Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 218 (2023) 67–73

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

A Hybrid Classifier-based Ontology driven image Tag Recommendation framework for social image tagging

Vaibhava Lakshmi R[1], Gerard Deepak [2], Santhanavijayan A[3], Radha S[4]

[1]*Madras Institute of Technology, Anna University, Chennai,600 044, India*
[2]*Manipal Institute of Technology, Manipal-576 104, India*
[3]*National Institute of Technology, Tiruchirapalli-620 015, India*
[4] *Vellore Institute of Technology, Chennai-600 127, India*

Abstract

With the burgeoning of social network sites, social image tagging, where images are annotated with tags, is gaining traction. But, in the domain of social media, this task becomes complex as images are associated with plenitude of object information. In this paper, we have put forth a Hybrid Classifier-based Ontology driven image Tag Recommendation framework for social image tagging which is based on a knowledge-centric, metadata generating, semantically-inclined and ontology-driven strategy. The COIL-100 dataset is used for the implementation. The proposed HCOntoTR approach is evaluated using sundry performance measures by comparison with baseline models like UIT, HRSDL and RSSVM. The proposed HCOntoTR model gives finer results set against the baseline models, with a precision of 94.49 % which surpasses the others, owing to the amalgamation of a profusion of knowledge from heterogeneous knowledge sources namely the Wikidata and Google's Knowledge API.

Keywords: Tag recommendation; Social image tagging; Ontology; Knowledge sources; Hybrid classifier

1. Introduction

With the advent of Internet technology and communal network platforms, social media has become a sought-after means for sharing multimedia information. Image sharing refers to the posting of photos on the Internet [1]. Since the outset of social networks, image sharing has become a customary online activity. It is a vogue that people are taking enormous number of photos every day and uploading in their social profiles. Most of the social media users post pictures about the highlights of their days' memorable activities or occasions. Hence, the number of photos taken or obtained can soon create a great problem on their organization and retrieval from sundry sources. For enormous social image resources, making personalized recommendations to users based on their images of interest has become a

10.1016/j.procs.2022.12.402

significant field of research for social media websites [2]. Currently, tags are recommended to users by sundry social tagging systems based on the tags that other users use for same items. Different meanings of items can be described by tags and hence, the need for users to create tags manually, can be eliminated [3]. Tags provided by users can be obscure owing to the labeling process which consumes great stretch of time. Such kind of inexplicit tags, known as *defective tags,* significantly impact the images' usefulness [4],[5]. Taking into consideration this issue, several solutions based on matrix factorization exist, whose training procedure is expensive. Training must be redone recurrently as and when new images are added, hence making it cumbersome for social networking sites which undergo ceaseless transformation of collections of tags and images. Hence, several data-driven techniques are coming up, and are being widely used for Tag recommendation [6].

1.1. Motivation

The explosion of social networks and Web 2.0 has created a humongous and gratifying source of information that has inspired researchers in a myriad of fields to utilize it. There is an eruptive growth in the amount of available images in social image sharing networks and hence, poses a problem to the user to store and retrieve images manually. Hence, one of the most daunting tasks in social media sites is retrieving images. Hence, annotating content with relevant labels is a good solution for accessing and retrieving images. Retrieval of images based on annotations allow users to add metadata information to images, using additional data such as tag relationship for ameliorating the quality of image retrieval.

1.2. Contribution

In this paper, we propose a Hybrid classifier-based framework for recommending image tags, which incorporates an ontology-driven and knowledge-centric strategy. The framework predominantly involves amalgamation of knowledge from Google's Knowledge API and Wikidata to derive the knowledge atomic structures that encapsulate information relevant to the annotations. Also, enrichment with domain ontology relevant to the dataset and metadata generation are done to add more background knowledge to the framework. Hence, inclusion of knowledge stores, ontologies and metadata for incorporating knowledge diminishes the rational gap betwixt the external world knowledge and knowledge amassed in the localized model.

1.3. Organization

This paper contains content as mentioned: Section II poses the related works and the extant methodologies. In section III, the architecture of the proposed framework is elucidated along with the experimental environment details. In section IV, results of the proffered methodology are discussed and compared with the baseline models. Finally, conclusions are put forth in section V.

2. Related works

Xueming Qianet. al., [1] have suggested way to retag social images with a variety of semantics. The tag's similarity to an image and the semantic corrections with the previously established tag are merged for a particular image to find the final list of tags. Jin Zanghet. al. , [2] have developed an adaptive approach to recommending social images, including building a user interest tree with deep functionality and a tag tree. Experiments using the NUS-WIDE dataset convey it transcends the enhanced techniques in respect of both precision and recall of personalized recommendations.

Shangzheng Liu et. al., [3] have developed a 3D tensor model that outlines three different entities in a social tag recommender system, and created a recommender model for three diverse datasets. Xing Xu et. al., [4] have used practical vocabulary to address the concerns of mobile users' social image tagging on social media. They introduced a new technique based on nonlinear matrix completion to mark improperly tagged images. Zechao Li et. al., [5] have proposed a Deep Collaborative Embedding (DCE) model for exposing a unified latent space of tags and images. The proposed architecture blends the conjunctive factor analysis and end-to-end learning in a coalesced framework for the impeccable compatibility of representation learning and uncovering of latent space.

Lambert rose net. al. , [6] have reviewed the latest methods for automatically annotating and enhancing social image tags, also considering the temporal patterns of their use.They have also discussed web video sequence tag suggestions and localization add-ons. Di Lu et. al., [7] have proposed a novel model based on Visual Attention that not just provides deeper visual comprehension on the decisions of the model, but also surpasses adept position-dependent rank aggregation technique to accrue numerous ranking results based on specification of the user preferences.

Zhao et. al., [8] have presented OSIR, a solution methodology to ease the assorted preference styles in searching online social media images through textual question inputs. Jianlong Fu et. al., [9] have sorted and assessed many Deep learning based image tagging techniques. They have also explained about the applications of image tagging and problems associated with them, including data collection, existing commercial systems and evaluation metrics. Yaxiong Wang et. al., [10] have come up with a novel technique using latent topic analysis for variegating the result retrieval with the NUS-Wide dataset. The proposed architecture's efficiency is implied by the outcomes.

3. Proposed architecture

Fig. 1. illustrates the architecture of the proposed framework. The model uses the COIL-100 dataset that is pre-processed in order to draw out the labels. The dataset is categorical, comprising of the image labels as well as the categories. If there is a category, it is considered a label in itself. The labels from the images are extracted. The labels are quite shallow and have a modicum of information. Hence, there is a need for enrichment of the labels in order to

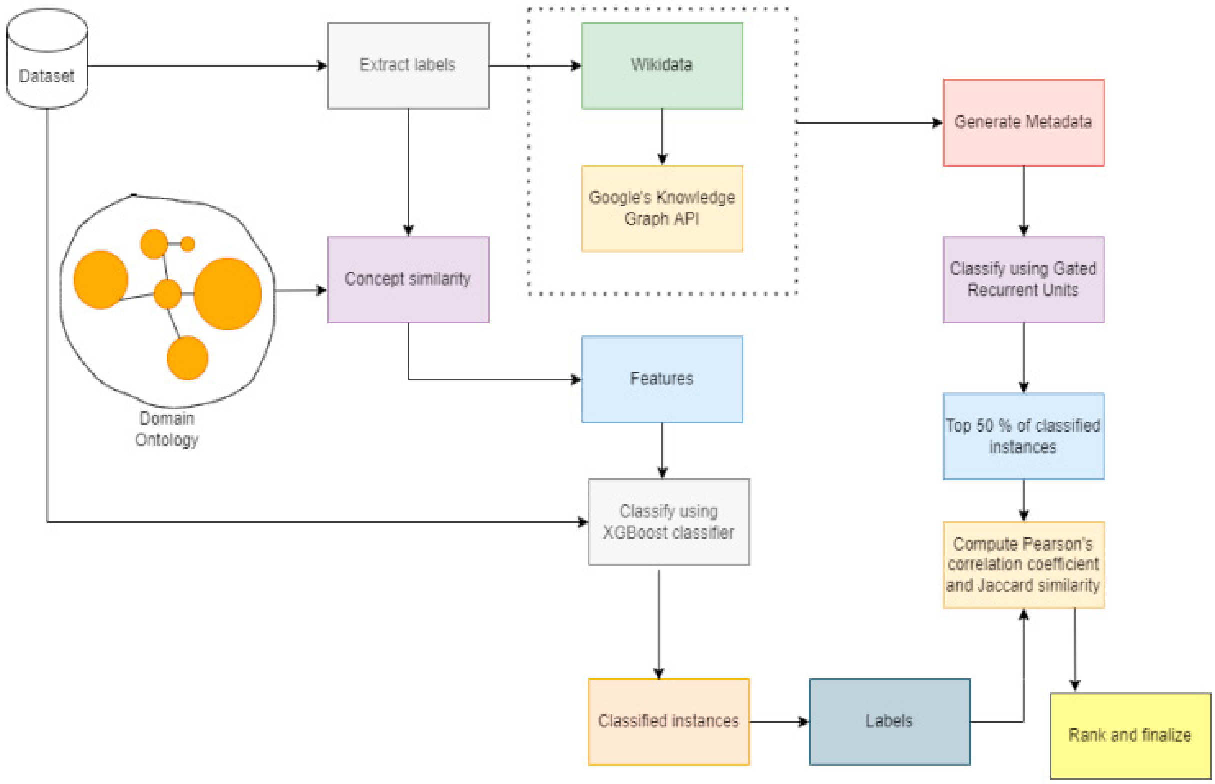

Fig 1. Ontology focused image tagging framework

make it more informative. In order to do so, instead of going for a topic modelling framework, and to lower the complexity of the framework directly, the labels are passed through Wikidata knowledge store via the Wikidata API and the Google's Knowledge Graph API. The Wikidata API yields Lexico-syntactic, hierarchical linked knowledge. The Google's Knowledge Graph API also yields knowledge graphs or knowledge sub-graphs which are light weight in nature and relevant to the label. However, these entities gleaned from Wikidata and the Google Knowledge Graph API are still paltry. Therefore, there is a need for generation of metadata. Metadata generation is done using metadata harvester RepoMMan. The RepoMMan architecture encompasses distinct layers including the RepoMMan and Fedora Web Services, BPEL processes and Adobe Flex Presentation layer. The metadata generated by RepoMMan is subjected to classification done using Gated Recurrent Units. The GRUs are deep learning classifiers. Being a constituent of a particular model of Recurrent Neural Networks, it aims to use links through a succession of nodes so as to do machine learning tasks associated with clustering and memory, like recognition of speech. Automatic feature selection is used to classify the metadata. Since, features are learnt directly from the metadata, GRUs or deep learning classifiers are quite feasible. GRU being a very powerful classifier, the metadata which is exponentially at a large scale of volume, is classified. The top 50% of the catergorized instances are taken from the output of the GRU classifier owing to the fact that the metadata is of exponentially large scale, and complexity must be avoided. Subsequently, domain ontology is used. This domain ontology is either statically modelled or it is modelled based on the domain relevant crawled data. The ontology is formulated using the Web Protege or automatic ontology generated like the Stardog. Ultimately, the generated domain ontology is aggregated with that of the extracted labels from the dataset by computation of concept similarity. The concept similarity is set to a threshold of 0.5 in order to ensure that most of the elements from the labels extracted and that of the domain ontologies get aligned. Concept similarity passed instances are selected as features and the same features are passed into the XGBoost classifier in order to classify the dataset. XGBoost, also known as Extreme Gradient Boosting, is a distributed Gradient Boosted Decision Tree machine learning library. The reason for using this machine learning classifier is that it is controlled manually. Hence, semi-automatic feature selection takes place instead of auto-handcrafted feature selection. So, feature controlling can steer classifying the data. All the classified instances, which are outcomes of the XGBoost classifier are taken and the labels are extracted. These labels are used to compute the Semantic Similarity using the Pearson's correlation coefficient and Jaccard Similarity with the outcome of the top 50 % of the classified instances from the GRUs. Pearson's correlation coefficient is used to determine the strength of the inter-dependence between two variables. The closer the two variables are to each other, the closer the value tends to be 1 or - 1.The mathematical representation is given as:

$$r = \frac{\mathrm{N}\sum xy - \sum x \sum y}{(\mathrm{N}\sum x^2 - (\sum x^2))(N\sum y^2 - (\sum y^2))} \tag{1}$$

Jaccard Similarity is a vicinage measure for computing the correspondence amongst entities. The mathematical representation is given as:

$$J(A,B) = \frac{|\mathrm{A} \cap \mathrm{B}|}{|A| + |B| - |A \cap B|} \tag{2}$$

The Pearson's correlation coefficient is set through a step deviation of 0.3 and the Jaccard Similarity is set to an empirical variation of 0.70. This is done mainly due to the reason that very high quality tags have to be formulated and must be exhibited in the framework. These generated tags are extracted and ranked based on the increase in the Jaccard similarity value. These finalised tags are aggregated with those of the categories or labels. Hence, the images are finally tagged.

3.1. Experimental environment

For the implementation of the proposed framework, the Columbia Object Image Library(COIL-100) dataset was obtained from the online Machine learning community Kaggle. There are 7200 images in this dataset, with 100 different objects having a wide variety of complex geometric and reflectance characteristics.The implementation was carried out using I7 processor with 16GB RAM. The textpreprocessing operations were done using the Python's NLTK library in Google's Colaboratory. Onto Colab was used for generation of ontologies. For static modelling of ontologies, Web Protégé was made use of. For reasoning and removal of ontological inconsistencies, HermiT Reasoner was used.

4. Results and performance evaluation

The performance of the proffered HCOntoTR (Hybrid Classifier based Ontology driven image Tag Recommendation) framework is evaluated using Precision, Recall, Accuracy and F-measure for quantifying the relevance of the results, and the FDR (False Discovery Rate) for determining the false positives count which are furnished by the system. In order to rate the performance of the HCOntoTR, it is baselined with UIT (User-Image-Tag), HRSDL and RSSVM models respectively. It is indicated from Table.1. that the proffered HCOntoTR furnishes highest precision, recall, accuracy and F-measure percentages of 94.49, 97.31, 95.90 and 95.87 respectively and lowest FDR of 0.16.

Table. 1. Comparison of performance of the proposed HCOntoTR model with the baseline models

Model	Average Precision %	Average Recall %	Average Accuracy %	Average F-Measure %	FDR
UIT	84.32	86.79	85.55	85.53	0.16
HRSDL	90.19	93.07	91.63	91.60	0.10
RSSVM	91.21	94.43	92.82	92.79	0.09
Proposed HCOntoTR	94.49	97.31	95.90	95.87	0.06

The reason why the HCOntoTR yields the highest precision, accuracy, recall and F-measure with a very low FDR value is mainly due to the fact that the HCOntoTR is knowledge centric which means it amalgamates a lot of knowledge from several sources namely the Wikidata and Google's Knowledge API. Using the same knowledge sources, atomic knowledge structures were obtained. These atomic knowledge structures encapsulate more information relevant to the labels. However, these just act as a map. In order to accumulate and culminate more knowledge, the metadata is further generated and the usage of both feature control Machine learning classifier to classify the dataset and Deep learning auto-handcrafted GRU to classify the metadata ensures that this framework accommodates two classifiers at two different ends. Also, the enrichment with domain ontology which is relevant to the dataset ensures much more background knowledge which is added into the framework. So, by inclusion of knowledge stores, ontologies and metadata for incorporating knowledge ensures that the semantic disparity betwixt the external world knowledge and the knowledge accumulated into the localised framework is met, i.e., the cognitive gap is reduced. Apart from that, the usage of Pearson's correlation coefficient and Jaccard similarity for determining the Semantic Similarity value and the usage of concept similarity for accumulating labels and the ontologies ensures that a high degree of varied relevance computation mechanisms are encompassed into the framework and as a result, the proposed model performs a cut above the models serving as basis. The reason why UIT model produces the least values for Precision, Accuracy, Recall and F-measure along with a high FDR is that the user image tag model uses semantic rules for accumulating the tags. Although it is socially aware, the main reason why this model lags is because it uses correlation with type additive graph construction. However, external knowledge is not taken into consideration and strong classifiers in the model are absent and hence, the UIT model [11] lags. More emphasise is given to the user based cluing mechanism which is not very appropriate. In [14], even though correlation between labels, feature learning and ranking are coalesced in the RSSVM (deep Ranking Structural Support Vector Machine) framework, due to lack of knowledge aggregation as used in the proposed model, it lags in performance owing to the semantic gap inspite of the incorporation of context information. Also, the ALPTR model [12] doesn't perform as expected, although it incorporates adversarial learning for tag recommendation with personalization. A chief concern is learning's reliance on dataset. The dataset dependency with a strong learning framework ensures overfitting at some instances in terms of learning and underfitting in terms of tagging. As a result, the joint training of user preference and visual encoding in this model brings about overfitting during learning and demonstrates underfitting during the tagging process because the external knowledge is completely neglected in the framework. The ILEDL model [13] recommends hashtag based on image label extraction and deep learning framework where CNN model is used. A

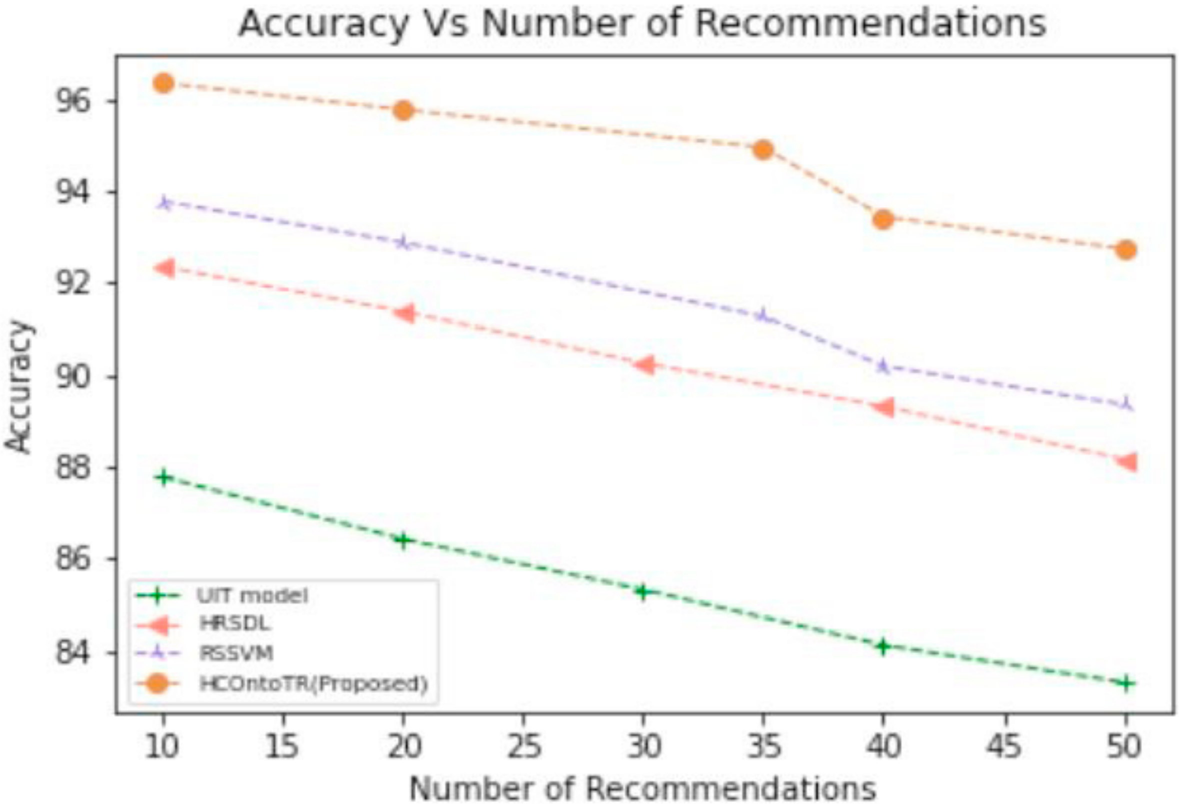

Fig 2. Precision percentages versus number of recommendations for several approaches

strong classification model alone is imbibed. Also, importance for external knowledge is not given. Hence, overtraining happens. So, image features are given the most significance. Text features are given less emphasise. Hence, the ILEDL model doesn't perform well as expected when compared to the proposed model. Fig 2. illustrates the performance versus count of recommendations distribution curve where the proposed HCOntoTR occupies the uppermost position which is immediately followed by the RSSVM model. Then comes the HRSDL in the hierarchy and lowest in the hierarchy is the UIT model.

5. Conclusion

Social image tagging focuses on eliminating tags which are noisy and recommend new pertinent tags. It plays a significant role in employing the accruing quantity of multimedia tagged by novices. We propose the knowledge centric HCOntoTR framework used for social image tagging in this paper. This Hybrid classifier based model uses XGBoost classifier and GRUs for classifying instances. The framework entails Semantic Intelligence, Knowledge aggregation and ontology-driven strategy. The performance evaluation of the HCOntoTR is done by baselining it with the UIT, HRSDL and RSSVM models. So, the incorporation of knowledge in the form of ontologies, knowledge stores and metadata ensures that the congnitive gap is reduced. Also, the usage of Pearson's correlation coefficient and Jaccard similarity for computing the Semantic Similarity and the usage of concept similarity for accumulating labels and the ontologies ensure that a high degree of varied relevance computation mechanisms are encompassed into the framework and as a result, the proposed framework furnishes finer outcome than the reference models. As a part of future work, for the auxiliary knowledge incorporated, we intend to do further reasoning. Also, further hybridization of novel classification strategies could be incorporated.

References

[1] Xueming Qian, Xian-Sheng Hua, Yuan Yan Tang, and Tao Mei. (2014) "Social image tagging with diverse semantics." *IEEE transactions on cybernetics* **44** (**12**): 2493–2508.

[2] Jing Zhang, Ying Yang, Li Zhuo, Qi Tian, and Xi Liang. (2019) "Personalized recommendation of social images by constructing a user interest tree with deep features and tag trees." *IEEE Transactions on Multimedia* **21** (**11**): 2762–2775.

[3] Shangzheng Liu, and Bin Liu. (2015) "Personalized Social Image Tag Recommendation Algorithm Based on Tensor Decomposition.", in *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)* : 1025–1028.

[4] Xing Xu, Li He, Huimin Lu, Atsushi Shimada, and Rin-Ichiro Taniguchi. (2016) "Non-Linear Matrix Completion for Social Image Tagging." *IEEE Access* **5** : 6688–6696.

[5] Zechao Li, Jinhui Tang, and Tao Mei. (2018) "Deep Collaborative Embedding for Social Image Understanding." *IEEE transactions on pattern analysis and machine intelligence* **41** (**9**): 2070–2083.

[6] Lamberto Ballan, Marco Bertini, Tiberio Uricchio and Alberto Del Bimbo. (2015) "Data-driven approaches for social image and video tagging." *Multimedia Tools and Applications* **74** (**4**): 1443–1468.

[7] Di Lu, Leonardo Neves, Vitor Carvalho, Ning Zhang, and Heng Ji. (2018) "Visual attention model for name tagging in multimodal social media.", in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* **1**: 1990–1999.

[8] Xuezhuan Zhao, Lishen Pei, Tao Li and Zheng Zhang. (2020) "Online social image ranking in diversified preferences." *EURASIP Journal on Image and Video Processing* **53** (**1**): 1–28.

[9] Jianlong Fu, and Yong Rui. (2017) "Advances in deep learning approaches for image tagging." *APSIPA Transactions on Signal and Information Processing* **6** (**11**): 1–13.

[10] Yaxiong Wang, Li Zhu and Xueming Qian. (2021) "Social image retrieval based on topic diversity." *Multimedia Tools and Applications* **80** (**8**): 12367-12387.

[11] Jing Zhang, Ying Yang, Qi Tian, Li Zhuo, and Xin Liu. (2017) "Personalized social image recommendation method based on user-image-tag model." *IEEE Transactions on Multimedia* **19** (**11**): 2439–2449.

[12] Erik Quintanilla, Yogesh Rawat, Andrey Sakryukin, Mubarak Shah, and Mohan Kankanhalli. (2020) "Adversarial learning for personalized tag recommendation." *IEEE Transactions on Multimedia* **23**: 1083–1094.

[13] Kim, Seon-Min, Cho, and Dae-Soo. (2020) "Design and Implementation of Hashtag Recommendation System Based on Image Label Extraction using Deep Learning." *The Journal of the Korea institute of electronic communication sciences* **15** (**4**): 709–716.

[14] Gang Chen, Ran Xu, and Zhi Yang (2018) "Deep ranking structural support vector machine for image tagging." *Pattern Recognition Letters* **105** : 30–38.