IMAG/e

TU/e Technische Universiteit Eindhoven University of Technology

Exploring the similarity of medical imaging classification problems

Veronika Cheplygina, Pim Moeskops, Mitko Veta,
Behdad Dashtbozorg, Josien P.W. Pluim

v.cheplygina@tue.nl
http://www.veronikach.com

Introduction

- You have experience with medical imaging problems A and B. Your colleague asks for advice about what classifier to use for problem C
- Your advice is based on similarity of C to A and B. How can we quantify this similarity?
- Let's represent problems A, B and C in the same feature space (meta-learning)

Method

- Assume we are given $\{(D_i, M_i)\}_1^n$, where D_i is a dataset from A, B or C and M_i is a meta-label, such as the best complex classifier.
- Represent each dataset D_i by (normalized) performances of k simple classifiers
- Embed the $n \times k$ representation in 2D. If clusters reflect labels M_i, could potentially predict best complex classifier for unseen problems

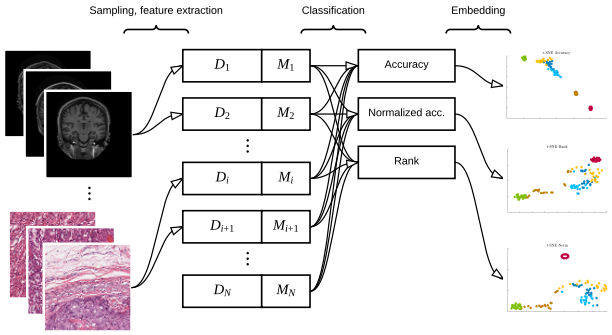

Figure 1: Overview of the method

Experiments

Dataset	Type	Images	Features
Tissue	Brain MR	20	768, CNN
Mitosis	Histopathology	12	200, CNN
MitosisNorm	Histopathology	12	200, CNN
Vessel	Retinal	20	29, Classical
ArteryVein	Retinal	20	30, Classical
Microaneurysm	Retinal	381	30, Classical

Figure 2: Six segmentation problems used in the experiments.

- Sample 20 $\times$ from 6 segmentation problems, for a total of $n = 120$ datasets. Define M_i as the segmentation problem, i.e. Tissue etc. (Ask me about the assumptions!)
- Classifiers ($k = 6$): nearest mean, {linear, quadratic} discriminant, logistic regression, 1-nearest neighbor and decision tree.
- Normalize accuracies for each of the n rows between 0 and 1, or by ranking between 1 and 6
- t-Stochastic Neighbor Embedding (t-SNE) and multidimensional scaling (MDS) for embedding

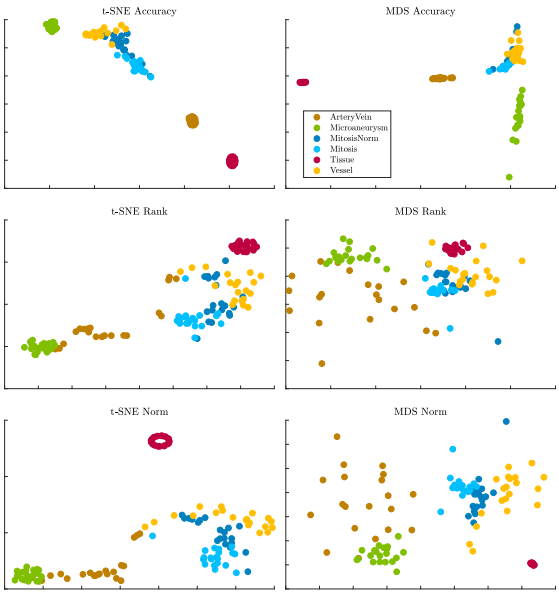

Figure 3: 2D embeddings with t-SNE and MDS for datasets from 6 classification problems

Discussion

- Clusters reflect the artificial labels M_i, representation is promising for quantifying similarity of datasets
- Further experiments needed with real M_i labels
- Feature extraction should be included in M_i, not predefined