Classification of spectra of emission-line stars using wavelet transform

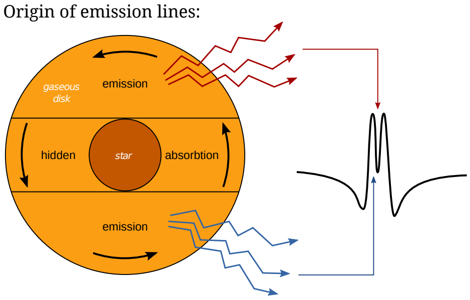
Pavla Bromová[1], David Bařina[1], Petr Škoda[2], Jaroslav Vážný[2], and Jaroslav Zendulka[1]

[1]Faculty of Information Technology, Brno University of Technology, Brno, Czech Republic
[2]Astronomical Institute of the Academy of Sciences, Ondřejov, Czech Republic

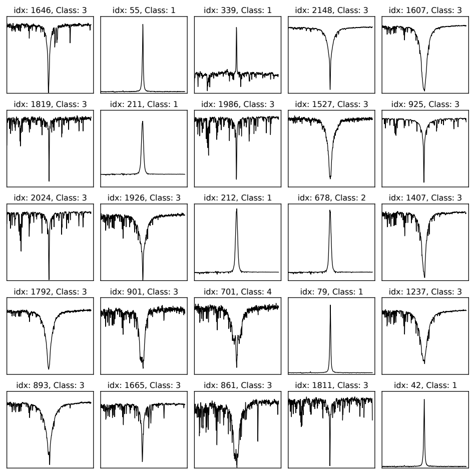
Introduction

Goal
Automatically identify emission stars spectra in large archives and classify their types based on a typical shape of their H_α emission lines.

Origin of emission lines:



Many different shapes of H_α emission lines:



Too much data
Due to the length of spectra we can't simply use all points (~2000) of each spectrum but we have to find a reduced representation of spectral features.

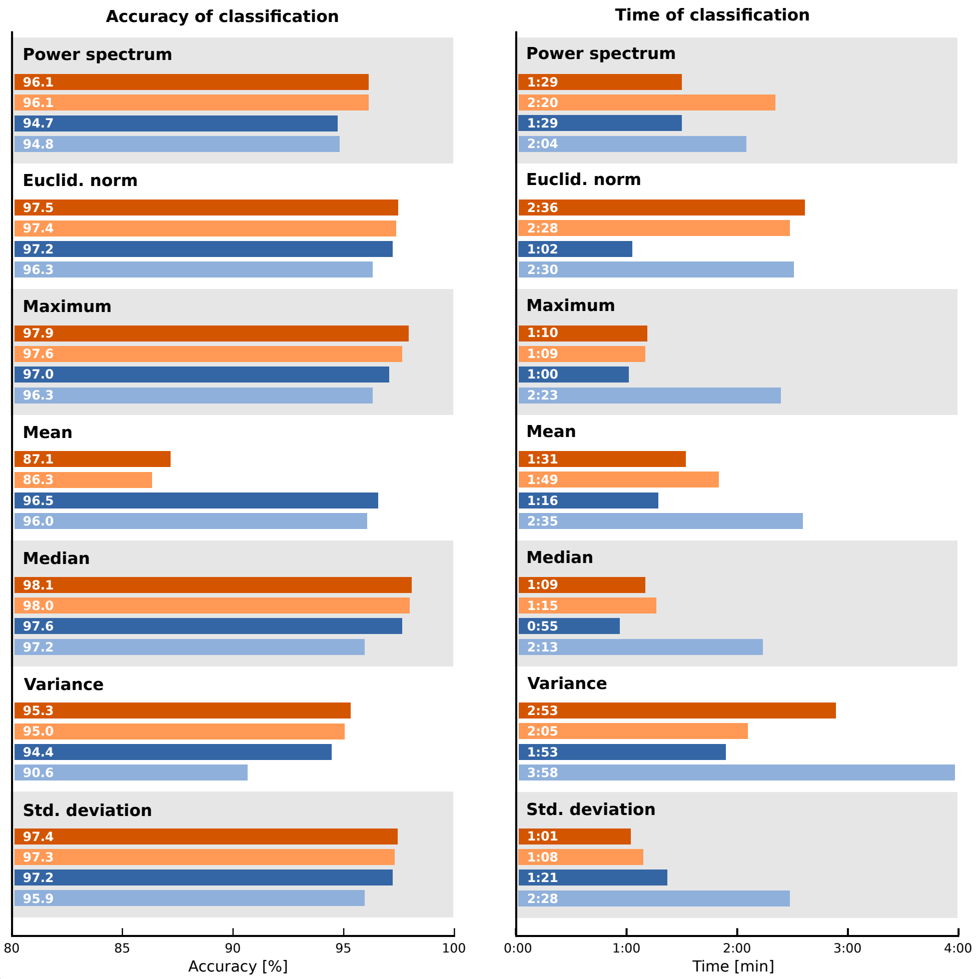
Different methods and parameters were used and compared (see Details in Methodology).

Results

Comparison of classification of data before and after feature extraction (FE):

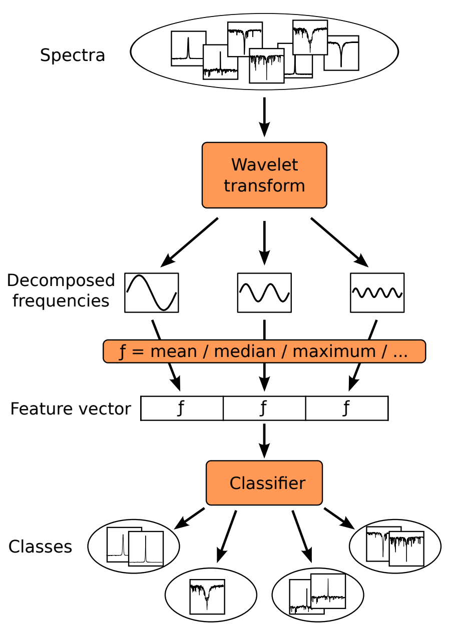
Feature vector	Length	Accuracy [%]	Time [min]
Before FE	1997	96.7	~330
After FE (median)	10	98.1	1.09

Comparison of different methods of feature extraction:



Methodology

Overview
Once the spectra are gathered from the spectrograph, the wavelet transform is applied on each spectrum, decomposing data to components of different frequencies. From each frequency band, an aggregate function (mean, median, etc.) is calculated resulting in a feature vector. All elements in one feature vector are computed using the same method. By this procedure, the length of a feature vector is reduced from ~2000 to 10 values.
The feature vector is the input of the classifier, which classifies it into one of the predefined classes representing the different shapes of emission lines.



Details
Different methods were used and compared:
1. Wavelet transform (WT):
   - Discrete (DWT)
   - Stationary (SWT)
2. In both WT, two biorthogonal wavelets were tested:
   - CDF 9/7
   - CDF 5/3
2. Different metrics as aggregate function (see Results)
3. Classification:
   - classifier: support vector machines, kernel: radial basis function
   - dataset: 2164 spectra of Be and normal stars, 4 classes: 408, 289, 1338, 129 samples
   - used 5-fold cross validation

Conclusion

Classification of the original data is very time-consuming. In (Bromová et al. 2013) we showed that classification of the original data with the same method takes ~330 min. with the accuracy 96.7%.
Proposed method reduces the number of attributes and the processing time to a small fraction and moreover increases the accuracy in many cases.

Median has the best accuracy in all cases and best processing time in case of SWT with wavelet CDF 9/7.
In overall accuracy, SWT outperforms DWT, and wavelet CDF 9/7 outperforms wavelet CDF 5/3.
In overall processing time, DWT with wavelet CDF 9/7 has the best results.

Future: 1. Compare different classifiers
2. Compare classification and clustering results
3. Find the best clustering model
4. Find new interesting stars

References

Bromová,P., Bařina,D., Škoda,P., Vážný,J., and Zendulka,J. Classification of Be stars using feature extraction based on discrete wavelet transform. Accepted to *Proceedings of the 12th annual conference Znalosti 2013*.

ibromova@fit.vutbr.cz