Solid Earth Discuss.,
doi:10.5194/se-2016-44-RC1, 2016

Solid Earth
Discussions
Open Access
EGU

Interactive comment on "Phase Segmentation of X-Ray Computer Tomography Rock Images using Machine Learning Techniques: an Accuracy and Performance Study" *by* Swarup Chauhan et al.

Anonymous Referee #1

Received and published: 10 April 2016

Image segmentation is the most crucial step in image processing of micro-CT images of porous rocks, because functional properties are usually not derived from grayscale data itself but from the segmented data. For instance, Lattice-Boltzmann simulations are conducted on the segmented pore space, statistical analysis are carried out for different material classes, etc. A multitude of different segmentation exists to date, which differ vastly in computational complexity, underlying rationale and so on. There are many review papers on image segmentation, that try to sort these existing segmentation algorithms according to their methodological approaches or rank them according to their suitability for a set of test images. The general conclusion is often, that no segmentation excels all others in all cases and it depends very much on the image content

which one is suited best. Chauhan et al. wrote yet another of these review articles and limit their focus on machine learning algorithms, which might be well established in remote sensing, life sciencces or other scientific disciplines, but have not yet gained much attention when it comes to micro-CT images of rocks. Therefore, the article could in principle be useful in closing that gap. However I cannot recommend its publication for several reasons.

First of all, the article by Chauhan et al. can be understood as a follow-up study to Chauhan et al. (2016): Computers & Geosci., doi:10.1016/j.cageo.2015.10.013. The general purpose of the current study is again to test the suitability of various machine learning algorithms (supervised or unsupervised) to segment micro-CT images on a set of different rock images. The overlap with the precursor study is high. In fact, the only salient difference between them is that four images have been used for testing instead of one and I'm wondering why this hadn't been done in this first place.

There are several other issues with this paper:

1. The choice of different validation methods for different methods hampers comparability among all methods. For one method you use MSE, for another purity and entropy and for yet another method ROC curves are computed. Moreover, these validation methods need to be explained in much more detail including formulas. This includes MSE, ROC curve, 10K cross validation, purity and entropy metric.

2. Wording is frequently mixed with unexplained jargon and often does not meet a sufficient standard to follow the line of argument. I list plenty of examples below.

3. Introduction is too short. The introduction is only half a page long and doesn't barely touch the current state of knowledge.

4. Preprocessing and/or postprocessing is not discussed. Therefore, the suggested work flow is far away from common practice of most scientists, working in the field. For instance, the sandstone image seems to be extremely noisy. Therefore, most

colleagues would probably use some noise filter as a preprocessing step, or use some spatial regularization during image segmentation, e.g. by applying a locally-adaptive method, or apply some post-processing, e.g. majority filter or morphological operators, to clean up the results.

5. Conclusions are weak. There is no real line of argument in the conclusions. The findings are mainly reported for each method independently and are sometimes too specific (explanation below) to be generalized into something useful. In turn, some conclusions are basically what seems to be common sense, e.g. feature vectors have to be chosen carefully or there should only be as many classes as there are real phases in the image (not more, not less).

5. The machine learning algorithms seem to be very impractical for realistic datasets due to excessive computation time. The dataset were actually quite small, up to 31 megapixels (MP). Real micro-CT datasets nowadays have dimensions up to 8000MP. I could not find a comment on how CPU:time scales with image size, but it would definitely render LS-SVM as one of the recommended methods useless for most practical applications. This would leave k-means and FCM (p9l7-9) as the only recommended ML methods, and these exists already for more then four decades. All in all, I'm not convinced why I should use machine learning algorithms for micro-CT image segmentation in the future.

Technical comments:

p1l22-23: Bad wording

p2l38: Bad wording: ... to obtain images of elements 1024 x 1024 x 1024 ...

p3l1: Bad wording: ... from the by applying Fourier ...

p3l6: Bad wording: ... rely of features ...

p5l10: Meaning unclear: "... minimum leaf size of five and learning rate of 0.1."

p5l12: Meaning unclear: "... apriori information in the form of most useful pixel values." How can a pixel value be useful? Do you mean, that a pixel value is most likely to be assigned to a specific material, because according to the manually chosen feature vectors it is more similar to the class statistics of this material than any other material?

p5l14-15: meaning unclear: "... a set of ten XCT images". Do you mean ten slices from a XCT image?

p5l26-27: meaning unclear: "... the know classes, despite number of classes are different from number of segmented classes."

p5l30: "... between output and targets." Be more specific. What are targets? Class assignments for manually selected pixels? Are feature vectors the same as targets?

p5l33-34: meaning unclear: "It shows a trade-off between sensitivity ..." This sentence is a stub. Trade-off between sensitivity and what?

p5l35: typo: prefect

p6l6-9: What about unresolved porosity below the image resolution? Shouldn't the image-derived porosities be all much lower than the experimental porosity values, because they do not capture very small pores?

p6l10-15: Okay, so a higher fuzziness parameter shifts the pore/matrix threshold towards lower gray values, so that the volume fraction of pores is reduced? But why exactly is this the case? Also, in the conclusions (p8l16-18) you state that somehow FCM can distinguish between pores and pore-throats, which is not true, because it would mean that the algorithm could distinguish between different pore sizes or functional units of the pore space. All FCM does, is to evaluate the histogram (only grayscale information, no spatial information) and depending on how you set the fuzziness parameter, partial volume voxels are assigned to pores or matrix. Similar statement in p8l16-17.

p6l21-22: meaning unclear: "Morphological and filtering operations were performed

based on the complexity of the segmented images". Which image was processed how?

p6l31: meaning unclear: "... and post processing of the unknown dataset." This is the first time you mention post processing. Do you mean with "post", that it is carried out after some tentative image segmentation is completed? What do you do specifically during post-processing? Why is it somehow linked to the size of feature vectors?

p7l2: meaning unclear: "As a consequence, the individual (weak) classification models". What do you mean by weak?

p7l5: meaning unclear: "... most appropriate class for each phase." Do you mean: ... for each pixel?

p7l11: meaning unclear: " ... cluster homogeneity is over-segmented ..." This makes no sense to me.

p7l28-30: "As the hand-picked ... quality and speed". This sentence can probably only be understood by an absolute insider. Which material in which rock did you hand-pick to represent class four? Why do you consider a mix of all phases and noise appropriate? For me this actually sounds like the method failed completely, when class four does not represent a single material.

p7l39: meaning unclear: " The initial growth of the leaf size ...".

p8l1: typo: chucks

p8l1-3: meaning unclear: "Using a for-loop with an increment of from one to ten, ... ith fold." Since you did not describe 10K cross validation in the first place, it is not possible to understand this sentence without background knowledge.

p8l8-14: This paragraph should be part of the discussion and not the conclusions.

p8l14: What is class six? Why is class six different from others?

p8l19-22: This conclusion is highly specific to the SOM method and cannot be generalized into something useful. You're basically saying the the parameters that you've chosen a priori worked well. But maybe a hexagonal topology and a Euclidean distance function would have also done the job?

p8l26-30: Here in the conclusions you refer for the first time to "scaling". Why would you want to scale your images, when you're only interested in the segmentation results at the scale at which you acquired the image. Often a segmentation result at a coarse scale is of very little use. Therefore, the sentence "Additionally, the accuracy ..." is hard to follow.

p8l35-39: Of course the class labels should contain only one real phase in your rock, not only parts of it and not many simultaneously. Why is that an important conclusion?

Fig. 1: directly copied from older paper. It's not clear from the text, why this figure needs to be added.

Fig. 2: Why only show ten grayscale bins, when the original data is at least 8-bit, i.e. 256 bins?

Fig. 5: The legend suggest that there are also classes/colors with half labels, e.g. 0.5, 1.5, 2.5? Is this really the case?

Fig. 6: Figure caption is out of context.

Fig. 9: Fonts are too small. ROC curves are not self-explanatory. What does "-inf vs. 1" for Berea sandstone mean? What is probability of false alarm and probability of detection? How do you derive total accuracy from the individual curves?

Interactive comment on Solid Earth Discuss., doi:10.5194/se-2016-44, 2016.