International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161
Available at http://inpressco.com/category/ijcet

Research Article

Supervised Social Image Understanding using Deep Matrix Factorization

Ms.Tambe Sonali and Prof. Madane Supriya

Department of Computer Engg. TCOER,Pune

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

The number of images associated with user-provided tags has increased in recent years,these tags are insufficient to describe contents of image,sometimes irrelevant and noisy. Proposed system, focus on the problem of social image understanding, which can perform tasks such as tag refinement, tag assignment, and various types image retrievals such as tag based image retrieval and content based image retrieval simultaneously.System discovers latent representations of images and tags using deep matrix factorization algorithm .latent representations are embedded in the latent subspace by simultaneously exploring the visual structure, the semantic structure and the tagging information. The visual structures and semantic structures are integrated to learn a semantic subspace without over-fitting the irrelevant, incomplete or subjective tags. In addition to expel the loud or repetitive visual highlights, an inadequate model is forced on the change grid of the first layer in the profound design. Broad examinations on true social image databases are led to the successful image understanding . Empowerment of results are accomplished, which shows the adequacy of the proposed strategy.

Keywords: *Tag Refinement, Tag Assignment, Tag Based Image Retrieval,Content Based Image Retrieval, Social Image Understanding,Deep Matrix Factorization*

Introduction

Social image understanding refers to the various operations needed to be performed for effective search on social media images. Proposed system focus on performing tasks such as image tagging,tag refinement,tag assignment,cross modal search(TBIR). Due to tremendous growth in web 2.0 applications and social media, users are sharing large amount 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 with them such as user-provided tags. These user defined tags describes the semantic content of images to some extent, which can be used to many tasks related to social image understanding, such as Content-Based Image Retrieval (CBIR) and Tag-Based Image Retrieval (TBIR)[1].

It is beneficial to collaboratively explore the rich information of community-contributed images which is frequently available along with social media images.Since users tag images arbitrarily , information is weakly supervised that means there are images without tagging information. The images associated with weakly supervised user-provided tags has increased in recent years. User-provided tags are inadequate, subjective and noisy,in order to use it for retrieval we need to refine it. Input to this system is images with context tags which is then represented in there latent image representations and tag representation using weakly supervised deep matrix factorization algorithm.Latent representation is embedded in the latent subspace by simultaneously exploring the weakly supervised tagging information,that is the visual structure, and the semantic structure of images and tags.

Literature Survey

1) Deep Collaborative Embedding for Social Image Understanding, Zechao Li, Jinhui Tang, and Tao Mei,2018. In this work author focused on the problem of learning knowledge from the community-contributed images with rich weakly-supervised context 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 Deep Collaborative Embedding (DCE) model is developed to uncover a unified latent space for images and tags. It uses weakly supervised image and its tag correlation ,image correlation and tag correlation to determine latent space simultaneously.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 same framework.The proposed model is scalable with new images as it is capable to embed new tags in the uncovered space.

2) Tag Based Image Searchby Social Re-ranking", Xueming Qian, Dan Lu, and Xiaoxiao Liu 2016. In this paper, author proposed a social re-ranking system for tag-based image retrieval with the consideration of image's relevance and diversity in the database.Image re-ranking is done according to semantic information, visual information and social clues. The initial results include images contributed by different social users . Usually each user contributes several images. First sorted these images by inter-user re-ranking. Users that have higher contribution to the given query rank higher. Then they implement intra-user re-ranking on the ranked user's image set, and only the most relevant image from each user's image set is selected. These selected images compose the final retrieved results. An inverted index structure is built for the social image dataset to improve search speed.

3) Long Xu, Jia Li, Weisi Lin, Yongbing Zhang, Lin Ma, Yuming Fang, "Yihua Yan, Multi-task Rank Learning for Image Quality Assessment", IEEE Transactions on Circuits and Systems for Video Technology, 2016. In practice, images are distorted by more than one distortions. For image quality assessment (IQA), the existing machine learning (ML) based methods generally established a unified model for all the distortion types, or each model is trained independently for each distortion type, which is therefore distortion aware. In distortion-aware methods, the common features among different distortions were not exploited. In addition, there were fewer training samples for each model training, which may result in over fitting. To address these problems, authors proposed a multi-task learning framework to train multiple IQA models together, each model is for each distortion type; however all the training samples are associated with each model training. Thus, the common features among different distortion types, and the said underlying relatedness among all the learning tasks are exploited, which would benefit the generalization ability of trained models and prevent over fitting possibly.

4) Z. Li, J. Liu, J. Tang, and H. Lu, "Projective Matrix Factorization with unified embedding for social image tagging", Computer Vision and Image Understanding, 2014. This paper presents a general formulation, named Projective Matrix Factorization with unified embedding (PJMF), by which social image retagging is transformed to the nearest tag-neighbor search for each image. Authors solve the proposed PJMF as an optimization problem mainly considering the following issues. First, it attempt to find two latent representations in a unified space for images and tags respectively and explore the two representations to reconstruct the observed imagetag correlation in a nonlinear manner. In this case, the relevance between an image and a tag can be directly modeled as the pair-wise similarity in the unified space. Second, the image latent representation is assumed to be projected from its original visual feature representation with an orthogonal transformation matrix.

Proposed Methodology

System uses novel Weakly-supervised Deep Matrix Factorization (WDMF) algorithm for social image tag refinement, assignment and retrieval, which identifies the latent image representations and tag representations which is further embedded in the latent subspace by collaboratively exploiting the tagging information, the visual structure of images and the semantic structure.Common embedding of representation is in such a way that semantically similar images and tags appear near by to each other. The proposed system can deal with the noisy, incomplete or subjective tags and the noisy or redundant visual features. The proposed system is formulated as a joint optimization problem with a well-defined objective function, which is solved by a gradient descent procedure with curvilinear search. Extensive experiments on two realworld social image databases are conducted to demonstrate the effectiveness of the problem

A. Architecture

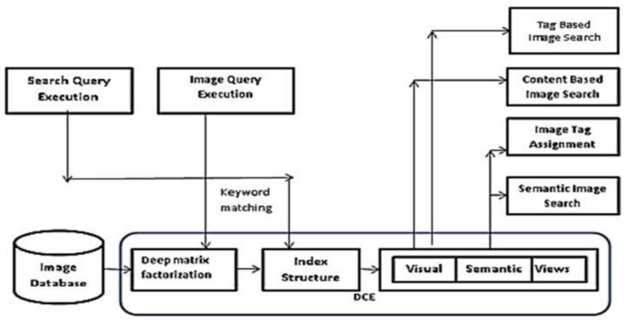

Fig. 1. System Architecture

B. Algorithm

1) Weakly Supervised Deep Matrix Factorization: Deep neural network with M layers uncovers hidden representations of images from visual features matrix X. WDMF 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 structure of hidden space and text space Thus WDMF learn 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

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

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

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

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

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

  - 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 proposed approach 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

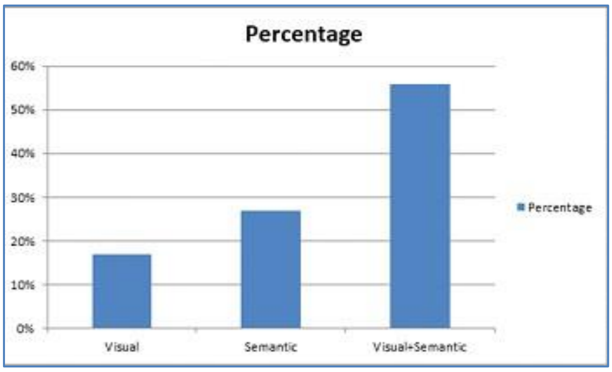

Fig. 2. Semantic Image Description Score

C. Experimental results

1) Tag Expansion: The keyword queries may be ambiguous for image retrieval. Therefore, it is necessary to expand queries to address the query ambiguity problem. The proposed DCE method can also expand query tags. The experiment is conducted a user study to evaluate the performance and compare the proposed DCE with the Word Net-based method and IQKS. The better the results are, the larger the scores are. The average scores are shown in Figure. The experimental results observed that the proposed method performs better

2) Semantic Tag Description: In the semantic image algorithm 1 and algorithm 2 is used for better performance. The overall image retrieval results improved by this semantic image description. The importance of the semantic description which illustrated by the two example retrievals. The users were given a target image together with six retrieved images to quantify the improvement. There are only two images which were similar contained in six images only in a visual sense, two in a semantic sense and two which are similar in both visual and semantic description. The user is unaware which algorithm is related to which image and the images are randomly presented. Next, most similar to the target image the user is asked to select the image from the six images.

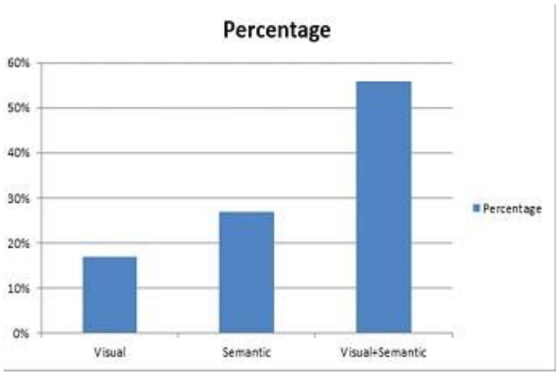

Fig. 3. Tag expansion result

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, tagbased 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 specific 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]. L. Xu et al., "Multi-task rank learning for image quality assessment", IEEE Trans. Circuits Syst. Video Technol., 2016, doi:10.1109/TCSVT.2016.2543099.

[3]. 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.

[4]. 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.

[5]. 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.

[6]. 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.

[7]. 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.

[8]. 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

[9]. 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