Global versus Individual Classification of Pain. Identifying Standard and Non-Standard Mimic Types Based on Similarity of Individual to Global Decision Trees

Bachelorarbeit

im Studiengang Angewandte Informatik
der Fakultät Wirtschaftsinformatik
und Angewandte Informatik
der Otto-Friedrich-Universität Bamberg

Verfasserin: Simone Schineller Matrikelnr. 1541728
Gutachterin: Prof. Dr. U. Schmid
Contents

List of Figures v
List of Tables vii

1. Introduction 1
   1.1. State of the Art 2
   1.2. Outline 3

2. Relationship between Global Classifiers and Standard Mimic Types 5
   2.1. Standard Mimic Types of Pain 5
   2.2. Classifiers for Pain 7
   2.3. Classification by Decision Trees 9
   2.4. Finding Standard Mimic Types based on Global Decision Tree Performance 12

3. Similarity Assessment 15
   3.1. Preconditions for Similarity Assessment 15
   3.2. Properties of Proximity Measures 16
       3.2.1. Distance Measures 16
       3.2.2. Similarity Measures 17
   3.3. Methods for Similarity Assessment 17
       3.3.1. Binary Variables 18
       3.3.2. Quantitative Variables 19
   3.4. Similarity Assessment for Decision Trees 20
       3.4.1. Method 20
       3.4.2. Properties 22
       3.4.3. Distances between features 23

4. Relationship between the Similarity and the Performance of Decision Trees 25
   4.1. Data 25
       4.1.1. Decision Trees 25
       4.1.2. Multiset Representation 27
   4.2. Similarity of Global and Individual Trees 30
       4.2.1. Example 30
4.2.2. Results ................................................................. 33
4.3. Performance Measurement ....................................... 34
   4.3.1. Method .............................................................. 34
   4.3.2. Results .............................................................. 34
4.4. Comparison between Similarity and Performance .......... 35

5. Conclusion ............................................................ 39

A. Appendix .............................................................. 41
   A.1. List of features .................................................... 42
   A.2. Decision trees ..................................................... 45
   A.3. Rapidminer output for performance measurement ......... 56

References ............................................................... 59
List of Figures

2.1. Typical facial expression of pain shown by neonates .......................... 6
2.2. Annotated Reference Points ......................................................... 8
2.3. Example of a decision tree ......................................................... 9
2.4. Facial expressions of pain .......................................................... 14

3.1. Decision Trees and their representation as multisets ......................... 20

4.1. Decision Tree for individual 09 ($DT_{09}$) ...................................... 25
4.2. Global Decision Tree ($DT_{gl}$) ..................................................... 26
4.3. Reference points and horizontal and vertical distances ....................... 30
4.4. Descending sorted similarity values for the individual decision trees ... 33
4.5. Descending sorted accuracy rates of the global decision tree ............... 35
4.6. Comparison of similarity and accuracy (sorted by similarity) ............... 36
4.7. Regression line for similarity and accuracy values ........................... 37

A.1. Individual decision tree 03 .......................................................... 46
A.2. Individual decision tree 09 .......................................................... 47
A.3. Individual decision tree 13 .......................................................... 48
A.4. Individual decision tree 16 .......................................................... 49
A.5. Individual decision tree 17 .......................................................... 50
A.6. Individual decision tree 19 .......................................................... 51
A.7. Individual decision tree 20 .......................................................... 52
A.8. Individual decision tree 21 .......................................................... 53
A.9. Individual decision tree 32 .......................................................... 54
A.10. Global decision tree ................................................................. 55
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.</td>
<td>Counts of binary variables for two objects</td>
<td>18</td>
</tr>
<tr>
<td>4.1.</td>
<td>Parameter values for learning the decision trees</td>
<td>28</td>
</tr>
<tr>
<td>4.2.</td>
<td>Meaning of reference points abbreviations</td>
<td>29</td>
</tr>
<tr>
<td>4.3.</td>
<td>Meaning of shape abbreviations</td>
<td>29</td>
</tr>
<tr>
<td>4.4.</td>
<td>Similarity values for the individual decision trees</td>
<td>33</td>
</tr>
<tr>
<td>4.5.</td>
<td>Accuracy rates of the global decision tree</td>
<td>34</td>
</tr>
<tr>
<td>4.6.</td>
<td>Similarity and accuracy values</td>
<td>36</td>
</tr>
<tr>
<td>4.7.</td>
<td>Correlation values for similarity and accuracy</td>
<td>37</td>
</tr>
<tr>
<td>A.1.</td>
<td>Global decision tree performance for individual 03</td>
<td>56</td>
</tr>
<tr>
<td>A.2.</td>
<td>Global decision tree performance for individual 09</td>
<td>56</td>
</tr>
<tr>
<td>A.3.</td>
<td>Global decision tree performance for individual 13</td>
<td>56</td>
</tr>
<tr>
<td>A.4.</td>
<td>Global decision tree performance for individual 16</td>
<td>57</td>
</tr>
<tr>
<td>A.5.</td>
<td>Global decision tree performance for individual 17</td>
<td>57</td>
</tr>
<tr>
<td>A.6.</td>
<td>Global decision tree performance for individual 19</td>
<td>57</td>
</tr>
<tr>
<td>A.7.</td>
<td>Global decision tree performance for individual 20</td>
<td>57</td>
</tr>
<tr>
<td>A.8.</td>
<td>Global decision tree performance for individual 21</td>
<td>57</td>
</tr>
<tr>
<td>A.9.</td>
<td>Global decision tree performance for individual 32</td>
<td>57</td>
</tr>
</tbody>
</table>
1. Introduction

Pain is a highly unpleasant feeling, which most human beings experience during their lifetime. Because of the distress caused by pain, actions that relieve pain should be taken. To help somebody who suffers from pain, it is necessary to identify their pain. Unfortunately, this task is not always easy. This task is especially difficult in assessing the pain level of people who are not able to verbalize their pain. This inability to verbalize pain can have cognitive roots. Patients with dementia may not be capable to report their pain in a for pain relief required manner (Kunz et al., 2007). Very young children who have not yet developed the ability to speak cannot give verbal reports about their pain. Consequently, other approaches have to be applied to identify pain in persons who fail to verbalize their distress.

One possible way is to analyze the behavior of individuals who may suffer from pain. This includes the analysis of facial expressions. This approach is very promising as the facial expression provides dynamic and plastic information about a person’s feelings. Health care professionals and other persons who interact with people suffering from pain can use the facial expression of these in order to estimate their physical state. One problematic aspect of this approach is the potential subjectivity of those judgments. For instance, observers underestimate the pain of physically attractive people. Also observers’ experience in the interaction with individuals in pain can influence their estimation: People who are often exposed to pain displays tend to underestimate the distress a person suffers from (Prkachin & Craig, 1995).

The use of automatic tools for detecting pain in facial expressions avoids the described biases. There are different global and individual classifiers trained by
photographs or video sequences showing facial pain expressions. The main difference between global and individual classifiers is their training set and their application domain. Generally, global classifiers learn a concept from a training set that contains data of several persons, while individual classifiers are trained with data from a single individual. The latter can be employed to detect pain in the persons face that provided the data they were trained with. Global classifiers, however, have the purpose of detecting pain in the facial expressions shown by different individuals. Usually the classification performed by an individual classifier is better than the one performed by a global classifier but in some cases global classifiers perform better.

The main concern of this thesis is to identify the properties a person must have to be classified well by a global classifier in order to find standard mimic types. These standard mimic types could be used for the development of classifiers that can determine if different individuals actually suffer from pain.

1.1. State of the Art

One important tool for analyzing facial expressions is the Facial Action Coding Systems (FACS). Its 44 Action Units (AUs) represent the movement of one single facial muscle or a whole group of muscles. Trained FACS coders determine the activation and the activation intensity and duration of different AUs. Several studies identified 12 key AUs that are related to pain. Analyzing facial expressions using FACS is very complex and time-consuming and requires a special training of coders. For this reason, it is not applicable in clinical contexts and some scientific studies.

Therefore, less complex systems, which focus on pain related AUs, have been developed \cite{Turk2001}. One of these is the Neonatal Facial Coding System (NFCS). This system codes 10 movements of facial muscles, which are typical for the pain display of infants.

Another approach that avoids the effort of FACS coding is the automatic detection of AUs, which was implemented using Machine Learning methods. One system combines Ada Boost and Support Vector Machines. In this study 94.5\% agreement
with human FACS labels for 18 AUs were obtained \cite{Bartlett04}. There is also an approach for emotion classification. In this study several emotions were classified by a Neural Network. The classifier obtained a performance rate of 86\% at emotion classification \cite{Padgett96}.

Machine Learning methods were also applied to detect pain in facial expressions. The combination of Active Appearance Models and Support Vector Machines obtained 81.21\% accuracy \cite{Ashraf07}. Another method using Ada Boost and Support Vector Machines distinguished faked facial expressions of pain from genuine ones with 72\% accuracy, while humans’ performance was 52\% \cite{Bartlett07}.

1.2. Outline

First I will introduce the potential relationship between global classifiers and typical facial expressions for pain. In chapter 3 the similarity measure used for similarity assessment of decision trees will be explained. In chapter 4 I will describe the application of this measure. Furthermore, the similarities between global and individual decision trees will be presented and compared with performances of global decision trees. Chapter 5 will summarize the results and provide future prospects.
2. Relationship between Global Classifiers and Standard Mimic Types

A facial expression that is characteristic for a particular group of persons could be seen as standard mimic type for this group. The existence of standard mimic types for facial expressions of pain could simplify pain identification in many contexts. Health care professionals could be trained using these various types. This would improve their assessment of a patient’s pain and reduce the time needed to make this assessment. Moreover, classifiers apt to different mimic types of pain could be developed. These would probably have higher accuracy rates because of their focus on features that are relevant for the classification of the respective standard type.

2.1. Standard Mimic Types of Pain

There is the idea of a Primal Face of Pain (PFP) shown by neonates, which is universal and inborn. This expression is not influenced by developmental or sociocultural factors and is similar for different sexes and ethnicities. It consists of opening of the mouth, drawing in of the brows, closing of the eyes, and raising of the cheeks (Schiavenato et al., 2008) (cf. the expression illustrated in figure 2.1).

While neonates do not have the opportunity to learn behavioral patterns that could alter or mediate their facial expression of pain, the pain display of adults is influenced by many factors during development. Little variations in facial expressions of persons in pain may be a result of socialization in familial and different cultural contexts (Turk & Melzack, 2001). Learned coping strategies, socialization
2. Relationship between Global Classifiers and Standard Mimic Types

Figure 2.1.: Typical facial expression of pain shown by neonates

Taken from Schiavenato et al. (2008)

and cultural variation could effect the way of expressing pain (Grunau & Craig 1987). Although there are individual differences in the facial expression of adults, there are several constants. The facial display typically shown by adults experiencing pain includes brow lowering, orb tightness, levator contraction and eye closure. This display is similar to the one shown by infants. These actions may be shown independently from types of pain and are perhaps universal (Prkachin 1992). In the Western area this expression is typical for faces of humans in pain. Natives of New Guinea showed the described expression when they were asked to pose pain, so it is
possible that it is independent of cultural influences. This hypothesis is supported by the fact that facial displays of different emotions such as happiness or anger are quite constant across cultures. But as pain is not an emotion in the truest sense of emotion this topic needs further examination (LeResche, 1984). Facial expression of pain is probably not only independent of culture, but also sex independent. Women show stronger facial expression of emotions in general. However, in one study the comparison between facial pain displays of men and women did not find any significant differences (Kunz et al., 2006).

Taking all this into consideration one could think of different standard mimic types, which are similar to the described universal expression of pain. These types would then be a result of several factors influencing the innate expression shown by neonates.

2.2. Classifiers for Pain

All studies about pain classification that were mentioned in section 1.2 only made use of global classifiers and at most examined three different Machine Learning methods. In a study conducted by Siebers et al. (2009) the suitability of global and individual classifiers for the automatic detection of pain were investigated. The result was that most classifiers (Decision Tree, Support Vector Machine, Linear Regression, Naïve Bayes and k-nearest Neighbors) that were examined are suitable for pain classification. Individual classifiers performed better for most classifiers. As the dataset only contained 543 entries, it is not sure if these results are representative (Siebers et al., 2009).

Siebers (2011) described several Machine Learning methods for global and individual automated pain detection in facial expressions. These are Decision Trees, Naïve Bayes, Support Vector Machines, Perceptrons, Artificial Neural Networks, Classification by Linear Regression, AdaBoost, Bagging and Stacking. The raw data used in this study were image stills of videos taken during an experiment, in which pressure stimuli were applied to the upper edge of subjects' trapezius muscle. The experi-
A global classifier for a subject was learned from data from all other eight subjects (i.e. the global classifier for individual 09 was learned from the data of the individuals 1-8). Individual classifiers were learned from the data of each individual.

One result of this study was conclusion that all examined learning algorithms
can be used to classify facial expressions of pain but that the performance of the different classifiers varies. Furthermore, the results do not confirm the assumption that individual classifiers always perform better than global ones.

The fact that sometimes the global classifier for an individual has a higher accuracy than the corresponding individual classifier is remarkable as the training data used for the learning of an individual classifier are more specific for the respective individual. The global perceptron for one individual obtains 86.7% average accuracy while the individual classifier obtains only 71.4%, for example. The question arises if the facial expression of persons with higher accuracy of global classifiers could serve as a kind of standard pain expression. This question will be tackled by the examination of decision trees since this type of classifier is also comprehensible to persons who know few or nothing about Machine Learning methods.

2.3. Classification by Decision Trees

Figure 2.3.: Decision tree for the concept “Play Tennis”

Content taken from Mitchell, 1997

A decision tree is a type of classifier that is applied for many classification problems. It is especially suited for problems that fulfill the following properties.

I) Representation of instances by attribute-value pairs
II) Discrete valued output function

III) Training data may contain errors

IV) Training data may contain missing attribute values (Mitchell, 1997)

Decision trees are a possible way to classify pain based on images of facial expressions: The instances can be represented by the values of several properties of the face, such as the distance between the eyes (I). Since the classification task is to determine if a person is suffering from pain (pain = 1) or not (pain = 0) the target function has a discrete output value (II). There is also the possibility that some training examples are incorrect. This could be caused by persons who give an incorrect report about their pain (III). Missing attribute values can be a result of unsuitable images. The brows of a person could be covered by hair, for example (IV).

A decision tree represents the attributes of a training set by nodes. The branches of the tree represent the different values that an attribute can take. The decision tree in figure 2.3 for example, represents the attributes “Outlook”, “Humidity” and “Wind”. An instance is classified by starting at the root node. It is checked which branch represents the value the instance has for the attribute of the root node and the appropriate path is followed. This process is repeated for every subtree along the path until a leaf node is reached. Every path in a tree can be seen as a conjunction of attribute tests. The rightmost path of the tree for the concept “Play Tennis” represents the conjunction \( (Outlook = Rain \land Wind = Weak) \). The tree itself is a disjunction of conjunctions. The tree shown in figure 2.3 could be expressed by:

\[
(Outlook = Sunny \land Humidity = Normal) \lor (Outlook = Overcast) \lor (Outlook = Rain \land Wind = Weak)
\]

The main idea of decision trees is to place attributes that alone classify the training data best at the root. For each value this attribute can take, a branch is created and the training examples are sorted to one branch according to the value they take for the root attribute. The described process is repeated for every leaf node until all
attributes are part of the path that leads to this leaf or until all training examples associated with this leaf are members of the same class. A statistical measure called Information Gain is used to determine how well an attribute classifies the training set. The Information Gain is the expected reduction of the impurity of a collection of examples. The impurity, also called Entropy, is 1 if the collection contains an equal number of examples for each value of the target function. Its value is 0 if all examples have the same class.

A tree that overfits the training examples can be produced by the described algorithm if the training set is too small to be representative or if it contains noisy data. A hypothesis overfits the training examples if the accuracy for the training examples is higher than the accuracy for all instances (i.e. also instances that do not belong to the training set). There are two different types of approaches that avoid overfitting:

- stop the growing of the decision tree before it classifies the instances of the training set perfectly
- allow the production of an overfitting tree but post-prune it

As the second method is more successful in avoiding overfitting, a short description of two approaches that implement pruning are given (Mitchell, 1997):

1. Reduced error pruning: Every node of the tree is a potential candidate for pruning but nodes are only removed if the resulting tree does not perform worse over the set that is used for testing the performance. A node is removed by performing three steps. First, the subtree that is rooted at this node is removed. Then the respective node is made a leaf node. Finally, the node is assigned the most common classification of the training examples associated with this node.

2. Rule post-pruning: This approach consists of four steps. First, a decision tree that classifies the training data as well as possible and that might overfit the
training set is produced. In the next step the tree is converted into a set of rules by creating one rule for each path from the root node to a leaf node. Then every rule is generalized by removing preconditions that improve its estimated accuracy. Finally, the rules are sorted by their estimated accuracy. The rules are considered in this sequence for the classification of subsequent instances. (Mitchell, 1997)

2.4. Finding Standard Mimic Types based on Global Decision Tree Performance

A group of individuals who can be classified better by a global classifier than others should have the same standard mimic type. So, there could be groups of individuals sharing the same standard facial expression of pain. The individual classifiers for these persons should be more similar to the respective global classifiers than the individual classifiers of those persons for whom the performance of the individual classifier is worse. In this case the global decision tree and the individual decision tree should have more features in common. One could use these features for sorting individuals into different groups of standard mimic types.

If a global decision tree uses many features containing attributes of the region around the eyes, it could be reasoned that individuals who show strong facial activity in this region belong to the standard type of this tree. The main idea of this approach is to find a decision tree for each standard mimic type. If the expression of a new individual has to be classified, it has to be determined which standard mimic types the person shows. The next step would be the classification of the facial expression by the global decision tree belonging to the identified standard mimic type.

Figure 2.4a shows one pain display of the person whose data were classified best by the global decision tree that was learned from the data of the study described in section 2.1. This global decision tree was learned from all data of all nine individuals. For the individual shown in Figure 2.4a the global decision tree has an accuracy
of 95.0\%. Furthermore, this facial display is similar to the universal expression described above: The person shows orbit tightening, levator contraction and eye closure. As the brows are not completely visible, a statement about this part of the face cannot be made. Before calculating the similarity between the individual decision trees and the global decision tree an subjective estimation of the similarity between the expressions of the other eight individuals and the expression shown in figure 2.4a was made. For each individual one image, which was subjectively assessed as most intensive, was chosen. This images are ordered by descending similarity. This means figure 2.4b shows the expression that is most similar to the standard mimic type while Figure 2.4i shows the expression that is least similar.

One criterion for assessing the similarity of the facial expressions was the activation of facial areas. A facial expression that includes closing the eyes, for example, is considered to be more similar to the expression of individual 09. Also the intensity of a facial expression was taken into account. Since the expression shown in figure 2.4a is very intensive, while the expression of individual 20 looks neutral, the similarity between those two individuals is considered as relatively low.

Based on the similarity of facial expression and the performance of a global classifier the following question arises: Do the performance of the global decision tree for an individual and the similarity between the individual decision tree of this individual and the global decision tree correlate? This question will be tackled by comparing the similarity of one global decision tree with the corresponding performance rates this tree obtains for different individuals.
2. Relationship between Global Classifiers and Standard Mimic Types

Figure 2.4.: Facial expressions of pain
3. Similarity Assessment

Measures that quantify the similarity or distance (dissimilarity) between objects are called proximity measures. A similarity measure quantifies the similarity between two objects. The more similar the objects are, the higher the value is that is calculated by the similarity measure. A distance measure, however, indicates how dissimilar two objects are. Its value is the higher, the more dissimilar two objects are (Backhaus 2000).

3.1. Preconditions for Similarity Assessment

The main precondition for the use of a similarity measure is that pairs of objects can be ordered weakly regarding the similarity between objects. This means that for every four objects $o_i$, $o_j$, $o_k$ and $o_l$ which are elements of the considered set of objects $A$ a relation $(o_i, o_j) \succsim (o_k, o_l)$ exists. This relation indicates that the similarity between $o_i$ and $o_j$ is equal to or greater than the similarity between $o_k$ and $o_l$. The relation $\succsim$ has to fulfill the following properties:

- $\succsim$ is connex
- $\succsim$ is reflexive
- $\succsim$ is transitive
- For two pairs of objects of $A$ there has to be a pair of objects of $B$, with $B \subseteq A \times A$, that lies regarding to $\succsim$ between these pairs.
These conditions allow the weak ordering of a set of pairs of objects. Though, they can also be properties of relations that are used for the comparison of data different from similarity values. The expression \((o_i, o_j) \succeq (o_k, o_l)\) could, for instance, also mean that the sum of weights of \(o_i\) and \(o_j\) is as high or higher than the sum of the weights of \(o_k\) and \(o_l\). For this reason \(\succeq\) should have two more properties:

- \((o_i, o_j) \succeq (o_i, o_j)\)

- \((o_i, o_j) \succeq (o_j, o_i)\) and \((o_i, o_i) \succeq (o_j, o_j)\)

\(\text{[Eckes \\& Rossbach 1980]}\)

### 3.2. Properties of Proximity Measures

Since the idea of distance measures is closely related to similarity measurement, first of all the properties of this type of proximity measure are given.

#### 3.2.1. Distance Measures

A function \(d\) that calculates a real number \(d_{ij}\) between two objects \(o_i\) and \(o_j \in A\) is called distance function iff for all \(o_i, o_j, o_k \in A\) the following properties hold:

- \(d_{ij} \geq 0\): The distance between two objects is never negative.
- \(d_{ij} = 0\), iff \(o_i = o_j\): The distance between two objects is 0 if these objects are identical.
- \(d_{ij} = d_{ji}\): The distance between \(o_i\) and \(o_j\) equals the distance between \(o_j\) and \(o_i\).

If the function also fulfills the condition

- \(d_{ij} \leq d_{ik} + d_{jk}\): The distance between \(o_i\) and \(o_j\) is not greater than the sum of the distances between \(o_i\) and \(o_k\) and \(o_j\) and \(o_k\).

It is called metric. A distance function \(d\) is called ultrametric if the property
\[ d_{ij} \leq \max(d_{ik}, d_{jk}) \]

(Eckes & Rossbach, 1980) holds for its distance values. For an ultrametric distance function the distance between \( o_i \) and \( o_j \) is not greater than the maximum of the distance between \( o_i \) and \( o_k \) and \( o_j \) and \( o_k \).

### 3.2.2. Similarity Measures

A function \( s \) that assigns a real number \( s(o_i, o_j) \) (or shorter \( s_{ij} \)) to a pair of objects \( o_i, o_j \in A \) is called similarity function iff for all \( o_i, o_j, o_k \in A \) the following properties hold:

- \( s_{ij} \leq 1 \): The similarity between \( o_i \) and \( o_j \) is a maximum of 1.
- \( s_{ij} = 1 \) iff \( o_i = o_j \): The similarity between two objects is 1 if these objects are identical.
- \( s_{ij} = s_{ji} \): The similarity between \( o_i \) and \( o_j \) equals the similarity between \( o_j \) and \( o_i \).

If the function also fulfills the condition

- \( s_{ij} \geq s_{ik} * s_{jk} \): The similarity between \( o_i \) and \( o_j \) is equal to or greater than the product of the similarities between \( o_i \) and \( o_k \) and \( o_j \) and \( o_k \).

it is a metric similarity function. Ultrametric similarity functions have an additional property:

- \( s_{ij} \geq \min(s_{ik}, s_{jk}) \): The similarity between \( o_i \) and \( o_j \) is equal to or greater than the minimum of the similarities between \( o_i \) and \( o_k \) and \( o_j \) and \( o_k \).

(Eckes & Rossbach, 1980)

### 3.3. Methods for Similarity Assessment

The choice of a similarity measure depends on the type of the objects for which similarity values should be calculated. In the following different approaches for
3. Similarity Assessment

Table 3.1: Counts of binary variables for two objects

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>a+b</td>
<td>a+c</td>
</tr>
<tr>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>c+d</td>
<td>b+d</td>
</tr>
</tbody>
</table>

Table 3.1.: Counts of binary variables for two objects

Assessing the similarity of objects with binary and quantitative variables will be
given, in order to provide an idea of how the similarity between objects can be
assessed.

3.3.1. Binary Variables

Binary variables can only take two values. These values can either indicate the
presence or absence of a quality (values 0 or 1) or two equivalent characteristics such
as “high/low”. Table 3.1 shows a 2 x 2 table, which can be used to indicate how
many attributes two objects \( o_i \) and \( o_j \) have in common. Cell \( a \) gives the number
of attributes that \( o_i \) and \( o_j \) have in common, whereas cell \( d \) gives the number of
attributes that neither \( o_i \) nor \( o_j \) has. Analogical, the cells \( c \) and \( d \) give the numbers
of disagreements. There are two coefficients based on these quantities that are
commonly used. The matching coefficient \( \frac{a+d}{p} \) calculates the ratio of the number of
the matching variables to the number of all variables. This coefficient is equivalent
to the accuracy of classifiers since it can also be interpreted as ratio of correct
classifications to the whole number of classified instances. Jaccard’s coefficient \( \frac{a}{a+b+c} \)
calculates the same but ignores the number of negative matches \( d \). Not considering
\( d \) is reasonable when the presence of attributes is used for similarity assessment. It
seems unreasonable to rate the similarity of two objects high because there are many
attributes that they both do not have.

If variables can have more than two values, a score \( s_{jk} \) is allocated to each variable
3. Similarity Assessment

$k$, which indicates if the objects $o_i$ and $o_j$ have the same value for the respective variable. The sum of the scores is averaged by division by $p$ to obtain a value between 0 and 1: (Everitt 1993)

$$\frac{\sum_{k=1}^{p} s_{ijk}}{p}$$

3.3.2. Quantitative Variables

Quantitative variables can take real numbers as values. One approach for the assessment of the similarity between objects $o_i$ and $o_j$ with quantitative variables is Pearson’s product moment correlation coefficient. The calculation produces an “average variable value” for each object by averaging over the value of the different variables:

$$s(i, h) = \frac{\sum_{j=1}^{h} (x_{ij} - \bar{x}_i)(x_{hj} - \bar{x}_h)}{\left(\sum_{j=1}^{h} (x_{ij} - \bar{x}_i)^2\right)\left(\sum_{j=1}^{h} (x_{hj} - \bar{x}_h)^2\right)^{\frac{1}{2}}}$$ (Deimer, 1986)

The coefficient has been criticized for not working in practice by some authors. It was not able to separate four relatively distinctive groups which belonged to a data set that had been constructed from a bivariate normal population. The correlation index was also successfully applied for similarity assessment, though. It was used to measure similarity in a study that applied various clustering techniques to psychiatric data. Pearson’s product moment correlation coefficient was able to recover five groups in an artificially generated data set of hundred individuals measured on 84 variables (Everitt 1993). Another method for the similarity assessment of quantitative variables is the use of distance measures. Deimer (1986) describes several measures (e.g. Manhattan Distance, Euclidean Distance, Mahalanobis Distance, etc.) that are suitable for this problem (Deimer 1986).
3. Similarity Assessment

(a) Tree 1 and its multiset of nodes

Class K

Class L

Class N

≤ 1.5 ≤ 1.5

> 2.5 ≤ 2.5

≤ 3.5 > 3.5

Class M

Class N

MS₁={1, 2, 1, 4}

(b) Tree 2 and its multiset of nodes

Class K

≤ 1.5 > 1.5

≤ 2.5 ≤ 2.5

> 2.5 ≤ 3.5

Class L

Class M

Class N

MS₂={1, 2, 3, 4, 5}

Figure 3.1.: Decision Trees and their representation as multisets

3.4. Similarity Assessment for Decision Trees

3.4.1. Method

Since none of the described similarity measures seemed to be apt for assessing the similarity between decision trees, a similarity measure that is based on the minimal distance between trees has been developed.

Generally speaking, this measure calculates the minimal distance between two decision trees by summing up the minimal distances between the nodes of the respective trees. The nodes’ positions in the trees are ignored, as well as the labels of the path, which are equivalent to the values of the variables represented by the nodes. The position of a node is not considered in this case as the size of the examined decision trees does not vary enough. For this reason the different positions of nodes are not very meaningful. The labels of the paths are ignored since they are dependent on the proportions of an individual’s face. This is why they are different for each individual. Roughly speaking, the used distance measure only considers the
number of nodes and the attributes, which are represented by the nodes.

To find the minimal distances between the nodes of two trees the decision trees are represented as multisets of nodes, which can contain an element several times. Multisets are used as one label of a node can occur more than one time in a tree. Figure 3.1 shows two decision trees and their corresponding representation as multisets.

In the following these multisets are used in order to demonstrate how the minimal distances between the nodes of two decision trees are obtained.

Before calculating the minimal distance between the elements of two multisets one has to determine which multiset is the smaller on. This simplifies the process of finding the minimal distances between the multisets’ elements. The multiset that contains less elements is named \( A \), while the other one is named \( B \). If the sizes of the multisets are equal the sets can be named \( A \) and \( B \) arbitrarily. In this example \( A = MS_1 \) and \( B = MS_2 \) as \( |MS_1| = 4 \) and \( |MS_2| = 5 \). Then for each element of \( A \) the minimal distance is determined by finding an element in \( B \) for which the distance is minimal.

1. \( A = \{1, 2, 1, 4\} \quad B = \{1, 2, 3, 4, 5\} \)
   \[ d_{1,1} = 0 \]

2. \( A = \{2, 1, 4\} \quad B = \{2, 3, 4, 5\} \)
   \[ d_{2,2} = 0 \]

3. \( A = \{1, 4\} \quad B = \{3, 4, 5\} \)
   \[ d_{1,3} = 2 \]

4. \( A = \{4\} \quad B = \{4, 5\} \)
   \[ d_{4,4} = 0 \]

As \( B \) contains more elements than \( A \), one element of \( B \) is left. In order to account for the different number of nodes that trees can have, for each value that is left the maximal distance that two elements can have is added: \( (|B| - |A|) \times \text{max} \). The expression \( |B| - |A| \) gives the number of left elements, while \( \text{max} \) is the maximal
distance. The idea of this procedure is to “penalize” differences in the size of decision trees as trees that differ much in size are considered to be more dissimilar than trees that consist of an equal number of nodes. Therefore, the distance between \( MS_1 \) and \( MS_2 \) is:

\[
d_{1,1} + d_{2,2} + d_{1,3} + d_{4,4} + (5 - 4) \times max
\]

Generally, the distance between two multisets \( A \) and \( B \) is:

\[
d_{A,B} = \sum_{i=1}^{\|A\|} mindistance(a_i, b) + (\|B\| - \|A\|) \times max
\]

This sum is normalized by dividing it by the maximal distance that two multisets or rather decision trees can have to obtain distance values between 0 and 1. The distance between two trees is maximal if the distance between all elements of \( A \) and \( B \) is maximal. As \( B \) is the larger set and the different size of trees is penalized by adding the maximal distance between to nodes for each element of \( B \) that is left, the maximal distance between two trees is \( \|B\| \times max \).

Hence, the minimal distance \( mind \) between two multisets or rather decision trees \( A \) and \( B \) is:

\[
mind_{AB} = \frac{\sum_{i=1}^{\|B\|} mindistance(a_i, b) + (\|B\| - \|A\|) \times max}{\|B\| \times max}
\]

As the values for this distance measure lie between 0 and 1, the corresponding similarity values can be obtained by subtracting the distance values from 1: \( Everitt \) 1993:

\[
s_{ij} = 1 - mind_{ij}
\]

3.4.2. Properties

The described similarity measure has the following properties:

- \( s_{ij} \leq 1 \): As the minimal distance “\( mind \)” between two decision trees is normalized by division by the maximal distance between two trees, this value lies between 0 and 1. The similarity \( s_{ij} \) is calculated with \( 1 - d_{ij} \). So \( s_{ij} \) can be at most 1 if \( d_{ij} = 0 \).
3. Similarity Assessment

- $s_{ij} = 1$ iff $t_i = t_j$: If two decision trees $t_i$ and $t_j$ are identical, the minimal distance for each node is 0. The result of $|B| \times \max$ is also 0 in this case. As a consequence of this, the minimal distance between identical trees is 0. Hence, $s_{ij} = 1 - 0 = 1$.

- $s_{ij} = s_{ji}$: The similarity between two trees $t_i$ and $t_j$ equals the similarity between the trees $t_j$ and $t_i$ since the similarity measure is not dependent on the size of the trees. The smaller tree is always named $A$ and the larger tree is named $B$, so the similarity measure is symmetric.

3.4.3. Distances between features

The similarity values between the global and the nine individual decision trees were calculated using the described approach. Each tree was represented as multiset of features and each feature was represented as multiset of reference points. Hence, it was necessary to find reasonable distances between reference points as this similarity measure makes use of distances between elements of multisets and as the respective elements are reference points in this case.

One possible distance between reference points could be the Euclidean distances between the points in the schema (cf. fig 2.2). This approach has two disadvantages: First, it would be very time-consuming and error-prone to measure and store all distances. Second, what is more important, the area a point belongs to would be ignored. The distance between “m2” (point 2 of the mouth) and “n5” (point 5 of the nose), for example, is smaller than the distance between “m2” and “m0” (point 0 of the mouth). Hence, the fact that point “m2” and “m0” are part of the same facial area would not be included in the calculation.

The activity of facial areas plays a great role when detecting pain in facial expressions. The distances between reference points should reflect this by assigning greater distances to pairs of points of different facial areas. Therefore, these distance values are used:

- identical points: distance = 0
• points of the same facial area: distance = 0.5

• points of different facial areas: distance = 1

These distance values fulfill the properties described in section 3.2.1. The fact that the values lie between 0 and 1 allows these distances to be used for similarity assessment. (cf. section 3.4). The distance between two points is maximal if they lie in different facial areas and the distance value 1 is assigned to such a pair of points.

Based on these distance values the distances between features are calculated. First the features are represented as multisets of points. In the next step the smaller multiset is determined. This multiset is named $A$, while the other multiset representation is named $B$. Then for each element of $A$, or rather for each coordinate, one element of $B$, for which the distance to the respective element of $A$ is minimal, was determined. The respective elements are removed from $A$ and $B$. After determining the minimal distance for each element of $A$ these distances are summed up. For each element of $B$ that is left the maximal distance between two coordinates, i.e. 1, is added to this sum. The result of this computation is divided by the maximal distance between two features, i.e. $|B|*1$, in order to obtain distance values between 0 and 1.
4. Relationship between the Similarity and the Performance of Decision Trees

The similarity values of nine individual decision trees to one global tree were calculated using the similarity measure described in section 3.4. These values were compared with the accuracy rates the global decision tree obtained for each of the nine individuals.

4.1. Data

4.1.1. Decision Trees

The data that were used by Siebers (2011) served as training set for learning the nine individual decision trees. This data set contains 538 pain and non-pain
4. Relationship between the Similarity and the Performance of Decision Trees

Figure 4.2.: Global Decision Tree ($DT_{gl}$)

training examples. One tree for each individual was learned from the examples of the respective individual. The decision tree in Figure A.2, for instance, was learned from the data of individual 09. The other eight decision trees can be found in the Appendix in section A.2. The global decision tree (cf. Figure 4.2) was learned from the whole data set, i.e. from all 538 entries.

RapidMiner 5.1.006 was used for learning the decision trees. The used operator is named “Decision Tree”. The algorithm employed is similar to the C4.5 Algorithm that was developed by Quinlan (1993). Generally, the attribute that maximizes the discriminative power of a node is chosen. The measure that determines this power can be chosen by the user. The algorithm stops in one of the following cases:

- No attribute reaches the “minimal gain”
- The maximal depth of the tree is reached
The number of examples of a subtree is less than “minimal size for split”

The algorithm also includes pruning the tree in order to avoid overfitting. Several values of parameters can be set. These are:

- **criterion**: The measure that is used for determination of the descriptive power of an attribute. (Information Gain, Gain Ratio, Gini Coefficient)
- **minimal size for split**: The minimal number of instances available at a new node that allows a split.
- **minimal leaf size**: The minimal number of examples per leaf.
- **minimal gain**: Precondition that determines the minimal gain value needed for a split.
- **maximal depth**: The maximum depth of a decision tree.
- **confidence**: The confidence level that is used for the determination of pruning.
- **number of prepruning alternatives**: The number of trials for prepruning an alternative node if prepruning the original node prevented a split.
- **no prepruning**: Inhibits prepruning a tree.
- **no pruning**: Inhibits pruning.

The values of these parameters, which were taken from [Siebers 2011], are listed in table 4.1.

### 4.1.2. Multiset Representation

The examined decision trees were represented as multisets of features. This means that every node of a decision tree is an element of its corresponding multiset. The
Table 4.1.: Parameter values for learning the decision trees

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>criterion</td>
<td>Information Gain</td>
</tr>
<tr>
<td>minimal size for split</td>
<td>1</td>
</tr>
<tr>
<td>minimal leaf size</td>
<td>1</td>
</tr>
<tr>
<td>minimal gain</td>
<td>0.0</td>
</tr>
<tr>
<td>maximal depth</td>
<td>-1</td>
</tr>
<tr>
<td>confidence</td>
<td>0.01</td>
</tr>
<tr>
<td>number of prepruning</td>
<td>3</td>
</tr>
<tr>
<td>no prepruning</td>
<td>disabled</td>
</tr>
<tr>
<td>no pruning</td>
<td>disabled</td>
</tr>
</tbody>
</table>

global decision tree, for example, is represented by the following multiset:

\[
\{iz\_d\_re2\_re6, \\
ssmd\_diff\_diff\_a\_re2\_rec\_re4\_a\_re4\_rec\_re6\_\_diff\_a\_le2\_lec\_le4\_a\_le4\_lec\_le6, \\
iz\_diff\_a\_m0\_mv\_m1\_a\_m2\_mv\_m4, issmd\_diff\_a\_m2\_mc\_m5\_a\_m5\_mc\_m7, \\
z\_diff\_frac\_d\_m0\_n03\_d\_m0\_n04\_frac\_d\_m4\_n07\_d\_m4\_n06\}
\]

Every feature of this multiset is in turn represented as multiset of reference points. This means that a features type, such as “diff” (difference), is ignored. This approach seemed to be reasonable as the features are very specific. Many features that have different names express similar properties. For this reason only reference points (cf. figure 4.3) were taken into account. For example the reference point “c11” is point 11 of the contour of a face. The meanings of the reference points’ and shapes’ abbreviations are listed in table 4.2 and 4.3 (Siebers, 2011).

For example, the multiset representing the feature “iz\_d\_re2\_re6 ” is:

\{re2, re6\}
### Table 4.2.: Meaning of reference points abbreviations

<table>
<thead>
<tr>
<th>abbreviation</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>lb0–lb4</td>
<td>left eyebrow</td>
</tr>
<tr>
<td>rb0–rb4</td>
<td>right eyebrow</td>
</tr>
<tr>
<td>le0–le7</td>
<td>left eye</td>
</tr>
<tr>
<td>re0–re7</td>
<td>right eye</td>
</tr>
<tr>
<td>n00–10</td>
<td>nose</td>
</tr>
<tr>
<td>m0–m7</td>
<td>mouth</td>
</tr>
<tr>
<td>c0–c12</td>
<td>face’s contour</td>
</tr>
</tbody>
</table>

### Table 4.3.: Meaning of shape abbreviations

<table>
<thead>
<tr>
<th>abbreviation</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>leh</td>
<td>horizontal center between le0 and le4</td>
</tr>
<tr>
<td>reh</td>
<td>horizontal center between re0 and re4</td>
</tr>
<tr>
<td>mh</td>
<td>horizontal center between m0 and m4</td>
</tr>
<tr>
<td>lev</td>
<td>vertical center between le2 and le6</td>
</tr>
<tr>
<td>rev</td>
<td>vertical center between re2 and re6</td>
</tr>
<tr>
<td>mv</td>
<td>vertical center between m2 and m6</td>
</tr>
<tr>
<td>lec</td>
<td>intersection between the lines $\overline{04}$ and $\overline{26}$ of the left eye</td>
</tr>
<tr>
<td>rec</td>
<td>intersection between the lines $\overline{04}$ and $\overline{26}$ of the right eye</td>
</tr>
<tr>
<td>mc</td>
<td>intersection between the lines $\overline{04}$ and $\overline{26}$ of the mouth</td>
</tr>
</tbody>
</table>
4.2. Similarity of Global and Individual Trees

4.2.1. Example

The calculation of the similarity between the global decision tree and the individual decision tree of individual 09 exemplifies how the described measure was applied.

Multiset representation of the global decision tree

\[ DT_{gl} = \{ F_{gl1} = \{ re2, re6 \}, F_{gl2} = \{ re2, rec, re4, rec, re6, le2, le4, le6, lec, le6 \}, F_{gl3} = \{ m0, mv, m1, m2, mv, m4 \}, F_{gl4} = \{ m2, mc, m5, m5, mc, m7 \}, F_{gl5} = \{ m0, n03, m0, n04, m4, n07, m4, n06 \} \} \]

Multiset representation for the individual decision tree 09

\[ DT_{i09} = \{ F_{i091} = \{ re2, rec, re4, rec, re6, le2, lec, le4, le6 \}, F_{i092} = \{ m0, n03, m0, n04, m1, n07, m4n06 \}, F_{i093} = \{ m1, m7, m2, m6, m3, m5, m2, m6 \} \} \]
1. Finding the larger decision tree multiset

\[ |DT_{gl}| = 5 \quad |DT_{i09}| = 3 \]

\[ \Rightarrow A = DT_{i09} \quad B = DT_{gl} \]

2. Finding minimal distance for each feature in tree \( A \)

To find the minimal distance for each feature in \( A \), every element of \( A \) is compared with the elements of \( B \). A distance value \((0, 0.5, 1)\) is assigned to each pair. For each element in \( A \), one pair with minimal distance is chosen. All pairs that contain the corresponding element of \( B \) are removed.

- \( F_{i09}, F_{gl} \)
  a) Finding the larger feature
  
  \[ |F_{i09}| = 12 \quad |F_{gl}| = 2 \]
  
  \[ \Rightarrow A = F_{gl} \quad B = F_{i09} \]

  b) Finding the minimal distance for each reference point in feature \( A \)

| \( F_{gl} \)  | re2 | re6 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| \( F_{i09} \) | re2 | re6 | rec | re4 | re4 | rec | le2 | lec | le4 | le4 | lec | le6 |                      |
| distances    | 0   | 0   | -   | -   | -   | -   | -   | -   | -   | -   | -   | -   |                      |

\[ \sum 0 \]

\( \text{mind}_{AB} = \sum_{i=1}^{[B]} \text{mindistance}(a_i, b_i) + ([B] - |A|) \times \max \]

\[ \text{mind}_{F_{gl}, F_{i09}} = 0 + 10 \times 1 = 10 \]

\[ \text{mind}_{F_{gl}, F_{i09}} = \frac{0 + 10 \times 1}{12 \times 1} = 0.834 \]

- \( \text{mind}_{F_{gl}, F_{i09}} = 0 \)
- \( \text{mind}_{F_{gl}, F_{i09}} = 1 \)
- \( \text{mind}_{F_{gl}, F_{i09}} = 1 \)
- \( \text{mind}_{F_{gl}, F_{i09}} = 1 \)
4. Relationship between the Similarity and the Performance of Decision Trees

Minimal distances for each feature in tree A

<table>
<thead>
<tr>
<th></th>
<th>$F_{i09_1}$</th>
<th>$F_{i09_2}$</th>
<th>$F_{i09_3}$</th>
<th>$F_{i09_4}$</th>
<th>$F_{i09_5}$</th>
<th>$F_{i09_6}$</th>
<th>$F_{i09_7}$</th>
<th>$F_{i09_8}$</th>
<th>$F_{i09_9}$</th>
<th>$F_{i09_{10}}$</th>
<th>$F_{i09_{11}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$DT_{i09}$</td>
<td>$F_{gl_1}$</td>
<td>$F_{gl_2}$</td>
<td>$F_{gl_3}$</td>
<td>$F_{gl_4}$</td>
<td>$F_{gl_5}$</td>
<td>$F_{gl_6}$</td>
<td>$F_{gl_7}$</td>
<td>$F_{gl_8}$</td>
<td>$F_{gl_9}$</td>
<td>$F_{gl_{10}}$</td>
<td>$F_{gl_{11}}$</td>
</tr>
<tr>
<td>distances</td>
<td>0</td>
<td>0</td>
<td>0.4375</td>
<td>-</td>
<td>-</td>
<td>$\sum$ 0.4375</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Summing the minimal distances up and normalizing the sum

$$mindist_{DT_{i09}, DT_{gl}} = \frac{0 + 0 + 0.4375 + 2 \times 1}{5} = 0.4875$$

$$\Rightarrow sim_{DT_{i09}, DT_{gl}} = 1 - 0.4875 = 0.5125$$
4. Relationship between the Similarity and the Performance of Decision Trees

<table>
<thead>
<tr>
<th>decision tree</th>
<th>$DT_{i03}$</th>
<th>$DT_{i09}$</th>
<th>$DT_{i13}$</th>
<th>$DT_{i16}$</th>
<th>$DT_{i17}$</th>
<th>$DT_{i19}$</th>
<th>$DT_{i20}$</th>
<th>$DT_{i21}$</th>
<th>$DT_{i32}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>similarity</td>
<td>.39</td>
<td>.51</td>
<td>.40</td>
<td>.30</td>
<td>.41</td>
<td>.47</td>
<td>.32</td>
<td>.22</td>
<td>.27</td>
</tr>
</tbody>
</table>

Table 4.4.: Similarity values for the individual decision trees

Figure 4.4.: Descending sorted similarity values for the individual decision trees

4.2.2. Results

Table 4.4 and figure 4.4 show the similarities between the different individual decision trees and the global decision tree.

Generally, the similarity values are quite low. Only $DT_{i09}$ has a value higher than 0.5. These low values are a result of the varying decision tree sizes. While the global decision tree consists of 5 nodes, the number of nodes ranges from 2 to 7 for the individual trees. Moreover, the probability that two trees have identical features is not very high as 30 different features occur in the compared trees. (cf. Appendix section A.1) Although the similarity values are relatively low, there are noticeable differences as the values range from 0.5125 to 0.2208.
4.3. Performance Measurement

4.3.1. Method

For performance measurement the RapidMiner operator “Performance” was employed. There is only one parameter that can be set. This parameter “use example weights” was enabled. Therefore, example weights were used for performance calculation if this had been possible. The operator “Performance” returns three values:

- class precision: ratio of correctly positive(negative) classified examples to the sum of correctly positive(negative) classified and wrongly positive(negative) classified examples

- class recall: ratio of correctly positive(negative) classified examples to the sum of correctly positive (negative) classified and wrongly negative (positive) classified examples

- accuracy: rate of correct class predictions

The operator was used to measure the performance of the global decision tree for the data of the nine individuals. This means that in order to obtain the performance for individual x, the performance of the global decision tree on the data of individual x was measured.

4.3.2. Results

Table 4.5 and figure 4.5 show the accuracy rates of the global decision tree for the different individuals. In section A.3 of the Appendix the complete RapidMiner output can be seen. While the accuracy rate is quite high for individual 09 (95.0 %)

<table>
<thead>
<tr>
<th>individual</th>
<th>03</th>
<th>09</th>
<th>13</th>
<th>16</th>
<th>17</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>accuracy</td>
<td>.83</td>
<td>.95</td>
<td>.90</td>
<td>.88</td>
<td>.82</td>
<td>.78</td>
<td>.65</td>
<td>.82</td>
<td>.91</td>
</tr>
</tbody>
</table>

Table 4.5.: Accuracy rates of the global decision tree
Figure 4.5.: Descending sorted accuracy rates of the global decision tree

it is relatively low for most individuals. Especially for individual 20 the classification performance of the global decision tree is relatively bad (65.0%). It is remarkable that the expression of this individual also was considered to be at least similar to the expression of individual 09. (cf. section 2.4).

4.4. Comparison between Similarity and Performance

The accuracy rates of the global decision tree for the data of the nine individuals were compared with the similarity values between the global decision tree and the nine corresponding individual decision trees. Figure 4.6 and table 4.6 show the accuracy rates and similarity values for each of the nine individuals. As can be seen, particularly from figure 4.6, there is no relationship between accuracy rates and similarity values. Only for individual 09 both accuracy and similarity have the highest value. Additionally, for individual 03 accuracy rate and similarity both have the
fifth highest value. Also the linear regression line (cf. fig. 4.7) illustrates the fact that there is no relationship between the calculated similarity values and the performance of the global decision tree in this case. To obtain the correlation between distance and similarity several measures were calculated using SPSS Statistics 19. The results are listed in table 4.7. These values make it clear that there is no correlation between similarity and accuracy in this case, as every calculated correlation value is very low.
Figure 4.7.: Regression line for similarity and accuracy values

<table>
<thead>
<tr>
<th>correlation measure</th>
<th>Spearman $\rho$</th>
<th>Kendall $\tau$</th>
<th>Goodman $\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>value</td>
<td>.029</td>
<td>.028</td>
<td>.084</td>
</tr>
</tbody>
</table>

Table 4.7.: Correlation values for similarity and accuracy
5. Conclusion

The main concern of this thesis, which was to identify standard mimic types by comparing global and individual decision trees with respect to their similarity, could not be accomplished. Unfortunately, a relationship between similarity and performance, which was the most important precondition of this approach for finding standard mimic types, could not be detected. There may be several reasons for this.

First, the used similarity measure is possibly to imprecise. Perhaps a feature’s type (e.g. “diff”) is not considered as much as necessary. There is also the possibility that including the position of a node in the calculation could produce more accurate results.

Moreover, the fact that only nine decision trees were examined could have distorted the results. Another reason for not finding a relationship between the similarity and the performance of the different trees could be the design of the examined classifiers. Maybe, the features used for classification are too specific or examining another type of classifier, such as perceptrons, would improve the results.

Since for at least individual 09 the calculated similarity value and the corresponding accuracy rate had both the highest value there is still the possibility that similarity and accuracy correlate. Future work could use another similarity measure, which for example does not ignore a feature’s type or a node’s position in a decision tree. One could also repeat the study using a larger data set, i.e. more than nine decision trees. Another approach could be to examine classifiers different from decision trees or examine decision trees that use attributes that are less specific than the employed features.
A. Appendix
A.1. List of features

Points

- $c_{11} x^*$

Differences

- $\text{diff}_{a m2 m7} a_{m4 m5}$
- $\text{diff}_{a m2 mv} a_{m6 mv7}$
- $\text{diff}_{d c12 n09} d_{c12 n07}$
- $\text{diff}_{\frac{d c03 m0 d c06 m0} d_{c09 m4 d c06 m4}}$

Quotients

- $\frac{a_{le0 lec le2 a_{le0 lec le6}}}{a_{m0 mc m7 a_{m1 mc m2}}}$

Mean values

- $z_{m7 x}$
- $z_{\text{diff}_{a m0 mc m1 a_{m0 mc m2}}}$
- $z_{\text{diff}_{a m1 mc m5 a_{m2 mc m7}}}$
- $z_{\text{diff}_{a m1 mc m6 a_{m3 mc m6}}}$
- $z_{\text{diff}_{a re0 rev re4 a_{le0 lev le4}}}$
- $z_{\text{diff}_{\frac{a re0 rev re6 a_{re2 rev re4}} a_{re4 rev re6}}}$
- $z_{\text{diff}_{\frac{a re0 rev re6 a_{re2 rev re4}} a_{re4 rev re6}}}$
- $z_{\text{diff}_{\frac{a re0 rev re6 a_{re2 rev re4}} a_{re4 rev re6}}}$

Features marked with * appear in the learned decision trees
• ssmd_diff_a_m0_mv_m5_a_m2_mv_m4*
• ssmd_diff_a_m1_mv_m6_a_m6_mv_m7
• ssmd_diff_a_m2_mc_m3_a_m3_mc_m5*
• ssmd_diff__diff_a_re2_rec_re4_a_re4_rec_re6__diff_a_le2_lec_le4_a_le4_lec_le6*
• ssmd_diff__frac_a_re0_rev_re6_a_re2_rev_re4__frac_a_le0_lev_le6_a_le2_lev_le4
• ssmd_frac_a_m0_mv_m3_a_m2_mv_m4*
• ssmd_frac_a_m1_mv_m3_a_m1_mv_m4
• ssmd_frac_a_m1_mv_m7_a_m4_mv_m5*
• ssmd_frac_a_m2_mh_m6_a_m3_mh_m6*
• iz_leh_x
• iz_re7_y*
• iz_d_c12_m07
• iz_d_re2_re6*
• iz_diff_a_m0_mh_m2_a_m6_mh_m7*
• iz_diff_a_m0_mv_m1_a_m0_mv_m3
• iz_diff_a_m0_mv_m1_a_m2_mv_m4*
• iz_diff_a_m0_mv_m4_a_m0_mv_m7
• iz_diff_a_m1_mh_m4_a_m3_mh_m4
• iz_diff_a_m2_mc_m7_a_m4_mc_m5*
• iz_diff__frac_a_re2_rev_re4_a_re4_rev_re6__frac_a_le2_lev_le4_a_le4_lev_le6
• iz_frac_a_m0_mh_m1_a_m1_mh_m5
• iz_frac_a_m2_mc_m3_a_m4_mc_m7
• iz_frac_a_m2_mv_m7_a_m4_mv_m7
• iz_frac_d_re0_re4_d_re2_re6*
• issmd_m1_y*
• issmd_re7_y
• issmd_d_c12_m4*
• issmd_d_n00_n01*
• issmd_d_re2_re6
• issmd_diff_a_m0_mc_m7_a_m1_mc_m5
• issmd_diff_a_m2_mc_m5_a_m5_mc_m7*
• issmd_frac_a_m0_mh_m1_a_m1_mh_m4*
• issmd_frac_a_m0_mh_m1_a_m1_mh_m5
• issmd_frac_a_m0_mh_m6_a_m4_mh_m7
• issmd_frac_a_m1_mc_m2_a_m2_mc_m7
A.2. Decision trees
Figure A.1.: Individual decision tree 03
Figure A.2.: Individual decision tree 09
Figure A.3.: Individual decision tree 13
Figure A.4.: Individual decision tree 16
Figure A.5.: Individual decision tree 17
Figure A.6.: Individual decision tree 19
Figure A.8.: Individual decision tree 21
Figure A.9.: Individual decision tree 32
Figure A.10.: Global decision tree
A.3. Rapidminer output for performance measurement

<table>
<thead>
<tr>
<th>accuracy 83.33 %</th>
<th>true false</th>
<th>true true</th>
<th>class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. false</td>
<td>24</td>
<td>4</td>
<td>85.71 %</td>
</tr>
<tr>
<td>pred. true</td>
<td>6</td>
<td>26</td>
<td>81.25 %</td>
</tr>
<tr>
<td>class recall</td>
<td>80.00 %</td>
<td>86.67 %</td>
<td></td>
</tr>
</tbody>
</table>

Table A.1.: Global decision tree performance for individual 03

<table>
<thead>
<tr>
<th>accuracy 95.00 %</th>
<th>true false</th>
<th>true true</th>
<th>class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. false</td>
<td>29</td>
<td>2</td>
<td>93.55 %</td>
</tr>
<tr>
<td>pred. true</td>
<td>1</td>
<td>28</td>
<td>96.55 %</td>
</tr>
<tr>
<td>class recall</td>
<td>96.67 %</td>
<td>93.33 %</td>
<td></td>
</tr>
</tbody>
</table>

Table A.2.: Global decision tree performance for individual 09

<table>
<thead>
<tr>
<th>accuracy 90.00 %</th>
<th>true false</th>
<th>true true</th>
<th>class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. false</td>
<td>25</td>
<td>1</td>
<td>96.15 %</td>
</tr>
<tr>
<td>pred. true</td>
<td>5</td>
<td>29</td>
<td>85.29 %</td>
</tr>
<tr>
<td>class recall</td>
<td>83.33 %</td>
<td>96.67 %</td>
<td></td>
</tr>
</tbody>
</table>

Table A.3.: Global decision tree performance for individual 13
### Table A.4.: Global decision tree performance for individual 16

<table>
<thead>
<tr>
<th></th>
<th>true false</th>
<th>true true</th>
<th>class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. false</td>
<td>28</td>
<td>5</td>
<td>84.85 %</td>
</tr>
<tr>
<td>pred. true</td>
<td>2</td>
<td>25</td>
<td>92.59 %</td>
</tr>
<tr>
<td>class recall</td>
<td>93.33 %</td>
<td>83.33 %</td>
<td></td>
</tr>
</tbody>
</table>

### Table A.5.: Global decision tree performance for individual 17

<table>
<thead>
<tr>
<th></th>
<th>true false</th>
<th>true true</th>
<th>class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. false</td>
<td>27</td>
<td>8</td>
<td>77.14 %</td>
</tr>
<tr>
<td>pred. true</td>
<td>3</td>
<td>22</td>
<td>88.00 %</td>
</tr>
<tr>
<td>class recall</td>
<td>90.00 %</td>
<td>73.33 %</td>
<td></td>
</tr>
</tbody>
</table>

### Table A.6.: Global decision tree performance for individual 19

<table>
<thead>
<tr>
<th></th>
<th>true false</th>
<th>true true</th>
<th>class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. false</td>
<td>22</td>
<td>5</td>
<td>81.48 %</td>
</tr>
<tr>
<td>pred. true</td>
<td>8</td>
<td>25</td>
<td>75.76 %</td>
</tr>
<tr>
<td>class recall</td>
<td>73.33 %</td>
<td>83.33 %</td>
<td></td>
</tr>
</tbody>
</table>

### Table A.7.: Global decision tree performance for individual 20

<table>
<thead>
<tr>
<th></th>
<th>true false</th>
<th>true true</th>
<th>class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. false</td>
<td>23</td>
<td>14</td>
<td>62.16 %</td>
</tr>
<tr>
<td>pred. true</td>
<td>7</td>
<td>16</td>
<td>69.57 %</td>
</tr>
<tr>
<td>class recall</td>
<td>76.67 %</td>
<td>53.33 %</td>
<td></td>
</tr>
</tbody>
</table>

### Table A.8.: Global decision tree performance for individual 21

<table>
<thead>
<tr>
<th></th>
<th>true false</th>
<th>true true</th>
<th>class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. false</td>
<td>24</td>
<td>5</td>
<td>82.76 %</td>
</tr>
<tr>
<td>pred. true</td>
<td>6</td>
<td>25</td>
<td>80.65 %</td>
</tr>
<tr>
<td>class recall</td>
<td>80.00 %</td>
<td>83.33 %</td>
<td></td>
</tr>
</tbody>
</table>

### Table A.9.: Global decision tree performance for individual 32

<table>
<thead>
<tr>
<th></th>
<th>true false</th>
<th>true true</th>
<th>class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. false</td>
<td>27</td>
<td>2</td>
<td>93.10 %</td>
</tr>
<tr>
<td>pred. true</td>
<td>3</td>
<td>26</td>
<td>89.66 %</td>
</tr>
<tr>
<td>class recall</td>
<td>90.00 %</td>
<td>92.86 %</td>
<td></td>
</tr>
</tbody>
</table>
References


Erklärung
Ich erkläre hiermit gemäß § 17 Abs. 2 APO, dass ich die vorstehende Bachelorarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

October 19, 2011

(Unterschrift)