



An effective character multi role identification model using k- relation graph clustering in visual image datasets

1.Mr.S.Jagadeesan.,M.Sc.,M.C.A.,M.Phil.,M.E., Assistant Professor

2. Mr. C.Gowthaman Final M.C.A.

Department of MCA, Nandha Engineering College,(Autonomous), Erode-52.

jagadeesan12398@gmail.com, cgowthamanmca@gmail.com

Abstract— In this paper proposes a global face-name graph matching based framework for robust movie character identification. The proposed system two schemes are considered. There are connections as well as differences between them. Regarding the connections, the proposed two schemes both belong to the global matching based category, where external script resources are utilized. This research study to improve the robustness, the ordinal graph is employed for face and name graph representation and a novel graph matching algorithm called Error Correcting Graph Matching (ECGM) is introduced. Regarding the differences, scheme 1 sets the number of clusters when performing face clustering. The face graph is restricted to have identical number of vertexes with the name graph. While, in scheme 2, no cluster number is required and face tracks are clustered based on their intrinsic data structure. Auto face identification of characters in films has drawn most research interests and led to many interesting applications. Since huge variation in the appearance of each character is found, it is a challenging problem. Existing methods evaluates promising results in clean environment, the performances are limited in complex movie scenes due to the noises generated during the face tracking and face clustering process. This study presents two schemes of global face-name matching based framework for robust character identification. In this paper main contributions of this study include the first noise insensitive character relationship representation is incorporate, next study introduces an edit operation based graph matching algorithm, next complex character changes are handled by simultaneously graph partition and graph matching and beyond existing character identification approaches. The proposed schemes demonstrate state-of-the-art performance on movie character identification in various movies.

Index Terms—lossless image compression, data folding,KDD Process,ECGM

INTRODUCTION

Data Mining

Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data set. These tools can include statistical models, mathematical algorithm and machine learning methods. Consequently, data mining consists of more than collection and managing data, it also includes analysis and prediction. Classification technique is capable of processing a wider variety of data than regression and is growing in popularity.

Data mining refers to extracting or “mining” knowledge from large amounts of data. The term is actually a misnomer. Remember that the mining of gold from rocks or sand is referred to as gold mining rather than rock or sand mining. Thus, data mining should have been more appropriately named “knowledge mining from data,” which is unfortunately somewhat long. “Knowledge mining,” a shorter term may not reflect the emphasis on mining from large amounts of data. Nevertheless, mining is a vivid term characterizing the process that finds a small set of precious nuggets from a great deal of raw material (Figure 1.1).

Thus, such a misnomer that carries both “data” and “mining” became a popular choice. Many other terms carry a similar or slightly different meaning to data mining, such as knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology,

and data dredging. Data mining is a synonym for another popularly used term, Knowledge Discovery for Data, KDD process. Knowledge discovery as a process is depicted in Figure 1.2 and consists of an iterative sequence of the following steps:

- Data cleaning: also known as data cleansing, it is a phase in which noise data and irrelevant data are removed from the collection.
- Data integration: at this stage, multiple data sources, often heterogeneous, may be combined in a common source.
- Data selection: at this step, the data relevant to the analysis is decided on and retrieved from the data collection.
- Data transformation: also known as data consolidation, it is a phase in which the selected data is transformed into forms appropriate for the mining procedure.
- Data mining: it is the crucial step in which clever techniques are applied to extract patterns potentially useful.
- Pattern evaluation: in this step, strictly interesting patterns representing knowledge are identified based on given measures.

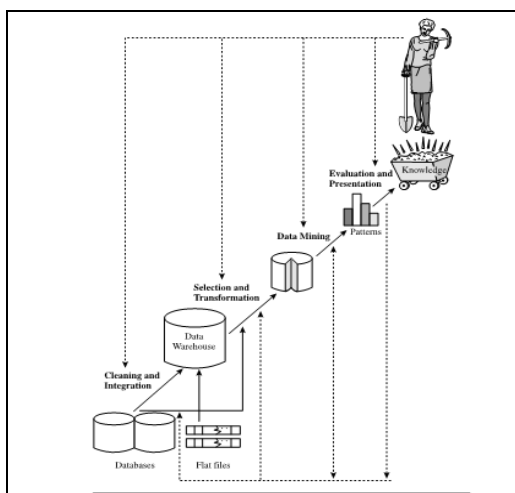


Fig 1.2 Data Mining steps in KDD Process

Knowledge representation: is the final phase in which the discovered knowledge is visually represented to the user. This essential step uses visualization techniques to help users understand and interpret the data mining results.

➤ It is common to combine some of these steps together. For instance, data cleaning and data integration can be performed together as a pre-processing phase to generate a data warehouse. Data selection and data transformation can also be combined where the consolidation of the data is the result of the selection, or, as for the case of data warehouses, the selection is done on transformed data

Data mining derives its name from the similarities between searching for valuable information in a large database and mining rocks for a vein of valuable ore. Both imply either sifting through a large amount of material or ingeniously probing the material to exactly pinpoint where the values reside. It is, however, a misnomer, since mining for gold in rocks is usually called “gold mining” and not “rock mining”, thus by analogy, data mining should have been called “knowledge mining” instead. Nevertheless, data mining became the accepted customary term, and very rapidly a trend that even overshadowed more general terms such as knowledge discovery in databases (KDD) that describe a more complete process.

Auto face identification of characters in films has drawn most research interests and led to many interesting applications. Since huge variation in the appearance of each character is found, it is a challenging problem. Existing methods evaluate promising results in clean environment, the performances are limited in complex movie scenes due to the noises generated during the face tracking and face clustering process. This study presents two schemes of global face-name matching based framework for robust character identification.

The contributions of this study include:

- A noise insensitive character relationship representation is incorporated.
- The study introduces an edit operation based graph matching algorithm.
- Complex character changes are handled by simultaneously graph partition and graph matching.
- Beyond existing character identification approaches, we further perform an in-depth sensitivity analysis by introducing two types of simulated noises.

The proposed schemes demonstrate state-of-the-art performance on movie character identification in various movies. The project has been developed using Visual Studio .Net 2005 as front end and SQL Server 2000 as back end. C# is used the coding language.

The explosion of movie and TV provides a huge amount of digital video data. It led to the need of more efficient techniques of video content analyzing and organization. Automatic video annotation is one of the key techniques to the proposed schemas. Automatic video explanation is one of the key techniques. In this paper our focus is on explaining characters in the digital videos and TV's which is termed as movie character identification. The main goal is to identify the faces of the characters in the digital video and label them with their corresponding names in the cast. The textual cues like lists, scripts, sub-titles and closed captions are usually demoralized. In a movie, characters are the focus point of interests for the audience. Their occurrences provide as a lots of clues about the structure of the movie and its contents. Automatic character identification is essential for semantic movie index and retrieval, scene segmentation, summarization and also for other applications.

Taking the Bite Out Of Automated Naming Of Characters

Mark Everingham et al [1] investigated the problem of automatically labelling appearances of characters in TV or film material with their names. This is tremendously challenging due to the huge variation in imaged appearance of each character and the weakness and ambiguity of available annotation. However, they demonstrated that high precision can be achieved by combining multiple sources of information, both visual and textual. The principal novelties that we introduce are: (i) automatic generation of time stamped character annotation by aligning subtitles and transcripts; (ii) strengthening the supervisory information by identifying when characters are speaking. In addition, they incorporate complementary cues of face matching and clothing matching to propose common annotations for face tracks, and consider choices of classifier which can potentially correct errors made in the automatic extraction of training data from the weak textual annotation. Results are presented on episodes of the TV series "Buffy the Vampire Slayer". The objective of this work is to label television or movie footage with the names of the people present in each frame of the video. As has previously been noted [2] such material is extremely challenging visually as characters exhibit significant variation in their imaged appearance due to changes in scale, pose, lighting, expressions, hair style, etc. There are additional problems of poor image quality and motion blur. They build on previous approaches which have matched frontal faces in order to "discover cast lists" in movies [3] or retrieve shots in a video containing a particular character [2] based on image queries. The main novelty we bring is to employ readily available textual annotation for

TV and movie footage, in the form of subtitles and transcripts, to automatically assign the correct name to each face image.

Alone, neither the script nor the subtitles contain the required information to label the identity of the people in the video – the subtitles record what is said, but not by whom, whereas the script records who says what, but not when. Knowledge that a character is speaking then gives a very weak cue that the person may be visible in the video. A key to the success of our method is the novel use of visual speaker detection to leverage cues from the text – visually detecting which (if any) character in the video corresponds to the speaker. This gives us sufficient annotated data from which to learn to recognize the other instances of the character. In addition to effective exploitation of cues from textual annotation, success depends on robust computer vision methods for face processing in video. They proposed extensions to our method for connecting faces in video [4], which provides robust face tracks, and a novel extension of the "pictorial structure" method [5] which gives reliable localization of facial features in presence of significant pose variations. This paper is an extended version of [2]. Previous work on the recognition of characters in TV or movies has often ignored the availability of textual annotation. In the "cast list discovery" problem [3], faces are clustered by appearance, aiming to collect all faces of a particular character into a few pure clusters (ideally one), which must then be assigned a name manually. It remains a challenging task to obtain a small number of clusters per character without merging multiple characters into a single cluster. Other work has addressed finding particular characters specified a priori by building a model of a character's appearance from user-provided training data, and efficient retrieval of characters based on example face images [4]. Assigning names given a combination of faces and textual annotation has similarities to the "Faces in the News" labelling. In this work, faces appearing in images accompanying news stories are tagged with names by making use of the names appearing in the news story text. A clustering approach is taken, initialized by cases for which the news story contains a single name and the accompanying image contains a single (detected) face.

Character-Based Movie Summarization

Jitao Sang et al [2] explained about a movie summary which is helpful for movie producer to promote the movie as well as audience to capture the theme of the movie before watching the whole movie. Most exiting automatic movie summarization approaches heavily rely on video

content only, which may not deliver ideal result due to the semantic gap between computers calculated low-level features and human used high-level understanding. They incorporate script into movie analysis and propose a novel character-based movie summarization approach, which is validated by modern film theory that what actually catches audiences' attention is the character. They first segment scenes in the movie by analysis and alignment of script and movie. Then they conduct sub story discovery and content attention analysis based on the scene analysis and character interaction features. Given obtained movie structure and content attention value, we calculate movie attraction scores at both shot and scene levels and adopt this as criterion to generate movie summary. The promising experimental results demonstrate that character analysis is effective for movie summarization and movie content understanding. The proliferation of movie content requires efficient and effective techniques for data organization and management. Summarization is one of such key techniques to obtain a brief and to the point representation of the voluminous video data. A decent movie summary is helpful for the producer to promote the movie as well as audience to capture the theme of the movie before watching the whole content. The aim of movie summarization is to select portions that most attract audiences' attention from the original movie. Nevertheless, defining which movie segments is attractive and how to efficiently integrate them into summary is subjective and still remains an open issue and deserves further study. Most recent work on movie summarization heavily relies on video content only.

However, due to the semantic gap between low-level features and high-level understanding, computer calculated audiovisual features cannot characterize attractive movie content at the semantic and affective level. According to modern film theory, "All films are about nothing - nothing but character", which reveals characters are important for movie summarization. From audiences' perspective, movie is attractive and catches their attention because they want to know about the story of characters. The occurrence and interaction of characters provide meaningful representation of the movie structure and content. Motivated by this, they investigate an alternative way to movie summarization based on character analysis. They first utilized character relation to exploit the movie structure including scene segmentation and sub story discovery.

The content attention is evaluated according to three specifically designed character interaction features. Given obtained movie

structure and content attention description, they calculated movie attraction scores at both shot and scene levels and use this as criterion to generate movie summary. Compared with low-level or mid-level audiovisual features, character features exhibit high-level meaning and thus can be considered as a more natural representation of movie semantics.

Their main contribution is to incorporate character analysis into semantic movie summarization. They proposed several novel approaches for character-based movie summarization, including 1) a scene segmentation method using analysis and alignment of character co-occurrence in movie and script. 2) A character based story flow graph to conduct sub story discovery. 3) A summarization strategy enabling both in formativeness and enjoyability of the generated summary. They presented how to formulate movie structure through character based scene segmentation and sub story discovery. Movie attraction evaluation and summarization strategy are proposed.

Learning from Ambiguously Labeled Images

Timothee Cour et al [3] describes the accessed only to partially labeled image and video collections data. For example, personal photo collections often contain several faces per image and a caption that only specifies who is in the picture, but not which name matches which face. Similarly, movie screenplays can tell us who is in the scene, but not when and where they are on the screen. They formulated the learning problem in this setting as partially-supervised multiclass classification where each instance is labeled ambiguously with more than one label. They theoretically showed that effective learning is possible under reasonable assumptions even when all the data is weakly labeled.

Motivated by the analysis, they propose a general convex learning formulation based on minimization of a surrogate loss appropriate for the ambiguous label setting. They applied their framework to identify faces culled from web news sources and to naming characters in TV series and movies. They experiment on a very large dataset consisting of 100 hours of video, and in particular achieve 6% error for character naming.

Photograph collections with captions have motivated recent interest in weakly annotated images. Using a screenplay, they can tell who is in the scene, but for every face in the images, the person's identity is ambiguous. Learning accurate face and object recognition models from such imprecisely annotated images and videos can improve many applications, including image retrieval and summarization. They investigate theoretically and empirically when effective learning from this weak supervision is possible. To put the ambiguous labels learning problem into perspective, it is useful to lay out several related learning scenarios.

In semi-supervised learning, the learner has access to a set of labeled examples as well as a set of unlabeled examples. In multiple-instance learning, examples are not individually labeled but grouped into sets which either contain at least 1 positive example, or only negative examples. In multi-label learning, each example is assigned multiple binary labels, all of which can be true. Finally, in our setting of ambiguous labeling, each example again is supplied with multiple potential labels, only one of which is correct.

There have been several papers that addressed the ambiguous label framework. Proposes several nonparametric, instance-based algorithms for ambiguous learning based on greedy heuristics, uses expectation maximization (EM) algorithm with a discriminative log linear model to disambiguate correct labels from incorrect.

Additionally, these papers only report results on synthetically-created ambiguous labels and rely on iterative non-convex optimization. They provided intuitive assumptions under which they can expect learning to succeed.

Essentially, they identify a condition under which ambiguously labeled data is sufficient to compute a useful upper bound on the true labeled error. They proposed a simple, convex formulation based on this analysis and show how to extend general multiclass loss functions to handle ambiguity. They showed that their method significantly outperforms several strong baselines on a large dataset of pictures from newswire and a large video collection.

Video-Based Face Recognition on Real-World Data

Johannes Stalkamp et al [4] describe the classification sub-system of a real-time video-based face identification system which recognizes people entering through the door of a laboratory. Since the subjects are not asked to cooperate with the system but are allowed to behave naturally, this application scenario poses many challenges. Continuous, uncontrolled variations of facial appearance due to illumination, pose, expression, and occlusion need to be handled to allow for successful recognition. Faces are classified by a local appearance-based face recognition algorithm. The obtained confidence scores from each classification are progressively combined to provide the identity estimate of the entire sequence.

They introduced three different measures to weight the contribution of each individual frame to the overall classification decision. They are distance to-model (DTM), distance-to-second-closest (DT2ND), and their combination. Both a k-nearest neighbor approach and a set of Gaussian mixtures are evaluated to produce individual frame scores.

The users must be in focus of attention, not the computer system which supplies certain functionality. In such environments, it is essential that the computer system is able to identify the people it is dealing with. A feasible approach to biometric identification is the use of facial features. Alongside speech, it is a very natural approach and mimics human recognition. The nature of the cue inherently allows for unobtrusive, interaction-free recognition, as the visibility of the face does not need any specific action. This property is indispensable in smart environments, as the necessity of individuals to explicitly cooperate with the system (e.g. by putting a finger on a fingerprint scanner) violates the central idea of such environments, since it interrupts a user's actions and moves the computer system into the focus of attention.

Their central goal is to build a real-time capable face recognition system (FRS) for uncontrolled environments. It is supposed to handle robustly real-life situations with all the challenges they bring along that make the task harder. The central challenge is to achieve unobtrusive recognition, i.e. to create a system that operates in the background and does not need specific user interaction. This is essential to grant users the

freedom to behave naturally. As a consequence of this freedom, difficulties arise from varying pose, like out-of-plane rotations, and different facial expressions. Accessories and facial hair can cause partial occlusions. Daylight leads to very different illumination depending on the time of day, time of year and weather conditions.

In spite of this hardly controllable natural influence, even the artificial light sources are withdrawn from the system's control if unobtrusive recognition as postulated above is to be implemented. Since the users, i.e. the persons to be recognized, are not supposed to be restrained by the system, they are free to switch on and off any light sources that might be available. This leads to a wide variety of illumination configurations in terms of light intensity, direction and even color.

Names and Faces in the News

Tamara L. Berg et al [5] describe the good face clustering is possible for a dataset of inaccurately and ambiguously labelled face images. Our dataset is 44,773 face images, obtained by applying a face finder to approximately half a million captioned news images. This dataset is more realistic than usual face recognition datasets, because it contains faces captured "in the wild" in a variety of configurations with respect to the camera, taking a variety of expressions, and under illumination of widely varying color. Each face image is associated with a set of names, automatically extracted from the associated caption. Many, but not all such sets contain the correct name. They clustered face images in appropriate discriminant coordinates.

They use a clustering procedure to break ambiguities in labelling and identify incorrectly labelled faces. A merging procedure then identifies variants of names that refer to the same individual. The resulting representation can be used to label faces in news images or to organize news pictures by individuals present. An alternative view of their procedure is as a process that cleans up noisy supervised data.

They demonstrated how to use entropy measures to evaluate such procedures. It is Straight forward to obtain enormous datasets of images, with attached annotations. Examples include: collections of museum material ; the Corel collection of images; any video with sound or

closed captioning; images collected from the web with their enclosing web pages; or captioned news images. Exploiting partially supervised data is a widely studied theme in vision research. Image regions may usefully and fairly accurately be linked with words, even though the words are not linked to the regions in the dataset originally. For example, models based around templates and relations may be learned from a dataset of motorcycle images where one never specifies where in the image the motorcycle lies. They showed that faces and names can be linked in an enormous dataset, despite errors and ambiguities in proper name detection, in face detection and in correspondence.

Face recognition is well studied, and cannot be surveyed reasonably in the space available. Early face recognition is done in and is reviewed. Their problem is slightly different from face recognition, in that it is more important to identify discriminant coordinates — which can be used to distinguish between faces, even for individuals not represented in the dataset — than to classify the faces. As a result, they focused on adopting the kPCA/LDA methodology, rather than on building a multi-class classifier.

Their current work is a necessary precursor to real world face recognition machinery: building large and realistic sets of labelled data for recognition. They can leverage past work by using it to determine what features might be useful for identifying similar faces. The general approach involves using unambiguously labelled data items to estimate discriminant coordinates. They then use a version of k -means to allocate ambiguously labelled faces to one of their labels. Once this is done, they clean up the clusters by removing data items far from the mean, and re-estimate discriminant coordinates. Finally, they merged clusters based on facial similarities.

On Affine Invariant Clustering and Automatic Cast Listing in Movies

Andrew Fitzgibbon describes a distance metric for clustering and classification algorithms which is invariant to affine transformations and includes priors on the transformation parameters. Such clustering requirements are generic to a number of problems in computer vision. They extended existing techniques for affine-invariant

clustering, and show that the new distance metric outperforms existing approximations to affine invariant distance computation, particularly under large transformations. In addition, they incorporate prior probabilities on the transformation parameters. This further regularizes the solution, mitigating a rare but serious tendency of the existing solutions to diverge. For the particular special case of corresponding point sets they demonstrate that the affine invariant measure they introduced may be obtained in closed form.

As an application of these ideas they demonstrate that the faces of the principal cast of a feature film can be generated automatically using clustering with appropriate invariance. This is a very demanding test as it involves detecting and clustering over tens of thousands of images with the variances including changes in viewpoint, lighting, scale and expression. Clustering and classification problems abound in the applied sciences, in applications from citation indexing to the study of gene function. The task of clustering is to divide a large amount of data into disjoint subsets or classes, such that some measure of distance is minimized within classes, and maximized between classes.

In computer vision, several recent advances have incorporated clustering algorithms for the canonicalization of large data sets: selecting exemplars building unsupervised object recognizers text on generation and learning in low-level vision. To overcome this, the distance measures must include a model of the noise process underlying the measurement errors, and the clustering algorithms must employ sophisticated search techniques in order to minimize the distance. In some problems—particularly those arising in vision applications—there is another common source of variation, caused when the observed data undergo a parameterized transformation.

EXISTING SYSTEM

The existing system includes clustering faces using K-means, where the number of clusters is set as the number of distinct speakers. Co-occurrence of names in script and ace clusters in video constitutes the corresponding face graph and name graph. It modifies the traditional global matching framework by using ordinal graphs for

robust representation and introducing an ECGM-based graph matching method.

For face and name graph construction, the existing system represents the character co-occurrence in rank ordinal level, which scores the strength of the relationships in a rank order from the weakest to strongest. The affinity graph used in the traditional global matching is interval measures of the co-occurrence relationship between characters.

Drawbacks

- While continuous measures of the strength of relationship hold complete information, but it is highly sensitive to noises.
- Same character name for given face is tracked even if movies vary.
- Noise removal process is not discussed.
- The sequential statistics for the speakers is not carried out.

PROPOSED SYSTEM

In addition to existing system implementation, the proposed system includes identifying face/names for different movies with different character names also. K Means clustering is applied for clustering process. And also, different character names for given face is tracked even if movies vary. Median filter concept is applied to remove the noise either before/after the clustering process.

Advantages

- Not sensitive to noises, since noise removal can be applied either before or after clustering process.
- Different character name for given face is tracked even if movies vary.
- Noise removal process is carried out.

CONCLUSION

The distance metric developed in this paper is used in a standard clustering algorithm to extract the principal cast list from the several thousand faces detected in a typical feature film. Clustering algorithms for computer vision must allow for, or be invariant to, such transformations. Our contribution in this paper is to develop a set of distance functions which take account of affine transformation of the data. These distance functions may either be invariant to the transformation or

contain priors based on the transformations parameters. Closed form and numerical iterative solutions are given for these functions. This enables a Bayesian maximum a posteriori (MAP) cluster estimation.

REFERENCES

- [1] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, Vol. 2, No. 3, pp. 27:1-27:27, 2011, software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [2] R. Chaudhari and A. M. Patil, "Content based image retrieval using color and shape features," *International Journal of Advanced Research in Electrical, Electronics and instrumentation Engineering*, Vol. 1, No. 5, pp. 386-392, 2012.
- [3] R. Choudhary, N. Raina, N. Chaudhary, R. Chauhan, and R. H. Goudar, "An integrated approach to content based image retrieval," *Proc. International Conference on Advances in Computing, Communications and Informatics*, pp. 2404-2410, 2014.
- [4] W. S. Chow and M. K. M. Rahman, "A new image classification technique using tree-structured regional features," *Neurocomputing*, Vol. 70, No. 4-6, pp. 1040-1050, 2007.
- [5] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color texture regions in images and video," *IEEE Transactions of Pattern Analysis and Machine Intelligence*, Vol. 23, No. 8, pp. 800- 810, 2001.