Automatic evaluation of examination tasks in the form of function plot

Wojciech Bieniecki, Sebastian Stolinski, Jacek Stando

Abstract - In this paper the approach to automatic exam sheets evaluation is presented. In our task the exam is written on paper forms and scanned. The bitmaps are processed to extract the specific task area and then the task is evaluated with aid of image processing and analysis. In this work we present the algorithm for automatic evaluation of function plot. The algorithm has been verified against the set of 47 sheets. Simultaneously the works have been graded by a qualified teacher.

Keywords – e-evaluation, image analysis.

I. INTRODUCTION

E-evaluation is a new idea in the field of didactics which aim is to enable automatic marking of the exams with aid of artificial intelligence and computer recognition systems, especially OCR and image understanding. E-evaluation software is a system which enables evaluation of examination answer sheets by the examiner on a computer screen rather than reading paper documents. In our previous paper we worked out the requirements for such a system [1]. The advantage of the e-evaluation system is better organization of the examination session, because it performs segmentation of the answer sheets into individual tasks, distributes the tasks to selected examiners, specialized in their domain (for instance, physics), evaluates of the tasks and sends them back to the examination board and verifies the evaluation, gathers statistics, and presents the results.

The e-evaluation method has been widely introduced in Great Britain and the USA. The experience gained by Examination Boards like AQA, OCR and EDTEXCEL in Great Britain and ETS in the USA suggests that introducing e-evaluation improves the quality and reliability of the exams.

Studying the impact of changes in the process of evaluation to its quality has shown that if only this change does not involve exam preparation, it can keep its original accuracy and increase its reliability [4, 9]. For example, Williams and van Lent [9] claim that introducing e-evaluation can improve exam quality by:

- providing complete anonymity for task assessment;
- suppress the halo effect – a way of solving by one student does not affect the way later tasks are evaluated;
- steadily spreading potential errors in assessing – tasks to assess to the examiners are assigned at random.

Since 2007 in Poland the project „E-EVALUATION” is carried out by the Central Examination Board [3].

Tasks that appear in examination papers may be formulated as “closed” or “open” [6]. A closed task is either a single or multiple choice test, or it consists of filling gaps with specific words or numbers. One of the incorrect answers presented as a choice in a multiple-choice test is known as distracter and the correct answer is called verstracter.

Computer-aided evaluation of such a task is easy. In case of electronic form it is simply comparing the result to the key answer. In case of a paper form, the scanned sheet is processed to recognize some marks (shapes "X" or "V") put by a student on a form.

Much more difficult is automatic evaluation of the open task, because the set of proper solutions may be infinite. Moreover, incomplete or inaccurate solutions should be graded with a reduced score.

One of the problems that has been discussed in the literature is recognition of the mathematical formulae. Most often pattern recognition methods are utilized for this purpose. The survey over the algorithms is presented in [2]. There exist only few commercial solutions of this problem. One of these systems is xMath Journal for TabletPC (with on-line recognition feature) which is a kind of scientific calculator.

Most of the algorithms for mathematical formulae recognition work in two steps:

1. detection of individual symbols by image segmentation and object indexation with use of pattern recognition methods;
2. structural analysis of relations (for example spatial) between the selected segments.

A classical approach known from OCR is used for recognition of some symbols. Initially the segmentation is performed and the classifier assigns the symbols to individual objects. In some publications [10] the hidden Markov models method is applied to reduction of processing steps. The structural analysis of the expression (and the correction in case of recognition errors) is feasible by application of two-dimensional context-free grammar [5, 7, 8].

The type of the open task that we take into account in this paper is drawing a plot of a given math function. The type of task often appears in final high-school or maturity exams of mathematics. The student should draw the plot in a printed coordinate system. Because the plot is drawn manually, the teacher checks if the drawn curve crosses some characteristic lattice points. In a case of complex plot (few curves or lines) the curve may be discontinuous and the student should mark the endings of the segments.

II. IMAGE SEGMENTATION ALGORITHM
For a segmentation of scanned exam sheets we use the pattern matching method, which uses the cross-correlation algorithm. As a query pattern a small extract of the image was used (with the word Zadanie (in Polish „task”) which appears in the top of the region of every task).

The algorithm runs in the following steps:
1. Find in the sheet image \((1...M, 1...N)\) all local maxima for pattern matching. Assign the coordinates of the points as \((x_1, y_1) \ldots (x_k, y_k)\).
2. For each local maximum \((x_i, y_i)\), find another local maximum \((x_j, y_j)\) for which \(x_j > x_i\). If the point was found, it means that the image region containing the task reaches the value of \(x_j\). If not, it means that \((x_i, y_i)\) is in the right column and the region reaches the right edge of the image \((M)\).
3. For each location \(o\) the local maximum \((x_i, y_i)\), seek the nearest local maximum \((x_j, y_j)\) for which \(y_j > y_i\) and \(|x_j-x_i|<\text{eps}\), \(\text{eps}>0\). If the point is found, the region reaches the value \(y_j\). If not - the \((x_i, y_i)\) is the last task in the page and the region reaches the bottom edge of the image \((N)\).

Two task were given for the group of students.

Draw the plot of the linear function given by \(y = 2x\). The score for the task was 0 or 1.

Draw the plot of the square function given by \(y = x^2 - 4x + 4\). The solution was scored from 0 to 2.

The plots have to be drawn in the coordinate system shown in Fig. 1.

III. PLOT EXTRACTION AND EVALUATION ALGORITHM FOR LINEAR FUNCTION

The algorithm uses three images for each evaluation:
- \(I_1\) – the image of empty coordinate system (Fig. 1)
- \(I_2\) – scanned image of the model solution (the teacher draws the plot on the examination sheet)
- \(I_3\) – scanned image of the student solution

Steps of the algorithm:
1. With aid of the cross correlation method find the best match \(o I_1\) and \(I_2\) images. Obtain \(I_t = I_1 - I_2\). Using mathematical morphology filters, remove the distortions (for instance, fragments of the coordinate system remained from imperfect scanner geometry) from \(I_t\) and leave only the plot curve (Fig. 2). Use a thinning algorithm to obtain 1-pixel thick line.

2. Do the same processing for \(I_3\) and \(I_1\) images. Obtain \(I_t\) image. Label \(S_i\) as a connected component in \(I_t\) image containing the image curve.

3. For \(I_t\) do the dilation (the structuring element is a circle \(r = 10\ \text{px}\)). Assign it as \(I_{rd}\) and \(S_{rd}\) as a connected component in the image.

4. Calculate, how much pixels from reference plot \(S_{rd}\) matches the pixels of test plot \(S_t\).

We use the formula:

\[
P_1 = \frac{|I_{rd} : I_{rd} \rightleftarrows 1|}{|I_{rd} & I_t : I_{rd} & I_t \rightleftarrows 1|}
\]

\[
P_2 = \frac{|I_{rd} & I_t : I_{rd} & I_t \rightleftarrows 1|}{|I_t|}
\]

\[
P = \frac{P_2}{P_1}
\]
It may happen, that the number of pixels $P_2$ is very small (the student drew only a few dots, not the whole plot). We compare the number of pixels in test and reference dilated plots. We set the condition experimentally:

$$P_1 \leq 100 \cdot P_2$$  \tag{2}

If the condition is not satisfied, the plot is not valid (we put the negative number). In all other cases $P$ is in range $(0, 1)$ and it describes the compliance rate between reference and query plots.

IV. PLOT EXTRACTION AND EVALUATION ALGORITHM FOR LINEAR FUNCTION

The evaluation of a square function is a bit more complex, because the student cannot draw the curve as accurately as a straight line. During the real exam he receives 1 point for calculation of the parabola extreme and the second point if the plot crosses at least three “crate” points. In case of the function $y = x^2 + 4x + 4$ we have the minimum at $(2, 0)$ and the crate points are $(1, 1), (3, 1), (0, 4), (4, 4)$.

The exemplary comparison of the grades given by the teacher and calculated by the computer system is presented in Tables 1-2.

![Fig. 5 Reference plot for square function. Crate points with neighborhoods are marked.](image)

### TABLE 1

**EXEMPLARY SAMPLES FOR THE ALGORITHM EVALUATION (LINEAR FUNCTION)**

<table>
<thead>
<tr>
<th>Sample code</th>
<th>Grade (teacher)</th>
<th>Compliance of images</th>
<th>Grade at threshold 0.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>800023</td>
<td>1</td>
<td>0.94</td>
<td>1</td>
</tr>
<tr>
<td>800030</td>
<td>1</td>
<td>0.77</td>
<td>1</td>
</tr>
<tr>
<td>800031</td>
<td>0</td>
<td>0.14</td>
<td>0</td>
</tr>
<tr>
<td>800037</td>
<td>1</td>
<td>0.66</td>
<td>1</td>
</tr>
<tr>
<td>800043</td>
<td>0</td>
<td>-0.68</td>
<td>0</td>
</tr>
</tbody>
</table>

### TABLE 2

**EXEMPLARY SAMPLES FOR THE ALGORITHM EVALUATION (SQUARE FUNCTION)**

<table>
<thead>
<tr>
<th>Sample code</th>
<th>Teacher Compliance of images Grade</th>
<th>Ext</th>
<th>Cr. E 1 2 3 4 ext crate</th>
</tr>
</thead>
<tbody>
<tr>
<td>800023</td>
<td>1</td>
<td>1</td>
<td>0.52 0.58 0.42 0.63 0.63 1 1</td>
</tr>
<tr>
<td>800024</td>
<td>1</td>
<td>1</td>
<td>0.45 0.62 0.49 0.62 0.57 1 1</td>
</tr>
<tr>
<td>800025</td>
<td>1</td>
<td>1</td>
<td>0.51 0.55 0.39 0.66 0.40 1 1</td>
</tr>
<tr>
<td>800026</td>
<td>0</td>
<td>0</td>
<td>-0.55 0.09 0.05 0.12 0.14 0 0</td>
</tr>
<tr>
<td>800027</td>
<td>1</td>
<td>0</td>
<td>0.40 0.60 0.54 0.71 0.29 1 1</td>
</tr>
</tbody>
</table>

For all 47 samples the average error rates have been calculated (Tables 3-4). The error occurs if the teacher’s score is not the same as the computer’s. In the case of real exam, if the grade is under the pass level the work is evaluated once again. The undesirable situation is when we accept the improper task solution (overestimation).

### TABLE 3

**AVERAGE ERROR RATES FOR LINEAR FUNCTION (47 SAMPLES)**

<table>
<thead>
<tr>
<th>Threshold 0.25</th>
<th>Threshold 0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard error</td>
<td>Relative error</td>
</tr>
<tr>
<td>Compatible results</td>
<td>44 6.38%</td>
</tr>
<tr>
<td>Overestimated score</td>
<td>0 0%</td>
</tr>
<tr>
<td>Underestimated score</td>
<td>13 27.6%</td>
</tr>
</tbody>
</table>

### TABLE 4

**AVERAGE ERROR RATES FOR SQUARE FUNCTION (47 SAMPLES AT THRESHOLD 0.25)**

<table>
<thead>
<tr>
<th>Threshold 0.25</th>
<th>Threshold 0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard error</td>
<td>Relative error</td>
</tr>
<tr>
<td>E Compatible results</td>
<td>34 27.6%</td>
</tr>
<tr>
<td>f Overestimated score</td>
<td>0 0%</td>
</tr>
<tr>
<td>r Underestimated score</td>
<td>1 2.12%</td>
</tr>
</tbody>
</table>

V. THE EXPERIMENT

The experiment has been carried out on 47 students. Each student received a print-out with two empty coordinate systems and had to draw a plot of the linear function and a plot of the square function. All pages were scanned with the resolution 300 DPI and in 8-bit grayscale mode. Then the teacher assessed all the solutions, as during a normal exam. He also drew the plots that were used as reference data (the page was also scanned).
VI. CONCLUSION

The presented algorithm works fine in a case of linear function. Some errors occurred for solutions with strike-throughs and amendments.

In a case of square function the error rate is worse, but still the rate of overestimated solutions is quite acceptable.

What should improve the results is changing the exam sheets and introducing new method for grading the solutions.

Our future research will include evaluation of complex (spline) plots where the function is discontinuous and graphical solution of equation set.

REFERENCES