

## Accepted Manuscript

Deep multi-view representation learning for social images

Feiran Huang, Xiaoming Zhang, Zhonghua Zhao, Zhoujun Li,  
Yueying He



PII: S1568-4946(18)30458-7  
DOI: <https://doi.org/10.1016/j.asoc.2018.08.010>  
Reference: ASOC 5038

To appear in: *Applied Soft Computing Journal*

Received date: 12 December 2017  
Revised date: 13 June 2018  
Accepted date: 6 August 2018

Please cite this article as: F. Huang, et al., Deep multi-view representation learning for social images, *Applied Soft Computing Journal* (2018), <https://doi.org/10.1016/j.asoc.2018.08.010>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Deep Multi-View Representation Learning for Social Images

Feiran Huang<sup>a</sup>, Xiaoming Zhang<sup>b,\*</sup>, Zhonghua Zhao<sup>c</sup>, Zhoujun Li<sup>a</sup>, Yueying He<sup>c</sup>

<sup>a</sup>State Key Laboratory of Software Development Environment, Beihang University, Beijing, 100191, China

<sup>b</sup>School of Cyber Science and Technology, Beihang University, Beijing, 100191, China

<sup>c</sup>National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing, 100029, China

---

## Abstract

Multi-view representation learning for social images has recently made remarkable achievements in many tasks, such as cross-view classification and cross-modal retrieval. Since social images usually contain link information besides the multi-modal contents (e.g., text description, and visual content), simply employing the data content may result in sub-optimal multi-view representation of the social images. In this paper, we propose a Deep Multi-View Embedding Model (DMVEM) to learn joint embeddings for the three views including the visual content, the associated text descriptions, and their relations. To effectively encode the link information, a weighted relation network is built based on the linkages between social images, which is then embedded into a low dimensional vector space using the Skip-Gram model. The learned vector is regarded as the third view besides the visual content and text description. To learn a joint representation from the three views, a deep learning model with three-branch nonlinear neural network is proposed. A three-view bi-directional loss function is used to capture the correlation between the three views. The stacked autoencoder is adopted to preserve the self-structure and reconstructability of the learned representation for each view. Comprehensive experiments are conducted in the tasks of image-to-text, text-to-image, and image-to-image searches. Compared to the state-of-the-art multi-view embedding methods, our approach achieves significant improvement of performance.

*Keywords:* multi-view learning, image embedding, representation learning, stacked autoencoder

Download English Version:

<https://daneshyari.com/en/article/10139438>

Download Persian Version:

<https://daneshyari.com/article/10139438>

[Daneshyari.com](https://daneshyari.com)