Research paper

A transfer learning method for automatic identification of sandstone microscopic images

Na Li[a,b], Huizhen Hao[a,b,c], Qing Gu[a,b,*], Danru Wang[a,b], Xiumian Hu[d]

[a] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
[b] Department of Computer Science and Technology, Nanjing University, Nanjing, China
[c] School of Communication Engineering, Nanjing Institute of Technology, Nanjing, China
[d] School of Earth Sciences and Engineering, Nanjing University, Nanjing, China

ARTICLE INFO

ABSTRACT

Classification of sandstone microscopic images is an essential task in geology, and the classical method is either subjective or time-consuming. Computer aided automatic classification has been proved useful, but it seldom considers the situation where sandstone images are collected from separated regions. In this paper, we provide a method called Festra, which uses transfer learning to handle the problem of interregional sandstone microscopic image classification. The method contains two parts: one is feature selection, which aims to screen out features having great difference between the regions, the other is instance transfer using an enhanced TrAdaBoost, whose object is to mitigate the difference among thin section images collected from the regions. Experiments are conducted based on the sandstone images taken from four regions in Tibet to study the performance of Festra. The experimental results have proved both effectiveness and validity of Festra, which provides competitive prediction performance on all the four regions, with few target instances labeled suitable for the field use.

1. Introduction

Identification of rock microscopic thin section images is an essential task in geology. Traditionally, the identification work is done manually by geologists, which is quite time-consuming. Furthermore, by visual estimation of the microscopic image, the mineral contents and textures are difficult to be quantitatively analyzed, and the results may not be repeatable between different identifiers. Currently, computer-aided automatic methods (Albar et al., 2013; Młynarczuk et al., 2013; Ślipek and Młynarczuk, 2013; Marmo et al., 2005; Zuo et al., 2009; Singh et al., 2010; Goncalves and Leta, 2010; Chatterjee, 2013; Wei et al., 2014a) have been developed for rock thin section identification.

Sandstone is one of the main reservoirs for oil, gas and groundwater. For example, sandstones are the main reservoirs of the famous Daqing, Shengli, Dagang and Karamay oil fields in China. Hence, sandstone identification is essential in both oil and gas exploration, which requires ocean and continental drilling, where the timely identification plays a crucial role. The automatic identification of sandstone microscopic images can satisfy the requirements of on-site grinding and real-time thin sections identifying, which will facilitate the drilling process.

Automatic identification of sandstone images requires the prior knowledge learned from the identified sandstones with similar compositions. However, due to diverse geological environments, the textural differences of the sandstones formed can be great among different regions. Sandstone images collected from an unfamiliar region are difficult to classify using the knowledge of identified images from known regions since the images may be quite different from each other. This problem can be defined as the interregional sandstone microscopic image classification. To solve this problem, we develop a transfer learning method Festra, which applies both feature and instance transfer to minimize the textural differences among images from multiple regions.

The main contributions of the paper are four-folds:

- We study the problem of automatic classification of sandstone microscopic images collected from multiple distinct regions, and formally define it as the interregional sandstone image classification.
- We propose the method Festra which combines the state-of-art transfer learning technologies to handle the interregional sandstone classification problem.
- We perform experiments on sandstone images collected from four regions in Tibet, and the results demonstrate both effectiveness and potentials of Festra.

* Corresponding author at: State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China.
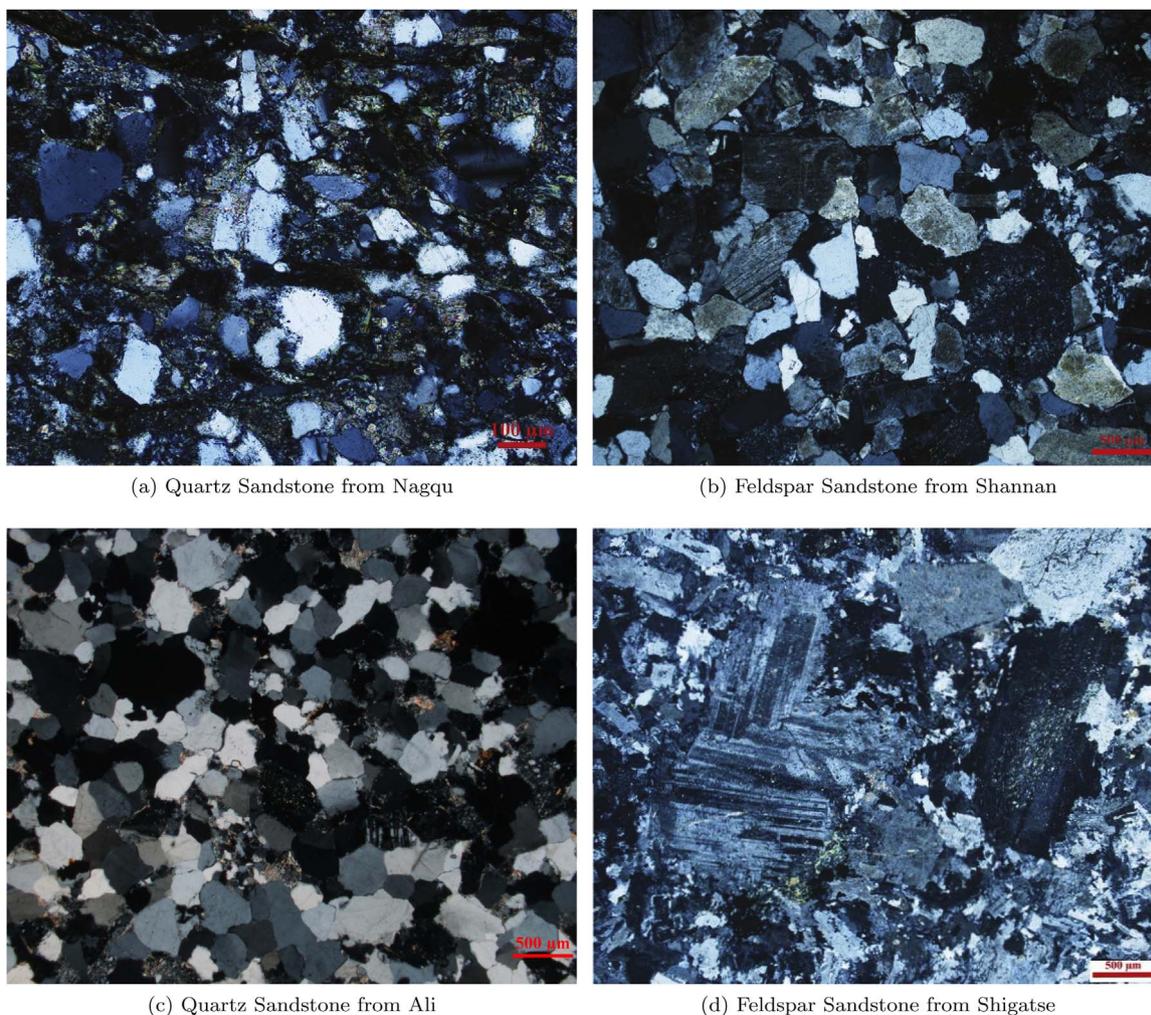E-mail address: guq@nju.edu.cn (Q. Gu).

(a) Quartz Sandstone from Nagqu

(b) Feldspar Sandstone from Shannan

(c) Quartz Sandstone from Ali

(d) Feldspar Sandstone from Shigatse

Fig. 1. The sandstone microscopic images of two types taken from separated regions.

- We extract the color, texture and grain features from the images, and make empirical studies on each feature group to compare the effects, which can provide suggestions for field use.

Although currently, Festra classifies sandstones at the thin section level, it can be extended to work at the mineral level for comprehensive sandstone identification, which is part of our future work.

2. Interregional sandstone microscopic image classification

Due to diverse geological environments and diagenesis, the textural differences of sandstones of similar types can be great between separated regions. Fig. 1 gives four sandstone images collected from different regions in Tibet, which belong to two sandstone types. As shown in Fig. 1, the image of feldspar sandstone taken from Shannan has great difference from the image of the same type taken from Shigatse. In addition, the image of the quartz sandstone taken from Ali and the image of the same type taken from Nagqu are totally different. Moreover, the resolutions or magnifications of the images may be different, and the thickness of the distinctive thin sections leads to different sharpness or brightness. The classification is rather intricate, and hard to determine even by experienced geologists.

In this paper, we study the interregional sandstone microscopic image classification, which can be formally defined as follows: given the sandstone microscopic images collected from the target region, a small number of which have been manually identified (i.e. labeled). The number may depend on the time and geologists available. The problem is to automatically classify the unknown sandstone images using the small fraction of labeled images, plus identified images collected from other fully explored regions. The labeled sandstone images from the target region are essential, but the number of them may be too small to train a useful classifier. Hence, the labeled sandstone images collected from other regions must be used, but those images are collected from variant regions. The difference between the target images and those images may be too great to make them directly helpful. A transfer learning method shall be developed to make effective use of them.

Transfer learning is to handle the problem of cross-domain classification (Ammar et al., 2015; Matasci et al., 2015). Transfer learning methods aim to leverage abundant labeled data in the source domain to build an accurate classifier for the target domain. The two domains may have distinct feature space distributions or labeling criteria (Pan and Yang, 2010).

3. Festra–the method

In this section, we describe the details of our proposed method Festra (A Feature based transfer learning method), which combines the feature transfer and instance transfer technologies, in order to solve the interregional sandstone microscopic image classification. In the following, we firstly give the formal notations, then describe the framework of Festra and the image features, and finally explain the major parts of Festra.

3.1. The problem and notations

Let the sandstone microscopic images collected from the new region compose the target domain **T**, and the labeled sandstone images from the known regions compose the source domain **S**. Let $\mathbf{T}_{(lab)} \in R^{k \times l}$ be the labeled sandstone images in the target domain, where k is the number of the labeled and l is the number of image features. Similarly, the set of unlabeled sandstone images in the target domain is denoted as $\mathbf{T}_{(unk)} \in R^{n \times l}$, and the set of labeled images in the source domain is denoted as $\mathbf{S} \in R^{m \times l}$, where n and m are the numbers of sandstone images in $\mathbf{T}_{(unk)}$ and **S** respectively.

Based on above notations, the function of Festra is to select the suitable sandstone images from **S** to complement $\mathbf{T}_{(lab)}$, and screen out harmful image features, so that the labeled image set is both sufficient in size and coherent in feature distributions to build the classifier for automatically classifying unlabeled sandstone images in $\mathbf{T}_{(unk)}$.

3.2. The framework

Festra has two parts: the first is feature selection (hereinafter referred to as FS) which selects features based on distribution similarity, the second is instance transfer which uses an enhanced TrAdaBoost (Dai et al., 2007) (E-TrAdaBoost) dealing with heterogeneous instances (images). TrAdaBoost is a boosting method for transfer learning, which iteratively selects suitable instances from source domain to assist training an accuracy classifier for target domain by adjusting the weights of labeled instances in both source and target domain iteratively. Fig. 2 presents the framework of Festra.

3.3. The image features

Given a sandstone microscopic image, we extract the color, texture and grain features, which are commonly used in visual sandstone thin section identification in geology. Since the resolutions or magnifications of the images maybe different, we take the relative values rather than the absolute ones for the features.

The color features are extracted from the gray level image of a sandstone thin section. All the color features are extracted based on the histogram of the gray levels (See Fig. 3), which include Mean, Variance,
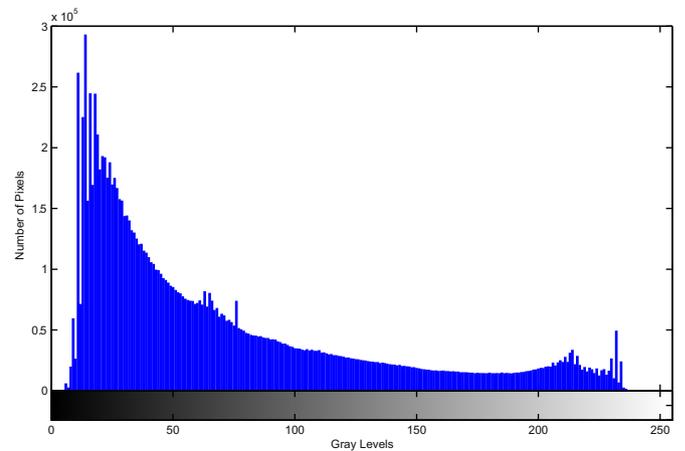

Fig. 3. The histogram of the gray levels of the quartz sandstone thin section image (Fig. 1a).

Median, Mode, Range, Mean absolute deviation, Interquartile range, Smoothness, Uniformity, Entropy, Skewness, Kurtosis and Fractal dimension features (Chatterjee and Bhattacherjee, 2011; Gonzalez et al., 2010). Table 1 shows the detailed descriptions of these statistics. Totally, 13 color features are extracted.

We use the gray-level co-occurrence-based features as texture features. To compute the gray-level co-occurrence-based features, two steps are required. Firstly, four co-occurrence matrices are computed from the gray level image along the 0°, 45°, 90° and 135° directions respectively, where the distance is 1. Secondly, from each of the matrices where elements are normalized to [0,1], four statistics are computed including Energy, Entropy, Contrast and Correlation (See Table 2) (Freund and Schapire, 1995). From the four co-occurrence matrices, a total of 16 features are extracted.

The grain features include two parts, one is image-based features, the other is grain-based features. The image-based features (Singh et al., 2010) are computed from the whole image, including Binary threshold, Number of Canny edge pixels, Number of perimeter pixels, Number of white areas and Number of pixels of white areas. Binary
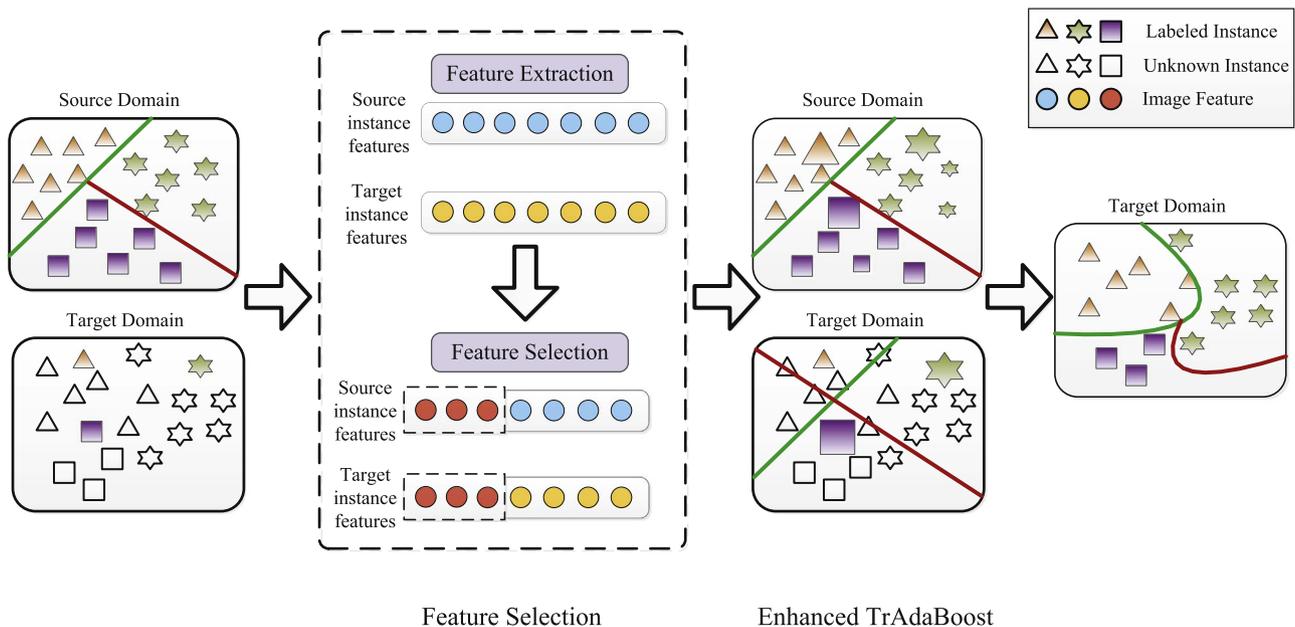

Feature Selection Enhanced TrAdaBoost

Fig. 2. The framework of Festra, which mainly involves two parts: Feature selection and Enhanced TrAdaBoost. In the figure, the source domain contains labeled data, while the target domain has few data labeled (Colored objects). Feature selection can select suitable features (Red circles), while Enhanced TrAdaBoost trains an useful classifier iteratively by adjusting the weights (Object sizes) of the labeled data in each iteration.

Table 1
The detailed descriptions of the color features.

Statistics	Formula	Remarks
Mean	$\mu = \frac{\sum_{i=0}^{255} x_i * y(x_i)}{N}$	Mean measures the average gray level of the image. In the formula, x_i is the gray levels, while $y(x_i)$ is the corresponding pixel number in the histogram; N is the total number of pixels. The same below.
Variance	$\sigma^2 = \frac{\sum_{i=0}^{255} (x_i - \mu)^2 * y(x_i)}{N-1}$	Variance measures the dispersion degree of the gray levels in an image.
Median	$\sum_{i=0}^{m-1} y(x_i) < \frac{N}{2}$ and $\sum_{i=0}^{m} y(x_i) \geq \frac{N}{2}$	Median measures the median value of the histogram of the gray levels. m is the value of this statistic in the formula.
Mode	$M_0 = \operatorname{argmax}_{x_i} y(x_i)$	Mode measures the most frequent gray level, which reflects the concentration degree of the gray level image. Function " argmax " returns x_i, at which the value of $y(x_i)$ is greatest.
Range	$r = \max(x_i) - \min(x_i)$	Range is range of the gray levels in an image. Function " max "(" min ") returns the greatest (smallest) value.
Mean absolute deviation	$MAD = \frac{\sum_{i=0}^{255} |x_i - \mu| * y(x_i)}{N}$	MAD is the average distance between each gray level and the mean.
Interquartile range	$IQR = Q_3 - Q_1$	Interquartile range is a basic robust measure of gray scale. In the formula, Q_3 is the 3rd quartile, and Q_1 is the 1st quartile within the gray level range.
Smoothness	$smoothness = 1 - \frac{1}{1+\sigma^2}$	Smoothness measures the degree of derivative continuous of the gray level histogram.
Uniformity	$U = \sum_{i=0}^{255} y^2(x_i)$	Uniformity measures similarity of gray level spins in the histogram.
Entropy	$E = -\sum_{i=0}^{255} y(x_i) \log y(x_i)$	Entropy measures the degree of randomness of the gray levels.
Skewness	$s = \frac{N * \sum_{i=0}^{255} y(x_i * (x_i - \mu)^3}{(N-1)*(N-2)*\sigma^3}$	Skewness measures the degree of asymmetry and direction of the gray level histogram.
Kurtosis	$Kurt = \frac{(N-1)*\sum_{i=0}^{255} y(x_i * (x_i - \mu)^4}{(N-2)*(N-3)*\sigma^4}$	Kurtosis is a descriptor of the shape of the histogram of gray levels.
Fractal dimension	See paper Chatterjee and Bhattacherjee (2011)	Fractal dimension measures the complexity and irregularity of patterns in the image.

threshold is the threshold value which separates the gray levels into two classes (above / below the threshold) where the inter-class variance reaches maximum. Number of Canny edge pixels is the number of white pixels in the binary image which resulted after the Canny edge detection (Canny, 1986). The rest three image-based features are computed from the binary image, obtained by converting the gray level image into binary using Ostu (Gonzalez et al., 2010), which takes Binary threshold as the threshold. A perimeter pixel is the one which is white and there is black pixels in its direct neighbors. Number of white areas and Number of pixels of white areas are computed by counting the continuous regions of white points in the binary image.

The grain-based features (Wei et al., 2014b) are computed using the three biggest grains, resulted from the superpixel-based segmentation (A method we developed for segmentation of sandstone thin section image, see Fig. 4). From each grain, four features are computed, which include Area of object, Convex deficiency, Extent and Solidity. Area of object means the number of the pixels in the grain. Convex deficiency is the percent difference of the grain area to the area of convex hull, which is the smallest convex set that contains the grain(See Fig. 5). Extent is the area of grain divided by area of its bounding box, which is the smallest rectangle containing the grain. Solidity is the area of grain divided by area of its convex hull. Totally 17 grain features are extracted.

3.4. Feature selection

In this part, suitable features are selected which have similar
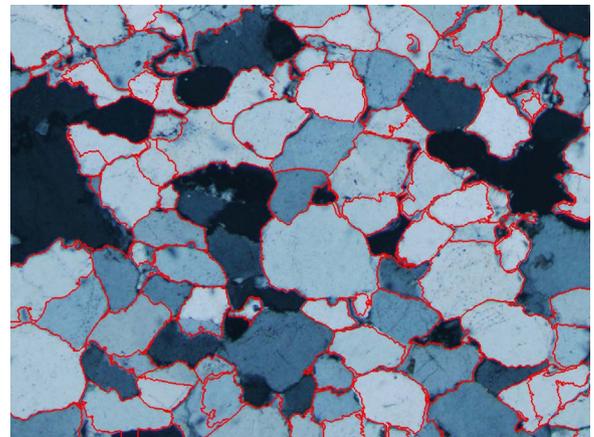

Fig. 4. The segmentation result of a quartz sandstone thin section image collected from Nagqu.

distributions between the source and target domains, so that the modified source domain is useful to train classifiers for the target domain.

The following three steps are involved in feature selection:

Step 1. Each feature is linearly normalized to [0,1] separately on data from either the source or target domain, where the maximum value is scaled to 1, and the minimum scaled to 0.

Step 2. For each feature, the frequency histogram of bins 0.1 is

Table 2
The detailed descriptions of the texture features.

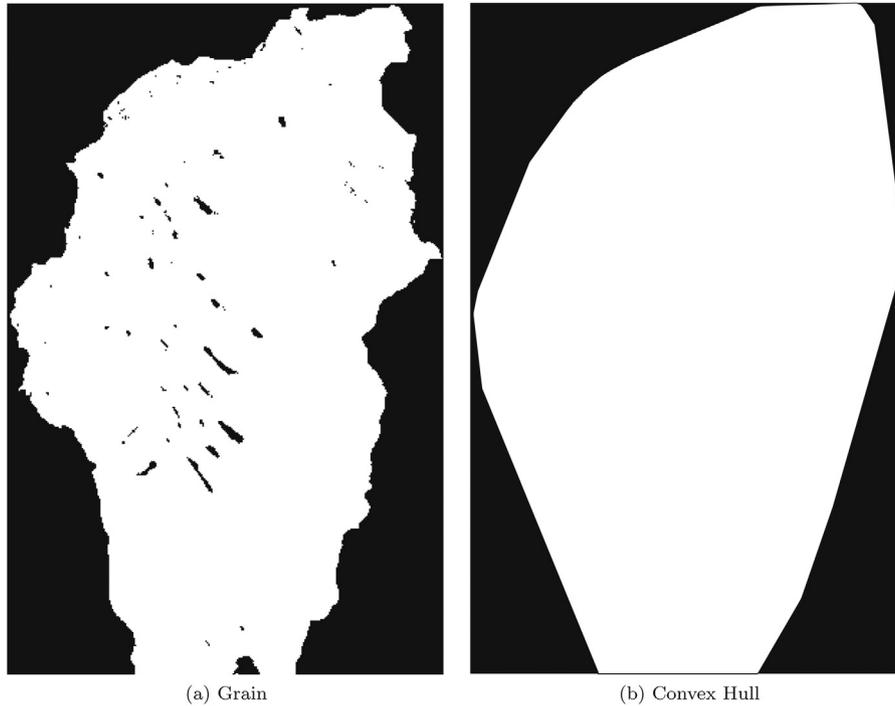| Statistics	Formula	Remarks
Energy	$E_0 = -\sum_i \sum_j M^2(i,j)$	Energy measures the uniformity of the texture in the image. In the formula, $M(i,j)$ is the co-occurrence matrices, where i and j are the row number and column number in the matrices respectively.
Entropy	$E = -\sum_i \sum_j M(i,j) \log M(i,j)$	Entropy measures the randomness of gray level distribution, it indicates the complexity of the textures in the image.
Contrast	$con = -\sum_i \sum_j (i-j)^2 M(i,j)$	Contrast measures the sharpness of the image and the depth degree of the texture grooves.
Correlation	$cor = -\sum_i \sum_j \frac{(i-\mu)(j-\mu)M(i,j)}{\sigma^2}$	Correlation measures the sum of correlations between each pixel and its neighbors in the image. In the formula, μ is the mean of the matrices, and σ^2 is the variance.

(a) Grain   (b) Convex Hull

Fig. 5. The binary images of a grain and its convex hull.

computed on the source and target data separately. Thus, each feature is transformed into a vector of 10 frequency values.

Step 3. The Cosine similarity measure is computed for each feature between the pair of vectors from both domains. After that, the features are ranked in descending order by the similarity measure, and the high rank features are selected to rebuild the source and target domain.

3.5. Enhanced TrAdaBoost

E-TrAdaBoost is an enhanced version of TrAdaBoost (Dai et al., 2007). For interregional sandstone microscopic image classification, E-TrAdaBoost makes the following enhancements.

- The sandstone microscopic images belong to more than two classes. E-TrAdaBoost requires to identify arbitrary number of classes instead of binary classification.
- TrAdaBoost requires the base classifiers be trained according to the weights of the instances. This may limit the choice of effective base classifiers. E-TrAdaBoost will make instance weighting optional.

Algorithm 1. E-TrAdaBoost.

Input:
Source domain data $\mathbf{S}$, target domain data $\mathbf{T}_{(unk)} \cup \mathbf{T}_{(lab)}$ and base classifier C
Output:
the predicted class for each instance in $\mathbf{T}_{(unk)}$
1: **Use sampling to balance $\mathbf{S}$ and $\mathbf{T}_{(lab)}$ among the classes**
2: **for** each target class **do**
3: **Assign the label of target class as 1, the rest classes as 0**
4: Initialize weights of instances in $\mathbf{S}$ and $\mathbf{T}_{(lab)}$:
$$\mathbf{W} = (w_1, w_2, \ldots, w_{m+k})$$
assure that sum of the total weights is 1
5: **for** each $t \in [1, N]$
6: **Train the classifier C on $\mathbf{T}_{(lab)} \cup \mathbf{S}$ with $\mathbf{W}$**
7: Apply TrAdaBoost (Dai et al., 2007), modify $\mathbf{W}$ accordingly

8: **end for**
9: **Compute the hypothesis $h(i)$ for each instance i in $\mathbf{T}_{(unk)}$:**
$$h(i) = \prod_{t=\lceil \frac{N}{2} \rceil}^{N} \beta_t^{-C_t(i)}$$
10: **end for**
11: **return** the predicted class leading to the maximum $h(i)$

Algorithm 1 gives a formal description of E-TrAdaBoost, which includes the following facilities.

Imbalanced data preprocessing. We reduce the possible data imbalance by over-sampling the minority classes and under-sampling the classes with too many instances.

Multiple class handling. E-TrAdaBoost deals with one target class at a time, where the class is labeled as 1 while the rest classes labeled as 0. Both the modified $\mathbf{S}$ and $\mathbf{T}_{(lab)}$ are relabeled before the training process. After that, voting is used to determine the final class of a target instance.

Instance weighting. If the base classifier C does not use instance weights, an instance selection process is applied to select instances according to $\mathbf{W}$ to build a training set for C. The total number of instances selected is unchanged for either class 1 or class 0, but the individual instance is selected randomly according to its weights. This means that an instance with greater weight (usually from $\mathbf{T}_{(lab)}$) may be selected multiple times, while an instance with lesser weight (usually from $\mathbf{S}$) may be filtered out from the training data. Such handling will increase the proportion of instances selected from the target domain, while decrease the proportion of misclassified instances from the source domain.

Confidence-based prediction. The base classifiers trained from $N/2$ to N rounds are combined to compute the hypothesis value of each unlabeled instance in $\mathbf{T}_{(unk)}$ according to the current target class (labeled as 1). Given instance i, its hypothesis value $h(i)$ indicates the confidence level of instance i belonging to the current target class. Hence, after all the target classes are treated, for instance i, its predicted class is the one which corresponds to the greatest hypothesis value.

4. Experiment setup

In this section, we introduce the corpus, the performance measures and experiment design to study the following research questions.

RQ 1. Whether Festra has better performance for interregional sandstone microscopic image classification, compared with other commonly used classification methods?

RQ 2. What is the effect of the number of labeled target instances on the prediction performance of Festra?

RQ 3. What is the effect of the proportion of image feature selection on the prediction performance of Festra?

RQ 4. Whether a specific group of image features leads to a better prediction performance than others for interregional sandstone microscopic image classification?

4.1. Corpus

The sandstone microscopic thin section images collected from four separated regions in Tibet compose the corpus. Totally 279 sandstone images are acquired, which include 78 images from Shigatse, 49 images from Nagqu, 89 images from Shannan, and 63 images from Ali. Each image is taken from a distinct sandstone thin section. All the sandstones are classified into 3 classes (Dott, 1964): Feldspar sandstone, Lithic sandstone, and Quartz sandstone. Table 3 lists the number of images of each class as well as their geological time and formations taken from each region. In general, the resolution of the images is 1280×1024 pixels, and thickness of the thin sections is 0.03 mm. Minor differences exist between images taken from different regions, since they are collected by different teams.

The regions are selected to be representative, and the geological times are different. Three regions are formed in singular geological time, mostly the Cretaceous, while Nagqu contains two geological times, the Jurassic and the Cretaceous. In addition, from Ali, the sandstones are taken from the Cretaceous Duoni Formation and Daxiong Formation in Coqen. From Nagqu, the sandstones are taken from the Late Cretaceous Jingzhushan Formation in Bangoin, and the Jurassic Sewa Formation, Suorika Formation and Biluocuo Formation in Shuanghu. From Shigatse, the feldspar sandstones and lithic sandstones are taken from the Forearc Basin in the Cretaceous, and the quartz sandstones taken from the Cretaceous Gucuo Formation and Jidula Formation at the northern margin of India. From Shannan, the sandstones are taken from the Paleogene Zongzhuo Formation. Fig. 6 shows the simplified tectonic map showing the locations of the four regions.

4.2. The performance measure

During the experiments, we compute the mean accuracy (hereinafter referred to as *mean − acc*) (Long et al., 2015) and F1 (Powers, 2011) as the performance measures, which are suitable for datasets with imbalanced class distributions. *mean − acc* measures the mean prediction accuracy of the three classes, and is computed using the following formula:

$$mean - acc = \frac{\sum_{i=1}^{N} acc_i}{N} \tag{1}$$

where acc_i denotes the proportion of correct predictions for the ith class, and $N=3$ denotes the number of sandstone classes.

F1 measure is computed using the following formula:

$$F1 = \frac{2TP}{2TP + FN + FP} \tag{2}$$

where TP, FN and FP are the sample number of true positives, false negatives and false positives respectively. Both *mean − acc* and F1 range from 0 to 1, and better classifier has greater values.

4.3. Experiment design

We conduct four experiments each using one of the four regions as the unfamiliar target, and treating the rest as known regions. Under each case, we take sandstone images taken from the target region as the target domain, and the others as the source domain. We assume that only a small number of sandstone microscopic images are labeled (identified) in the target domain, and the images in the source domain are already labeled.

For the methods, we implement Festra-40, Festra-80 and E-TrAdaBoost, which means the proportion of features selected are 40%, 80% and 100% (i.e. no feature selection) respectively. We also remove E-TrAdaBoost from Festra, which leaves only the feature selection part (FS), to study its effect on interregional sandstone microscopic image classification. For example, FS-40 and FS-80 represent the cases where the proportion of features selected are 40% and 80% respectively. The base classifiers are used as the baselines during the experiments.

We choose one of four commonly used classifiers, Support Vector Machine (SVM) (Hearst et al., 1998), Logistic Regression (LR) (Hosmer et al., 2000), Decision Tree (DT) (Swain and Hauska, 1977) and Naïve Bayes (NB) (Murphy, 2006) as the base classifier, to study the effects of different models on sandstone classification.

To study the research questions, we design four experiments. Firstly, we randomly take one image (if exist, refer to Table 3) from each of the three classes in the target region as the labeled instance, and combine the images in the source domain to compose the training set. We train all the methods on the training set to study whether Festra has better performance for predicting unlabeled images in the target domain. Secondly, we vary the number of target instances labeled to investigate its effects on the prediction performance of Festra. For each target region, the number of target instances labeled from each of the three classes is set from 1 to 13 step by 1. Refer to Table 1, if a certain class does not have enough images (e.g. Feldspar sandstone or Quartz sandstone), the number of labeled target instances will be half of the available ones. Thirdly, we vary the proportion of features selected to investigate its effects on the prediction performance of Festra. For each region, we vary the proportion from 10% to 100% step by 10%. Fourthly, we compare the base classifier and E-TrAdaBoost using one of the three categories (color, texture, grain) of image features respectively, to investigate the effects of a specific

Table 3
The number of sandstone microscopic images taken from the four regions in Tibet.

| Region	Places	Age	Units	Number of Images			
				Feldspar Sandstone	Lithic Sandstone	Quartz Sandstone	Total
Ali	Coqen	Cretaceous	Duoni,Daxiong	0	41	22	63
Nagqu	Bangoin	Cretaceous	Jingzhushan	0	40	0	40
	Shuanghu	Jurassic	Sewa, Suorika, Biluocuo	0	5	4	9
Shigatse	Ngamring, Tingri, Zhongba, Shigatse	Cretaceous	Ngamring, Padana, Quxia, Jialazi, Weimei, Jidula	3	45	30	78
Shannan	Nangarze	Paleogene	Zongzhuo	8	75	6	89
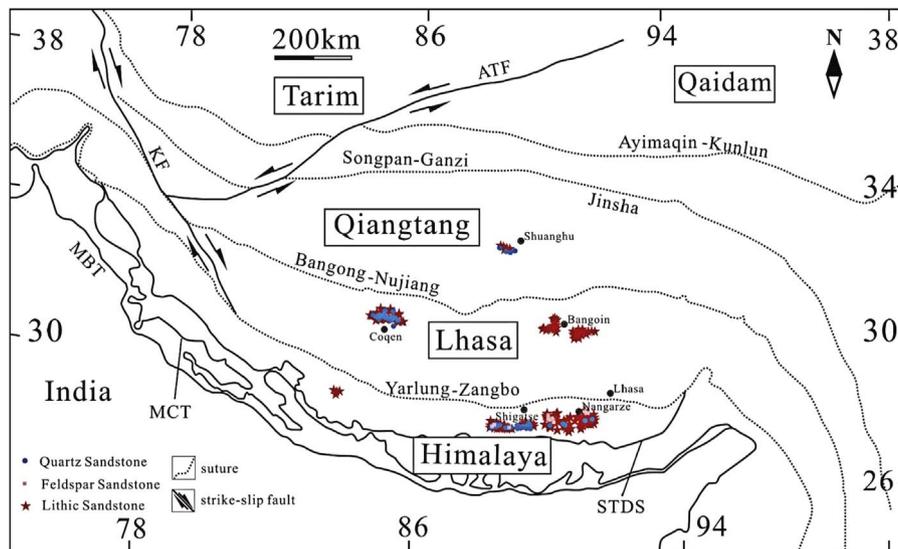
Fig. 6. Simplified tectonic map of the Tibetan Plateau. MBT - Main Boundary Thrust; MCT - Main Central Thrust; STDS - Southern Tibetan Detachment System; KF - Karakorum Fault; ATF - Alty Tagh Fault.

category of features for interregional sandstone image classification. Start from the second experiment, the base classifier used is Naïve Bayes. In the third and fourth experiments, the number of labeled target instances from each class is 1.

For cross-validation, to reduce the randomness, we repeat each experiment 50 times, each with randomly selected target instances as labeled. The mean performance measures are computed for comparison.

5. Results and analysis

As described in Section 4.3, experiments are conducted taking in turn each of the four regions in Tibet as the target. Here we describe the experiment results of each target region respectively.

5.1. Results and analysis of ali

To study the prediction performance of Festra on Ali, we conduct experiments using the sandstone images taken from Ali as the target domain, while the rest as the source domain. Table 4 and Fig. 7 present the performance measures resulted from the four experiments.

Table 4 compares the performance measures among all the methods. From the table, no matter which of the base classifiers is used, Festra always has the best prediction performance. When the base classifier is Naïve Bayes (NB), Festra-40 and Festra-80 can obtain 5.52% and 5.82% improvement in mean-acc as well as 1.92% and 2.33% improvement in F1 measure respectively by adding E-TrAdaBoost, compared to the counterparts using only feature selection. Similar improvements can be seen when using the other three base

classifiers. In addition, feature selection can also help improve the prediction performance. Festra-40 performs the best in most case, while Festra-80 performs the best in the rest. For example, using DT, compared with the base classifier, FS-40 and FS-80 improve the performance by 3.32% and 1.5% in mean-acc as well as 0.31% and 1.73% in F1 measure. Hence, both feature selection and E-TrAdaBoost are useful, and their combination is valuable for sandstone image classification on Ali.

Fig. 7(a) presents in line charts the effects of the number of labeled target instances on the prediction performance of the methods. According to the figure, when the number of labeled target instances is small (less than four from each class), Festra has the best performance. When the number of the labeled target instances is greater, the merits of feature selection and E-TrAdaBoost are diminished. The reason may be that as the number of labeled target instances increases, effects of the source domain will decrease, which makes Festra less effective.

Fig. 7(b) presents in line charts the effects of the proportion of selected image features on the prediction performance of feature selection. According to the figure, as the proportion of selected features increases, the prediction performance of either FS or Festra presents similar trends. Both increase at first, then become steady after reaching the peak. With feature selection only (FS), the prediction performance reaches the peak point at proportion range 40–70%, and by adding E-TrAdaBoost (Festra), the optimal proportion shifts backward to 40%. Hence for Ali, Festra can make a good classification of sandstone images with only 40% of the original image features.

Fig. 7(c) presents in bar charts the prediction performance of the methods when a specific group of image features is used. According to

Table 4

The mean-acc and F1 measure of all the methods using one of the four base classifiers on Ali. The highest measure of each method combination (column) is marked in bold.

Methods	Mean-acc				F1 Measure			
	NB	DT	LR	SVM	NB	DT	LR	SVM
Base	74.47%	75.49%	76.68%	85.78%	0.8325	0.8115	0.8114	0.9045
FS–40	77.82%	78.81%	78.80%	86.07%	0.8544	0.8146	0.8160	0.9094
FS–80	76.44%	76.99%	78.10%	87.07%	0.8532	0.8288	0.8146	0.9106
E-TrAdaBoost	81.65%	78.96%	80.46%	87.80%	0.8586	0.8290	0.8185	0.9130
Festra–40	**83.34%**	**80.34%**	**81.05%**	89.19%	0.8736	**0.8532**	**0.8443**	0.9156
Festra–80	82.26%	79.38%	80.86%	**89.93%**	**0.8765**	0.8305	0.8372	**0.9167**

(a) mean-acc for RQ 2   (b) mean-acc for RQ 3   (c) mean-acc for RQ 4
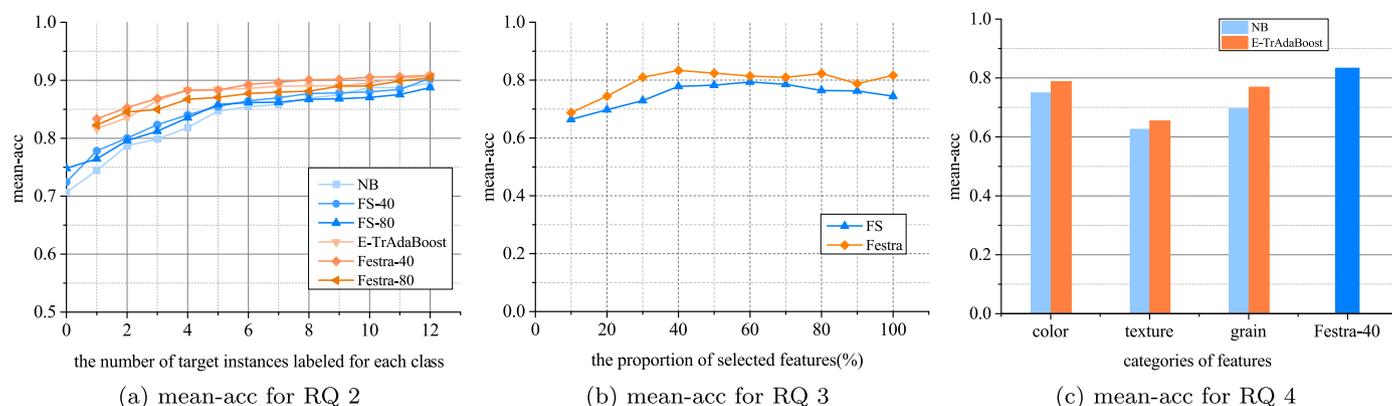
Fig. 7. The results of Ali.

the figure, the color features perform best on Ali. Both NB and E-TrAdaBoost using the color features outperform those using the texture or grain features. In addition, by adding feature selection (Festra-40), the performance can be further improved with similar number of features, which suggest that feature selection can provide a good means of combining features from multiple categories.

Based on the above analysis, Festra has the best prediction performance on the Ali samples, when the number of labeled target instances is small. It requires only 40% of the image features to get a good performance. In addition, the color features are more useful for sandstone image classification on Ali than the texture and the grain features.

5.2. The results of nagqu

Table 5 and Fig. 8 present the performance measures of the methods using the sandstone images taken from Nagqu as the target domain, while the rest as the source domain. As shown in Table 5, both feature selection and E-TrAdaBoost improve the prediction performance. No matter which of the base classifiers is used, Festra always outperforms other methods, and Festra-40 performs the best in most cases. In Fig. 8(a), when the number of labeled target instances is small, both feature selection and E-TrAdaBoost can help improve the performance. When the labeled target instances increase, both become less effective. In Fig. 8(b), when the proportion of selected features is within 30–60%, Festra performs the best. In Fig. 8(c), the color features perform better than the other two groups, while Festra can make use of both texture and grain features to improve the classification performance.

5.3. The results of shigatse

Table 6 and Fig. 9 present the performance measures of the

methods using the sandstone images taken from Shigatse as the target domain. As shown in Table 6, both feature selection and E-TrAdaBoost are helpful for improving the prediction performance, and Festra performs the best in all cases. Festra-40 performs the best when using DT or SVM as the base classifier, while Festra-80 outperforms the others when using NB or LR. In Fig. 9(a), when the number of labeled target instances is small (less than five from each class), E-TrAdaBoost is more effective than feature selection. When the labeled target instances increase, the difference between all methods decreases. In Fig. 9(b), the prediction performance of either NB+FS or Festra becomes poor if too many features are selected. Festra works well when the proportion of selected features is around 40%. In Fig. 9(c), the color features perform poor on Shigatse, while Festra can still make good use of the three feature categories.

5.4. The results of shannan

Table 7 and Fig. 10 present the performance measures of the methods using the sandstone images taken from Shannan as the target domain. As shown in Table 7, both feature selection and E-TrAdaBoost improve the prediction performance of the four base classifiers. In Fig. 10(a), when the number of labeled target instances is small (less than four from each class), both feature selection and E-TrAdaBoost can help improve the prediction performance. When the labeled target instances increase, the merit of E-TrAdaBoost will decrease, while feature selection still performs well. In Fig. 10(b), the optimal proportion of selected features for Festra is around 40%. In Fig. 10(c), the color features outperform the other groups when using the base classifier alone, while Festra further improve the performance by making use of other feature groups.
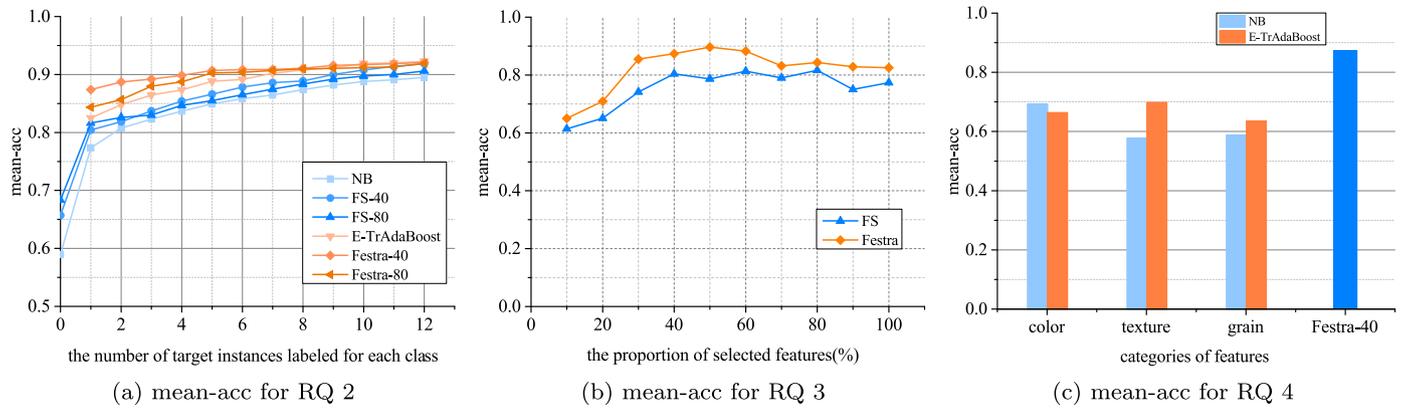
Table 5
The mean-acc and F1 measure of all the methods using one of the four base classifiers on Nagqu. The highest measure of each method combination (column) is marked in bold.

Methods	Mean-acc				F1 Measure			
	NB	DT	LR	SVM	NB	DT	LR	SVM
Base	77.36%	64.80%	86.64%	82.92%	0.8659	0.7636	0.8625	0.7949
FS−40	80.42%	70.23%	88.63%	83.95%	0.8706	0.7704	0.8731	0.8043
FS−80	81.64%	69.91%	89.60%	85.45%	0.8681	0.7512	0.8774	0.8126
E-TrAdaBoost	82.49%	71.66%	92.61%	85.54%	0.8672	0.7752	0.8754	0.8142
Festra−40	**87.38%**	**74.83%**	**93.18%**	85.52%	**0.8773**	**0.7882**	0.8713	0.8139
Festra−80	84.34%	71.07%	92.84%	**86.93%**	0.8682	0.7559	**0.8815**	**0.8203**

(a) mean-acc for RQ 2 (b) mean-acc for RQ 3 (c) mean-acc for RQ 4

Fig. 8. The results of Nagqu.

Table 6
The mean-acc and F1 measure of all the methods using one of the four base classifiers on Shigatse. The highest measure of each method combination (column) is marked in bold.

Methods	Mean-acc				F1 Measure			
	NB	DT	LR	SVM	NB	DT	LR	SVM
Base	52.77%	50.17%	52.85%	53.87%	0.6178	0.4999	0.5529	0.5696
FS−40	53.56%	51.15%	54.74%	55.44%	0.6232	0.5249	0.5597	0.5822
FS−80	54.12%	53.35%	57.99%	54.99%	0.6180	0.5159	0.5626	0.5907
E-TrAdaBoost	55.39%	54.09%	59.05%	55.91%	0.6357	0.5374	0.5664	0.5938
Festra−40	57.93%	**58.04%**	58.94%	**57.38%**	**0.6504**	**0.5688**	0.5691	**0.6123**
Festra−80	**58.18%**	55.89%	**61.67%**	56.40%	0.6392	0.5640	**0.5859**	0.6041

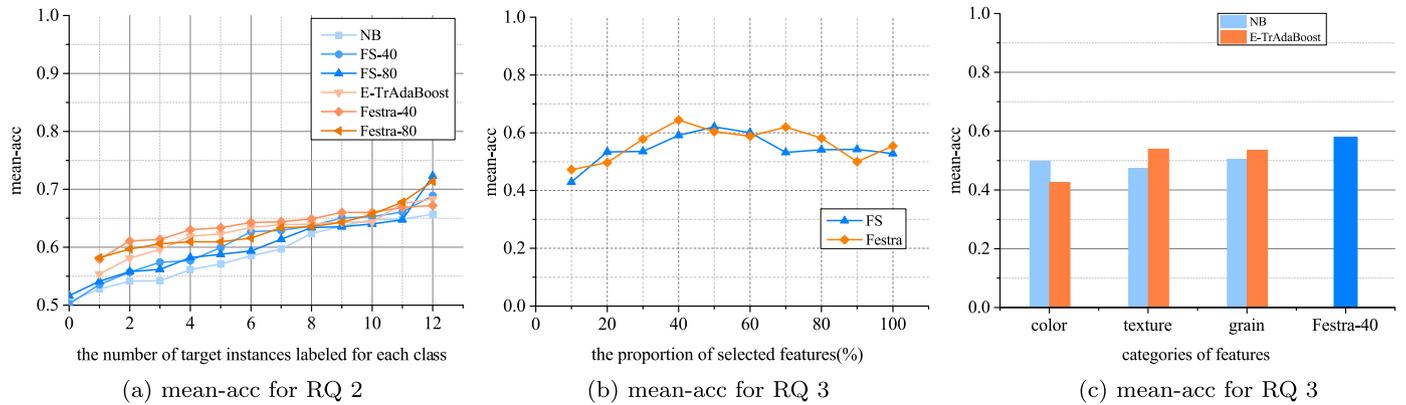(a) mean-acc for RQ 2 (b) mean-acc for RQ 3 (c) mean-acc for RQ 3

Fig. 9. The results of Shigatse.

Table 7
The mean-acc and F1 measure of all the methods using one of the four base classifiers on Shannan. The highest measure of each method combination (column) is marked in bold.

Methods	Mean-acc				F1 Measure			
	NB	DT	LR	SVM	NB	DT	LR	SVM
Base	52.67%	45.38%	50.29%	43.16%	0.5286	0.4889	0.6266	0.5025
FS−40	53.95%	50.02%	53.45%	44.35%	0.5304	0.4911	0.6333	0.5079
FS−80	54.61%	51.27%	55.34%	43.96%	0.5342	0.4970	0.6359	0.5093
E-TrAdaBoost	56.02%	50.64%	52.83%	44.61%	0.5405	0.4983	0.6388	0.5109
Festra−40	**58.30%**	**52.87%**	54.94%	**47.54%**	0.5439	**0.5152**	0.6390	0.5186
Festra−80	56.05%	51.47%	**60.75%**	45.26%	**0.5511**	0.4973	**0.6622**	**0.5239**

6. Discussions

Based on the experiment results, we can have the following conclusions: Festra has good the prediction performance for inter-regional sandstone microscopic image classification. Both feature selection and E-TrAdaBoost are useful, and their combination can further improve the prediction performance.

The purpose of both feature selection and E-TrAdaBoost is to reduce the difference in data distribution between the source and target domain. Feature selection may screen out features which have great difference between the target and source regions, while E-TrAdaBoost can mitigate the difference among thin section images by the weighting mechanism. During experiments, it is evident that both the number of labeled instances from the target domain, and the proportion of selected features will affect the prediction performance of Festra. In case of the number of labeled instances, Festra has good performance
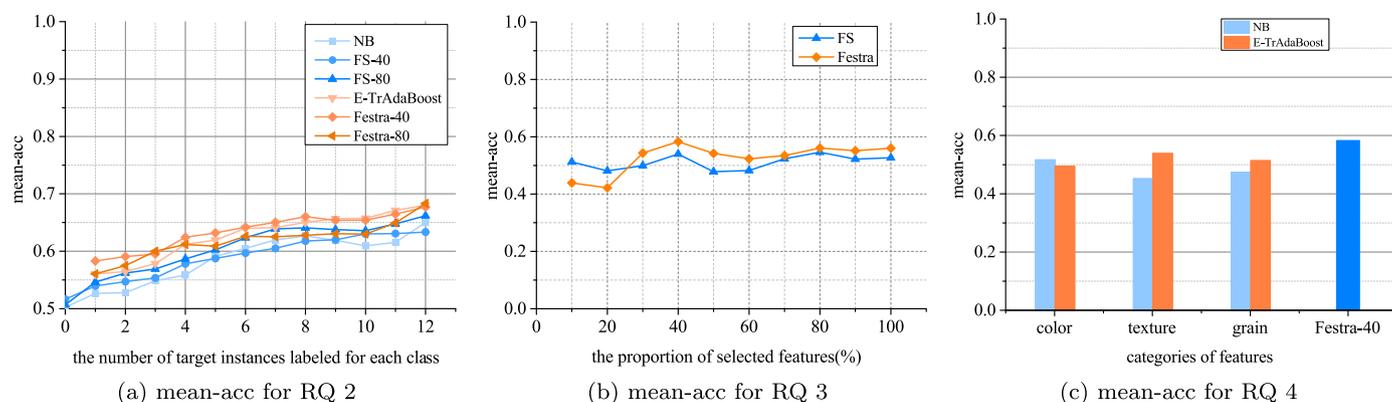
(a) mean-acc for RQ 2

(b) mean-acc for RQ 3

(c) mean-acc for RQ 4

Fig. 10. The results of Shannan.

Table 8
The F1 measure of Feastr-40 using different similarity measure on Shigatse.

Classifier	Similarity Measure			
	Cosine	Euclidean	Jaccard	Pearson
NB	0.6504	0.6431	0.6451	0.6518
DT	0.5688	0.5661	0.5642	0.5625
LR	0.5691	0.5696	0.5711	0.5679
SVM	0.6123	0.6043	0.6091	0.6137

when the number of target instances labeled is small, which is suitable for sandstone classification in the field. In case of the proportion of features selected, the optimal proportion varies little by Festra in different target regions. Specifically, the best proportion is 40% in Ali, Shigatse and Shannan, while within 30–60% in Nagqu.

The color, texture and grain features are extracted from the sandstone images during our experiments. The three groups are commonly used by visual sandstone thin section identification in geology. To investigate the effects of singular group of features for interregional sandstone microscopic image classification,we conduct experiments on each group separately. Based on the experiment results, on average the color features perform best on three regions except Shigatse. By adding feature selection and/or E-TrAdaBoost, the performance can be further improved by selecting features from all three groups. Hence, focusing on singular feature group is not enough, and a combination of diverse image features besides color, texture, and grain is helpful for interregional sandstone microscopic image classification.

The similarity measure is a key factor in the feature selection stage. In order to choose the suitable measure, we conduct experiments comparing different kinds of similarity measures (Cha, 2007), including Cosine similarity, Euclidean distance, Jaccard coefficient and Pearson coefficient. Table 8 lists the F1 measure of Festra-40 using one of the four similarity measures on Shigatse. From the table, we find that the performance difference is little among the similarity measures. The same is true in the other three regions and using other feature proportions. Hence, we can safely choose Cosine similarity as the similarity measure without losing generality.

As described in Section 4.1, the geological times of the four regions in Tibet are different. According to the experiment results, Festra works best in Nagqu, which contains two geological times: the Jurassic and the Cretaceous, and in Ali, which is formed in the Cretaceous with multiple formations. In Shigatse, where the sandstones are collected from formations in the Cretaceous, and in Shannan, which contains singular formation in the Paleogene, the merits of Festra decrease. Above suggests that compared with the difference in geological times, the regional difference has more effect on Festra. For interregional sandstone microscopic image classification, other methods are required besides Festra.

It is important to note that although in this paper, Festra is applied to sandstone classification at the thin section level, it can be extended to the mineral level for composition analysis and comprehensive sandstone identification, which is part of our future work.

7. Conclusions and future work

In this paper, we propose a transfer learning method Festra to handle the problem of interregional sandstone microscopic image classification. Festra includes both feature selection and E-TrAdaBoost, which combines the techniques of both feature and instance transfer. The object of feature selection is to screen out features which have great difference between the target and source regions, while E-TrAdaBoost is to mitigate the difference among thin section images collected from different regions. Hence, labeled instances from multiple regions can be used to train high-quality classifiers for prediction in the target region. We conduct experiments based on sandstone microscopic images collected from four separated regions in Tibet, which involve different geologic times and formations. The results demonstrate both effectiveness and potentials of Festra.

In our future work, on one hand, we plan to expand our dataset to include sandstone images taken from other regions to verify the validity of Festra. We also plan to apply Festra to other types of rock images to explore the potentials of the method. On the other hand, we plan to design extra image features to describe the characteristics of sandstone images, and try other feature selection methods (e.g. Principal Component Analysis) and use more comprehensive classifiers (e.g. Neural Networks and deep learning methods) to handle the interregional classification problem.

References

Albar, A., Osman, M.H., Abdullah, M.S., Ismail, B.N., 2013. Classification of intrusive igneous rocks using digital image processing: a binary approach. J. Eng. Sci. 9, 11–19.

Ammar, H.B., Eaton E., Ruvolo P., Taylor M.E., 2015. Unsupervised cross-domain transfer in policy gradient reinforcement learning via manifold alignment. In: Proceedings of AAAI.

Canny, J., 1986. A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. 6, 679–698.

Cha, S.H., 2007. Comprehensive survey on distance/similarity measures between probability density functions. City 1 (2), 1.

Chatterjee, S., 2013. Vision-based rock-type classification of limestone using multi-class support vector machine. Appl. Intell. 39 (1), 14–27.

Chatterjee, S., Bhattacherjee, A., 2011. Genetic algorithms for feature selection of image analysis-based quality monitoring model: an application to an iron mine. Eng. Appl. Artif. Intell. 24 (5), 786–795.

Dai, W., Yang, Q., Xue, G.R., Yu, Y., 2007. Boosting for transfer learning. In: Proceedings of the 24th international conference on Machine learning. ACM; p. 193-200.

Dott, R.H., Jr, 1964. Wacke, graywacke and matrix-what approach to immature

sandstone classification? J. Sediment. Res. 34, 3.

Freund, Y., Schapire, R.E., 1995. A desicion-theoretic generalization of on-line learning and an application to boosting. In: European conference on computational learning theory. Springer; p. 23-37.

Gonçalves, L.B., Leta, F.R., 2010. Macroscopic rock texture image classification using a hierarchical neuro-fuzzy class method. Mathematical problems in engineering 2010.

Gonzalez, R.C., Woods, R.E., Eddins, S.L., 2010. Digital Image Processing Using MATLAB®. McGraw Hill Education.

Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Scholkopf, B., 1998. Support vector machines. IEEE Intell. Syst. their Appl. 13 (4), 18–28.

Hosmer, D., Jr, Lemeshow, S., Sturdivant, R.X., 2000. Introduction to the logistic regression model. Appl. Logist. Regres..

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. p. 3431–3440.

Marmo, R., Amodio, S., Tagliaferri, R., Ferreri, V., Longo, G., 2005. Textural identification of carbonate rocks by image processing and neural network: methodology proposal and examples. Comput. Geosci. 31 (5), 649–659.

Matasci, G., Volpi, M., Kanevski, M., Bruzzone, L., Tuia, D., 2015. Semisupervised transfer component analysis for domain adaptation in remote sensing image classification. IEEE Trans. Geosci. Remote Sens. 53 (7), 3550–3564.

Młynarczuk, M., Górszczyk, A., Ślipek, B., 2013. The application of pattern recognition in the automatic classification of microscopic rock images. Comput. Geosci. 60, 126–133.

Murphy, K.P., 2006. Naive bayes classifiers. University of British Columbia.

Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. data Eng. 22 (10), 1345–1359.

Powers. D.M., 2011. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation.

Singh, N., Singh, T., Tiwary, A., Sarkar, K.M., 2010. Textural identification of basaltic rock mass using image processing and neural network. Comput. Geosci. 14 (2), 301–310.

Ślipek, B., Młynarczuk, M., 2013. Application of pattern recognition methods to automatic identification of microscopic images of rocks registered under different polarization and lighting conditions. Geol., Geophys. Environ., 39.

Swain, P.H., Hauska, H., 1977. The decision tree classifier: design and potential. IEEE Trans. Geosci. Electron. 15 (3), 142–147.

Wei X.S., Qin X.H., Rong C.L., Nan J.X., Cheng G.J. 2014a. Image classification recognition for rock micro-thin section based on probabilistic neural networks. In: Applied Mechanics and Materials; vol. 602. Trans Tech Publ; p. 2147–2152.

Wei, X.S., Rong, C.L., Nan, J.X., Cheng, G.J., Liu, Y., 2014b. Rock classification based on image processing and neural networks. In: Applied Mechanics and Materials; vol. 568. Trans Tech Publ, p. 685–690.

Zuo, R., Cheng, Q., Xia, Q., Agterberg, F., 2009. Application of fractal models to distinguish between different mineral phases. Math. Geosci. 41 (1), 71–80.