MULTI-TASK FEATURE SELECTION FOR ADVANCING PERFORMANCE OF IMAGE SEGMENTATION

HAN LIU[1], HUIHUANG ZHAO[2]

[1]School of Computer Science and Informatics, Cardiff University
Cardiff CF24 3AA, United Kingdom
[2*]College of Computer Science and technology, Hengyang Normal University
Hengyang 421008, China
E-MAIL: liuh48@cardiff.ac.uk, happyday.huihuang@gmail.com(*corresponding author)

Abstract:

Image segmentation is a popular application area of machine learning. In this context, each target region drawn from an image is defined as a class towards recognition of instances that belong to this region (class). In order to train classifiers that recognize the target region to which an instance belongs, it is important to extract and select features relevant to the region. In traditional machine learning, all features extracted from different regions are simply used together to form a single feature set for training classifiers, and feature selection is usually designed to evaluate the capability of each feature or feature subset in discriminating one class from other classes. However, it is possible that some features are only relevant to one class but irrelevant to all the other classes. From this point of view, it is necessary to undertake feature selection for each specific class, i.e, a relevant feature subset is selected for each specific class. In this paper, we propose the so-called multi-task feature selection approach for identifying features relevant to each target region towards effective image segmentation. This way of feature selection requires to transform a multi-class classification task into n binary classification tasks, where n is the number of classes. In particular, the Prism algorithm is used to produce a set of rules for class specific feature selection and the K nearest neighbour algorithm is used for training a classifier on a feature subset selected for each class. The experimental results show that the multi-task feature selection approach leads to an significant improvement of classification performance comparing with traditional feature selection approaches.

Keywords:

Machine learning; Multi-task learning; Multi-task feature selection; Image segmentation;

1. Introduction

Image segmentation has a wide array of applications ranging from object detection, image classification scene understanding and so on [1]. Deep Convolutional Neural Networks (DCNNs) have pushed the performance of computer vision systems to soaring heights on a broad array of high-level problems, including image segmentation and object detection [2]. In the setting of machine learning, image segmentation is partially defined as a multi-class classification problem. In this context, an image is divided into several target regions and each of them is considered as a class towards recognizing instances that lie in this region. In order to train classifiers that can effectively identify the region to which an instance belongs, it is crucial to extract and select relevant features.

In traditional machine learning, feature selection is typically done in a single-task manner, which means that all features extracted from different target regions of images are simply used together to form a feature set and each feature or feature subset is evaluated in terms of their capability of discriminating between different classes. However, as argued in [3], it is highly possible that some features are only relevant to one class but are irrelevant to all the other classes. From this point of view, single-task feature selection is likely to result in high dimensionality and sparsity of a feature set, since some selected features may be relevant in general but not for specific classes, i.e. inclusion of such features would increase the dimensionality and cause missing values for instances that belong to the classes for which the features are irrelevant.

In this paper, we propose a multi-task feature selection approach to identify a set of relevant features for each specific class. In particular, we employ the Prism algorithm [4] to learn a set of rules for each class, in order to identify the relevance of features or feature subsets for each class. Furthermore, a KNN classifier is trained on each feature subset selected for a specific class to investigate the impact of multi-task feature selection on the classification performance for each class, in comparison with the case of single-task feature selection, i.e. a KNN classifier is trained on a single feature subset.

The rest of this paper is organized as follows: Section 2 provides an overview of image segmentation and a review of feature selection approaches; In Section 3, we present our proposed approach multi-task feature selection and justify its significance in image segmentation. In Section 4, we report an experimental study using a UCI data set on image segmentation and the results are discussed critically and comparatively. Section 5 provides a summary of the contributions of this paper and some further directions are suggested towards advancing this research area in the future.

2. Relate Work

In this section, we review some related work on image segmentation. Also, existing approaches of feature selection are reviewed in order to identify the advantages and disadvantages of these approaches.

2.1. Overview of image segmentation

Over the past few years the breakthroughs of deep learning in image classication have been quickly transferred to the semantic segmentation task, segmentation techniques can be generally categorized into two frameworks, non-deep learning methods [5, 6, 7, 8, 9] and deep learning methods.

For the first framework, most of the successful semantic segmentation systems developed in the previous decade relied on hand-crafted features combined with at classiers. For example, [10] proposed an learned model, which is used for automatic visual understanding and semantic segmentation of photographs based on Boosting. [11] proposed semantic text on forests, which are ensembles of decision trees that act directly on image pixels, and therefore do not need the expensive computation of filter-bank responses or local descriptors. [12] conducted a study with color image segmentation using pixel wise support vector machine classification. The performance of these methods has always been compromised by the limited expressive power of the features.

Recently, various methods [13, 14] based on Fully Convolutional Networks (FCNs) [15] demonstrate astonishing results on several semantic segmentation benchmarks. [1] proposed a DeepLab system which had three challenges in the application of DCNNs to semantic image segmentation, including reduction of feature resolution, use of object multiple scales and reduction of localization accuracy due to DCNN invariance.

Multiscale Combinatorial Grouping (MCG) was proposed in [2] for image segmentation. During the approach, there was a high-performance hierarchical segmenter that makes effective use of multiscale information, and a grouping strategy that combines the multiscale regions into highly-accurate objects.

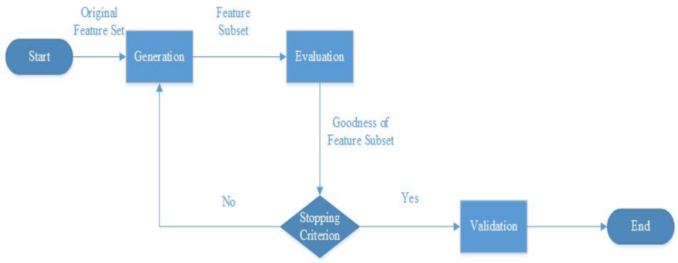

FIGURE 1. Feature Section Process [3, 17]

2.2. Review of feature selection approaches

As introduced in [16], the feature selection process involves four main steps: generation, evaluation, stopping criterion and validation, as illustrated in Fig. 1. In particular, the generation step is aimed at generating a candidate feature subset. In the evaluation stage, a heuristic function is used to evaluate the subset of features selected in the generation stage. A stopping criteria is then used to decide whether it is necessary to stop the feature selection process. If yes, the selected feature subset is validated in the last stage. Otherwise, the feature selection process needs to be repeated through generation and evaluation of a candidate feature subset.

In general, feature selection techniques belong to two approaches, namely, filter and wrapper. The main difference between the approaches is in terms of the way of feature evaluation. The filter approach employs heuristics to rank the features according to their importance, whereas the wrapper approach employs a learning algorithm to train classifiers on different subsets of features and then check the performance of these classifiers for evaluating the corresponding feature subsets.

In terms of the performance of feature selection, the filter approach is aimed at independent evaluation of features regardless of the fitness of the employed learning algorithm. In other words, a set of features is evaluated and the relevant ones are selected without considering that the selected feature subset is suitable or not for the chosen learning algorithm to train a model. According to experimental results reported in [16], the filter approach leads to a low level of time complexity. However, when the selected feature subset is used for a selected algorithm to train a classifier, the classification accuracy may be low due to the case that the selected feature subset is not suitable for training of classifiers by using this algorithm [18].

In contrast, the wrapper approach is aimed at evaluation of features through measuring the accuracy of the classifiers trained on different subsets of features. In other words, a number (n) of different feature subsets are provided and a learning algorithm is used to train n classifiers on these feature subsets. The feature subset, which leads to the best performing classifier, is selected. According to the experimental results reported in [16], the wrapper approach leads to very high accuracy but the time complexity is very high, which mainly results from the case that all the possible combinations of features forming different feature subsets need to be examined.

3. Multi-task Feature Selection

As mentioned in Section 1, multi-task feature selection is aimed at selecting a feature subset for each specific class. In this section, we present our proposed approach to illustrate how to evaluate the relevance of a feature or feature subset for a class. Also, the significance of this approach for image segmentation is justified.

3.1. Procedure

Our proposed approach is essentially wrapper based, since it needs to employ a learning algorithm to train a classifier towards feature evaluation and selection. In particular, we propose to use the Prism algorithm to learn a set of rules for each class. The procedure of this algorithm is illustrated in [4] and shown as below:

For each class c used in turn for original training set T:
Step 1: Calculate the probability $P(class = c|A_i = v_{ij})$ given each attribute-value pair $A_i = v_{ij}$;
Step 2: Select the attribute value pair $A_m = v_{mn}$ that provides the max probability and create a subset T' of instances that meet $A_m = v_{mn}$
Step 3: repeat Steps 1 and 2 for the subset T' until all instances in T' belong to the same class
Step 4: repeat Steps 1-3 for original training set T

In the above context, the learning of each single rule is essentially to select one or more feature-value pairs to be added as rules terms to form the antecedent (the left hand side) of the rule. Each set of rules learned for a specific class would have the same rule consequent (the right hand side of a rule). Therefore, each feature, which is selected alongside one of its possible values as a term of at least one rule for a class, would be judged as relevant for this class. For example, there are two rules that are learned from a data set that contains four features a, b, c and d and two classes 0 and 1, and the two rules are represented as follows:

- if a=1 and b=1 then class=0;
- if c=1 and d=1 then class=1;

The above example would indicate that the two features a and b are relevant for class 0 and the other two features c and d are relevant for class 1.

Since the feature selection is undertaken in a multi-task manner, it is necessary to train classifiers in the same manner, i.e. multi-task learning of classifiers. In particular, if a data set contains more than two classes, we need to transform the multi-class classification task into n binary classification tasks, where n is the number of classes. For example, if a data set contains three classes A, B and C, we need to recreate three data sets by selecting one class as the positive class and renaming the other class labels to reflect the negative class, i.e. the three data sets would show the pair of class labels (A and $\neg A$), (B and $\neg B$) and (C and $\neg C$), respectively.

Following the transformation of the classification task and recreation of the data sets, a classifier is trained on each data set for a specific class by using a learning algorithm. In this paper, we would employ the K nearest neighbours (KNN) algorithm due to its nature of instance based learning and its popular application in image classification. In the classification state, each classifier would provide a single output. If two or more classifiers output their corresponding positive classes, e.g. the first classifier outputs A rather than $\neg A$ and the second classifier outputs B rather than $\neg B$, then the one with higher confidence would be selected as the final classification. Measure of confidence can be achieved by evaluating the overall accuracy of a classifier, the prevision/recall/F-measure for a class or the prediction confidence on an instance.

3.2. Justification

As described in Section 2.2, feature selection can be achieved through two approaches, namely, filter and wrapper.

Both approaches are actually designed in a single-task manner, since they generally evaluate the capability of features in discriminating one class from the other classes, without considering the case that a feature may be only relevant to one or some but not all of the classes.

In general, a feature may be irrelevant for most of the classes but is still relevant to one or a few classes. For example, in the context of image segmentation, the features are extracted from different target regions of images. When each target region is defined as a class, it would be very normal that features extracted from one region are not relevant to other regions.

In the setting of single-task feature selection, the above kind of features is likely to judged as irrelevant and thus to be removed prior to training of classifiers, leading to the case of incorrect classification of instances that belong to the class for which the removed feature is relevant. However, multi-task feature selection is essentially designed to evaluate the relevance of features for each specific class, so the above case can be avoided and instances of each class can be identified more effectively given more relevant features for discriminating the class from other classes.

On the other hand, a feature may be relevant to the majority of the classes but it is irrelevant to one or a few classes. In the setting of single-task feature selection, this kind of features is likely to be kept for training classifiers, which would increase the sparsity of a data set, i.e. missing values would be present on instances of the class for which some features are irrelevant. However, multi-task feature selection can achieve to keep only features that are relevant for a specific class, i.e. all the selected features are relevant only for the target class (corresponding to a target region), so the dimensionality and sparsity would be much reduced in comparison with single-task feature selection.

4. Experimental setup and results

In this section, we report an experimental study using the image segmentation data set retrieved from the UCI repository [19]. This data set contains 19 features and 2310 instances. The instances are normally distributed over 7 classes, namely, brickface, sky, foliage, cement, window, path and grass, i.e. each of the 7 classes contains 330 instances. The 7 classes essentially represent 7 target regions of images and the 2310 instances are randomly drawn from 7 outdoor images [19].

The experimental study consists of three parts. The first two parts are aimed at comparing the KNN algorithm with several popular learning algorithms, which include C4.5 (the most popular algorithm of decision tree learning), Naive Bayes (NB) and multi-layer perceptron (MLP), in order to show how well KNN can achieve to effectively recognize instances of each target region, in comparison with other learning algorithms. In particular, the first part of the experimental study does not involve feature selection, all the algorithms are used to train classifiers on the original features. In contrast, the second part involves filter based single-task feature selection by using the Correlation-based Feature Subset Selection (CFSS) method [20] in order to show how single task feature selection can impact on the performance of image segmentation when different learning algorithms are employed. The third part of the experimental study is designed to show the effectiveness of our proposed multi-task feature selection on boosting the performance of image segmentation, while KNN is used to train classifiers.

TABLE 1. Feature selection results

Feature	b	s	f	c	w	p	g
region-centroid-col	1	0	1	1	1	0	0
region-centroid-row	1	0	1	1	1	1	1
region-pixel-count	0	0	1	0	1	0	0
short-line-density-5	0	0	1	1	1	0	0
short-line-density-2	0	0	1	1	1	0	0
vedge-mean	1	0	1	1	1	0	0
vegde-sd	1	0	1	1	1	0	0
hedge-mean	0	0	1	1	1	0	0
hedge-sd	1	0	1	1	1	1	0
intensity-mean	1	1	1	1	1	1	1
rawred-mean	1	0	1	1	1	0	0
rawblue-mean	0	0	1	1	1	0	0
rawgreen-mean	1	0	1	1	1	1	0
exred-mean	1	0	1	1	1	0	0
exblue-mean	0	0	1	1	1	0	0
exgreen-mean	1	0	1	1	1	0	0
value-mean	0	0	1	0	1	0	0
saturation-mean	0	0	1	1	1	0	0
hue-mean	1	0	1	1	1	1	1

All the experiments are conducted by using 10-fold cross validation. In terms of parameters setting, the K value of the KNN algorithm is set to 3 and the MLP classifiers are trained through 100 iterations with 2 hidden layers and 10 units in each layer. In addition, in the setting of multi-task feature selection, 7 new data sets are recreated by renaming the class labels of the original data set, i.e. the 7 new data sets contain the 7 pairs of class labels ('brickface' and 'not brickface'), ('sky' and 'not sky'), ('foliage' and 'not foliage'), ('cement' and 'not cement'), ('window' and 'not window'), ('path' and 'not path') and ('grass' and 'not grass'), respectively. As introduced in

TABLE 2. Classification accuracy on segment data

Features	MLP	C4.5	NB	KNN
original	0.831	0.959	0.763	0.963
reduced	0.788	0.965	0.816	0.929

TABLE 3. F-measure for each class

Class	All	STFS	MTFS
brickface	0.976	0.900	0.979
sky	1	1	1
foliage	0.920	0.864	0.916
cement	0.944	0.925	0.934
window	0.910	0.837	0.889
path	0.992	0.998	0.995
grass	0.998	0.979	0.995
overall	0.963	0.929	0.958

Section 3, the above setting is aimed to transform a multi-class classification task into 7 binary classification tasks, such that feature evaluation can be done in a class specific way.

According to the results reported in [3], there are 247 rules learned from the image segmentation data set by using the Prism algorithm. For each class, different features are used to form the antecedent of these relevant rules, as shown in Table 1. In particular, '1' indicates that the the corresponding feature is selected alongside one of its possible values for forming the antecedent of at least one rule for the target class, whereas '0' indicates that the feature is not selected and is thus considered irrelevant for the class. Also, each class is expressed by using the initial of its label, e.g. 'b' represents the 'brickface' class.

The results for the first two parts of experimental studies are shown in Table 2, which indicate that the KNN algorithm outperforms all the other ones when the original features are used for training classifiers. However, after the dimensionality is reduced to 8 through using the CFSS method in a single-task manner, the performance of MLP and KNN drops, whereas the one of C4.5 and NB is improved. The results support the argument made in Section 2.2 that the filter approach does not take into account the fitness of the selected features for a specific learning algorithm. Moreover, the negative impact of the feature selection approach on the KNN performance could indicate that some important features have potentially been removed, i.e. the way of single task feature selection may results in removal of some features that are relevant for one or some classes but not for the majority of the classes. Given that KNN essentially needs to examine all the given features for training a classifier, the above way of feature selection is likely to affect the classification performance.

In order to overcome the limitations of single-task feature selection, the third part of the experimental study is thus conducted and the results are shown in Table 3. In particular, the second column shows the performance on the original features, the last two columns show the performance when single-task and multi-task feature selection are adopted, respectively.

The results indicate that the adoption multi-task feature selection leads to advances in performance (F-measure) for 5 out of 7 classes, in comparison with the case of single-task feature selection. Comparing with the case of using original features, although the results do not show an improvement of the overall F-measure, it still achieves to improve the F-measure for 2 classes. Moreover, the dimensionality is significantly reduced, which contributes to effective reduction of the computational complexity without loss of accuracy, especially given the case that KNN achieves perfect classification for the 'sky' class on the original features but the multi-task feature selection approach only provides one feature (intensity-mean) for KNN to train a classifier leading to perfect classification for this class.

5. Conclusions

In this paper, we proposed a multi-task feature selection approach through class-specific feature evaluation. In particular, we employed the Prism algorithm to learn a set of rules for each class, in order to identify which features are relevant for which classes. Furthermore, the KNN algorithm was used for training a classifier on each feature subset selected for a specific class. We compared the performance of multi-task feature selection with the one of filter based single-task feature selection, while KNN was used for training classifiers. The experimental results show that the performance for recognizing instances of most classes is significantly improved and thus indicate the effectiveness of multi-task feature selection on image segmentation. In addition, comparing with the case of using original features, the adoption of multi-task feature selection contributes to significant reduction of computational complexity without loss of accuracy. In future, we will investigate more broadly the impact of multi-task feature selection on image classification. We will also investigate the use of optimization techniques towards selection of an optimal set of features for each class.

Acknowledgements

This work was supported by National Natural Science Foundation of China (61503128), Science and Technology Plan

Project of Hunan Province (2016TP102), Scientific Research Fund of Hunan Provincial Education Department (14B025, 16C0311), and Hunan Provincial Natural Science Foundation(2017JJ4001).

References

[1] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[2] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[3] H. Liu, M. Cocea, and W. Ding, "Multi-task learning for intelligent data processing in granular computing context," *Granular Computing*, vol. 3, no. 3, In press.

[4] H. Liu and M. Cocea, "Induction of classification rules by gini-index based rule generation," *Information Sciences*, vol. 436-437, pp. 227–246, 2018.

[5] L. Hertz and R. W. Schafer, "Multilevel thresholding using edge matching," *Computer Vision, Graphics, and Image Processing*, vol. 44, no. 3, pp. 279–295, 1988.

[6] X. Xu, Z. Yang, and Y. Wang, "A method based on rank-ordered filter to detect edges in cellular image," *Pattern Recognition Letters*, vol. 30, no. 6, pp. 634–640, 2009.

[7] H. Cai, Z. Yang, X. Cao, W. Xia, and X. Xu, "A new iterative triclass thresholding technique in image segmentation," *IEEE transactions on image processing*, vol. 23, no. 3, pp. 1038–1046, 2014.

[8] M. J. Carlotto, "Histogram analysis using a scale-space approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 1, pp. 121–129, 1987.

[9] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern recognition*, vol. 19, no. 1, pp. 41–47, 1986.

[10] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context," *International Journal of Computer Vision*, vol. 81, no. 1, pp. 2–23, 2009.

[11] J. Shotton, M. Johnson, and R. Cipolla, "Semantic texton forests for image categorization and segmentation," in *Computer vision and pattern recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[12] X.-Y. Wang, T. Wang, and J. Bu, "Color image segmentation using pixel wise support vector machine classification," *Pattern Recognition*, vol. 44, no. 4, pp. 777–787, 2011.

[13] J. Dai, K. He, and J. Sun, "Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1635–1643.

[14] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr, "Conditional random fields as recurrent neural networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1529–1537.

[15] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[16] M. Dash and H. Liu, "Feature selection for classification," *Intelligent Data Analysis*, vol. 1, pp. 131–156, 1997.

[17] H. Liu, M. Cocea, and W. Ding, "Decision tree learning based feature evaluation and selection for image classification," in *International Conference on Machine Learning and Cybernetics*, Ningbo, China, 9-12 July 2017, pp. 569–574.

[18] I. Guyon, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[19] M. Lichman, "UCI machine learning repository, http://archive.ics.uci.edu/ml," 2013.

[20] M. A. Hall and L. A. Smith, "Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper," in *Proceedings of the Twelfth International Florida Artificial Intelligence Research Society Conference*, Orlando, Florida, 1-5 May 1999, pp. 235–239.