Fraud Detection in Selection Exams Using Knowledge Engineering Tools

Marcus de Melo Braga
Post-Graduate Program in Knowledge Engineering
Universidade Federal de Santa Catarina (UFSC)
Florianópolis, Brazil
marcus@egc.ufsc.br

Mario Antonio Ribeiro Dantas
Department of Informatics and Statistics
Universidade Federal de Santa Catarina (UFSC)
Florianópolis, Brazil
mario@inf.ufsc.br

Abstract—This paper proposes a method for fraud detection in automated selection exams, using knowledge engineering tools for identifying groups of answers with a strong indication of fraud, based on probabilistic evidence. Founded on an analysis of the wrong answers of the various candidates, the proposed method enables identification of suspicions, and evidence of fraud attempts through finding candidates with a significant number of identical wrong answers. This method of using knowledge engineering tools can be employed in various types of selection examinations, adapting to the characteristics and features of each one and contributing to the achievement of fair exams and free of fraud attempts.

Keywords: knowledge engineering tools; fraud detection; selection exams; multiple-choice questions.

INTRODUCTION

The need for detection of various types of fraud in several fields of application has grown in recent years, possibly due to the great developments in information and communication technologies (ICT) that provide citizens with diverse resources and technological facilities for access to databases of public and private organizations.

The growing complexity of equipment and services provided by new technologies hinders the identification of possible failure points or some susceptible to fraud attempts. In several areas of activities, there is a lack of study and research that enables the creation of methods and techniques to prevent fraud and to reduce financial losses caused by this type of crime. Knowledge Engineering (KE) provides us with tools that can assist in detecting and preventing fraud, helping us to fight this type of crime.

The application of knowledge engineering tools in fraud detection is already common in many fields of study. It occurs mainly in the following areas: mobile communications (cellular), finance, electronic commerce, healthcare, credit cards and government organizations [1], [2]. A systematic review on the subject of this research has not identified a study published in the databases Scopus (www.scopus.com) or ISI Web of Knowledge (apps.isiknowledge.com) associating the keywords "detection," "fraud," "exam," and "examination." Based on the review conducted in the two largest databases, we can state that we did not find any study of application of KE tools to detect fraud in selection examinations. Therefore, we propose a new possibility for application of these tools.

Some related studies apply others' software techniques for fraud detection. In [3], artificial immune systems are used to identify fraud attempts in online credit cards operations. Data-mining techniques can also be applied for fraud detection in e-Commerce [4], mobile communication networks [5], and in conventional telephony [6]. Other studies specifically propose statistical methods for fraud detection in several areas [7], [8].

The model presented in this paper for the application of KE tools in fraud detection in admission exams is structured into eight sections: Section II makes a brief introduction to selection examinations processes, presenting their main characteristics. Section III presents the main types of fraud that can occur in a contest aimed at giving an overview of the method to be proposed in this paper. Section IV examines the various types of questions that can be adopted in an examination. Section V presents a statistical study of the probability analysis of identical wrong answers (IWA) to a specific type of question: multiple-choice. In Section VI, we present a proposed method for fraud detection in automated selection processes. Section VII presents and discusses the results, and finally in Section VIII, we give conclusions and suggestions for future work.

II - SELECTION EXAMS

A. Definition

A selection examination process can be defined as an activity consisting of several sub-processes or steps, with the aim of selecting candidates through the application of questionnaires, tests, and other selection tools to fill a limited number of vacancies.

B. Phases of a selection exam

The phases, steps or sub-processes of a selection examination are necessary to meet statutory deadlines for publication, call for entries, preparation of tests, allocation of material and human resources, realization of exams, test
reading, test processing, and disclosure of results for the subsequent filling of vacancies by admission or hiring. Figure 1 shows the complete cycle of the phases of a selection process.

The limitation of the vacancies is precisely the main motivation of the contest; if there were no limits to the vacancies or the number of qualified candidates for completion was not significant, selection would have been unnecessary.

The process initiates with the disclosure activities. The purpose of the announcement phase is to enroll the largest number of candidates, giving chances to all interested in attending the event. After complying with legal requirements for the announcement, the next phase opens up the inscription or registration for all applicants who may be qualified for the selection exam. Once the inscriptions are finished, in the next phase, we make the arrangements for the preparation of the tests that will be applied to candidates; this is done through enjoining experts or specialized companies for the authoring of tests and examinations. At the phase of resource allocation, the candidates are distributed by each exam place; all the materials and human resources are allocated for the next phase, such as rooms, equipment, proctors, and coordinators. In the test application stage, there are major activities related to the security aspects of a selection process, including review of all the candidates with metal detectors during their access to the exams places, planning and proctoring the application of the tests, and the electronic monitoring of all the examination locations.

Once the tests are completed, all material is collected in a safe place – usually the institution’s information technology department – where the test reading phase is done with automatic readers, scanning responses of all candidates’ tests for later assessment. In the test processing phase, all test answers are processed using the templates and automated processing systems for getting the results of the tests, creating the scores for the candidates, and generating their respective places (classifications). Finally, after verifying and auditing the whole process of assessment, final reports are issued for disclosure of the results toward the registry or hiring of selected candidates.

III. FRAUD IN SELECTION EXAMS

Currently, security is one of the most relevant points in managing selection processes, given the considerable number of technological tools that tempt people to perform activities of fraud, especially in contests where there is a great competition of candidates per position.

Security issues in exams can now be handled from the time of enrollment by consulting special database lists maintained by various Brazilian governmental institutions that contain the names of all candidates already identified in prosecution of fraud attempts, throughout the national territory. Despite having their entries approved, such candidates can be identified from the time of enrollment and allocated in special rooms to be monitored in their activities during application of the tests.

At the stage of resource allocation, some security measures can also be taken, trying to find candidates with relatives in the same contest and separating them in different rooms or places for application of the tests, avoiding their proximity.

However, it is at the phase of test application in which security issues should have greater weight because it is precisely at that point that the largest number of fraud attempts occurs. All measures already mentioned – such as the use of metal detectors and electronic monitoring of electronic transmissions – are key in this phase.

At the stage of test reading, it is also necessary to adopt preventive measures, with rigid control of access to the processing data center and precautions against possible reading processing errors.

Finally, in the phase of disclosure of results, safety measures must be adopted with respect to the risks of information leakage or premature disclosure of the results, avoiding harming the reliability of the whole process.

Our practical experience in conducting selection exams over two decades shows that occurrences of fraud attempts are fairly common on any type of competition or selection process.

The attempts of fraud range from traditional and frequent acts of trying to cheat or plagiarize during the test, copying the responses of a competitor in the same room, and even the most daring attempts, carried out by means of electronic transmission. All these attempts must be considered and scrutinized by the authorities responsible for conducting the selection process, and all the omissions in these security cases are inadmissible. Absolutely nothing can justify the absence of mechanisms for its prevention or precaution.

IV. TYPES OF QUESTIONS

For a better understanding of the proposal for fraud detection presented in this article, it is necessary to describe the types of questions usually adopted in examination tests.

At first, the main types of questions used in tests can be grouped into three categories:
- Multiple-choice questions.
- True/false questions.
- Open questions.

Open questions significantly hinder the attempts of fraud since they require a response not coded and difficult to be transmitted—although not impossible. Open questions are the most difficult to process when adopted in contests with a large number of participants. Another disadvantage that this type of question presents is how to standardize criteria for evaluation, aiming to avoid disparities in the assessment by reducing the subjectivity and inconsistency among the different evaluators.

True/false questions inhibit the adoption of the fraud detection strategy proposed in this paper. With this type of test question, it becomes difficult to identify the "signature" or the statistical evidence of fraud, as will be shown below. For this type of question in particular, other strategies should be established for the automatic detection of fraud, taking advantage of their peculiarities.

Multiple-choice questions are still frequently adopted in selection processes for their tradition and simplicity [9]. One of their main advantages is the ease of representation on printed answer sheets that can be read by automatic readers. This procedure enables the processing of the large volumes of data collected in medium and large examinations.

The method for fraud detection proposed in this study assumes that the type of questions adopted in the contest is multiple-choice. This type of question presents an interesting feature that can be exploited for the automatic detection of fraud: We observe that when two or more candidates try to cheat on a test, the probability that their errors are equal is high; that is, the wrong answers are the same and the alternative letters answered incorrectly are also equal. Taking this observation as a basis, the method proposed in this paper makes a study of the statistical probability of simultaneous occurrence of errors on the same questions and the same alternatives to substantiate the mathematical evidence of a fraud attempt, further proposing KE tools to identify the occurrence of this problem, detecting and fighting some fraud activities in selection exams.

In the following section, we create a probability analysis of the simultaneity of the same wrong answers with the same wrong alternative letters in a test of multiple-choice questions using a total of 40 questions for a case study.

V. PROBABILITY ANALYSIS

To illustrate the method proposed in this paper, we assume that the test being examined has 40 multiple-choice questions with 5 alternatives (ABCDE). A sample of the possible answers can be seen in Figure 2.

The first line of the figure shows the position of each of the 40 questions in the test paper. The bottom line, displayed with highlighted background, corresponds to the template of these multiple-choice questions. The 10 lines between are examples of candidates' responses.

To further facilitate visualization, we can eliminate all the correct answers of the candidates, since they are of no interest to our study. Figure 3 shows only the wrong answers, already highlighted.

The research question for the analysis of probabilities is: What is the probability that two or more candidates err in a multiple-choice test with 40 questions (such as ABCDE) in the same question numbers and answering the same wrong alternative letters? Our hypothesis is that after a certain number of wrong answers exactly alike, the probability of occurrence is close to zero, providing formal evidence that there is a strong suspicion of a fraud attempt.

This probabilistic evidence substantiates the fraud detection method proposed in this study, applying KE tools for identifying high similarities in the wrong answers of the candidates in a selection exam that meets these characteristics and type of question.

C. Calculating the probability

For didactic purposes, it is necessary to define the concept of IWA. In this study, IWA are those special occurrences when two or more candidates miss the same questions, answering the same wrong alternative letters, resulting in a remarkable coincidence. One can intuitively predict that this coincidence has a significantly low probability of occurrence (i.e., it is almost impossible), and this fact sets up a strong indication of anomaly. We can interpret this occurrence as a "signature" of
the fraud attempt, a feature that makes it unique, singular, and – rightfully so – it should be investigated.

The probability of the simultaneous occurrence of IWA in an examination test can be expressed as follows:

Let:

- \( a \) - the number of alternatives in each question and
- \( k \) - the number of IWA,

then the probability \( P \) can be expressed as:

\[
P = \left( \frac{1}{a^k} \right)
\]

This equation represents a simplification of the problem and is based on the following assumptions:

- The questions are independent; that is, their responses do not interfere with the responses to other questions.
- The probability of each alternative (ABCDE) is equal.

Without this simplification, the solution would be far more complex, requiring a more detailed study of all proposed questions of a specific test to determine the probability of each question and each alternative individually. The proposed equation also implies that the probability of IWA does not depend on the size of the universe (i.e., number of candidates who have taken the test).

This equation demonstrates clearly that as the number of IWA increases, the value of its probability decreases significantly. We can perform the calculation in a small interval to analyze the behavior of the calculated probability. Table 1 shows the calculation of probabilities for the range of 1 to 10 IWA, considering a test with multiple-choice questions with five alternatives ("ABCDE," with \( a = 5 \)).

<table>
<thead>
<tr>
<th>IWA</th>
<th>( P ) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.000000000000</td>
</tr>
<tr>
<td>2</td>
<td>0.160000000000</td>
</tr>
<tr>
<td>3</td>
<td>0.006400000000</td>
</tr>
<tr>
<td>4</td>
<td>0.000256000000</td>
</tr>
<tr>
<td>5</td>
<td>0.000102400000</td>
</tr>
<tr>
<td>6</td>
<td>0.000004096000</td>
</tr>
<tr>
<td>7</td>
<td>0.000001638400</td>
</tr>
<tr>
<td>8</td>
<td>0.000000065536</td>
</tr>
<tr>
<td>9</td>
<td>0.000000000260</td>
</tr>
<tr>
<td>10</td>
<td>0.000000000001</td>
</tr>
</tbody>
</table>

From Table 1, we conclude that, if the number of IWA is equal to or greater than 3, the probability is already minimal (0.0064%).

VI. PROPOSED METHOD

We have applied a SQL (Structured Query Language) script and relational databases as a traditional Software Engineering tool for helping to detect some IWA. Nevertheless, the results using this technique were not always accurate due to different combinations of IWA and the absence of some IWA in some candidate’ responses.

There are many Knowledge Engineering tools that can be applied to detect this kind of fraud attempt with better accuracy. Among others, the main KE tools that can be used in this case are:

- Artificial neural networks (ANN).
- Artificial immune systems (AIS).
- Data mining.
- Case-based reasoning (CBR).

The proposed method is founded on applying case-based reasoning (CBR) as a data-mining tool. CBR is an artificial intelligence technique applied in solving problems and achieving learning, based on past experience such as the use of known cases to solve new cases [10].

CBR methodology fits the purpose of fraud detection proposed in this paper by presenting some features relevant to the problem presented. Among these characteristics, we can highlight the following [11]:

- A CBR system should be able to handle incomplete and highly specific queries.
- It can suggest appropriate cases, even when not all attributes are provided.
- It presents retrieved cases in a rational way; that is, avoiding excessive responses and respecting a limit of retrieved cases specified by the user.

These characteristics favor the application of CBR methodology in this type of fraud detection, since it allows retrieval of similar cases differently from traditional information retrieval, done through use of queries in relational databases.

The proposed method consists in designing a system of CBR for the identification of candidates with IWA above the minimum value of statistical probability (IWA = 3) for the retrieval of similar cases in an IWA file. The system starts out with a file containing all the wrong answers of several candidates, making an automatic sequential search of all similar cases until there is a recovery of more than one case where this similarity occurs. All cases recovered with high similarity will be listed for further investigation of attempted fraud.

As the volume of responses in a test can be quite high, it is necessary to adopt some criteria to restrict the sample studied in order to optimize queries. The first criterion to be adopted is to restrict the sample studied by eliminating the responses of candidates whose number of hits in the test is less than a minimum passing grade. Thus, all responses that have a
number of errors exceeding a certain percentage will be automatically discarded since these candidates failed to pass the competition. This measure considerably reduces the number of cases to be searched.

Another measure of restraint refers to the retrieval of cases in the database. It makes no sense to retrieve cases in which the number of IWA is less than a certain percentage; for example, 60%. The idea is to eliminate instances in which the number of IWA does not reach a certain percentage, which undermines the suspicion of fraud. Thus, we consider only the responses of the candidates whose percentage of IWA is equal to or exceeds 60%.

Looking closely at Figure 4, we note that there are some instances of IWA between the lines, as can be seen in Table 2.

![Figure 4. Relevant similar cases in the case base.](image)

**TABLE II. OCCURRENCES OF IWA IN THE FIGURE 4**

<table>
<thead>
<tr>
<th>Lines with IWA</th>
<th>N° of IWA</th>
<th>% of IWA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 and 8</td>
<td>5</td>
<td>100%</td>
</tr>
<tr>
<td>6 and 10</td>
<td>4</td>
<td>29%</td>
</tr>
<tr>
<td>7 and 9</td>
<td>1</td>
<td>10%</td>
</tr>
</tbody>
</table>

Although we have identified three instances (2-8, 6-10, 7-9) in which there is similarity among the cases of wrong answers, not all are relevant. If we consider the percentage of IWA in each of the events, we see that only one of them is of interest: line 2 with line 8 (Fig. 4).

The other events are not depicted with evidence of fraud since the percentage of IWA is less than a significant percentage. We can establish a minimum threshold value for this percentage (e.g., 60%), whereas in some cases the candidate who received a given template (crib note) can recognize that one of the answers is wrong and increased his or her number of correct answers or may even have tried to answer some questions on his or her own (chance), and it still is wrong, leading to wrong answers on the same issues but with different alternatives from the original source of cheat or crib.

The proposed method provides an algorithm to scan the responses meeting the restriction criteria predetermined. The process is sequential and begins with the first line, retrieving all its similar cases, then moving on to the next line (2) and doing new queries only from this starting point, reducing in every turn the universe to be searched and optimizing the process of the sequential search. Some CBR tools allow a case to be used as a query, facilitating the research process.

In a real application, we need the candidate identification number in addition to the answers of the multiple choice questions. This will enable us to identify those candidates who have a certain number of IWA.

To automate the fraud detection in an IWA file of multiple-choice questions, it is necessary to make a data-mining tool using an automatic scanning procedure on the IWA case base. As some CBR systems do not allow the creation of scripts to program this automatic scanning, one way to implement such a data-mining procedure would be through the creation of a similarity function, "Sim ()" in a relational database software, enabling the creation of an SQL script to perform the automatic scanning of all the IWA stored in the database, calculating their similarities, identifying which answers show a high similarity and, finally, retrieving them for inspection.

The similarity metric that best fits the proposed model is the average of similarities implemented in the CBR tool. Several simulations were conducted to find the best option for the case studied, including the codification of the alternatives as numeric values (ABCDE as 12345). None of them has better results than the model adopted.

**TABLE III. VALUES OF THE SIMILARITIES OF THE RETRIEVED CASES**

<table>
<thead>
<tr>
<th>Case</th>
<th>2</th>
<th>8</th>
<th>4</th>
<th>6</th>
<th>10</th>
<th>5</th>
<th>1</th>
<th>3</th>
<th>7</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sim()</td>
<td>1.00</td>
<td>1.00</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

The query on the case base done with the data shown in Figure 5 gives us the results returned by the CBR system used in this study, as presented in Figure 6.
Figure 6. Cases retrieved by the CBR tool.

Even when the query was submitted with one more wrong answer purposely introduced in the first column, it is clear that case numbers 2 and 8 were recovered with maximum similarity. The remaining cases that did not show a significant number of IWA had low similarity, proving the efficiency and effectiveness of the proposed model.

VIII. CONCLUSIONS AND FUTURE WORK

The methodology of CBR showed excellent performance in the detection of IWA, identifying all the cases in the case base used as a test. While there are some differences between the values of the query with the values found in the case base, the CBR software used in this study correctly retrieved all cases that showed the highest similarity.

We conclude that the method proposed in this study can be applied in detecting and preventing fraud in selection examinations that use multiple-choice questions, allowing the adoption of countermeasures in a timely manner, thanks to its ease of modeling and application.

One limitation of this study is the lack of a more detailed analysis of the behavior of the CBR software with respect to its performance with a case base with thousands of answers from a large group of applicants. However, based on other authors’ similar experiences with using bases of more than 200,000 cases [12], the CBR systems have a satisfactory performance. Thus, it is expected that the CBR solution here proposed presents a good performance even in the case of larger bases.

Future research could explore the application of the model proposed in this study for other kinds of questions and tests, adapting the methodology of CBR to the specificity of these applications. Others investigators can also expand this study, making the analysis of the behavior of other CBR software or others KE tools for a greater volume of data.

ACKNOWLEDGMENT

We want to thank Dr. Marcelo Sobottka of the Department of Mathematics at the Federal University of Santa Catarina (Brazil) for his valuable help in determining the probability of IWA in a test of multiple-choice questions.

REFERENCES


