E-ISSN: 2321–9637
Volume 1, Issue 5, December 2013
International Journal of Research in Advent Technology
Available Online at: http://www.ijrat.org

ACTION AND HORROR GENRE IDENTIFICATION OF MOVIES USING ARTIFICIAL NEURAL NETWORK

Afsar Baig M[1] , Vilas Naik[2]
[1] *Department of Computer Engineering* [2] *Department of Computer Science and Engineering*
[1] *Anjuman-e-Islam Kalsekar Technical Campus, New Panvel, Navi Mumbai*
[2] *Basaveshwar Engineering College, Bagalkot*

[1] *mab_sgp@yahoo.co.in*

ABSTARCT:
Motion-pictures play a significant role in fulfilling people's entertainment needs. There is a substantial amount of multimedia data in the world, diverse in content and origin. In order to make efficient use of this data it should be labeled or indexed in some manner. Such labeling would make it easier for individuals to retrieve the type of content desired. Today, due to advancements in Internet technology, consumers have access to an abundant amount of movies from various on-line services. This has created the need for automatic content-driven movie retrieval. Although considerable advancements have been made in the areas of video retrieval and collaboration. Movie genres identification still act as a key task in such systems. The proposed method for Action and Horror genre classification uses Artificial Neural Network (ANN). The system consists of an artificial neural network trained with values of color pixels. The system is tested with a dataset of 100 sample movie videos taken from a public domain YouTube. The system achieved over all accuracy of 90% in classifying the video into action and horror genres.

Keywords: *Action; Horror; Genre; Artificial Neural Network.*

1. INTRODUCTION

The automatic labeling of video footage according to genre is a common requirement when dealing with indexing of large and heterogeneous collection of video materials. This task may be addressed, either globally, or locally. Global-level approaches aim at classifying videos into one of several main genres, e.g. cartoons, music, news, sports, documentaries, etc. But the need to classify into sub-genres like movies(Action, Horror, etc) exists. In this paper focus is done on how a video sample can be classified into Action or Horror video and video genre classification is consequently interpreted as a typical machine learning problem that involves two fundamental steps: feature extraction and data classification. Especially the choice of a suitable task specific feature set is critical for the success of such a classification approach and an ideal feature set should contain as many genre specific cues as possible. Due to advancements in Internet technology, consumers have access to an unprecedented amount of movies from archives and various online services. This has created the need for automatic content-driven, user constrained choice of videos. Although considerable advancements have been made in the areas of video retrieval, annotation and indexing. The most needed task is to automate the classification process for classifying the movies to match the user preferences. Being able to automatically classify movies into variant genres by human intelligence has motivated the automation of the same. This will simplify 1) indexing multimedia databases to help search for particular types of movies, 2) automatically identifying movies for consumers through user preference modeling, 3) facilitating automatic movie content altering and summarization.

2. RELATED BACKGROUND

There have been a number of interesting studies on automatic genre classification of video files. Z Rasheed et.al. four low-level visual features, average shot length, color variance, key-lighting, and motion content, in order to classify trailers into four genres: action, comedy, drama, and horror. The results showed that the features, which were inspired by cinematic practices, led to a good genre classification performance on a set of 100 Hollywood movie trailers.[6] One of the contributions of this work is to revisit these features for a much larger trailers dataset.

Howard Zhou et.al Experimented and demonstrated to show the usefulness of introducing scene features for movie genre prediction.[7] However, the results leave room for improvement. First, the method does not consider the dynamic components of scenes, since it lacks a representation of the action and movement within shots. Second, although the database was constructed such that erroneous entries are all removed, noise still exists. The trailers have a wide variety of aspect ratios and resolutions, making direct feature comparison is difficult; additional compression artifacts within some videos significantly distort the extracted scene features. Lastly, some movies were filmed in black and white, making their scene features quite different from those present in the modern day trailers. The presented framework for automatic classification of movie genres using features from scene analysis. The results demonstrated that a temporally-structured feature based on this intermediate level representation of scenes can help improve the classification performance over the use of low-level visual features alone. In the future, the suggestion is to build upon static scene analysis to include scene dynamics, such as action recognition and camera movement estimation, to help achieve higher-level dynamic scene understanding.

Alex Black stock et.al gives the extensive tests they performed both with feature sets and the two classifiers;[8] the work determined that the classifier and feature sets are behaving as they should. However, the similarity of the accuracy across these datasets is still unnerving and indicative of something else that is wrong. The assumption was made that the data has an inherent limitation. Half of the labels applied to the 399 scripts were labeled with at least one of Drama, Thriller, Comedy, Action and Crime. Thus, even with a completely wild guess, the classifier still has a pretty good shot at being right if it guesses some of those labels. The confirmation was done that the classifier makes more-or-less unique guesses for each movie. Thus, the common problem was avoided in classification - believing that our results are accurate when in fact the classifier makes the same guess every time and that guess happens to be correct for the majority of the test data. Unfortunately, this seems to be the extent of movie scripts freely available on the Internet. Despite this limitation, the classifier was still able to perform particularly well in multiway classification. Although the classifier was charged with giving each movie several genre classifications, it was up to 55% accurate at nailing every single genre for a given movie. Considering the relatively small size of training and test sets, this is a very acceptable performance level for a classifier. With n different labels, a multiway classifier is expected to be correct 1/n times.

Anna Bosch
et.al presented a scene classifier that learns categories and their distributions in unlabelled training images using pLSA, and then uses their distribution in test images as a feature vector in a supervised nearest neighbor scheme in [9]. In contrast to previous approaches, significantly superior performance was obtained. The study of the influence of various descriptor parameters have shown that using dense SIFT descriptors with overlapping patches gives the best results for man-made as well as for natural scene classification. Furthermore, discovered topics correspond fairly well with different objects in the images, and topic distributions are consistent between images of the same category.

Jan C. van Gemert,
et.al presents scene category classification by learning the occurrence of proto-concepts in images.[10] The proto-concepts represent by using color invariance and natural image statistics properties. By exploiting similarity responses as opposed to strict selection of a codebook vocabulary, it was able to generalize the proto-concepts to be applicable in general image collections. The demonstrated applicability of the approach in a) learning 50 scene categories from a large collection of news video data; b) a collection of 101 categories of web images; and c) two large collections of photo-stock images, comprising 89 categories, where categories are learned from one and categorized from the other.

Zeeshan Rasheed and Mubarak Shah deviced a method to perform a high level classification of movies into genres using the previews is presented in [11]. In the future the plan is to extend this work to analyze complete movies in order to explore the semantics from the shot level to the scene level. The plan is to utilize the grammar of movie making to present the higher level description of the entire stories.

E-ISSN: 2321–9637
Volume 1, Issue 5, December 2013
International Journal of Research in Advent Technology
Available Online at: http://www.ijrat.org

Jan C. van Gemert,
et.al presents a preliminary examination on a new database of music collected from film scores in four genres (Action, Romance, Horror, and Drama) utilizing timbral and a select set of rhythm features.[12] Initial results suggest that the music from Action genres is the most clearly distinguishable (particularly from Drama and Romance) with Drama and Romance being the least distinct. For the purposes of conducting a preliminary analysis, all of the music tracks within a single film genre were broadly labeled by the movie genre. However, it is clear that such a labeling scheme is likely too broad as several tracks within a specific genre may exhibit characteristics of music from another genre (e.g., an Action sequence of music in a Drama movie or vice versa). The Future work will involved a closer examination of each track to determine the most appropriate groupings of the data and should serve to improve classification accuracy. Additionally, while rhythm features were considered on a small scale, the study was dominated by timbral features.

Matthew R. Boutell,
et.al presented an extensive comparative study of possible approaches to training and testing in multi-label classification.[13] In particular, the contribution follows: Cross-training, a new training strategy to build classifiers. Experimental results show that cross-training is more efficient in using training data and more effective in classifying multi-label data. Then, C-Criterion using threshold selected by MAP principle is effective for multi-label classification. Other classification criteria were proposed as well which may be better suited to different tasks where higher precision is more important than high recall and further, α-Evaluation, their novel generic evaluation metric, provides a way to evaluate multi-label classification results in a wide variety of settings.

M.J. Roach,
et.al showed discriminatory properties of different types of dynamics within video sequences have been presented.[14] The motion measures are content-dependent and are therefore the ultimate filter, in that they are based on the observed content. Thus it can be complementary to any form of static labeling e.g. meta-data. The results show that the dynamic feature extraction methods reports have good discriminatory properties and justify being part of an overall classification system possibly including static and audio features. Using just ~30 second randomly chosen clips, the system has a classification error rate of about 6% applied to the 3 video classes sport, cartoon and news. Genre is an area of inquiry that has been given significant attention in a variety of academic fields, with a particular emphasis found in above mentioned studies. Their exist different issues related to a video genre such as how genres are created, how they can be defined, how they are perceived and identified, how they are disseminated, how they change, how they are interrelated and how we make use of them. Furthermore, genres often encapsulate multiple discrete clusters. Hence, A method for identifying the Action & Horror genre for movie videos is proposed using Artificial Neural Network.

3. PROPOSED METHOD

The proposed method employs Feed Forward Network of Artificial Neural Network for training the system which automatically classifies the given video into the action or horror genre.
The method consists of two phases, namely:

(1) Genre Classifier Construction
(2) Genre Identification

The overall algorithm for the proposed methodology for training and testing phase is presented below.
Step 1: A video sequence of 60 seconds is given as an input to the system, wherein the video frames from every 5 second interval is selected for feature extraction.
Step 2: The object based features such as fire, gun, fist fight, etc are extracted from the above selected video frames.
Step 3: The extracted features are given as input to the FeedForward Network with 2 input nodes and 3 hidden nodes which will create a Knowledge base.
Step 4: The knowledge collected from repeating the above process collectively creates knowledge base which forms the Video genre classifier.
Step 5: The test video sequence of same 60 seconds is given as input for testing the above generated classifier.
Step 6: The object based features such as fire, gun, fist fight, etc are extracted from the above selected video frames and given as input for the classifier.

E-ISSN: 2321–9637
Volume 1, Issue 5, December 2013
International Journal of Research in Advent Technology
Available Online at: http://www.ijrat.org

Step 7: The test video feature are compared by the classifier with the knowledge base and the given video is identified as either Action or Horror video.

3.1. Feature Extraction

The video frames from the given input video sequence at a regular interval of 5 seconds are selected. The object base features such as fire, gun, fist fight, explosion, etc are recognized. These are stored in a vector of size 5 X 5. This forms a knowledge base for that particular video sequence. This process is repeated for all the training sample videos. The feature vectors are given as input for the Artificial Neural Network for the classifier construction. The selected video frames are then divided into various pixel blocks/zones with the size 5 X 5 as shown in **figure 3.1**. The sum of all pixels in the zone is computed which represent one feature/value of an image. Totally 36(30 * 30 size of the image) features are computed in the same manner for an entire image and features are stored in the feature vector ***X***. The feature vector ***X*** is described in equation (1).

Count of pixels representing zone 5 X 5

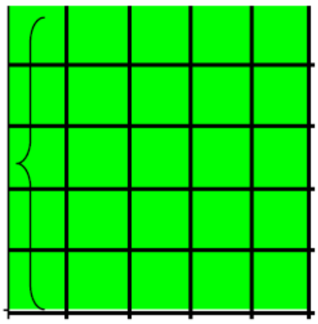

Figure 3.1: 5 X 5 image block

$$X = [f_i] \quad 1 <= i <= 36 \qquad \text{... (1)}$$

Where, ***fi*** is feature vector of ***i***th zone.

These features are used in FeedForward Network to train the system and the resultant will be the Genre classifier.

The sample Knowledge base for each feature is shown below.

Table 3.1 Knowledge base for features

Feature	Feature values
Fire	0 2 0 0 0 0 4 3 6 5 3 0 5 2 4 0 3 2 0 1 0 0 0 5 0 0 0 0 0 5 1 5 5 5 5 0
Gun	2 5 5 5 5 2 3 0 0 0 0 3 0 0 6 1 0 5 2 0 0 1 0 5 3 5 0 0 0 5 0 0 5 5 5 2
Blood	2 5 0 5 1 0 5 0 8 0 0 0 2 0 5 0 0 0 0 2 3 0 2 0 0 1 0 0 6 0 0 0 0 4 2 2
Fist fight	0 0 0 0 0 0 1 6 5 5 5 0 5 0 0 0 5 5 5 0 0 1 7 5 3 5 4 0 0 5 4 4 0 4 5 1
Sparks & explosion	4 5 5 5 5 3 5 0 0 0 0 0 5 3 0 0 0 2 7 0 0 0 0 1 5 0 1 3 0 0 3 5 4 2 5 3
Dark object	0 0 0 0 0 4 1 1 5 5 5 6 5 5 5 2 0 5 2 0 0 5 5 6 0 1 5 0 0 5 3 5 5 5 5 4
Dark background	0 0 0 0 0 0 0 0 8 0 0 0 2 0 5 0 0 0 0 2 3 0 2 0 0 1 0 0 6 0 0 0 0 4 2 2
Distorted image	0 0 0 0 0 0 1 6 5 5 5 0 5 0 0 0 5 5 5 0 0 1 7 5 3 5 4 0 0 5 4 4 0 4 5 1

3.2. The Genre Classifier Construction phase

A set of training Video are taken from a public domain like YouTube. From these videos the video frames are extracted at a regular interval and from these 6 features are extracted. The features extracted from Action videos are Fire, Gun, Blood, Fist fight, Sparks and explosion. The features extracted from Horror videos are Blood, dark object, dark background and distorted images. The Artificial Neural Network (ANN) used for training the system for genre identification. The architecture of the FeedForward Network of ANN is explained below.

3.3. Architecture of FeedForward Network

The type of Architecture implemented in the project is FeedForward Network, this performs Parallel Distributed Computing, Feedforward structure is shown in **Figure 3.1.**[16] This neural network is formed in three layers, called **the input layer, hidden layer, and output layer**. Each layer consists of one or more **nodes**, represented in this diagram by the small circles. The lines between the nodes indicate the flow of information from one node

E-ISSN: 2321–9637
Volume 1, Issue 5, December 2013
International Journal of Research in Advent Technology
Available Online at: http://www.ijrat.org

to the next. In this particular type of neural network, the information flows only from the input to the output (i.e., from left-to-right). The nodes of the input layer are **passive**, meaning they do not modify the data. They receive a single value on their input, and duplicate the value to their multiple outputs. In comparison, the nodes of the hidden and output layer are **active**. This means they modify the data as shown in Figure 3.4. The hidden layer extracts the features of the input and output. The active node performs storing and mapping work.

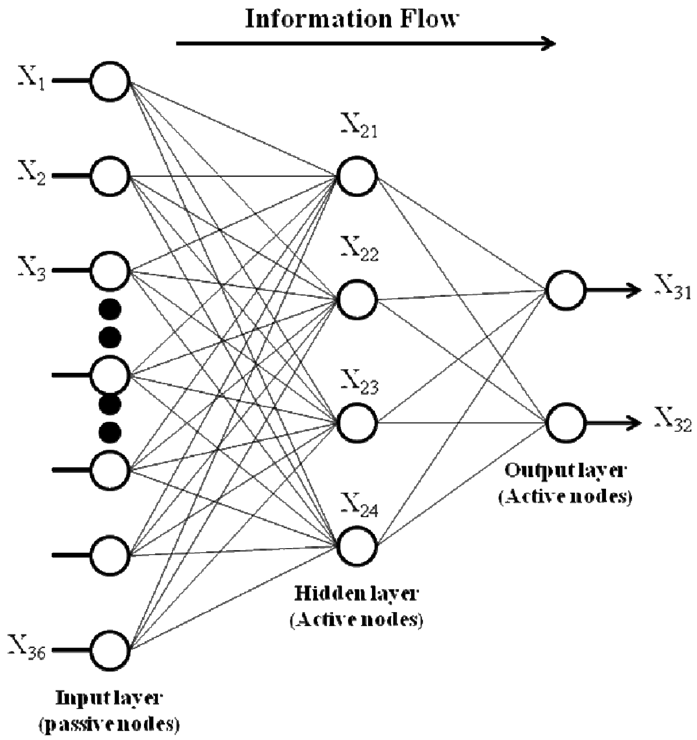

Figure 3.2 FeedForward network

The variables: $X1_1$, $X1_2$. . . . . $X1_{15}$ hold the data to be evaluated (as in **Figure 3.2**). For $X1_1$, $X1_2$. . . . . $X1_{15,}$ example, they may be pixel values from an image, samples from an audio signal, etc. In this case, its video frames from video files. They may also be the output of some other algorithm. The binary string input example: 'a' – [1 1 0 0 0 0 1], etc. is used in the project. Each value from the input layer is duplicated and sent to all of the hidden nodes. This is called a **fully interconnected** structure. As shown is **Figure 3.2**

$$Sum=\Sigma w_i x_i$$

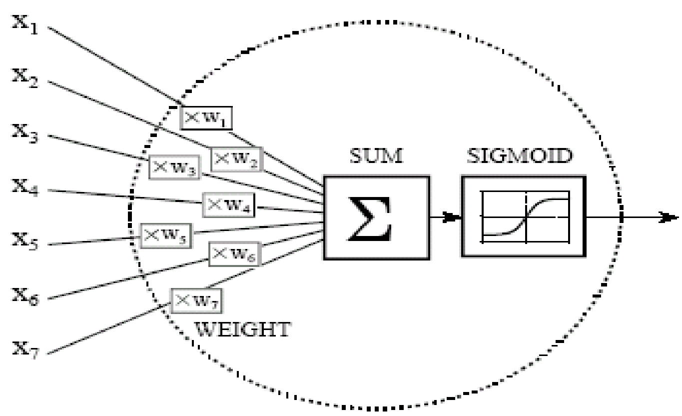

Figure 3.3 Neural Network Active Node

The values entering a hidden node are multiplied by **weights**, a set of predetermined number is stored in the program. The weighted inputs are then added to produce a single number. This is shown in the diagram by the symbol, **Σ.** Before leaving the node, this number is passed through a nonlinear mathematical function called a **Sigmoid**. This is an **"S "** shaped curve that limits the node's output. That is, the input to the sigmoid is a value between - ∞ to + ∞, while its output can only be between 0 and 1.

3.4. Error Back-Propagation Algorithm (EBP)

In the back-propagation algorithm (Rumelhart and McClelland, 1986) is used in layered feed-forward ANNs.[17] This means that the artificial neurons are organized in layers, and send their signals "forward", and then the errors are propagated backwards. The network receives inputs by neurons in the input layer, and the output of the network is given by the neurons on an output layer. There may be one or more intermediate hidden layers. The back-propagation algorithm uses supervised learning, which means that we provide the algorithm with examples of the inputs and outputs we want the network to compute, and then the error (difference between actual and desired results) is calculated. The idea of the back-propagation algorithm is to reduce this error, until the ANN learns the training data the training begins with random weights, and the goal is to adjust them so that the error will be minimal. The trained neural network itself operates in a feed forward manner. The weight adjustment enforced by the learning rules, propagate exactly backward from the output layer through the so-called "hidden layers" towards the input layer. The machine learning process of the neural network is an iterative process to obtain minimum error.

3.5 The Genre Identification

The test video samples are given as input to the system one by one. The frames from the video sequence are selected at regular intervals for feature extraction. The object based features such as fire, gun, fist fight, explosion, blood, dark object and distorted images etc are extracted from the above selected video frames and stored in vectors in the method explained in section 3.1. The Feature vector of the test video looks similar to the vector shown below.

T = [2 5 5 5 5 2 3 0 0 0 0 3 0 0 6 1 0 5 2 0 0 1 0 5 3 5 0 0 0 5 0 0 5 5 5 2]

The above generated vector is compare with the knowledge base shown in Table 3.1 and it is observed that the second row vector matches with the test video vector. Thus the object Gun is identified and the video is said to be of action genre. The classifier uses the knowledge base as shown in Table 3.1 and identifies each test sample videos into either Action or Horror videos. The important thing to be noted is the Artificial Neural Network creates the knowledge base during the process of training, hence, the vectors may differ at each training process and same with the genre identification process.

4 RESULTS AND DISCUSSION

A collection of 100 videos has been made from a public video domain like youtube and metacafe. The collection is a mixture of trailers, movie scenes. The videos are made to be of equal length(60 seconds) and same video format(.avi) by using freely available software, Videocutter. The videos are passed one by one through classifier and the video genre is classified into either Action or Horror. The recall and precision are 0.92 and 46 out of 50 for the action videos respectively. And The recall and precision are 0.88 and 44 out of 50 for the horror videos respectively. Overall the system performance can be showed as a recall and precision of 0.90 and 90 out of 100 videos tested.

CONCLUSION

The work carried out is to develop a method to identify the genre of the video specifically into Action and Horror. The FeedForward Network of Artificial Neural Networks is used to train the system forming the classifier for the system. The system is developed in MATLAB.This classifier is capable of identifying the genre of the video into either Action or Horror genre. The experiments were carried for a set of 100 videos downloaded from YouTube. And the results reveal that the developed system is 90% efficient in identifying the genre of the video. The system is 92% and 88% accurate in identifying Action and Horror genres respectively. The horror videos are usually of dark background which might have made the feature extraction difficult for the horror video which resulted in less accuracy. From the results achieved, the developed system can be used for

E-ISSN: 2321–9637
Volume 1, Issue 5, December 2013
International Journal of Research in Advent Technology
Available Online at: http://www.ijrat.org

genre identification for Action and Horror in the video archives. In future the system can be modified to identify multi-level genres as well.

REFERENCES

[1] Swales, John M, 1990, "Genre Analysis in Multimedia", Cambridge University Press: Cambridge, UK.
[2] Kessler, Brett, Nunberg, Geoff & Schütze, Hinrich,1997, "Automatic detection of genre", Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Meeting of the European Chapter of the Association for Computational Linguistics: pages 32-38.
[3] Karlgren, Jussi, 2000, "Stylistic Experiments for Information Retrieval", Doctoral thesis, Stockholm University.
[4] Muralidhar Talupur, Suman Nath, Hong Yan, 2003, "Classification of Music Genre", School of Music, Alaska, A report for 15781, Manual on music genre.
[5] Cory McKay, Ichiro Fujinaga, 2006, "Automatic genre classification using large high-level musical feature sets", International Conference on Information and Multimedia Technology (ICIMT) Sydney, Australia
[6] Z. Rasheed, Y. Sheikh, and M. Shah, "Use of computable features for classification", IEEE Trans. Circuits and Systems for Video Technology, 15:52-64, 2003.
[7] Howard Zhou, Tucker Hermans Asmita V. Karandikar, James M. Rehg, 2010, "Movie Genre Classification via Scene Categorization", International Conference On Informatics And Creative Multimedia (ICICM), Kuala Lumpur, Malaysia
[8] Alex Black stock, Matt Spitz, 6/4/2008, "Classifying Movie Scripts by Genre with a MEMM Using NLP-Based Features", ICTIMM.
[9] Anna Bosch, Andrew Zisserman and Xavier Munoz, 2007, "Scene Classification via pLSA", Computer Vision and Robotics Group, International Conference on Fuzzy Computation Theory and Applications(FCTA).
[10] Jan C. van Gemert, Jan-Mark Geusebroek, Cor J. Veenman, Cees G.M. Snoek, Arnold W.M. Smeulders, 2007, "Robust Scene Categorization by Learning Image Statistics in Context" Intelligent Systems Lab Amsterdam, 1st Balkan Conference in Informatics, Amsterdam
[11] Zeeshan Rasheed, Mubarak Shah, 2005, "Movie Genre Classification By Exploiting Audio-Visual Features Of Previews", Computer Vision Lab, University of Central Florida, Orlando, International conference on Computer Vision, Canes.
[12] Aida Austin, Elliot Moore II, Udit Gupta, 2002, " Characterization of movie genre based on music score",International Conference on Multimedia Processing, Communication and Information Technology - MPCIT
[13] Matthew R. Boutell, Jiebo Luo, Xipeng Shen, Christopher M. Brown, 2002, "Learning multi-label scene classification", International Journal of Multimedia Information Retrieval.
[14] M.J. Roach, J.S.D. Mason, 2010, "Multi-genre identification using Low level sigmoid values", Journal of Multimedia, Volume 2.
[15] Yuan-Kai Wang and Chih-Yao Chang, "Movie Scene Classification Using Hidden Markov Model", Department of Electronic Engineering, Fu Jen Catholic University,16th IPPR Conference on Computer Vision, Graphics and Image Processing (CVGIP 2003)
[16] Sovierzoski, M.A., Argoud, F.I.M. ; de Azevedo, F.M., 2008, "Evaluation of ANN Classifiers During Supervised Training with ROC Analysis and Cross Validation", BMEI, International Conference on Biomedical Engineering and Informatics, (Volume:1)
[17] Makram Ebeid, S. ; Lab. Electron & Appl. Phys., Limeil Brevannes, France ; Sirat, J.A. ; Viala, J.R.,1989, "A rationalized error back-propagation learning algorithm", IJCNN., International Joint Conference on Neural Networks.