CLASSIFICATION OF FIBERS USING IR-SPECTRA

Karl Kaupmees, Oliver Meikar, Edgar Sepp

INTRODUCTION

Identification of textile fibers is important in industry (quality control), forensic science (identification of fibers on crime scene), but also in conservation and archaeology (identification of historical textile fibers).

Common methods for fiber identification are microscopic observation, burning test and various solubility tests. Infrared spectroscopy (IR) has many advantages for fiber identification, because it offers highly characteristic information, is easy, fast, non-destructive and relatively inexpensive.

However the analysis of IR spectra is tedious and requires a trained scientist and some peak-analyzer software. The main difficulties are spectral inhomogeneities of repeated measurements and the intrinsic similarity between the spectra of some fibers.

Hence, a better method to analyze the IR spectra of textiles is needed.

OBJECTIVES

To meet the increasing demand from the academic and industry, we aimed build a classifier that can identify fibers by their IR spectra. The initial aim is to be able to detect pure fibers at 80% probability.

For successful classification, we aimed to:
• reduce the complexity of the dataset
• engineer additional features from the dataset
• generate data normalization tools

DATASET AND METHODOLOGY

Our dataset contains of 438 IR spectra of 12 pure textile fibers both, natural and synthetic.

<table>
<thead>
<tr>
<th>Cat. count</th>
<th>Name</th>
<th>1.5</th>
<th>2.0</th>
<th>2.5</th>
<th>3.0</th>
<th>3.5</th>
<th>4.0</th>
<th>4.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.11</td>
<td>silk</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.12</td>
<td>polyacrylic</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.13</td>
<td>cotton</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.14</td>
<td>linen</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.15</td>
<td>viscose</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.16</td>
<td>polyester</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.17</td>
<td>wool</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.18</td>
<td>jute</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.19</td>
<td>elastane</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.20</td>
<td>polyamide</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.21</td>
<td>acetate</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.22</td>
<td>polystyrene</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Each spectra contains 1700 measured datapoints.

RESULTS

• Feature Engineering

Right: Correlations between original data and features.

Down: 3 randomly sampled spectra from all classes and correlations between them. Big numbers show random forest (RF) classification accuracy on test data when using only this feature.

• Models

All features, glo loc [std], ang [norm], kNN
Diff. from mean [standardized], SVM
All features, glo loc [std], ang [norm], RF
Diff. from mean [standardized], RF
All features, glo loc [std], ang [norm], SVM
Diff. from mean [standardized], SVM
All features [raw], kNN
Diff. from mean [standardized], kNN
All features [raw], RF
Diff. from mean [standardized], RF
All features [raw], SVM
Diff. from mean [standardized], SVM
Slope/angle of the graph [normalized], RF
Slope/angle of the graph [normalized], SVM
Slope/angle of the graph [standardized], RF
Slope/angle of the graph [standardized], SVM
Diff. from local mean [standardized], RF
Diff. from local mean [standardized], SVM
Diff. from global mean [standardized], RF
Diff. from global mean [standardized], SVM
Original data [standardized], RF
Original data [standardized], SVM

CONCLUSIONS

• As the amount of data is small the best performing models can be considered more or less equal in performance.
• Feature Engineering helped to improve the classification.
• Most useful feature seems to be difference from local mean.
• Feature engineering could be developed further and additional filtering of data applied to concentrate on areas with more information and thus providing better separation.
• Separating different spectra works well in general, accuracy above 80%, but there are still difficulties with more similar fibers: linen, cotton, viscose.

ACKNOWLEDGEMENTS

We would like to thank researchers Signe Vahur and Piillaar Pärts from the Cultural Heritage research group, University of Tartu, for providing the data as well as domain knowledge throughout the project.