Irrelevant Testers Removal for Recognition Task

Lucjan Janowski, VQEG, Atlanta 2010

November 16, 2010
Testers
Testers
Red Paint
Standard Solution

• Each subject judges the same movies
Standard Solution

- Each subject judges the same movies
- The score is a value from at least 1-5 range
Standard Solution

- Each subject judges the same movies
- The score is a value from at least 1-5 range
- Pearson correlation starts to be reasonable
Each subject judges the same movies
The score is a value from at least 1-5 range
Pearson correlation starts to be reasonable
If a subject does not correlate with other subjects he/she is removed
Each subject judges the same movies

- The score is a value from at least 1-5 range
- Pearson correlation starts to be reasonable
- If a subject does not correlate with other subjects he/she is removed
- What does it mean “a subject does not correlate?”
Standard Solution

- Each subject judges the same movies
- The score is a value from at least 1-5 range
- Pearson correlation starts to be reasonable
- If a subject does not correlate with other subjects he/she is removed
- What does it mean “a subject does not correlate?”
- VQEG assumes 0.85 - it is a very weak assumption
Recognition Task

- Each subject judges the same movies - *very difficult to obtain*
Recognition Task

- Each subject judges the same movies - very difficult to obtain
- The score is a value from at least 1-5 range - more probable is 0-1 answer
Recognition Task

- Each subject judges the same movies - \textit{very difficult to obtain}
- The score is a value from at least 1-5 range - \textit{more probable is 0-1 answer}
- Pearson correlation starts to be reasonable - \textit{not true}
Each subject judges the same movies - very difficult to obtain
The score is a value from at least 1-5 range - more probable is 0-1 answer
Pearson correlation starts to be reasonable - not true
If a subject does not correlate with other subjects he/she is removed - yes, but correlation has to be defined
Correlation

We cannot use correlation, why?
Correlation

We cannot use correlation, why?
Recognition Task

- Each subject judges the same movies - *very difficult to obtain*
- The score is a value from at least 1-5 range - *more probable is 0-1 answer*
- Pearson correlation starts to be reasonable - *not true*
- If a subject does not correlate with other subjects he/she is removed - *yes, but correlation has to be defined*
Plate Recognition

The experiment was designed in such a way

- Three different views are shown. The original, cropped, and cropped and rescaled
Plate Recognition

The experiment was designed in such a way

- Three different views are shown. The original, cropped, and cropped and rescaled
- Each view is shown in the original resolution and four times lower i.e. four pixels are changed to one
Plate Recognition

The experiment was designed in such a way

- Three different views are shown. The original, cropped, and cropped and rescaled
- Each view is shown in the original resolution and four times lower i.e. four pixels are changed to one
- For each type of sequence (particular view and resolution) five different bit rates are used to code the sequence
Plate Recognition

The experiment was designed in such a way

- Three different views are shown. The original, cropped, and cropped and rescaled
- Each view is shown in the original resolution and four times lower i.e. four pixels are changed to one
- For each type of sequence (particular view and resolution) five different bit rates are used to code the sequence
- We used constant QP value for sequences compression. The reason is that I frames have to be coded with low QP
Plate Recognition

The experiment was designed in such a way

- Three different views are shown. The original, cropped, and cropped and rescaled.
- Each view is shown in the original resolution and four times lower i.e. four pixels are changed to one.
- For each type of sequence (particular view and resolution) five different bit rates are used to code the sequence.
- We used constant QP value for sequences compression. The reason is that I frames have to be coded with low QP.
- The original sequence (not compressed) is not used since the plates reading is too easy in this case.
Plate Recognition

The experiment was designed in such a way

- Three different views are shown. The original, cropped, and cropped and rescaled
- Each view is shown in the original resolution and four times lower i.e. four pixels are changed to one
- For each type of sequence (particular view and resolution) five different bit rates are used to code the sequence
- We used constant QP value for sequences compression. The reason is that I frames have to be coded with low QP
- The original sequence (not compressed) is not used since the plates reading is too easy in this case
- The consequence of the above conditions is that the total number of HRC is 30
Plate Recognition

The experiment was designed in such a way

- Three different views are shown. The original, cropped, and cropped and rescaled
- Each view is shown in the original resolution and four times lower i.e. four pixels are changed to one
- For each type of sequence (particular view and resolution) five different bit rates are used to code the sequence
- We used constant QP value for sequences compression. The reason is that I frames have to be coded with low QP
- The original sequence (not compressed) is not used since the plates reading is too easy in this case
- The consequence of the above conditions is that the total number of HRC is 30
- We have 30 SRCs i.e. each subject sees each HRC and SRC only once!
Plate Recognition Interface

PLATE NUMBER: I DON'T KNOW COLOR:

BRAND:

- Audi
- BMW
- Citroen
- Daewoo
- Fiat
- Ford
- Honda
- Hyundai
- Kia
- Mazda
- Mercedes
- Nissan
- Opel
- Peugeot
- Renault
- Rover
- Seat
- Skoda
- Subaru
- Suzuki
- Toyota
- Volkswagen
- Volvo

SEND
The simplest subject quality metric is overall detection probability which is

![Graph showing overall detection probability vs tester ID]
SRC Detection

SRC strongly influences the overall detection probability
Assumption

If one can read plate numbers for particular PVS than for all PVSes generated from the same SRC, view, and resolution but with lower or equal QP, the plate numbers should be read correctly.
Assumption

If one can read plate numbers for particular PVS than for all PVSes generated from the same SRC, view, and resolution but with lower or equal QP, the plate numbers should be read correctly.

The assumption means that a partial order can be defined for the PVSes set.
If one can read plate numbers for particular PVS than for all PVSes generated from the same SRC, view, and resolution but with lower or equal QP, the plate numbers should be read correctly.

The assumption means that a partial order can be defined for the PVSes set.

It is not obvious so I have investigated this assumption manually.
Assumption Investigation
Assumption Investigation
Subject Quality Metric

\[ Sq_i = \sum_{j=1}^{30} ssq_{i,j} \]  

where \( ssq_{i,j} \) is subject sequence quality
Subject Quality Metric

\[ \text{Sq}_i = \sum_{j=1}^{30} \text{ssq}_{i,j} \]  \hspace{1cm} (1)

where \( \text{ssq}_{i,j} \) is subject sequence quality and is given by

\[ \text{ssq}_{i,j} = \begin{cases} 
0 & \text{if } \text{rec}_{i,j} = 1 \\
\text{n} & \text{if } \text{rec}_{i,j} = 0 
\end{cases} \]  \hspace{1cm} (2)

\text{rec}_{i,j} \text{ is the } j\text{th sequence and } 0 \text{ otherwise,}
\text{S} \text{ is the set of all subjects,}
\text{A} \text{ is a set of all sequences with the same resolution and view but higher or equal QP than } j\text{th sequence.}
Subject Quality Metric

\[ Sq_i = \sum_{j=1}^{30} ssq_{i,j} \]  

(1)

where \( ssq_{i,j} \) is subject sequence quality and is given by

\[ ssq_{i,j} = \begin{cases} 
0 & \text{if } rec_{i,j} = 1 \\
n & \text{if } rec_{i,j} = 0 
\end{cases} \]  

(2)

where

\[ n = \sum_{k \in S, l \in A_j} rec_{k,l} \]  

(3)

\( rec_{k,l} \) is 1 if \( k \)th subject recognized \( j \)th sequence and 0 otherwise, \( S \) is all subjects set, \( A_j \) is a set of all sequences with the same resolution and view but higher or equal QP than \( j \)th sequence.
Sq Metric’s Results

The histogram shows the distribution of Sq values for different numbers of subjects. The x-axis represents the number of subjects, while the y-axis represents the number of subjects with a particular Sq value. The highest number of subjects (5) have a Sq value of 2, indicating a significant concentration in this metric range.
## The Out layers Errors

<table>
<thead>
<tr>
<th>ID</th>
<th>Entered number</th>
<th>Original number</th>
<th>Possible error explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>KR102L</td>
<td>KR1002L</td>
<td>typo error</td>
</tr>
<tr>
<td>18</td>
<td>KR99ES</td>
<td>KR992ES</td>
<td>typo error</td>
</tr>
<tr>
<td>18</td>
<td>KR9764S</td>
<td>KR9764S</td>
<td>unjustified error</td>
</tr>
<tr>
<td>18</td>
<td>KR97645</td>
<td>KR9764S</td>
<td>similar character</td>
</tr>
<tr>
<td>40</td>
<td>KR308</td>
<td>KR3084M</td>
<td>probably typo error</td>
</tr>
<tr>
<td>40</td>
<td>KR439HS</td>
<td>KR439HA</td>
<td>typo error</td>
</tr>
<tr>
<td>40</td>
<td>RR2492K</td>
<td>KR2492K</td>
<td>typo error</td>
</tr>
<tr>
<td>40</td>
<td>KR3527</td>
<td>KR3527L</td>
<td>probably typo error</td>
</tr>
<tr>
<td>40</td>
<td>KR97645</td>
<td>KR9764S</td>
<td>similar character</td>
</tr>
<tr>
<td>48</td>
<td>KR6966N</td>
<td>KR6986N</td>
<td>probably typo error</td>
</tr>
<tr>
<td>48</td>
<td>KR450GF</td>
<td>KR150GF</td>
<td>probably typo error</td>
</tr>
<tr>
<td>48</td>
<td>KR249ZK</td>
<td>KR2492K</td>
<td>similar character</td>
</tr>
<tr>
<td>48</td>
<td>KR925JG</td>
<td>KR9253G</td>
<td>similar character</td>
</tr>
<tr>
<td>48</td>
<td>W67045W(albo)</td>
<td>W67045W</td>
<td>additional information</td>
</tr>
</tbody>
</table>
Generalization

A single character error can be justified. The solution is Levenshtein distance.

\[ Sql_i = \sum_{j \in A_j} ssq_{i,j} \] (4)

where \( ssq_{i,j} \) is \( i \)th subject quality according to \( j \)th sequence and is given by

\[ ssq_{i,j} = \begin{cases} 0 & \text{if } leb(i,j) \leq leb(j) \\ leb(i,j) - leb(j) & \text{if } leb(i,j) > leb(j) \end{cases} \] (5)

where \( leb(i,j) \) is Levenshtein distance of sequence scored by subject \( i \) and having lower or equal QP than sequence \( j \) and \( leb(j) \) is Levenshtein distance of \( j \)th sequence.
Sql Metric’s Results

![Sql Metric's Results](chart_image)

Lucjan Janowski, VQEG, Atlanta 2010
Irrelevant Testers Removal for Recognition Task
\textbf{Sq and Sql results}

- \textit{Sq} helps to reveal subjects making typo errors

Lucjan Janowski, VQEG, Atlanta 2010
Sq and Sql results

- Sq helps to reveal subjects making typo errors
- Sql shows that some subjects are willing to score “not recognized” very easily
Sq and Sql results

- Sq helps to reveal subjects making typo errors
- Sql shows that some subjects are willing to score “not recognized” very easily
- Subjects make some strange errors probably by almost not seeing the sequence
Sq and Sql results

- Sq helps to reveal subjects making typo errors
- Sql shows that some subjects are willing to score “not recognized” very easily
- Subjects make some strange errors probably by almost not seeing the sequence
- It is difficult to say that a subject should be removed, more likely some of his/her answers should be
**Sq and Sql results**

- **Sq** helps to reveal subjects making typo errors
- **Sql** shows that some subjects are willing to score “not recognized” very easily
- Subjects make some strange errors probably by almost not seeing the sequence
- It is difficult to say that a subject should be removed, more likely some of his/her answers should be
- It is our goal to build a better interface
True Subjects Set Quality
Plate Recognition Interface
Both $Sq$ and $Sql$ metrics can be generalized.
Both $Sq$ and $Sql$ metrics can be generalized.

In $Sq$ case it is easy since we only need a partial order in the experiment. We should build the experiment in such way that there is partial order!

Manual investigation is still needed but it is limited to the marked sequences and subjects. More answers to a single PVS make the metrics more precise. Moreover, we can remove assumption and use only the answers for the same sequence.
Both $Sq$ and $Sql$ metrics can be generalized

In $Sq$ case it is easy since we only need a partial order in the experiment. We should build the experiment in such way that there is partial order!

In $Sql$ case we need a quality metric also, i.e. we have to be able to measure the error strength.
General Metrics

- Both $Sq$ and $Sq/l$ metrics can be generalized.
- In $Sq$ case it is easy since we only need a partial order in the experiment. We should build the experiment in such a way that there is partial order!
- In $Sq/l$ case we need a quality metric also, i.e. we have to be able to measure the error strength.
- Manual investigation is still needed but it is limited to the marked sequences and subjects.
General Metrics

- Both $Sq$ and $Sql$ metrics can be generalized.
- In $Sq$ case it is easy since we only need a partial order in the experiment. We should build the experiment in such a way that there is partial order!
- In $Sql$ case we need a quality metric also, i.e., we have to be able to measure the error strength.
- Manual investigation is still needed but it is limited to the marked sequences and subjects.
- More answers to a single PVS make the metrics more precise. Moreover, we can remove assumption and use only the answers for the same sequence.
Any questions/suggestions?
Lucjan Janowski
janowski@kt.agh.edu.pl