieval, and tag expansion. A model of Deep Collaborative Embedding (DCE) developed to discover a unified latent space for tags and images. It uses weakly supervised images and its tag correlation, image correlation, and tag correlation to determine latent space simultaneously. The system embeds images and tags in a unified latent space under the factorization framework by exploring the weakly supervised tag information, visual structure, and semantic structure simultaneously. In latent space, correlations between images-tags are directly modeled as the pairwise similarity, which allows various types of image retrieval tasks in the same framework. The proposed model is scalable with new images as it is capable to embed new tags in the uncovered space.

2) Weakly Supervised Deep Matrix Factorization for Social Image Understanding[2], Zechao Li and Jinhui Tang,2017

In this paper author proposed framework to refine the initial tags and assign tags to new images by discovering the vector image representations and tag representations which is embedded in the latent subspace by exploring the tagging information, visual structure,and semantic structure.A hierarchical model learns hidden representations of images which minimize semantic gap. It can naturally include new images into the subspace using deep architecture. Both semantic structure and visual structures are integrated to learn a latent subspace without overfitting the noisy, incomplete, or subjective tags with the help of factor analysis model. A sparse model is used on the transformation matrix to remove noisy and irrelevant visual features of the

3) Tag Based Image Search by Social Re-ranking [3], Xueming Qian, Dan Lu, and Xiaoxiao Liu 2016

In this paper, the author proposed a tag-based image re- trieval system using social re-ranking. Image re-ranking is done according to semantic information, visual infor- mation, and social clues. The initial evaluation include images contributed by different social users. Generally, every user uploads multiple images. Firstly these images are sorted by inter-user re-ranking method, where Users that have a high contribution to the given query ranking higher. Then in the second step intra-user, re-ranking on the ranked user's image set, and only the most relevant images from each user's image set are selected for the final retrieved results[3]. To boost the speed of image search inverted index is implemented on the image dataset.

4) Unsupervised Feature Selection via Non negative Spectral Analysis and Redundancy Control[4], Zechao Li and Jinhui Tang 2015

In most of the image processing and pattern recognition problems, the visual contents of images are described by highly-dimensional features, which are usually redundant and noisy. authors proposed an unsupervised feature selection process, non-negative spectral analysis along with constrained redundancy, by jointly having non-negative spectral clustering and redundancy analysis.

5) Projective Matrix Factorization with unified embedding for social image tagging, Computer Vision and Image Understanding[5],Z. Li, J. Liu, J. Tang, and H. Lu, 2014.

It contains Matrix Factorization with the unified embedding (PJMF) method, by using this social image re-tagging is transformed to the nearest neighboring tag search. Authors solve the proposed PJMF as an optimization problem for solving issues such as, To find two latent representations in a unified space for tags and images and it also explores the latent representations to reconstruct the observed image-tag correlation in a nonlinear manner. Here the relevance between an image and tag can be modeled as the pair-wise similarity in a unified space. Another benefit is that the image latent representation is supposed to be projected from representation of original visual feature by utilising an orthogonal transformation matrix.

6) A multi-view embedding space for modeling Internet images, tags, and their semantics,Y. Gong, Q. Ke, M. Isard, and S. Lazebnik 2014,

The paper focuses on the problem of modeling social images and correlated tags so these co-relations can be used in tasks as like image-to-image search, tag-to- image search, and image-to-tag search. It uses CCA i.e canonical correlation analysis, a widely used mapping approach for visual features and textual features in com- mon embedding space. The linear Canonical Correlation Analysis maximizes the correlation of the two views [5]. It includes a third view obtaining high-level image semantics, represented by a single category or multiple concepts. The difference between the two-view CCA and three-view embedding is that In two-view embedding space which is produced by maximization of the cor- relations between visual features and the associated tag features, images of various classes are mixed. Whereas the three-view embedding provides better separation between the classes, but the implementation of three- view CCA is costly.

III. PROPOSED METHODOLOGY

The system uses a Deep Matrix Factorization (DMF) algo- rithm for social image based tag refinement, assignment, and retrieval, which identifies the latent based image representa- tions and latent tag representations which is further embedded in the latent subspace by jointly utilizing , the visual structure of images ,the tagging information and the semantic structure. Common embedding of representation is in such a way that semantically similar images and tags appear nearby to each other. The proposed system deals with noisy, incomplete, or subjective tags and the noisy or redundant visual features. The proposed system is expressed as a joint optimization problem with a an objective function, which follows gradient descent procedure . To give the effectiveness of the system experiments are conducted on real-world databases.

A. Architecture

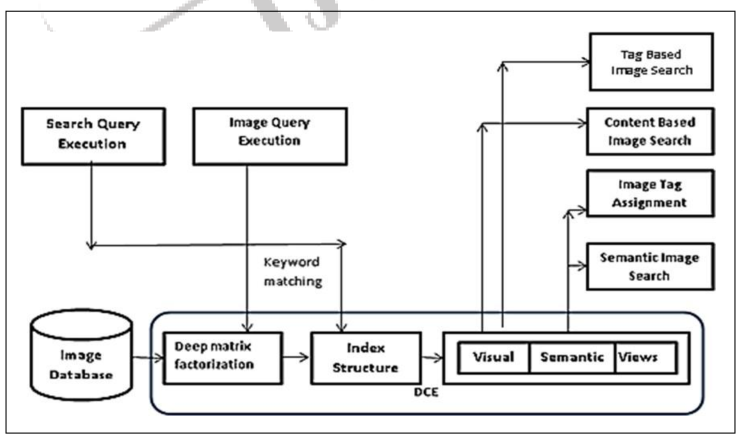

Fig. 1. System Architecture

B. Algorithm

1) Deep Matrix Factorization: Matrix factorization is a useful latent factor learning model. It decomposes a matrix as a product of two-factor matrices. This makes it easier to calculate more complex matrix operations, there are many ways to decompose a matrix, hence there is a range of different matrix decomposition techniques. MF can be found in many applications in the field of bioinformatics, computer vision, text processing, recommender systems, and others. DMF identifies latent space in which tags and images are embedded. It uncovers the hidden features from a visual structure in a progressive way. Progressive learning reduces the semantic gap. It is made up of a deep neural network with multiple layers of linear transformations. The transformation from the original visual space to the latent space is analyzed as a product of multiple factors. This framework preserves the visual and semantic Structure of images. Visually Similar images have similar representation in latent space. Latent features of semantically relevant tags are similar Deep neural network with M layers uncovers hidden representations of images from visual features matrix X. DMF factorizes Image tagging matrix F into M+1 matrices i.e V, Um,.., U1 V is latent tag feature matrix at each iteration. Um is Image representation matrix at mth layer The transformation from the original visual space to the latent space is analyzed as a product of multiple factors. i.e Um =Vm-1 x Wm Cost function measures the semantic difference between the structure of hidden space and text space Thus DMF learns a better unified subspace in which images and tags are embedded.

Input:Visual feature matrix X, the tagging matrix F, the number of network layers M, learning rate η, 0 ¡ ε ¡ 1 and 0 ¡ ρ_1 ¡ ρ_2 ¡1

- 1: Calculate T, L and M according to X and F;
- 2: Initialize V and W_m (1 ≤ m ≤ M); Set D as Identity Matrix
- 3: **repeat**
- 4: // Forward Propagation
- 5: **for** m = 1, 2, 3,..., M
- 6: **Do** forward propagation to get U_m;
- 7: **end**
- 8: // Computing Gradient
- 9: Compute Gradient

$$\frac{\delta\vartheta}{\delta v} = EU^T + \beta LV + \lambda_1 V$$

- 10: **for** m = M, M-1, ..., 1
- 11: Compute Z_m

$$Z_m = G_m W_m^T - W_m G_m^T$$

- 12: $\tau = 1$
- 13: **repeat**
- 14: $\tau = s\tau$
- 15: Compute Y_m (τ)
- 16: **until** Armijo Wolfe Conditions satisfied
- 17: **end**
- 18: // Back Propagation
- 19: Update V

$$V = V - \eta \frac{\delta\vartheta}{\delta v}$$

- 20: **for** m = 1, 2, ..., M
- 21: Update W_m
- 22: **end**
- 23: Update Diagonal Matrix D;
- 24: **Until** convergence criterion satisfied
- 25: Compute S and C

Output: Latent matrix V and Transformation matrix W_m

2) Semantic Based Image Retrieval:

- step1: Parser splits the input queries into tokens using string tokenizer .The split keywords are called tokens.
- step2: Stopword elimination to be perform to eliminate Stopwords.
- step3:Stemming is performed to find the root phrase of the keyword. If the keyword is "searching", it will reduce the word as "search".
- step4: Processed keywords are targets.
- step5:Then, objectives are given as data to the WordNet. WordNet is a dictionary which is used to find synonyms of the targets. In this regard, the system is linked with the WordNet software.
- step6:Using the WordNet dictionary, the keywords are classified as either noun or verb or adjective. If the keyword is noun or verb, it is classified as subject. If it is adjective, it is labeled as predicate.
- step7:Apply the subject, predicate and object concept, with indexed structure

C. Experimental results

1) Semantic Image Retrieval Results: Semantic Image Retrieval Results: In the proposed system semantic image retrieval algorithm is used for better performance. The overall image retrieval results improved by this semantic image description. We use the standard evaluation criteria used in most prior work on text-image retrieval tasks. We measure rank-based performance by Mean average precision.

$$Precision = TP/(TP + FP) \quad (1)$$

Mean average precision is percentage of test samples for which the correct result is ranked within the top retrieved results to the query sample.

Average precision(AP) is the standard measure used for retrieval benchmark. It corresponds to the average of the precision at each position where a relevant image appears. It is observed that in proposed system (DMF+wordnet) performs significantly better compared with the other methods for tag- based image retrieval, which shows that the proposed method can rank the relevant tags at the top positions.The qualitative results show that the proposed method achieves better quality compared to other approaches.

[7]L. Wu, R. Jin, and A. K. Jain, "Tag completion for image retrieval", IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 3, pp. 716727, Mar. 2013.
[8]Z. Lin, G. Ding, M. Hu, J.Wang, and X. Ye, "Image tag completion via image specific and tag-specific linear sparse reconstructions", in Proc. IEEE Comput. Vis. Pattern Recognit., Jun. 2013, pp. 16181625.
[9]Z. Li, J. Liu, C. Xu, and H. Lu, "Mlrank: Multi-correlation learning to rank for image annotation", Pattern Recognit., vol. 10, no. 46, pp. 27002710, 2013
[10]J. Tang, Z.-J. Zha, D. Tao, and T.-S. Chua, "Semantic-gap-oriented active learning for multi-label image annotation, IEEE Trans. Image Process., vol. 21, no. 4, pp. 23542360, Apr. 2012.

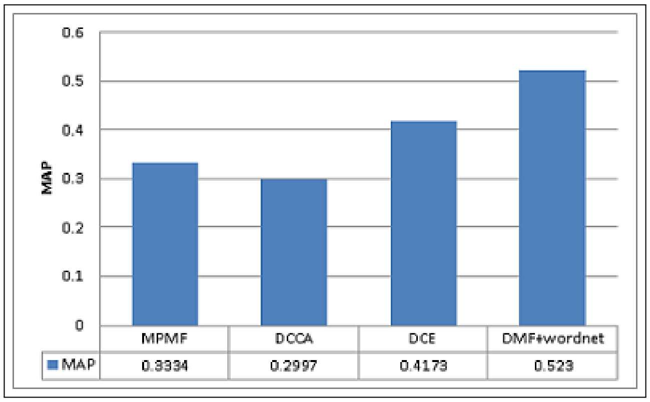

Fig. 2. Semantic Image retrieval Score

D. *Conclusion*

In this system we propose a novel Deep Embedding model for social image understanding tasks in single framework. It incorporates the end-to-end learning and cooperative calculate investigation one brought together structure for the ideal similarity of representation learning and inactive space revelation. To cooperatively investigate the rich logical data of social images, it factorizes correlation matrices at the same time
. A refined tagging matrix with non negative and discrete properties is specifically figured out how to deal with the noisy tags. The proposed strategy is connected to social image tag refinement and assignment, content-based image recovery, tag-based image recovery and semantic based image retrieval.

E. *Future Scope*

In future we can improve proposed method by utilizing other types of metadata e.g comments,location,social media groups, while learning the latent space embedding.If the amount of noisy tags associated with social images is high compared to clean relevant tags the objective of latent subspace learning may hamper. In such case improvement in proposed method can be added by designing loss functions or layers spe- cific to noise reduction, providing a more principled way for learning the latent space embedding in presence of significant noise.

References

[1] Zechao Li, Jinhui Tang, and Tao Mei, "Deep Collaborative Embedding for Social Image Understanding", in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, pp. 1-14.
[2] Zechao Li and Jinhui Tang,"Weakly Supervised Deep Matrix Factorization for SocialImage Understanding",in IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 26, NO. 1, JANUARY 2017, pp. 1-13.
[3]L. Xu et al., "Multi-task rank learning for image quality as- sessment", IEEE Trans. Circuits Syst. Video Technol., 2016, doi:10.1109/TCSVT.2016.2543099.
[4]Z. Li and J. Tang, "Unsupervised feature selection via non negative spectral analysis and redundancy control", IEEE Trans. Image Process., vol. 24, no. 12, pp. 53435355, Dec. 2015.
[5]Z. Li, J. Liu, J. Tang, and H. Lu, "Projective matrix factorization with unified embedding for social image tagging, Comput. Vis. Image Understand., vol. 124, pp. 7178, Jul. 2014.
[6]Y. Gong, Q. Ke, M. Isard, and S. Lazebnik, "A multi-view embedding space for modeling Internet images, tags, and their semantics, Int. J. Comput. Vis., vol. 106, no. 2, pp. 210233, 2014.