Connect Measurements of Car Body Parts to Adjustments in Production Line

Chenyan Gao
Abstract

Connect Measurements of Car Body Parts to Adjustments in Production Line

Chenyan Gao

The production of a car body part involves assembling each ingoing part into an assembly fixture that dictates the inherent geometry between the ingoing parts. The geometry of samples of finished body parts is continuously measured and adjustments to the assembly fixture are done on the basis of measurement results. The work of the thesis is to use Case-based reasoning technique to develop a system connecting the measurements of car body parts to adjustments done in the production line.

Case-based reasoning solves the target issue with previous successful solutions to the similar problems. In this thesis, the Euclidean distance is selected to measure the similarity between the target case and source case. Correlation analysis is done to find the important attributes, measurement points. We present the results we found, describe the system implementation and discuss the future work.

Key words: Case-based reasoning, production line, measurement data, adjustment data, Euclidean distance and correlation analysis.
Acknowledgement

It has been a pleasure that doing the thesis as a case study in Volvo Cars. First of all, I would like to thank my supervisor Erik Olsson and reviewer Roland Bol for giving me helps, advices. I also give my thanks to Alf Andersson for everything he did when I visited the Volvo factory. Finally, thanks to the engineers in production line for providing the adjustment log and the nice pictures of the production line. Without them, I couldn’t finish the thesis alone.
# Table of Contents

**ABSTRACT** .................................................................................................................. 3

**ACKNOWLEDGEMENT** ................................................................................................ 5

1 **INTRODUCTION** ........................................................................................................... 9
   1.1 CASE-BASED REASONING ....................................................................................... 9
       1.1.1 Overview ........................................................................................................... 9
       1.1.2 Case-based cycle .............................................................................................. 9
   1.2 THESIS MOTIVATION AND PURPOSE ................................................................... 11
   1.3 THESIS OVERVIEW ................................................................................................. 12

2 **PROBLEM DESCRIPTION** ............................................................................................ 13
   2.1 CAR PRODUCTION LINE ......................................................................................... 13
   2.2 DATA DEFINITION .................................................................................................. 13
       2.2.1 Crossmember .................................................................................................. 14
       2.2.2 Car coordinate ................................................................................................. 14
       2.2.3 Measurement point ......................................................................................... 15
       2.2.4 Adjustment point ............................................................................................ 16
   2.3 PROCESS OF MEASURING AND ADJUSTING BODY PARTS ..................................... 17

3 **CASE-BASED REASONING** ........................................................................................ 19
   3.1 CASE REPRESENTATION ......................................................................................... 19
   3.2 RETRIEVAL METHOD .............................................................................................. 19
       3.2.1 Nearest neighbour retrieval .............................................................................. 19
       3.2.2 Weighted nearest neighbour retrieval .............................................................. 20
   3.3 ADAPTATION .......................................................................................................... 20
   3.4 LOCKHEED — CLAVIER .......................................................................................... 22

4 **DATA ANALYSIS** ......................................................................................................... 23
   4.1 CREATING CASE DATABASE ............................................................................... 23
       4.1.1 Measurement and result data ........................................................................... 23
       4.1.2 Adjustment data .............................................................................................. 23
       4.1.3 Case database .................................................................................................. 24
   4.2 PLOT THE MEASUREMENT DATA TO FIND SIMILARITY ....................................... 26
   4.3 FIND THE SIGNIFICANT MEASUREMENT POINTS ..................................................... 29
       4.3.1 Calculating the differences ................................................................................ 30
       4.3.2 Correlation analysis ......................................................................................... 30
       4.3.3 Calculating the coefficients in Matlab ............................................................... 32
1 Introduction

In this section I start with the background of case-based reasoning, and then briefly describe the motivation and purpose of the thesis with relative technologies. At last, a short overview of the rest chapters is presented which provides an outline to the readers.

1.1 Case-based reasoning

1.1.1 Overview

Case-based reasoning (CBR) is one of the most promising recent technologies for problem solving. CBR is an intelligent-systems method that solves new problems by adapting previously successful solutions to similar problems, which increases efficiency and reduces cost [1].

CBR technology can be traced back to the work of Roger Schank and his students at Yale University in the early 1980s, and in the next decade, it appealed to the international community and developed rapidly. So far, CBR technology has been applied to a number of successful deployed systems, such as manufacturing applications (Lockheed’s CLAVIER\(^1\) [2] [3]) and help desk applications (Compaq SMART\(^2\) system [1]).

1.1.2 Case-based cycle

Compared to the rule induction algorithms of machine learning, CBR is a strategy of lazy generalization, which does not generalize cases until the target problem given. The rule induction algorithms always derive a set of rules before the testing time. Respecting the feature of CBR that no explicit model is required to elicit, CBR is applied to approach problems with rich and complex domains.

Case-based reasoning is described as a four-step cyclical process consisting of the four RE:s [1]:

- RETrieve the most similar case(s);

\(^1\) The first CBR commercial application

\(^2\) Support Compaq customer service
All the successful history data are collected to create a case database, from which search for the cases with the highest similarity to the current problem. This is the first step to retrieve cases. The assessment of similarity involves several well known algorithms: nearest neighbour, induction, knowledge guided induction and template retrieval. After the nearest cases are found, the CBR system reuses the solution to the new problem. It should be noted that the solution to the retrieved case should not be directly applied to the target problem but do some proper adaptation. The revised solution applies some formulate or rules which take into consideration the prominent differences between the retrieved case and the given case. In the end of the CBR process, the confirmed solution is retained as a part of new case in the database.

---

3 Adapted from Aamodt & Plaza, 1994
1.2 Thesis motivation and purpose

Volvo Car Corporation is a worldwide vehicle company in Sweden, which sell over 300,000 cars to over one hundred countries every year. Most of the Volvo car body components are manufactured in Volvo Cars Body Components in Olofström and delivered for final assembly to the plants in Gothenburg and Gent. The plant in Olofström produces approximately forty million car body parts per year.

The production line of car body parts is continuously adjusted on the basis of sample measurements of finished parts. A finished body part is made up from several ingoing parts. The production of a body part involves mounting each ingoing part into an assembly fixture that dictates the inherent geometry between the ingoing parts. Then the industrial robots spot welded these ingoing parts together. The geometry of samples of finished body parts is continuously measured and adjustments to the assembly fixture are done according to measurement results. Measurement results are stored in a computer and any performed adjustments to the assembly fixture are written down by hand in a notepad.

The task of the thesis is to build a case-based system to develop methods and algorithms to connect measurements of car body parts to adjustments done in the production line. Now all the adjustments are made by the experienced engineer in Volvo Cars. The goal of the system is to predict adjustments based on the history measurement results, which is expected to replace the work of the engineer or at least partly replaced.

The examples of Lockheed’s CLAVIER and BROADWAY [4] show the possibility that using case-based reasoning to make a prediction of ingoing parts’ adjustments. Investigate a set of selected production measurements and their resulting adjustments using the method of similarity metrics presented by Skalk [4] and correlation analysis [5]. However, it may be a problem that how to make proposed adjustments as accurate as possible.
1.3 Thesis overview

Some related work on car production line and some data definition will be introduced in chapter 2. Chapter 3 and 4, the two critical parts of the thesis are followed. Chapter 3 will depict the method Case-based reasoning from case representation, retrieve method and adaptation. Lockheed’s CLAVIER as a successful commercial CBR system will be introduced in the last section of chapter 3. The data analysis in chapter 4 will include the creation of case database and the whole process of finding the significant measurement points using correlation analysis. In chapter 5, it will introduce the system implementation. In the last chapter, it will discuss some complications that should be noticed in the thesis, then draw the conclusion and discuss the future work.
2 Problem Description

This chapter introduces car production line, some important notion and the process of measuring and adjusting ingoing body parts.

2.1 Car production line

Nowadays a large scale of mechanization saves rather large amounts of labour so that most of work is done by industrial robots in Volvo Cars. First, the worker uploads several ingoing parts to the platform where the welding spots have been adjusted by an experienced engineer. A set of robots then automatically welded these ingoing parts together in sequence. The finished body parts are piled up on the shelf by the mechanical arm, and shipped to the warehouse.

![Figure 2.1: The production line and the platform](image)

2.2 Data definition

Some important data definitions in this thesis are given in this section.

---

4 Volvo holds the copyrights on all the photos in this thesis
2.2.1 Crossmember

Crossmember [6] is a frame in the head of Volvo S80. It is on the crossmember that the headlights are mounted. The part is symmetrical and composed of nine ingoing parts: crossmember, right and left side of bracket bonnet lock, right and left side of stay rear inner, right and left side of stay rear outer and right and left side of bracket bottom. There are many measurement points and adjustment points marked on the crossmember, which should be focused on in this thesis.

![Crossmember Diagram](image)

Figure 2.2: The diagram of the crossmember

2.2.2 Car coordinate

The following figure illustrates the car coordinate [7] directions: X-axis, forward/backward; Y-axis, in/out; Z-axis, high/low. These directions are used when the engineer makes the adjustments. Each adjustment should be recorded the adjustment id, direction and value. The detailed adjustment data will be discussed in section 4.1.2.
2.2.3 Measurement point

Measurement points lie in the different ingoing parts of the crossmember, which are marked by the car designer. Each measurement point has three values respectively on x, y and z axis. According to the measurement points’ data, the designer can know the shape of the body part and make some alteration to the shape to meet some requirements. All the data of measurement points are measured by an instrument and stored in a computer. In this thesis, it involves as many as 290 measurement points.

The following graph illustrates the measurement points in the ingoing part bracket bottom.


2.2.4 Adjustment point

Adjustment points are marked as the welding spot to get the different ingoing parts of the crossmember together. As it is inevitable that the ingoing parts will be bent or twisted in the process of welding, the measurement points consequently deviate from the initial positions. That implies, adjustment points have influence on the measurement points. The experienced engineer is able to alter one or more adjustment points to control the deviation of the measurement points. The goal of the thesis is to find the relationship between the measurement point and the adjustment point.

The engineer numbered these adjustment points and wrote down all the adjustment data in a notepad. Each record includes adjustment point ID, the direction of the movement, the alteration and the date. In section 4.1.2, the example of adjustment data will be given.
2.3 Process of measuring and adjusting body parts

It is a cyclical process of measuring and adjusting body parts. A finished body part is continuously adjusted based on sample measurements of finished parts. The finished body parts from the production line are sent to measurement room in the factory, where a Coordinate Measurement Machine (CMM) automatically measures these feature points and saves the results in the computer.
On getting the measurement data, the engineer makes some calculations and then adjusts the position of references and support points in the assembly fixture by adding shims, to meet the requirements of the final part. The adjustment data are written down in a notebook. The newly produced body parts are sent to be measured again and new measurement data are saved. Every measurement point has its own tolerant range to check the deviation of the point from the required position. If it is not out of the range, the adjustment is successful; otherwise it needs second even third adaptation.
3 Case-based Reasoning

Case-based reasoning is an intelligent method of applying the previous solutions to the new problems. Lockheed’s CLAVIER is testimony to the successful application of case-based reasoning to industrial systems. In this chapter, the method is further discussed, focusing on case representation, retrieval method and adaptation. A brief introduction of CLAVIER system will be given in the last section of the chapter.

3.1 Case representation

As the name implies, case is the important part in the method of case-based reasoning. All the steps of the case-based cycle are related to the case. A case can be a kind of an event, a story or a record, in any type. However, it must typically consist of two parts: the problem space and the solution space. The content of a case contains the past lesson and the condition in which the lesson can be used. The definition of a case for this thesis will be discussed in the next chapter.

3.2 Retrieval method

To retrieve the most similar case is the first step in the process of CBR cycle. In general, nearest-neighbour retrieval and inductive retrieval are two prevalent techniques in current commercial CBR tools.

Inductive retrieval involves induction process, which is a technique to derive rules or construct decision trees from the past data. It is commonly used in operation research to help identify a strategy [8]. Inductive retrieval seems to be not suitable for the thesis to solve a numerical problem, therefore, nearest neighbour is selected as the retrieval algorithm.

3.2.1 Nearest neighbour retrieval

Nearest neighbour classification is a method for classifying an object based on the testing examples. An essential part of nearest neighbour classification is nearest neighbour retrieval, which works as follows: the first step is to create a database
saving example objects. When a new object is given, the system then queries the database to find the most similar one [9].

Finding out the nearest neighbour of an object means that we need a way to measure how similar a target object to a source object. In the project, I calculate the Euclidean distance between the target case and each source case in the database.

The distance of target case T from case A:

\[
\text{Dist}_A = \sqrt{\sum_{k=1}^{n} (T_k - A_k)^2}
\]

In the above formula, \(n\) is the number of measurement points (attributes), and \(T_k\) and \(A_k\) are, respectively, the kth measurement point of case T and A.

### 3.2.2 Weighted nearest neighbour retrieval

The concept of nearest neighbour retrieval is basically that simple, calculating the distance between two cases. To be more accurate and realistic, weighted nearest neighbour retrieval is introduced. It needs the weights in case the scales of the attributes differ. The formula presented in the former section changes to:

\[
\text{Weighted Dist}_A = \sqrt{\sum_{k=1}^{n} W_k(T_k - A_k)^2}
\]

where \(W_k\) is the weight of the kth measurement point.

I intend to find the significant attributes from as many as 290 measurement points and weight them. In the chapter 4, the attention will be focused on the data analysis, seeking to find the connection between measurement points and adjustment points. If no correlation exists between the kth measurement and adjustments in a case we will set \(W_k\) as 0, otherwise we will set \(W_k\) as 1.

### 3.3 Adaptation

Once the nearest case is retrieved, the CBR system will intend to reuse the solution suggested by the matching case. The solution is possible to be sufficient in most situations. However, in other instances the solution from the nearest case
is not close enough to the required solution, since the current case differs from the retrieved case. There is a need to adapt the retrieved solution in these circumstances.

Figure 3.1: The problem and solution

Adaptation is to look for dominant differences between the retrieved case and the target case, and take them into account when applying formulas. The adaptation formulas are usually not easy to define, so adaptation is rarely performed in most successful CBR systems, which, in other words, are just case-retrieval systems. In this project, we leave adaptation to the engineer.
3.4 Lockheed — CLAVIER

Lockheed’s CLAVIER is one of the first commercial CBR applications. Lockheed is the US aerospace company producing aircraft element. Each element has require curing correctly in large autoclaves, if not, it have to be discarded. Each part has its own heating characteristics, which unfortunately, are not completely understood by the operators of Lockheed. Lockheed turned to CBR as a last resort after the failure of using mathematical modeling and expert systems methods.

Operators used to depend on previous successful parts layouts to know how to layout the autoclave. Now CLAVIER acts as a collective memory, which suggests operators on layouts of composite parts in the autoclave curing ovens. Every layout contains such information [3]:
- geometric shape of the parts and their relative positions on a table
- tables, and their relative positions in the autoclave, and
- the settings of the autoclave (e.g. start and finish time, pressure and temperature).

When the best case retrieved is not totally matched with the target problem, the system will substitute the nonmatching part with a similar part to create a new layout.

CLAVIER has been in regular use for around twenty years since September 1990. Its success gives confidence to me to use CBR to solve the problem in this thesis.
4 Data Analysis

In this chapter, the focus then should be on the data analysis after the elaboration of problem solving method CBR. First step is to create a case database, followed by the data analysis to verify the dependence of adjustment points on measurement points. In this thesis, each adjustment point is treated as a separate problem and I will analyze them individually.

4.1 Creating case database

4.1.1 Measurement and result data

The car body parts are kept measuring in the past three years. The measurements of 290 feature points before and after an adjustment are all recorded. The nearest measurement after an adjustment is considered as the result, which compose the problem space with the measurement. The problem space describes the state in which the adjustment should be made.

4.1.2 Adjustment data

The solution space of the case implies the adjustment written down by the engineer. In figure 4.1, you can see that a solution includes the adjustment point id, the movement direction, and the value of the movement. The first record in the figure depicts that adjustment point 27 should be moved forward for 0.3 millimeters. (The definitions of directions have been given in the section 2.2.2 Car Coordinate.)
4.1.3 Case database

A case comprises the problem description and the solution. Since the measurements and the adjustments data both have the timestamps, they are matched based on their timestamps, considering the closest measurement done after an adjustment. In this thesis, a case consists of three parts: measurements, adjustment data and the outcome (measurement results). The adjustment of a case involves one or more adjustment points. A case in the database is as:

\[[Mb, Adjustment, Ma]\]

where Mb is the measurements of 290 measurement points before an adjustment, and Ma, the outcome, is the measurements of these measurement points after the adjustment.

When a new problem is given, Ma is the requirements of the final part. The adjustment of the problem is the solution to the nearest case of the problem. The
distance between the problem and a case is calculated based on the formula in section 3.2.2:

$$\text{Weighted Dist}_A = \sqrt{\sum_{k=1}^{n} W_k (T_k - A_k)^2}.$$  

$T_k$ is the difference of $M_A k$ and $M_B k$ of the target problem. $A_k$ is the difference of $M_A k$ and $M_B k$ of case A. $M_A k$ and $M_B k$ are respectively the kth measurement of Ma and Mb.

I exported the measurement data in the past three years from the software CM4D into the excel file. CM4D [10] is an application that helps product manufacturers generate rich graphical reporting for quality analysis. CM4D collects data from inspection devices and imports the data in a specified format into its own database in which it is coordinate with the associated 3D image. CM4D then creates charts and graphs to display the data and 3D model in an easy-to-read way.

The exported data not only record measurement value and timestamp, but also contain the part identity: Fixture A or Fixture B. I manually matched the measurement data to the adjustment data to create a case database. All the cases are saved in the file caseDB.xls. Case database consists of three sheets respectively saving measurements, adjustments and results, which are related by the case id. There are 56 cases in total involving 290 measurement points and 30 coordinate points. However, several adjustment data are excluded from the case database due to the lack of measurements corresponding to them.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Feature</td>
<td>1B</td>
<td>2A</td>
<td>3A</td>
<td>4B</td>
<td>5A</td>
<td>6B</td>
<td>7A</td>
<td>8A</td>
<td>9B</td>
<td>10A</td>
</tr>
<tr>
<td>2</td>
<td>FR0037KL</td>
<td>-0.20</td>
<td>-0.16</td>
<td>-0.14</td>
<td>-0.32</td>
<td>-0.30</td>
<td>-0.26</td>
<td>-0.19</td>
<td>-0.30</td>
<td>-0.23</td>
<td>-0.19</td>
</tr>
<tr>
<td>3</td>
<td>FR0039KL</td>
<td>0.05</td>
<td>0.10</td>
<td>0.15</td>
<td>0.01</td>
<td>0.07</td>
<td>0.09</td>
<td>0.02</td>
<td>0.15</td>
<td>0.13</td>
<td>0.09</td>
</tr>
<tr>
<td>4</td>
<td>FR0039KF</td>
<td>-0.37</td>
<td>-0.33</td>
<td>-0.29</td>
<td>-0.25</td>
<td>-0.21</td>
<td>-0.39</td>
<td>-0.38</td>
<td>-0.24</td>
<td>-0.36</td>
<td>-0.34</td>
</tr>
<tr>
<td>5</td>
<td>FR0071CL</td>
<td>0.12</td>
<td>0.15</td>
<td>0.16</td>
<td>0.08</td>
<td>0.08</td>
<td>0.11</td>
<td>0.09</td>
<td>0.12</td>
<td>0.08</td>
<td>0.26</td>
</tr>
<tr>
<td>6</td>
<td>FR0071CR</td>
<td>0.25</td>
<td>0.05</td>
<td>0.03</td>
<td>-0.06</td>
<td>0.02</td>
<td>-0.05</td>
<td>0.00</td>
<td>0.06</td>
<td>-0.10</td>
<td>0.09</td>
</tr>
<tr>
<td>7</td>
<td>FR0179CL</td>
<td>0.10</td>
<td>0.32</td>
<td>0.05</td>
<td>0.21</td>
<td>0.58</td>
<td>0.24</td>
<td>0.68</td>
<td>0.56</td>
<td>0.16</td>
<td>0.36</td>
</tr>
<tr>
<td>8</td>
<td>FR0179CR</td>
<td>0.02</td>
<td>0.07</td>
<td>-0.35</td>
<td>-0.40</td>
<td>0.01</td>
<td>-0.50</td>
<td>-0.10</td>
<td>-0.19</td>
<td>-0.43</td>
<td>-0.12</td>
</tr>
<tr>
<td>9</td>
<td>FR0179CY</td>
<td>-0.34</td>
<td>-0.12</td>
<td>-0.25</td>
<td>-0.47</td>
<td>-0.20</td>
<td>-0.54</td>
<td>-0.31</td>
<td>-0.62</td>
<td>-0.40</td>
<td>-0.54</td>
</tr>
<tr>
<td>10</td>
<td>FR0179CR</td>
<td>-0.19</td>
<td>-0.38</td>
<td>0.04</td>
<td>-0.25</td>
<td>-0.59</td>
<td>-0.43</td>
<td>-0.73</td>
<td>-0.72</td>
<td>-0.13</td>
<td>-0.76</td>
</tr>
</tbody>
</table>

Figure 4.2 (a): Excerpt of measurements
I think the measurement data with the similar adjustments of the same coordinate point are supposed to have some characteristics in common. These characteristics are possible to imply the significant measurement points, which are affected by the coordinate point. I then intend to plot the measurement data to find the similarity.

For instance, there are six cases: 5, 6, 7, 25, 49 and 50 where adjustment point 11 is altered.

The adjustment data:

<table>
<thead>
<tr>
<th>Ad point 11</th>
<th>Case 5</th>
<th>Case 6</th>
<th>Case 7</th>
<th>Case 25</th>
<th>Case 49</th>
<th>Case 50</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.4</td>
<td>0.3</td>
<td>-0.4</td>
<td>-0.8</td>
<td>0.2</td>
<td>-0.2</td>
</tr>
</tbody>
</table>
The above chart lists the cases related to coordinate point 11, from which we can know that the alterations in case 5 and case 7 are same, and value are similar in case 6 and case 49. I plot the measurement data of case 5 and 7, and that of case 6 and 49, respectively in Figure 4.3 and Figure 4.4.

In Figure 4.3, the blue line illustrates the measurement data of case 5, while the data of case 7 is in red. The X axis and Y axis are measurement point ID and measurement value respectively. It can be noted that there are some gaps in the lines, for example, from point 30 to 70, cause these points were not measured in the two cases. Moreover, the trends of these lines are roughly the same, especially around point 200 and point 250.
Figure 4.4 describes the measurement data of case 6 and case 49. As seen in the graph, the blue line has some gaps which means some measurement point were not measured in case 6, however the red line representing case 49 is continuous. In another word, 290 measurement points were all measured in case 49. We can induce that the points absent in case 6 are not significant points in terms of adjustment point 11, whose alteration has not influence on these absent points. The trend of the blue line is similar to that of the red line around point 200 and point 250, which we also can find in Figure 4.3.
In contrast to selecting similar samples in the former figures, two different cases are chosen in Figure 4.5. The adjustment data of case 49, 50 are respectively 0.2 and -0.2. Compared with the red line representing case 50, the blue line fluctuates in the opposite direction from measurement point 240 to 260.

It can be summarized that measurement points from 240 to 260 are affected by coordinate point 11. But I could not sure the exact points in this range are significant. In addition, coordinate point 11 has no influence on quite a few points such as points from 30 to 80.

Though some potential significant points are found by plotting the measurement data, I cannot affirm the exact important ones. Correlation analysis is thus used to verify the existence of these relationships in the next section.

4.3 Find the significant measurement points

The engineer infers which coordinate point is supposed to be modified and the adjustment value from the measurement data and required results. So in all cases with similar adjustment coordinates, I consider a relation should exist between measurement points and resulting measurement points. Among 290
measurement points there are some significant attributes, which are affected by the alteration of the adjustment points.

In this case, weighted nearest-neighbour is more accurate and realistic about distance calculation compared to nearest-neighbour. Hence I decide to use weighted nearest-neighbour retrieval. In this section, I calculate the difference between the measurements and results, and then use correlation analysis to look for the connection between the difference and adjustments. Some candidates of measurement points will be returned, which the alteration of coordinate points may have effects on. Next, I analyze the candidates further to remove the point that changes inconsistently in similar cases. Last, the results are the significant points.

The coordinate point 11 is taken as the example to explain in detail how to find the significant points in the following sections.

### 4.3.1 Calculating the differences

Before calculating the differences between the measurements and results, I first import the case database into Matlab from excel file. All the measurement data and result data are stored in two matrices. Next, pick up all the similar cases about coordinate point 11 and merge their measurements and results into two new matrices, respectively: Mea\(_{11}\) and Rslt\(_{11}\). Mea\(_{11}\) is a 290*10 matrix, the same size as matrix Rslt\(_{11}\). The differences matrix D\(_{11}\) is got after subtracting Mea\(_{11}\) from Rslt\(_{11}\).

### 4.3.2 Correlation analysis

In general statistical usage, saying that correlation exists implies that there is a relationship between two mathematical variables or measured data values. There are several correlation coefficients. The Pearson correlation coefficient is special among these coefficients, which is sensitive to a linear relationship between two variables.
Figure 4.6: Several sets of \((x, y)\) points, with the correlation coefficient of \(x\) and \(y\) for each set.

The definition of correlation coefficient \(R\) is [5]:

\[
R = \frac{1}{n-1} \sum_{i=1}^{n} \frac{X_i - \mu_x}{\sigma_x} \ast \frac{Y_i - \mu_y}{\sigma_y},
\]

where \(\mu_x\) and \(\sigma_x\) denote the sample mean and the sample standard deviation respectively for the variable \(x\) and \(\mu_y\) and \(\sigma_y\) denote the sample mean and the sample standard deviation respectively for the variable \(y\).

\(R\) is in the range from -1 to 1. The case \(R=1\) as shown in Figure 4.6 (a), represents the maximum possible linear positive connection between two variables. If \(R=-1\), the maximum possible linear negative connection exists. The magnitude of \(R\) indicates the strength of the association, while the sign implies the direction of association.
4.3.3 Calculating the coefficients in Matlab

The function corrcoef in Matlab [10] can calculate possibility of correlation between each measurement point and adjustment point.

\[
[R, P] = \text{corrcoef}(X, Y)
\]

R is a matrix of correlation coefficients calculated from two vectors X and Y. P is also a matrix returning p values, is for testing the hypothesis of no correlation. If \(P(i,j)\) is small, for example, less than 0.05, then the correlation between \(X_i\) and \(Y_j\) is significant.

In terms of adjustment point 11, there are cases: 5, 6, 7, 25, 28, 29, 31, 32, 49 and 50.

The difference data:

<table>
<thead>
<tr>
<th></th>
<th>Case 5</th>
<th>Case 6</th>
<th>Case 7</th>
<th>Case 25</th>
<th>Case 28</th>
<th>Case 29</th>
<th>Case 31</th>
<th>Case 32</th>
<th>Case 49</th>
<th>Case 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mea point 1</td>
<td>0.15</td>
<td>-0.01</td>
<td>-0.05</td>
<td>0.03</td>
<td>-0.01</td>
<td>-0.03</td>
<td>0.02</td>
<td>-0.01</td>
<td>0.04</td>
<td>-0.03</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Mea point 290</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>-0.06</td>
<td>0.01</td>
<td>-0.15</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

\[
D_{11} = \begin{bmatrix}
0.15 & -0.01 & -0.05 & 0.03 & -0.01 & -0.03 & 0.02 & -0.01 & 0.04 & -0.03 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
NaN & NaN & NaN & NaN & NaN & NaN & -0.06 & 0.01 & -0.15 & -0.01
\end{bmatrix}
\]

The adjustment data:

<table>
<thead>
<tr>
<th>Ad point 11</th>
<th>Case 5</th>
<th>Case 6</th>
<th>Case 7</th>
<th>Case 25</th>
<th>Case 28</th>
<th>Case 29</th>
<th>Case 31</th>
<th>Case 32</th>
<th>Case 49</th>
<th>Case 50</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.4</td>
<td>0.3</td>
<td>-0.4</td>
<td>-0.8</td>
<td>-0.4</td>
<td>-0.4</td>
<td>-0.4</td>
<td>-0.4</td>
<td>0.2</td>
<td>-0.2</td>
</tr>
</tbody>
</table>

Saving these data in the vector \(\text{Ad}_{11} = [-0.4, 0.3, -0.4, -0.8, -0.4, -0.4, -0.4, -0.4, 0.2, -0.2]\);
I have collected all the cases about one adjustment point 11, saving the adjustment data in the vector Ad_{11}. The differences between measurements and results in these cases are saved in a matrix D_{11}, and then call the function:

\[ [R, P] = \text{corrcoef}(D_{11n}, Ad_{11}), \]

where n = 1, ..., 290, D_{11n} is the nth row of matrix D_{11}.

Then, matrix R and P is returned after calling the function corrcoef, and select the measurement point whose p value is less than 0.1. Since there are only 10 cases, the criteria of 0.05 is too low to set. So I set the p value as 0.1. These points founded may be the significant measurement points.

### 4.3.4 Candidates

The criterion is set as 0.1. The p-value lower than 0.1 is considered as the strong relationship existent with red marked in Figure 4.7. The higher the p-value is, the lower the possibility of relationship is.

Moreover, some feature points are measured in some case but not in another case, and both cases are corresponding to the same adjustment point. In terms of these feature points marked in yellow in the first column, they have no influence on the adjustment point. Their weights can be set as zero.

As displayed in the followed chart, measurement point FR0038KL/X may be affected by adjustment point 31 and 47, but not affected by adjustment point 11, 12, 27, 32, 43, 46 and 48.
4.3.5 Sifting

In the last section, I find some candidates which may be significant points, but it still needs sifting them to remove the point that changes inconsistently in similar cases.

For example, there are ten similar cases about coordinate point 11. Five of the ten cases have the same adjustment data. The differences of each candidate point in these five cases are supposed to be similar or even the same; otherwise the candidate is not the significant point and can be taken away.

The adjustment data and the difference data of candidate point 245 and 260 are shown in the following charts.

The adjustment data of coordinate point 11:

<table>
<thead>
<tr>
<th>Ad point 11</th>
<th>Case 5</th>
<th>Case 6</th>
<th>Case 7</th>
<th>Case 25</th>
<th>Case 28</th>
<th>Case 29</th>
<th>Case 31</th>
<th>Case 32</th>
<th>Case 49</th>
<th>Case 50</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.4</td>
<td>0.3</td>
<td>-0.4</td>
<td>-0.8</td>
<td>-0.4</td>
<td>-0.4</td>
<td>-0.4</td>
<td>-0.4</td>
<td>0.2</td>
<td>-0.2</td>
</tr>
</tbody>
</table>
The difference data of measurement point 245 and 260:

<table>
<thead>
<tr>
<th>Measuring point 245</th>
<th>Case 5</th>
<th>Case 6</th>
<th>Case 7</th>
<th>Case 25</th>
<th>Case 28</th>
<th>Case 29</th>
<th>Case 31</th>
<th>Case 32</th>
<th>Case 49</th>
<th>Case 50</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.28</td>
<td>0.58</td>
<td>0.2</td>
<td>-0.33</td>
<td>-0.12</td>
<td>0.56</td>
<td>0.12</td>
<td>0.31</td>
<td>0.62</td>
<td>0.27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measuring point 260</th>
<th>Case 5</th>
<th>Case 6</th>
<th>Case 7</th>
<th>Case 25</th>
<th>Case 28</th>
<th>Case 29</th>
<th>Case 31</th>
<th>Case 32</th>
<th>Case 49</th>
<th>Case 50</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.09</td>
<td>0.22</td>
<td>-0.26</td>
<td>-0.13</td>
<td>-0.04</td>
<td>-0.05</td>
<td>-0.06</td>
<td>-0.06</td>
<td>0.13</td>
<td>0.28</td>
</tr>
</tbody>
</table>

Figure 4.8 illustrates the difference data of measurement point 245 and 260 in the ten cases about coordinate point 11. The X-axis is the difference of measurements and results, and the Y-axis is the adjustment value. The blue dots in the diagram represent measurement point 245, while the red stars represent measurement point 260.

As we can see from the figure, the blue ones scatter when the adjustment value is -0.4, however the red ones cluster together in the same situation. That is to say, the difference data of point 260 are rather close in the cases where the adjustment data are the same in contrast to that of point 245. So the conclusion can be drawn that measurement point 260 is an important point in terms of coordinate point 11.
4.3.6 Results

In the previous sections, I took the adjustment point 11 for example to depict the process of analyzing the connection between measurement points and adjustment points. As there are more similar cases about the adjustment point 11 than that about any other adjustment points, I selected the adjustment point 11 as an instance. Besides, the analysis based on another adjustment points follows the same process, thus there is no need to repeat the analysis process again in the report.

The results are gained after sifting the candidate points. The results are listed as follows:
Adjustment coordinate 11 may have influence on measurement point FR0174CR/Y, FR0180CL/Y, FR0177CL/Y.

Adjustment coordinate 12 may have influence on measurement point FR0174CR/Y, FR0180CL/Y, FR0180CR/X, FR0177CL/Y.
Adjustment coordinate 27 may have influence on measurement point FR0073CR/X.

Adjustment coordinate 43 may have influence on measurement point FR0245AL/NOR, FR0206CL/Z, FR0210CL/Z, FR0040CL/Y, FR0040CR/Y.

Adjustment coordinate 46 may have influence on measurement point FR0210CL/Y.

Adjustment coordinate 47 may have influence on measurement point FR0073CR/X, FR0177CR/Z.

Adjustment coordinate 48 may have influence on measurement point FR0210CR/X, FR0180CR/X.

The numbers in the circle are coordinate point id in Figure 4.9, and the red circles label the affected area. If coordinate point 11 is adjusted, the measurement points in area A and B will be affected. Coordinate point 27 has influence on the points in area C.
5 The System Implementation

This chapter discusses the prototype and the system implementation. The first part presents how to implement the prototype in Matlab, and second part describes the web-paged CBR system implemented by the group in Mälardalen University.

5.1 Prototype implementation in Matlab

As discussed in the last chapter, the important attributes have been found. I weight these significant points, and save these weights in a vector:

\[
\text{Weight} = [w_1, w_2, ..., w_{290}],
\]

where \(w_n\) is the weight of the \(n\)th measurement point. If a certain point has no relationship with the adjustment points, its weight then is set as 0, otherwise, the weight is assigned to 1.

The measurement data and result data of the target problem are respectively stored in two vectors: Mea and Rslt. Then call the function:

\[
[\text{Distance},\text{rsltCase}] = \text{NNCase}(\text{testCase},\text{caseData},\text{caseClass},\text{weight}),
\]

where caseData is a matrix saving case database, testCase means the target case, caseClass is a vector storing case ID, and weight is the weight vector. The function calculates the Euclidean distances of target case from every case in turn, and the results are returned in the matrix Distance. The most similar case to the target case is saved in the variable rsltCase.

5.2 The system implementation

The group in Mälardalen University has implemented the CBR system [12] in C#, which is a web-paged application. The system uses the SQL Server 2008 as the database to save the previous successful cases. The measurement data of the target case have to be input by the engineers, and the weights are set according to the correlation analysis, as displayed in Figure 5.1. After pressing the Match button, the distances between the target case and each case in the database will be returned in the middle column and the right column will display any case
clicked on the middle column by the user with the proposed adjustments. The proposed adjustments will be shown in the right column by scrolling down the horizontal scroll bar. Since the system is still under development, the screenshot is the only screen shot of the system I got from the group.

Since the adaption of the proposed solution is complicated, the system leaves the work to the engineers. They will modify the prediction of the adjustment if it is not acceptable for the target case, and the system will save the modification as the new case in the database.
6 Conclusion and Future Work

6.1 Complications

In summary, there are some complications should be paid attention:
- measurement data for Fixture A or Fixture B
- matching the measurement data to the adjustments, and
- data formats and conversion.

The technology of case based reasoning is based on the correct data, and these issues are the prerequisite for getting the correct data.

At first I thought Fixture A and Fixture B were the same that incurred some mistakes in creating cases. After the visit of Volvo’s plant, I found I had some misunderstanding of the data. Fixture A and Fixture B, two models of crossmember, should be identical, but actually they are slightly different. The measurement data exported from CM4D contain the part identity, which is the essential information to distinguish Fixture A from Fixture B when creating the cases. But the previous measurement data I had got did not contain the part identity. So I had to collect the data again and recreate the database. The database now presented in the section 4.1 is the recreated database.

The finished body part is measured several times a day, and the data are saved in the computer with timestamp, while the adjustment data are just logged by the engineer, only recording the date. In this thesis, I thought the closest measurement done after an adjustment. However, mismatching the measurements with the adjustments is still a potential problem, which may consequently affect the results of the data analysis. This problem will be handled in the future work.

The last issue is the data formats and conversion. The data exported from the software CM4D is in the text format. The data are supposed to be converted into float format and saved as excel file. After the conversion, the data can be directly loaded into a matrix in the Matlab. Otherwise, the text data only can be stored in a cell array, which is hard to deal with. The issue is also will be solved in the future work. The measurement data may import into the CBR system from CM4D.
6.2 Conclusion

The work of this thesis is to build a Case-based system which can connect the measurements to the adjustment. The objectives of the system are:

- reuse the previous successful adjustment to give proposed solution
- predict the adjustment for the engineers
- secure the expertise of the engineers, and
- help train new operator.

Case-based reasoning is a strategy of lazy generalization, which generalizes cases until the target case given. I think that case-based reasoning is more suitable than the rule induction method for this thesis. There are two reasons: One, we have not got enough data to derive a set of rules or a model; two, knowledge elicitation is a difficult process, and now the factor that adjustment is very complex contributes to the rules generalization even more tough. However, case-based reasoning has its own shortcoming that fewer cases reduce the accuracy of the case retrieval.

Since the weighted nearest-neighbour is chosen as the retrieval method, the most work of the thesis is to find the significant points for weighting them from as many as 290 measurement points. Due to the inadequate data about some coordinate points, only a part of relationship between measurement points and coordinate points are found. As regards the validation of the results, it still needs more cases. The engineers gave some comments on the results that the relationships found possibly existed, but we should take into consideration that several adjustment points sometimes affect one measurement point because the adjustment is a complicated work. So the validation and further data analysis will leave as the future work.

6.3 Future work

The thesis is the first step of a big project at Volvo. There is some work to do in the future. The verification of the results needs to be done after collecting more new cases. Analyze the data further, especially focusing on the connection of several adjustment points to one measurement point which I ignored in the thesis. Besides, find some method to handle the data. I matched the measurements to
adjustments for creating cases manually, which cost a lot of time. The body parts are measured several times a day, while the adjustment data only record the date without the timestamp, it will incur some bad matching of the measurements, results and adjustments in a case. If the adjustment data are logged directly in the computer, the case creation can be done by some software. Meanwhile, it will avoid matching errors as fewer as possible. The last, continue to develop the CBR system mentioned in section 5.2.
References


http://www.mathworks.com, last access date: December 20\textsuperscript{th}, 2009.