

Multi-label Poster Classification into Genres Using Different Problem Transformation Methods

Miran Pobar and Marina Ivasic-Kos^(✉)

Department of Informatics, University of Rijeka, Rijeka, Croatia
{mpobar, marinai}@uniri.hr

Abstract. Classification of movies into genres from the accompanying promotional materials such as posters is a typical multi-label classification problem. Posters usually highlight a movie scene or characters, and at the same time should inform about the genre or the plot of the movie to attract the potential audience, so our assumption was that the relevant information can be captured in visual features.

We have used three typical methods for transforming the multi-label problem into a number of single-label problems that can be solved with standard classifiers. We have used the binary relevance, random k-labelsets (RAKEL), and classifier chains with Naïve Bayes classifier as a base classifier. We wanted to compare the classification performance using structural features descriptor extracted from poster images, with the performance obtained using the Classeme feature descriptors that are trained on general images datasets. The classification performance of used transformation methods is evaluated on a poster dataset containing 6000 posters classified into 18 and 11 genres.

Keywords: Multi-label classification · RAKEL ensemble method · Binary relevance · Classifier chains · Movie poster · Classemes

1 Introduction

The purpose of a movie poster is to attract the potential audience to go and see the movie and to promote a movie by emphasizing the key information on actors, plot, genre, mood, etc. To achieve the marketing goals, the poster should present in an attractive manner a focused and obvious message that is tailored to the plot of the movie.

Therefore, the design of movie posters usually follows a set of conventions on the layout, colors, fonts and composition. The conventions can be dependent upon genre and target audience. For example, the title can say a lot about a film, but even the title colors can suggest the genre since some colors are considered to be more suitable for certain genres, e.g. dark and red colors for horror, light blue, and pink for romance.

The challenge in the poster design is even greater for many movies that belong to several genres. For example, according to the data from The movie database (TMDB) [1], “Fist fight” only belongs to the Comedy genre, while “Logan” belongs to Action, Drama and Science Fiction genres. In fact, a movie can be classified into a large number of genres, with no constraints.

The issue of classifying movies into genres using the accompanying promotional materials (trailers or posters) is thus a typical multi-label classification problem that is recently attracting increasing attention in the research community. Multi-label classification problems occur in different domains, ranging from text classification (news, web pages, e-mails etc.), scene classification, video annotations, poster classification to the functional genomics classification (gene and protein function). Many different approaches have been developed and their comparison is given in [2].

The early attempts of classifying posters into genres took into account only one genre per movie, in order to simplify the problem into a classical single-label classification. In [3], low-level features are extracted from movie trailers and are used to classify 100 movies into four genres (drama, action, comedy, horror). In [4], a visual vocabulary is formed from low-level features obtained from a collection of temporally ordered static key frames. This visual vocabulary is used for genre classification on 1239 movie trailers. In [5] the same visual features were used as in [3] but only three kinds of the genre were considered: action, drama, and thriller.

Lately, suitable methods to deal with the multi-label problem were applied to poster classification. In [6], algorithm independent transformation methods were used, but at most two labels (genres) per poster were considered. The classification performance was tested for two out of two correct labels and for at least one of two correctly detected labels with Naive Bayes, distance-ranking classifiers and with random k-labelsets (RAKEL) and all methods have performed uniformly well. The experiment was conducted on a set of 1500 posters classified into six genres. The features used in the classification were low-level features based on color and edge combined with the number of faces detected on posters.

In [7], a much larger poster dataset was used containing 6000 movie posters. The authors have stated that the best results were obtained using Naive Bayes (NB) with 740-dimensional feature vector comprising GIST [8] and color-based features. In [9] movies are classified into genres using posters and synopsis. The posters are represented with multiple visual features and the synopsis is represented with vector space model. A test film is classified based on the 'OR' operation on the outputs of support vector machine (SVM) classifiers trained on posters and the synopsis.

Our goal was to determine the movie genres automatically from movie posters using both the global structural GIST features extracted from poster images and the classifier-based image descriptor Classemes [10], that are pre-trained on "abstract categories" of 2659 real world object classes.

To solve the multi-label poster classification problem, we have used the binary relevance, RAKEL method, and classifier chains with Naive Bayes classifier as a base classifier.

Apart from examining how is the classification performance related to different problem transformation methods, we want to compare the effect of different feature sets.

The rest of this article is organized as follows. Section 2 gives a brief overview of typical methods of transforming the multi-label classification problem. Section 3 provides details about the experimental setup, along with benchmark metrics for algorithm evaluation. In Sect. 4, the experimental results are presented and discussed, and in Sect. 5, the contribution of the paper is given.

2 Multi-label Problem Transformation Methods

Our aim was to develop a method that would automatically provide a list of relevant movie genres (labels) for a given, previously unseen poster. The assumption was that a movie could simultaneously belong to several genres, so we had to deal with the problem of multi-label classification.

A multi-label classification for an example e_j can be formally defined as:

$$\exists e_j \in E : \varphi(e_j) = Z_j, Z_j \subseteq C, |Z_j| \geq 2 \text{ and } \varphi : E \rightarrow \mathcal{P}(C).$$

where E is a set of examples, C is a set of class labels and the function φ is a classifier, so that exists at least one example e_j that is mapped into two or more classes from C .

The generality of multi-label classification makes it more difficult for implementation and training than the single-label classification.

Methods that can cope with the multi-label classification problem can be roughly divided into algorithm adaptation (AA) methods and problem transformation (PT) methods [11, 12].

The algorithm adaptation (AA) methods are those that adapt, extend or customize an existing classification algorithm to solve multi-label problems [10]. In recent years, several of such multi-label classifiers methods have been developed, such as ML-kNN derived from kNN, Adaboost.MH and Adaboost.MR, derived from the Adaboost method [12]. Since the PT methods performed significantly better than AA methods on the poster dataset in [7], we will here focus on PT methods.

In the PT multi-label transformation approach, the multi-label classification problem is transformed into more than one single-label classification problem, each focusing on one label of the multi-label case [12]. The aim is to transform the data so that any classification method, designed for single-label classification, can be applied. Typically boosting, kNN, decision trees, neural networks, and SVM [12 and references within] are used in this case. The most common transformation methods are label power set (LP) and binary relevance (BR) methods. Ensemble methods transform the data and fuse the results from multiple ordinary classifiers, thus can be considered as problem transformation methods. Some of the widely known ensemble methods are Random k label sets (RAKEL), classifier chains (CC), random forest based predictive clustering trees (RF-PCT), random decision tree (RDT) etc.

Binary relevance

In the BR method, for each class label, a binary classifier is independently trained. Each classifier then decides whether its corresponding class label should be assigned to an example or not. The overall classification result is the set of all assigned labels.

The original data can be transformed similarly as when a binary classifier is used to deal with multi-class problems. The original dataset of examples and corresponding class labels is split into $|C|$ datasets D_i ; so that each dataset D_i contains all examples belonging to the

C_i class and all the others are treated as negative examples and are labelled as $\neg C_i$. Considering for example the movies “Logan” that belongs to Action, Drama and Science Fiction genres, and “Fist fight” that only belongs to the Comedy genre, the corresponding data sets are: $D_{\text{Action}} = \{(Logan, Action), (First Fight, \neg Action), \dots\}$, $D_{\text{Drama}} = \{(Logan, Drama), (First Fight, \neg Drama), \dots\}$, $D_{\text{SF}} = \{(Logan, Science Fiction), (First Fight, \neg Science Fiction), \dots\}$. $D_{\text{Comedy}} = \{(Logan, \neg Comedy), (First Fight, Comedy), \dots\}$. Then, a single-label classifier is trained for each label $C_i \in C$. The overall classification result consists of labels that are outputs of binary classifiers: $\varphi(e_j) = \bigcup_i C_i$; $\varphi_i''(e_j) = C_i$; $\varphi_i'' : E \rightarrow \{C_i, \neg C_i\}$.

Rakel

RAKEL method randomly breaks up a set of labels into a number of smaller label sets and trains a multi-label classifier for each of them. Each classifier φ_j solves the single-label classification problem for classes $C_i \in C$ that are the subsets of one of m k -sized label sets $R_j, j = 1..m$, sampled randomly from all distinct k -label sets of C .

The label set $Y_i = \{C_1, C_m, \dots, C_r\}$ of each example e_i in the training set is replaced with a new label that corresponds to the intersection of labels in Y_i and in R_j . For example, if a label set R_j is defined as $R_j = \{action, comedy, drama\}$ an example-label set pair in the training set $(e_i, \{action, comedy, crime\})$ will be replaced with the pair $(e_i, action \& comedy)$. For classifying an unknown example e , the results of all classifiers φ_j are joined into final decision using a voting process.

Classifier chains

The classifier chain (CC) involves $|C|$ binary classifiers as in BR [13]. Classifiers are arranged as a chain $\varphi_1'', \dots, \varphi_{|C|}''$ where each classifier handles the binary relevance problem associated with a label $C_i \in C$. The feature set of each classifier φ_i'' in the chain is extended with the label predictions of all previous classifiers $\varphi_1'', \dots, \varphi_{i-1}''$. For instance, an example/single-label pair in the training set (e_i, C_i) will be transformed into the $((e_i, \varphi_1''(e_i), \dots, \varphi_{i-1}''(e_i)), C_i)$. The overall classification result of all classifiers φ_j are joined into final decision using a voting process.

The advantage of the BR method is that can predict new label combinations, unseen in the training data. Also, BR is computationally simple, because for $|C|$ labels, only $|C|$ classifiers must be trained. However, BR ignores the dependence between labels and cannot profit from the information about the mutual occurrence of labels in the training data.

RAKEL and CC methods overcome the assumption of BR approach that all labels are mutually independent. The classifier chain method passes label information between classifiers, allowing CC to take into account label correlations. RAKEL solves the problem of power set label explosion since it takes into account only those combinations that exist in the data set.

3 Experimental Setup

The classification experiments are performed on a poster dataset obtained from the TMDB. The aim was to compare the influence of different problem transformation methods that deal with the multi-label classification problem, therefore in all cases the

same base classifier is used. The methods have been tested with the GIST structural descriptor and with the Classemes classifier-based descriptor.

3.1 Data and Preprocessing Step

We have used poster dataset containing movie posters from 24 years since 1990 [7]. The database includes 6739 movie posters belonging to 18 genres: Action, Adventure, Animation, Comedy, Crime, Disaster, Documentary, Drama, Fantasy, History, Horror, Mystery, Romance, Science Fiction, Suspense, Thriller, War, and Western.

In order to get class balance in the dataset, a set of poster images for each of the selected genres was gathered by taking the top 20 most popular movies in each year in the range. Despite the efforts to collect the same number of posters for every genre, the collected data was not entirely balanced. The distribution of movies for each selected genre is shown in Table 1.

Table 1. Distribution of movies per selected 18 genres.

C_i	Action	Adventure	Animation	Comedy	Crime	Disaster	Documentary	Drama	Fantasy	History	Horror	Mystery	Romance	SF	Suspense	Thriller	War	Western
$ C_i $	1815	1081	734	1524	874	64	924	2610	762	675	888	745	934	908	390	1947	527	273

In addition, as every movie could belong to other genres besides the selected genres, some additional genre labels outside the set of 18 selected genres have appeared in the data (e.g. Noir, Musical and Sport). Hence, the total number of genres was 34, and a problem occurred that some genres did not have enough data to define a model.

We solve the problem of lack of data and of unbalanced data in two ways. In the first, we have considered only the data belonging to the 18 selected genres and discarded the additional genres, further referred to as 18G classification task. In the second, we have merged the genres for which there was insufficient data with similar genres per our judgment, such as Neo-Noir and Crime with Thriller, Music with Documentary. After the merge, the number of genres was reduced to 11, forming our 11G classification task. The way the genres are merged is detailed in Table 2.

Table 2. List of originally obtained, 11 merged and 18 selected genres

Merged genre (11G)	Selected genres (18G)	Additional genres outside the list of selected genres
Action/Adventure	Action	
	Adventure	
	Disaster	
	War	
Animation	Animation	
Comedy	Comedy	
Drama	Drama	Musical
Crime/Thriller	Crime	Film Noir, Neo-noir
	Thriller	
	Mystery	
	Suspense	
Family/Romance	Romance	Holiday, Family, Kids
Horror	Horror	
Western	Western	Road Movie
Science fiction	Science Fiction	
	Fantasy	
Documentary	History	Music
	Documentary	
Misc.		Eastern, Erotic, Sport, Short, Indie, Sports Film, Foreign, TV movie

3.2 Features

We wanted to examine if features extracted from a poster image, such as spatial structure, or higher-level descriptors like classemes, have the discriminative ability in terms of automatic genre classification.

To capture the information on the poster spatial layout, the GIST descriptor [8] was used. This descriptor refers to the dominant spatial structure of the image characterized by properties of its boundaries (e.g., the size, the degree of openness, perspective) and its content (e.g., naturalness, roughness) [8]. The spatial properties are estimated using global features computed as a weighted combination of Gabor-like multi scale-oriented filters. The dimension of GIST descriptor is $n \times n \times k$ where $n \times n$ is the number of samples used for encoding and k is the number of different orientation and scales of image components. GIST descriptor of each genre is implemented with 8×8 encoding samples obtained by projecting the averaged output filter frequency within 8 orientations per 8 scales. The size of GIST feature vector is 500.

Classemes [10] are classifier-based features that are pre-trained on “abstract categories” of real-world object classes. The abstract categories are constrained to be a super-category obtained by grouping a set of object classes that can be distinguished from other sets of categories and could be used as features for training linear models.

For the extraction of the classeme features, the LP-beta classifier [14] is used, which is defined as a linear combination of M nonlinear SVMs, each trained on a

different low-level representation of the image in a 1-vs-rest manner. The low-level representation of the image used for the classemes classifiers contains a set of 15 low-level features concatenated into 22860-dimensional vector capturing different visual cues concerning the color distribution, the spatial layout in the image as GIST, oriented and unoriented HOGs, SIFT and SSIM [15].

The classifiers were pre-trained on the first 150 images retrieved with bing.com image search per each of 2659 classes from the LSCOM ontology [16], so the size of the classemes feature vector is 2659.

The authors stated that classemes provide a general image representation, which can be used to describe and recognize arbitrary categories and novel classes not present in the training set that was used to learn the descriptor [10]. Even though the posters may differ from general images, the object categories of classemes may still provide a high-level representation of layout suitable for use with linear classifiers.

3.3 Classification Methods

For classifying unknown posters into movie genres, we have used the binary relevance, RAKEL method, and classifier chains with Naïve Bayes classifier as base single-label classifier of each method. All three methods are tested on both 11G and 18G classification tasks, with GIST and Classeme features.

In the BR and CC cases, 11 and 18 binary classifiers were trained, one for each genre in the 11G and 18G tasks. To classify an unknown poster, all binary NB classifiers are applied. Each classifier gives a decision whether a movie belongs to its genre or not, and the overall results are obtained by gathering the decisions of all classifiers. In the CC case, we performed 20 runs of training and testing, with the randomly picked order of classifiers in the chain in the each run. We report the average result of all runs.

We have applied the variant of the RAKEL with the Naïve Bayes classifier. RAKEL randomly selects m distinct label sets R_j . The number of label sets m was twice the number of labels in the set C in order to achieve a high level of predictive performance, so for the 11G task it was $m_{11G} = 22$ and for the 18G task it was $m_{18G} = 36$. The size k of label sets R_j is small to avoid problems of LP (in our case $k = 3$).

3.4 Evaluation Measures

Multi-label classification problem requires appropriate evaluation measures, that differs from those used in single-label classification, example-based and label based. Example based measures are calculated on the average differences between the ground truth and the predicted sets of labels in the test set.

We used the following example-based measures. Hamming loss is the symmetric difference of the predicted label set and the true label set, i.e. the fraction of the wrong labels to the total number of labels:

$$HammingLoss = \frac{1}{N} \sum_{i=1}^N \frac{|(Y_i \setminus Z_i) \cup (Z_i \setminus Y_i)|}{|C|} \quad (1)$$

In the best case, the value of Hamming loss is zero.

To define the evaluation measures, we assume that an example $e_j \in E, j = 1..N$ has the set of ground truth labels $Y_j = \{C_l, C_m, \dots, C_r\}, Y_j \subseteq C$ where E is a set of feature vectors, and C is a set of all labels. The set of labels for the example e_j that are predicted by a classifier is Z_j .

Accuracy is defined as the average ratio of correctly predicted labels and all predicted and true labels for each example:

$$Accuracy = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|}. \quad (2)$$

Precision is defined as the average ratio of correctly predicted labels and all predicted label for each example.

$$Precision = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i \cap Z_i|}{|Z_i|}. \quad (3)$$

Recall is defined as the average ratio of correctly predicted labels classifier and the ground truth labels for each example.

$$Recall = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i \cap Z_i|}{|Y_i|}. \quad (4)$$

F-Measure can be interpreted as the harmonic mean of precision and recall.

$$F1 = \frac{1}{N} \sum_{i=1}^N \frac{2|Y_i \cap Z_i|}{|Z_i| + |Y_i|} \quad (5)$$

These measures reach their best value at 1 and the worst at 0.

Label based recall, precision, and F1 measures are calculated for each label separately similarly as in binary classification. For average based results, two kinds of averaging operations called macro-averaging and micro-averaging can be used.

A micro-average evaluation measure gives equal weight to each example and is an average over all the example and label pairs. A macro-average evaluation measure gives equal weight to each label, regardless of its frequency, so is a per-label average [17].

4 Results and Discussion

We have tested the classification performance on 11 genres (11G) and 18 genres (18G) classification tasks. All experiments were run using 5-fold cross-validation.

In the following tables, example-based and label-based results are presented. The micro-averaged label-based measures were very similar to example-based measures, so here the macro-averaged measures are shown.

For the macro-averaged evaluation measures, all tested transformation methods have performed similarly on both tasks (Table 3). All transformation methods have achieved better results for the 11G task, which was expected because it is a simpler task. Classeme features have generally proved to be better poster data representatives, although the obtained results are only slightly better than those achieved with 5 times shorter GIST vector.

Table 3. Label-based macro-averaged evaluation results for 11G and 18G tasks.

	Measure	Classifier					
		BR+NB	CC+NB	RAKEL+NB	BR+NB	CC+NB	RAKEL+NB
		Gist features			Classeme features		
11G	Precision	0.29	0.30	0.30	0.30	0.31	0.30
	Recall	0.61	0.51	0.55	0.62	0.60	0.61
	F1 score	0.38	0.36	0.38	0.39	0.39	0.39
18G	Precision	0.21	0.21	0.21	0.21	0.22	0.22
	Recall	0.60	0.38	0.49	0.62	0.55	0.58
	F1 score	0.30	0.25	0.30	0.30	0.30	0.30

CC+NB has the same or worse results concerning label-based evaluation measures than BR+NB, Fig. 1. Passing the genre information between the classifiers in CC+NB chain has not been proved useful in case of label-based evaluations since the scores have not been improved in comparison with BR+NB. This may be explained by error propagation between classifiers in the chain, where a misclassification in one step may impact the classification in all further steps. BR+NB has in all cases achieved the best recall, while precision was very similar for all methods.

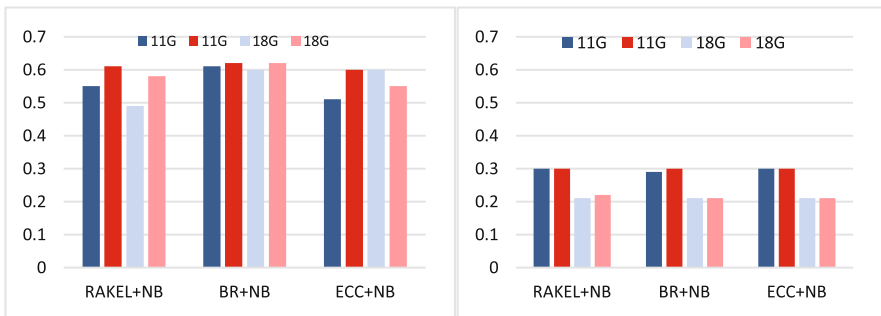


Fig. 1. Macro-averaged recall (left) and precision (right) scores.

Detailed example-based classification results are presented in Table 4 for both 11G and 18G tasks and for GIST and Classeme feature sets. All classifiers except RAKEL have obtained better F1 scores on the label-based than on example-based evaluation measures. RAKEL has achieved slightly better F1 scores for example-based measures in all cases. Simultaneously, RAKEL has significantly better results than other methods for all example-based measures except for Hamming loss.

Table 4. Example based evaluation results for 11 genres (11G) and 18 genres (18G) tasks.

	Measure	Classifier					
		BR+NB	CC+NB	RAKEL+NB	BR+NB	CC+NB	RAKEL+NB
		Gist features			Classeme features		
11G	Hamming loss	0.38	0.35	0.37	0.39	0.37	0.37
	Accuracy	0.21	0.21	0.27	0.20	0.20	0.29
	Precision	0.25	0.26	0.32	0.24	0.25	0.33
	Recall	0.47	0.41	0.56	0.46	0.44	0.63
	F1 score	0.31	0.31	0.39	0.30	0.30	0.41
18G	Hamming loss	0.38	0.29	0.32	0.39	0.34	0.35
	Accuracy	0.15	0.15	0.21	0.14	0.14	0.21
	Precision	0.17	0.18	0.24	0.15	0.16	0.23
	Recall	0.44	0.33	0.54	0.46	0.40	0.59
	F1 score	0.23	0.22	0.32	0.30	0.22	0.32

The highest example-based accuracy, precision, and recall are achieved also with RAKEL, Figs. 2 and 3. Considering that all methods performed similarly according to label-based measures and that RAKEL achieved significantly better example-based measures, we can conclude that RAKEL transformation method performs the best for both our tasks. It seems that information on the mutual occurrence of genres in the training data was significant in this case, so results using RAKEL were improved in comparison to BR. However, the results with CC are not improved, possibly due to propagation of errors through the chain.

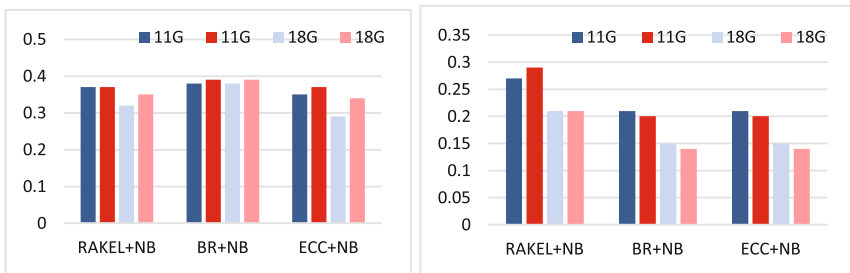


Fig. 2. Hamming loss (left) and accuracy (right) scores.

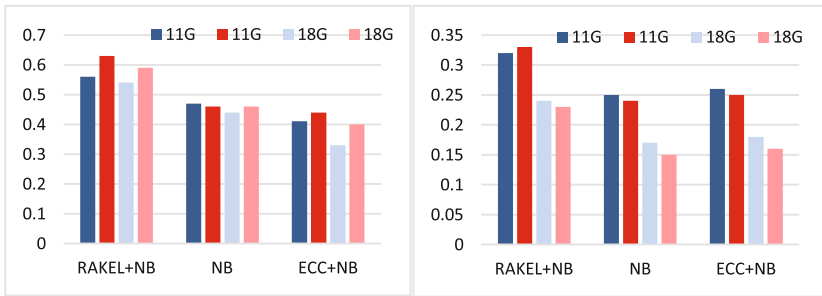


Fig. 3. Example-based recall (left) and precision (right) scores.

In comparison with published results in [7] for the same classification tasks, we have shown that introduction of RAKEL+NB and classeme features contribute the improvement of the classification performance regarding F1 for more than 20%.

5 Conclusion

In this paper, we compared the impact of different problem transformation methods on multi-label classification performance of poster images. The comparison was based on standard label-based and example-based multi-label evaluation measures. The considered problem transformation methods were binary relevance, RAKEL, and classifier chains. Naïve Bayes classifier was used as the base classifier in all cases. We have used a poster dataset of 6739 movie posters classified into 11 or 18 genres. The classification task was multi-label because each poster can be classified into more than one genre. The posters were represented with GIST and classeme features. The GIST feature vector is 5 times shorter than Classemes, but the achieved classification results are similar for both features.

For the macro-averaged evaluation measures, all tested transformation methods have performed similarly on both tasks and achieved better results for the 11G task. BR and CC have achieved better F1 scores on the label-based than on example-based evaluation measures, while the opposite was true for RAKEL. However, RAKEL has achieved significantly better example-based measures than other methods.

Considering that all methods performed similarly according to label-based measures and that RAKEL achieved significantly better example-based measures, we can conclude that RAKEL transformation method performs the best for both our tasks. It seems that information on the mutual occurrence of genres in the training data was significant in this case, so results using RAKEL were improved in comparison to BR. However, the results achieved with CC are not improved, possibly due to the propagation of errors through the chain.

In the future work, we plan to test a deep learning approach with a much larger poster dataset and different features.

Acknowledgment. This research was fully supported by Croatian Science Foundation under the project Automatic recognition of actions and activities in multimedia content from the sports domain (RAASS).

References

1. The movie database, March 2014. <http://www.themoviedb.org/>
2. Madjarov, G., Kocev, D., Gjorgjevikj, D., Džeroski, S.: An extensive experimental comparison of methods for multi-label learning. *Pattern Recogn.* **45**(9), 3084–3104 (2012)
3. Rasheed, Z., Sheikh, Y., Shah, M.: On the use of computable features for film classification. *IEEE Trans. Circ. Syst. Video Technol.* **15**(1), 52–64 (2005)
4. Zhou, H., Hermans, T., Karandikar, A.V., Rehg, J.M.: Movie genre classification via scene categorization. In: *Proceedings of the International Conference on Multimedia*, pp. 747–750. ACM (2010)
5. Huang, H.-Y., Shih, W.-S., Hsu, W.-H.: A film classifier based on low-level visual features. In: *9th IEEE Workshop on Multimedia Signal Processing, MMSP 2007*, pp. 465–468 (2007)
6. Ivašić-Kos, M., Pobar, M., Mikec, L.: Movie posters classification into genres based on low-level features. In: *Proceedings of International Conference MIPRO*, pp. 1448–1453, Opatija (2014)
7. Ivasic-Kos, M., Pobar, M., Ipsic, I.: Automatic movie posters classification into genres. In: Bogdanova, A.M., Gjorgjevikj, D. (eds.) *ICT Innovations 2014. AISC*, vol. 311, pp. 319–328. Springer, Cham (2015). doi:[10.1007/978-3-319-09879-1_32](https://doi.org/10.1007/978-3-319-09879-1_32)
8. Fu, Z., Li, B., Li, J., Wei, S.: Fast film genres classification combining poster and synopsis. In: He, X., Gao, X., Zhang, Y., Zhou, Z.-H., Liu, Z.-Y., Fu, B., Hu, F., Zhang, Z. (eds.) *ISCI 2015. LNCS*, vol. 9242, pp. 72–81. Springer, Cham (2015). doi:[10.1007/978-3-319-23989-7_8](https://doi.org/10.1007/978-3-319-23989-7_8)
9. Torresani, L., Szummer, M., Fitzgibbon, A.: Efficient object category recognition using classemes. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010. LNCS*, vol. 6311, pp. 776–789. Springer, Heidelberg (2010). doi:[10.1007/978-3-642-15549-9_56](https://doi.org/10.1007/978-3-642-15549-9_56)
10. Tsoumakas, G., Vlahavas, I.: Random k-label sets: an ensemble method for multi-label classification. In: *Machine Learning: ECML 2007*, pp. 406–417 (2007). Springer
11. Tsoumakas, G., Katakis, I.: Multi-label classification: an overview. *Int. J. Data Warehouse. Min.* **3**(3), 1–13 (2007)
12. Read, J., Pfahringer, B., Holmes, G., Frank, E.: Classifier chains for multi-label classification. *Mach. Learn. J.* **85**(3), 254–269 (2011). Springer
13. Oliva, A., Torralba, A.: Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int. J. Comput. Vis.* **42**(3), 145–175 (2001)
14. Gehler, P., Nowozin, S.: On feature combination for multiclass object classification. In: *2009 IEEE 12th International Conference on Computer Vision*, pp. 221–228. IEEE, September 2009
15. Shechtman, E., Irani, M.: Matching local self-similarities across images and videos. In: *CVPR 2007*, pp. 1–8. IEEE, June 2007
16. Naphade, M., Smith, J.R., Tesic, J., Chang, S.F., Hsu, W., Kennedy, L., Curtis, J.: Large-scale concept ontology for multimedia. *IEEE Multimed.* **13**(3), 86–91 (2006)
17. Yang, Y.: An evaluation of statistical approaches to text categorization. *Inf. Retrieval* **1**(1–2), 69–90 (1999)