Mapreduce Function in Hadoop for Mining Weakly Labeled Web Facial Images for Search Based Face Annotation

[1]Chithara J, [2]Prakash E P
[1]P.G. Scholar, [2]Assistant Professor
[1]Department of CSE
[1]SNS College of Engineering,Coimbatore, India

***Abstract*-Web mining is the application of data mining techniques to discover patterns from the Web. The main objective of web mining is to develop more intelligent tools for potentially help the user in finding, extracting, filtering and evaluating valuable information and resources. The popularity of digital images is rapidly increasing due to improving digital imaging technologies and convenient availability facilitated by the Internet. To improve the retrieval accuracy of content-based image retrieval systems, the research focus has been shifted from designing sophisticated low-level feature extraction algorithms to reducing the 'semantic gap' between the visual features and the richness of human semantics. Search-Based Face Annotation (SBFA) by mining weakly labelled facial images that are freely available on the World Wide Web (WWW). One challenging problem for search-based face annotation scheme is to effectively perform annotation by exploiting the list of most similar facial images and weak labels that are often noisy and incomplete. To overcome the above issues, a map reducing technique is developed to improve the efficiency and scalability of the images. Map reduce program is composed of a map procedure that performs filtering and sorting.**

***Keywords*-Face annotation, web facial images, search base face annotation, weak label search based facial annotation.**

I.INTRODUCTION

Digital photo albums are growing explosively in both number and size due to the rapid popularization of digital cameras and mobile phone cameras in the last decade. These large collections require the annotation of some semantic information to facilitate browsing manipulation and sharing of photos.

The large number of human facial images shared over the different social real world application some of this images are tagged properly but many of images are not tagged properly so the facial annotation are came.so the face annotation technology is important for photo management. Facial annotation also applied in video domain to identify the person who appeared in video.

The model base annotation has more limitations i.e.it is more time consuming and more costly to collect large amount of human labeled training facial image.It is more difficult to generalize the models when new persons are added in which retraining process is required and last the annotation performance is become poor when the number of person is very more.The "auto face annotation" is important technique which automatically gives name of relevant person.This technique is more beneficial to different real world application for(e.g.facebook)which annotates photos uploaded by the users for managing online album and searches the photos.Recently search base annotation are used for facial image annotation by mining the World Wide Web(WWW),where large number of Weakly-labeled facial images are freely available.The search-based face annotation paradigm aims to tackle the automated face annotation task by exploiting content-based image retrieval(CBIR) Techniques in mining number of weakly labeled facial images on the web.The main objectives of search-base face annotation is to assign correct name labels to a given query facial image.

II.RELATED WORK

Different studies are perform face annotation in mining weakly labeled facial images which are present over internet in this human name are treated as input query and aims is to refine the text-based search results by achieving consist facial images.

Face recognition algorithm

A straight forward idea for automatic/semi-automatic face annotation is to integrate face recognition algorithms which have been well studied in the last decade.Girgensohn et al.used face recognition technology to sort faces by their similarity to a chosen face or trained face model,reducing user workload to searching faces that belongs to the same person.however,despite progress made in recent years,face recognition continues to be a challenging topic in computer vision research.most algorithms

perform well under a controlled environment,while in the scenario of family photo management, the performance of face recognition algorithms becomes unacceptable due to difficult lighting/illumination conditions and large head pose variations.

Iterative Framework for face annotation

Recently,Riya developed an iterative framework for face annotation.In every iteration,the user was asked to manually label some faces,then the system used these labeled information to recognize faces that belong to the same person and proposed for user confirmation.Few technical details are available about iterative framework,but from experiments we can see that it still requires a lot of manual labeling to obtain final annotation results and also require user interaction for each iteration.

Pose Adaptive matching method

Pose adaptive matching method that uses pose-specific classifiers to deal with different pose combinations(e.g., frontal v.s. frontal,frontal v.s. left)of the matching face pair.It is comparable with the state-of-the-artmethods on the labeled face in wild(LFW) benchmark(achieve 84.54% recognition rate),while maintaining excellent compactness,simplicity,and generalization ability,across different datasets.But in this work,the face micro-pattern encoding is learned but pattern sampling is still manually designed.Automating this step with learning techniques may produce a more powerful descriptor for face recognition.

Graph based Approach

Ozkan and Duygulu proposed a graph-based model for finding the densest sub-graph as the most related result. Proposed a method to associate names and faces for querying people in large news photo collection In most cases the number of same faces of queried person will be large so the faces are more similar to each other. They proposed the graph based method to find the similar subset with possible set of faces with query person name. Similarity are represent by SIFT describers. Then apply a greedy graph algorithm.

Guillaumin et al. introduced a modification to incorporate the constraint that a face is only depicted once in an image. There are two scenarios of naming persons in database for finding face of person and assigning name to all faces The text based result is not greatly improved. To improve a resent graph based approach introduce the constraints when optimizing the objective function .generative models have previously been proposed to solve the multi-person naming task .by comparing generative and graph based methods the most significant method is graph based method .in future extends the graph based method to multi person naming

Guillaumin et al.proposed to iteratively update the assignment based on a minimum cost matching algorithm. In their follow-up work Guillaumin et al. they further improve the annotation performance by using distance metric learning techniques to gain more distinguish feature in low dimension space.

Content based image retrieval

Active learning has been shown as a key technique for improving content-based image retrieval (CBIR) performance. Among various methods, support vector machine (SVM) active learning is popular for its application to relevance feedback in CBIR. However, the regular SVM active learning has two main drawbacks when used for relevance feedback. First, SVM often suffers from learning with a small number of labeled examples, which is the case in relevance feedback. Second, SVM active learning usually does not take into account the redundancy among examples, and therefore could select multiple examples in relevance feedback that are similar (or even identical) to each other.

Search based face annotation

Dayong Wang, Steven C.H. Hoi et al. Propose an effective unsupervised label refinement for refining the web facial images. For improving the performance they also propose optimization algorithm to solve large-scale learning effectively i.e. clustering based approximation the propose system improve the performance of search based face annotation scheme. The work are different form all previous work by two things. To solve general content based face annotation problem using search based where face image as query image. They unsupervised label refinement algorithm which enhanced new label matrix. This work also related recent work of the WlRLCC method .The unified learning scheme .Adopted locality sensitive hashing . Adopted unsupervised face alignment technique .extract the GIST features. Despite the encouraging results,the work is limited in several aspects,First,assume each name corresponds to a unique single person.Duplicate name can be a practical issue in real –life scenarios.

III. PROPOSED APPROACH

The increasing volume of data in large networks to be analyzed imposes new challenges to an intrusion detection system. However, in order to cope with the increasing amount of data, new parallel methods need to be developed in order to make the algorithms scalable. In this paper, we propose an intrusion detection system based on a parallel particle swarm optimization

clustering algorithm using the MapReduce methodology. The MapReduce programming model has become an alternative parallel processing model for MPI especially for data intensive applications. Many advantages make the MapReduce methodology to be a good choice for parallelizing the data mining tasks such as easy implementation without having to know too many parallel programming details. In addition, MapReduce provides many solutions for node failure and load balancing. MapReduce usually divides the input data set into independent splits which depend on the size of the data set and the number of computer nodes used. MapReduce consists of two main functions: Map and Reduce functions. The Map function processes the input data records as (key, value) data pairs to generate intermediate output as (key, values list) data pairs, and then the Reduce function merges and aggregates all intermediate (values list) output coming from the Map function having the same intermediate key. built in order to deal with data-intensive applications. The Hadoop framework has its own distributed file system called the Hadoop Distributed File System (HDFS) that is used to support the management and processing of large scale data sets. Furthermore, the MapReduce in Hadoop is designed to work efficiently with HDFS by moving the computation process to the data and not the other way around to allow Hadoop to achieve high data locality.

MapReduce is a framework that allows developers to write functions that process data. There are two types of key functions in the MapReduce framework, the Map function which separates out the data to be processed and the reduce function which performs analysis on that data.

The real benefits of MapReduce start to occur when the framework for the execution of its functions is implemented in a large scale, shared nothing data cluster. The platform that implements a MapReduce framework can abstract the complexity of running distributed data processing functions across multiple nodes in the cluster.

A. *map reduce algorithm*

```
map(imageName, imagePixels):
foreach pixel p in imagePixels:
emitIntermediate(p, imageName);
done
reduce(pixel, values):
foreach imageName in values:
AddToOutputList(imageName);
done
emitFinal(FormattedImageListForPixel);
```

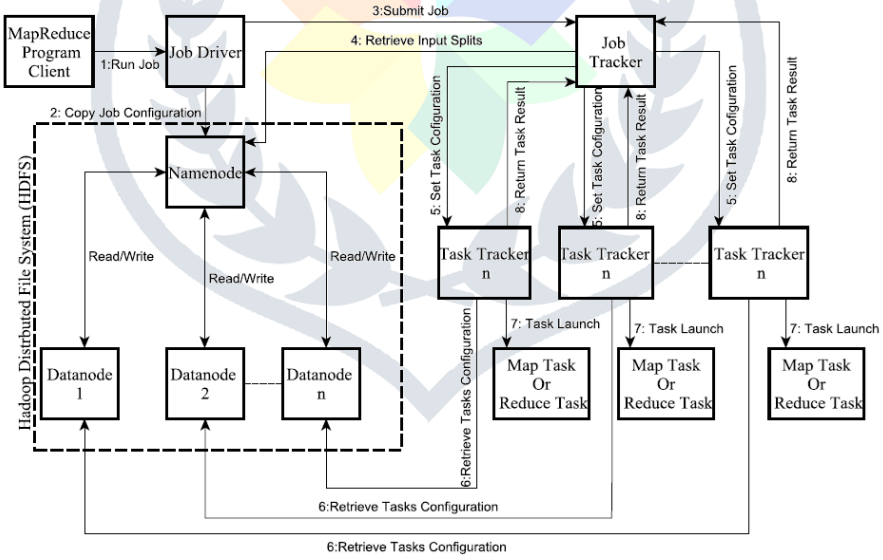
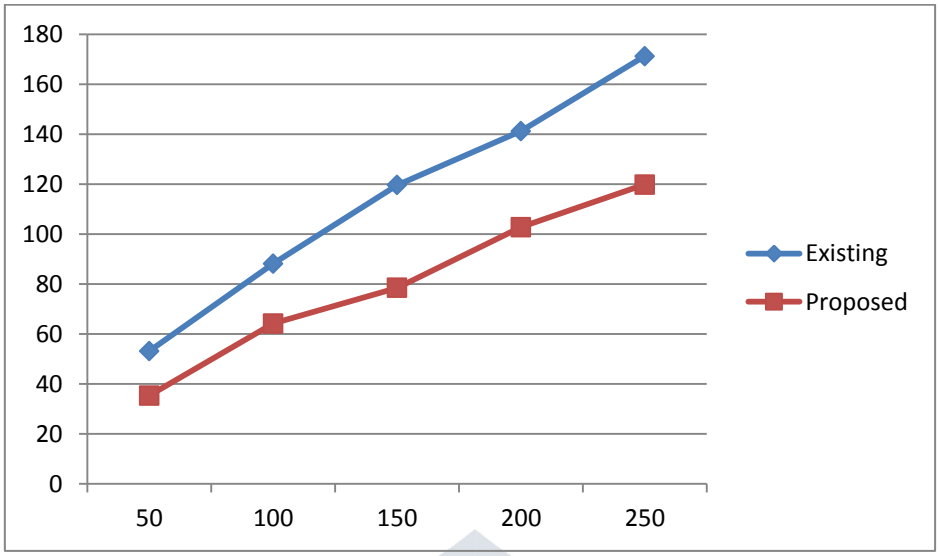


fig. 2.mapreduce function

IV. COMPARITIVE TABLE

Time Utilization	No of Database Images 50	No of Database Images 100	No of Database Images 150	No of Database Images 200	No of Database Images 250
Existing	53.2	88.2	119.6	141.2	171.2
Proposed	35.41	64.2	78.5	102.8	119.8



V. APPLICATION

Face annotation finds its application in the field of:

- Achieve relatively high performance without user interaction.
- When user interaction is included,reduce it to an acceptable level.
- Face annotation at macro scale and micro scale.
- Wild landmark face annotation.
- Online photo album management and also in video domain.

VI.CONCLUSION

A promising search-based face annotation framework is developed on tackling the critical problem of enhancing the label quality and a ULR algorithm. To further improve the scalability, develop a clustering-based approximation solution. Clustering-based approximation solution successfully accelerated the optimization task without introducing much performance degradation. From an extensive set of experiments, ULR technique achieved promising results under a variety of settings. Experimental results also indicated that the ULR technique significantly surpassed the other regular approaches. In future, plan carried out to improve the scalability and efficiency of the facial images. Map reduce program is composed of a map procedure that performs filtering and sorting.

REFERENCES

[1] Dayong Wang, Steven C.H. Hoi, Ying He, and Jianke Zhu., Jan (2014) "Mining Weakly Labeled Web Facial Images for Search- Based Face Annotation", IEEE Transactions On Knowledge and Data Engineering, Vol. 26, No. 1.

[2] Ahonen T.,Hadid A., and Pietikainen M., Feb (2004) "Face Recognition with Local Binary Patterns," Proc. European Conf. Computer Vision (ECCV), vol. 1, pp. 469-481.

[3] Akshay Asthana., Roland Goecke., Novi Quadrianto., Tom Gedeon., Dec (2009) "Learning Based Automatic Face Annotation for Arbitrary Poses and Expressions from Frontal Images Only," IEEE Conference on Computer Vision and Pattern Recognition.

[4] Cao Z., Yin Q., Tang X., and Sun J., Feb (2010) "Face Recognition with Learning-Based Descriptor," IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 2707-2714.

[5] Choi W.D., Neve K.N., Plataniotis and Ro Y.M., Feb (2011) "Collaborative Face Recognition for Improved Face Annotation in Personal Photo Collections Shared on Online Social Networks," IEEE Trans. Multimedia, vol. 13, no. 1, pp. 14-28

[6] Cui J., Wen F., Xiao R., Tian Y., and Tang X., Feb (2007) "EasyAlbum: An Interactive Photo Annotation System Based on Face Clustering and Re-Ranking," Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI), pp. 367-376.

[7] Fan J., Gao Y. and Luo H., April (2004) "Multi-Level Annotation of Natural Scenes Using Dominant Image Components and Semantic Concepts," Proc. 12th Ann. ACM Int'l Conf. Multimedia (Multimedia), pp. 540-547.

[8] Hanbury, Oct (2008) "A Survey of Methods for Image Annotation," J. Visual Languages and Computing, vol. 19, pp. 617-627.

[9] Jae Young Choi., Wesley De Neve, Yong Man Ro., and Konstantinos N. Plataniotis., Oct (2010) "Automatic Face Annotation in Personal Photo Collections Using Context-Based Unsupervised Clustering and Face Information Fusion" IEEE Transactions On Circuits And Systems For Video Technology, Vol. 20, No. 10.

[10] Lin Z., Ding G., and Wang J., Jun (2011) "Image Annotation Based on Recommendation Model," Proc. 34th Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 1097-1098.

[11] Mensink T., and Verbeek J J., Jan (2008) "Improving People Search Using Query Expansions," Proc. 10th European Conf. Computer Vision (ECCV), vol. 2, pp. 86-99.

[12] Rui X.,Li M.,Li Z., Ma W.-Y., and Yu N., Mar (2007) "Bipartite Graph Reinforcement Model for Web Image Annotation," Proc. 15th ACM Int'l Conf. Multimedia, pp. 585-594.

[13] Smeulders A.W.M., Worring M., Santini S., Gupta A., and Jain R., Dec. (2000) "Content-Based Image Retrieval at the End of the Early Years," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1349-1380.

[14] Steven C.H. Hoi., Dayong Wang., Yeu Cheng., Elmer Weijie Lin., Jianke Zhu., Ying He., Chunyan Miao., Jan (2013) "FANS: Face Annotation by Searching Large-scale Web Facial Images," ACM journal.

[15] Ibrahim Aljarah and Simone A. Ludwig., june 2013 "MapReduce Intrusion Detection System based on a Particle Swarm Optimization Clustering Algorithm" IEEE Congress on Evolutionary Computation