

Automatic data categorization by Multi label Classification Using Semi-Supervised Singular Value Decomposition

Pooja Mhaske¹, Aishwarya Thote², Shreyan Saraswat³, Mayur Songire⁴, Prof. A.V.Kolapkar⁵

^{1,2,3,4} Student, Computer, K. K. Wagh Institute Of Engineering Education & Research Nashik, Maharashtra, India

⁵ Professor, Computer, K. K. Wagh Institute Of Engineering Education & Research Nashik, Maharashtra,

ABSTRACT

In various domains including automatic multimedia data categorization, multi label problems are arised, and have generated significant interest in computer vision and machine learning. However, existing methods do not adequately address two key challenges i.e. exploiting correlations between labels and making up for the lack of labelled data or even missing labelled data. We propose use of a semi supervised singular value decomposition to handle these two challenges. The proposed model takes advantage of the nuclear norm regularization on the Singular Value Decomposition electively capture the label correlations. Proposed method can exploit the label correlations and obtain promising and better label prediction results than the state-of-the-art methods.

Keyword:- Singular value decomposition, multi-label, nuclear norm regularization.

1. INTRODUCTION

The explosive growth of online content such as images and videos nowadays has made developing classification system a very challenging problem. Such new classification system is usually required to assign multiple labels to one single instance: an image might be annotated by many semantic tags in image classification; one article can focus on several topics for text mining, each gene probably responses to multiple biological processes during gene function prediction.

In this project we are going to analyse the image data set which may be multi labelled or lack of labels. After taking the data set we are going to do labelling using the clarify API to overcome the challenges of lack of image processing. Then using all the data set we are going to do image processing and use the gathered data in matrix format and use SVD algorithm to get the efficient image objects searching in desktop application.

Multi label problems arises in image processing, automatic or manual multimedia data categorization, and it have already generated significant interest in related domains and computer vision. However, there are problems for exploiting correlations between labels and some lack of labelled data and unlabelled data. So for solving these problems this semi supervised SVD will work fine. This model takes advantage of the nuclear norm regularization on the Singular Value Decomposition and electively capture the label correlations. Proposed method can exploit the label correlations and obtain promising and better label prediction results than the state-of-the-art methods.

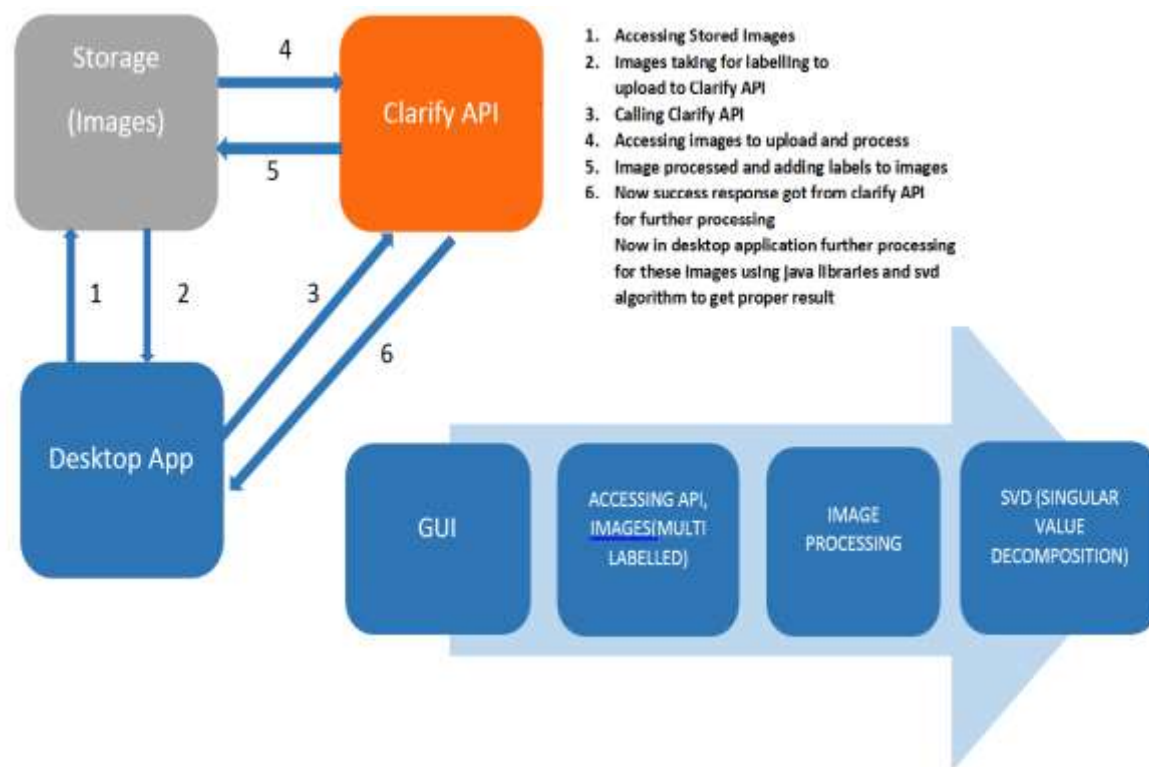
2. LITERATURE SURVEY

A semi-supervised Low-Rank Mapping (SLRM) model is used to find the mapping from feature space to label space given multi-label data. SLRM takes advantage of multi-labelled and unlabelled data, and has ability to handle missing labels by considering both label correlations and intrinsic structure among data. SLRM is easy to be extended by introducing different loss functions such as the least squares, the hinge and the logistics loss functions, and it is theoretically proved having an upper error bound which indicates the proposed model has good generalization properties. An efficient algorithm is presented to solve SLRM model under the alternating direction method of multipliers (ADMM) framework. Experiments have been conducted on a synthetic data as well as seven real-world multi-label multimedia data sets by comparing SLRM with existing methods (CPLST, FAIE, MLLOC, MC, MIML, and CRBM). The numerical results shows the effectiveness and efficiency of SLRM so we are going to use the same.

3. PROPOSED SYSTEM

In this project we are going to analyse the image data set which may be multi labelled or lack of labels. After taking the data set we are going to do labelling using the clarify API to overcome the challenges of lack of image processing. Then using all the data set we are going to do image processing and use the gathered data in matrix format and use SVD algorithm to get the efficient image objects searching in desktop application. System is designed to search images features like text without multi labelling, labelling or with lack of labels. Like Small data set containing limited labels, our system is only valid for image data set.

4. SYSTEM ARCHITECTURE DIAGRAM



5. CONCLUSIONS

Through this project we are designing the three layer model to classify the image data set. In this project we will be finding out the time required to search and categorization for given data set. The correlations among the images label are detected from process. Each detected possible correlation is processed to see the improvement. We are only aiming to improve the categorization of large or small data set. Data categorization result to user for utilizing multimedia data provides a good way to reduce the required labelled data with improving the classification performance. It will efficiently deal with large-scale data sets. We study the effectiveness of our approach hope to design a self-adaptive parameter for video data set.

6. ACKNOWLEDGEMENT

It gives us great pleasure in presenting the preliminary project report on 'Automatic data categorization by Multi label Classification Using Semi-supervised Singular Value Decomposition'.

We would like to take this opportunity to thank our internal guide Prof. A.V. Kolapkar for giving us all the help and guidance we needed. We are really grateful to them for their kind support. Their valuable suggestions were very helpful.

We are also grateful to Prof. S.S. Sane, Head of Computer Engineering Department, K. K. Wagh Institute of Engineering Education and Research for his indispensable support, suggestions.

7. REFERENCES

1. Z. Akata, F. Perronnin, Z. Harchaoui, and C. Schmid, Label-embedding for attribute-based classification, in Proc. CVPR, 2013, pp. 612620.
2. R. Cabral, F. Torre, J. Costeira, and A. Bernardino, Matrix completion for weakly supervised multi label image classification, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 1, pp. 121135, Jan. 2015.
3. D. Cai, X. He, J. Han, and T. Huang, Graph regularized non-negative matrix factorization for data representation, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 15481560, Aug. 2011.
4. G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, Robust recovery of subspace structures by low-rank representation, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171184, Jan. 2013.
5. M. Razaviyayn, M. Hong, and Z.-Q. Luo, A unied convergence analysis of block successive minimization methods for nonsmooth optimization, *SIAM J. Optim.*, vol. 23, no. 2, pp. 11261153, 2013.
6. F. Tai and H. Lin, Multi label classification with principal label space transformation, in Proc. ICML, 2010, pp. 25082542.
7. G. Tsoumakas, I. Katakis, and I. Vlahavas, Random k-labelsets for multi label classification, *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 7, pp. 10791089, Jul. 2011.

