Expert Systems With Applications 79 (2017) 236–243

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Gender detection and identifying one's handwriting with handwriting analysis

Murat Topaloglu*, Seher Ekmekci

Trakya University, Kesan Yusuf Çapraz School of Applied Sciences, Department of Computer Technology & Information Systems, 22800 Edirne, Turkey

A R T I C L E   I N F O

Article history:
Received 12 July 2016
Revised 28 February 2017
Accepted 1 March 2017
Available online 1 March 2017

Keywords:
Gender detection
Graphology
Data mining
Decision Tree
J48
ID3

A B S T R A C T

Today, graphology is seen as an experimental field of science that is dedicated to suggest ideas about diseases, profession choices, mood and characteristics of a person by investigating his/her handwriting. Graphology is in cooperation with medicine, psychology, sociology or other disciplines that are based on observation. Graphology is used for staff recruitment in business, diagnoses in medicine, identification of criminals in forensics, choosing a profession in education, guidance and counseling and other practices at every level of social structure. It is quite interesting that the number of the scientific studies on graphology is limited around the world, that there are no specific institutions providing education of graphology and that institutions except for a few international corporations do not benefit from graphology at all.

In terms of demographic properties, many statistical and mathematical analyses investigate similar and different variables. Especially, differences regarding gender have become subject to research. Therefore, detecting gender through handwriting can give pace to research in other disciplines. Moreover, the research can be useful in any field where gender detection is needed. This study fulfills two objectives. The first one is to find out whether a writer can identify his/her own handwriting. The second objective is to detect the gender of a writer of a text with the help of graphology and computer sciences. The impact of the study is reflected in the fact that findings can be used in fields where gender detection is needed, and that the detection is done with the help of expert and intelligent systems. At the end of the study, gender detection was performed for the individuals by making use of 133 attributes. Then, a decision tree and lists of rules were created with some algorithms. The purpose was to detect the gender of the person by making a character analysis of the handwriting with the help of decision tree formation methods in data mining. The analysis showed that it is possible to detect the gender of a person with the use of the specified attributes. The study reached a success level of 93.75% with ID3 algorithm.

1. Introduction

History begins with the utilization of writing (Munis, 2004). Writing is a unique and one of the most significant inventions of humanity. The analysis of writing is defined as graphology which is a method aimed at analyzing the character of a person by looking at the handwriting of a person. Research in this discipline is surprisingly promising. In the future, further developments will be achieved (Gökmener, 2009). Thanks to graphology, many people have been able to choose the occupation that they are skilled for, a great many diseases have been treated, a lot of problematic marriages have been prevented and employers are able to pick the employees with optimum performance (Altınköprü, 2005; Ketizmen, 2013). Although a quite many people are curious about graphology and find it interesting, they have no idea about where it is used exactly. In fact, graphology is in the foreground even if people are not aware of it (Altınköprü, 1999). Although graphoanalysis is a useful tool for a couple of reasons, the fact that the number of the studies is limited and that the criteria used for the analyses seem insufficient poses a problem.

Graphology can be used for people's choices of a profession in education. It can give clues about the children graduating from elementary school for their professional orientation (Dirks, 2000). The analyses are feasible for both children and adults. It is generally used for confirming the educational level of a child, monitoring the following periods, advising parents, educators and school management about his/her current level and future progress that fits her/his character best. Graphology is also proved to have some valid diagnosis properties (Dirks, 2000).

It is used in business to analyze the natural features of intelligence (e.g., to assess perception and discernment, the form and development of comprehension, to show selection skills for re-

* Corresponding author.
E-mail addresses: murattopaloglu@trakya.edu.tr, topaloglum@gmail.com (M. Topaloglu), seherekmekci@trakya.edu.tr (S. Ekmekci).

http://dx.doi.org/10.1016/j.eswa.2017.03.001
0957-4174/

sponses and regulating actions) (Gökmener, 2009). It is also utilized to understand a person's style of socialization, his/her level of self-control and whether s/he has got philanthropy, enthusiasm, sophisticated tendencies, etc. The effect of deliberate actions on handwriting can be analyzed to some extent with the use of graphology in legal system (Ataç et al., 2012).

Many mental and physical problems are known to influence handwriting. The examination of the notebooks kept for years simultaneously make it available to analyze the big changes in one's handwriting giving clues about one's character and state of health. The research shows that it is possible to get a check-up with the examination of one-year handwriting as a whole (Brewer, 1999; Kusiak, Kernstine, Kern, McLaughlin, & Tseng, 2000).

2. Data mining

Data mining is defined as the extraction of data from large-scale data. Thus, it is possible to show the relationship between the data and make forward-looking predictions when needed. For that reason, a new discipline was born to supply the need to access to intended, meaningful, usable data in large databases. The data help the decision processes of institutions with the development new strategies after it is analyzed with various statistical methods (Kaya & Köymen, 2008). Data mining is a multi-disciplinary system (Mackinnon & Glick, 1999). Data mining, in other words, is the discovery of the data in a database. Data mining is the data analysis techniques including statistics, artificial intelligence, machine learning, parallel and distributed processing and database management systems in the background where undiscovered and potentially useful and meaningful information is stored (Ayre, 2006; Berry & Linoff, 2000; Han, Kamber, & Pei, 2012).

3. Data mining models

The models used in data mining are divided into two groups, which are Predictive and Descriptive (Yao & Zhong, 1999). In predictive models, a model is developed from the recent and historical data which aims to make predictions of results values for unpredicted data sets (Özekes, 2003). In descriptive models, the goal is to define the current patterns that can be used for aiding the decision process.

4. ID3 and C5.0 algorithms

Decomposition process based on the common characteristics of the data is called classification. There are several classification methods available and decision tree is one of them. ID3 and C5.0 algorithms are decision tree formation methods based on entropy (Özkan, 2013). C5.0 has the property of pruning the unnecessary branches formed on the tree and it uses nominal attributes. In order to prevent overlearning, it uses "retrieval" ratio (Silahtaroğlu, 2013). This algorithm makes the formation of decision trees for the qualities with numerical values possible while ID3 algorithm classifies categorical qualities (numeric attributes) (Özkan, 2013). C5.0 algorithm is used in Weka and called "J48".

5. Related works

Hannad, Siddiqi, and El Kettani (2016) presented a new approach for the identification of the authors with the help of offline images of handwriting. By the proposed technique, the handwriting given in small pieces are examined and each piece is processed as a structure. The author of a script is analyzed with the histogram group calculated for all the pieces. The distance between two different scripts is compared with the use of the software.

In their research, Garz, Würsch, Fischer, and Ingold (2016) referred to new descriptive models which calculate the typefaces with geometrical arrangement and presented a simple and quick approach for author detection. The proposed descriptive models make the process easier compared the current methods which are also simple and productive and they can be applied to complex datasets.

Halder, Thakur, Phadikar, and Roy (2015) analyzed gradient images of 50 writers' handwriting to obtain a 64-dimensional feature of their writing styles. The classification tools of the Weka environment enabled them to recognize the individuality of characters, and in this way, they were able to identify writers with an accuracy of 99.12%.

In their study, Souza and Abe (2014) implemented the character detection system with an artificial neural network for the handwriting used for bank checks and numerical character analysis process.

Newell and Griffin (2014) also suggest that handwriting can yield some information about authors. The methods they employed, which were oBIF Column encoding and delta encoding, identified some texture-based characteristics in the sample's handwriting.

Vasquez, Travieso, and Alonso (2013) developed an off-line writer recognition system which managed to reach an accuracy level of 99.14% in a database of 100 writers. Their research design is noteworthy as it depends on graphology and forensic techniques instead of text features.

Herrera-Luna, Felipe-Riveron, and Godoy-Calderon (2011) presented a new approach for the detection of the author by looking at the handwriting. They utilized ALVOT algorithm to solve the problem.

Kaymaz, Gürsakal, and Eroğlu (2010) studied the signatures of business managers who sign very often in their busy schedules. They tried to explain the correlation between the fractal dimension of the signatures and the results of the personality tests. They concluded that there is no relation between the signature and personality.

Helli and Moghaddam (2010) developed a text independent method which obtained nearly 100% accuracy in detecting writers in a group of 100 participants. By using Gabor and XGabor filters, FRGs (feature relation graphs) are created for each person, and these FRGs enabled researchers to identify writers accurately.

Bangerter, König, Blatti, and Salvisberg (2009) studied the possibility of recruitment with the help of graphology.

In their study, Mutalib, Ramli, Rahman, Yusoff, and Mohamed (2008) developed fuzzy logic for handwriting feature extraction and then detected the feeling of the person. In this study, Mamdani extraction method can be used to determine whether the level of feeling control is very low, low, medium, high or very high.

In their study, Santana, Travieso, Alonso, and Ferrer (2008) presented an innovative biometric system for the handwriting graphology for writer detection. A Spanish word image database was developed in offline mode by 29 different people and it was found to be 92% successful.

The literature related to this study is shown in Table. 1.

6. System design and methodology

6.1. The aim of the study

People use handwriting quite often in daily life. Handwriting helps the detection of the personality traits in different areas. The characteristics of a person can be guessed by looking at his/her handwriting. The aim of this study is to investigate the possibility of the gender detection by analyzing handwriting and to observe

Table 1
Comparison of related works for handwriting analysis study till date.

Author	Features	Methodology	Conclusion
Hannad et al. (2016)	Texture based approach for identification of writers from offline images of handwriting	Texture descriptors including histograms of Local Binary Patterns (LBP), Local Ternary Patterns (LTP) and Local Phase Quantization (LPQ)	Writers can be identified with limited amount of text.
Garz et al. (2016)	Probability Density Function (PDF).	Difference-of-Gaussian (DOG) Interest Point detector	Simple and time-saving writer identification descriptors were listed
Halder et al. (2015)	64-dimensional feature	Gradient images of handwriting were analyzed for feature extraction	99.12% accuracy for LIBLINEAR with all writers
Souza and Abe (2014)	Recognizing handwritten numerical characters and magnetic ink character	Graphology and Graphoscopy techniques, and the analysis Paraconsistent Artificial Neural Networks.	High percentage of numeric character recognition was possible
Newell and Griffin (2014)	Oriented Basic Image Feature Columns (oBIFColumns) were used for writer identification	The Delta encoding	We demonstrate that the oBIF Column scheme on its own is sufficient to gain a performance level of 99% when tested using 300 writers from the IAM dataset.
Vasquez et al. (2013)	An off-line writer identification approach based on graphometrical and forensic features.	Graphology and forensic techniques	99.1% accuracy in identifying writers
Herrera-Luna et al. (2011)	A combination of line-level and word-level features were used.	ALVOT algorithms with a differentiated-weighting scheme	Descriptive features related to each writer were detected precisely
Kaymaz et al. (2010)	Fractal Dimension of Signature	Fractal dimensions of signatures	No relationship was observed.
Helli and Moghaddam (2010)	Gabor & XGabor Filter	FRG (feature relation graph)	When enough training data was provided, the study reached 100% accuracy.
Bangerter et al. (2009)	Using graphology in recruitment	Survey data is analyzed for statistical investigation	Graphology can be used to detect some desired personality traits
Mutalib et al. (2008)	Identifying emotion of a writer	Fuzzy technique	Handwriting of a person can yield their emotional status.
Santana et al. (2008)	A system for the handwriting graphology	Offline writer identification	Only 5 different graphological parameters were used and 99.34% accuracy was achieving

what results it will present and which criteria affect the analysis if it is possible.

6.2. The importance of the study

Through investigating and analyzing the data retrieved from the sample's handwriting, this study gathers graphology and expert & intelligent systems to detect gender. The current study helps to detect the gender of the author by analyzing the data collected from the handwritings and combining the science of graphology and modern technology.

6.3. Formation of the data set

Data collected from 39 females and 41 males, 80 participants in total, were analyzed. The participants were required to be at least high school graduates in order to participate in the study.

First step: collection of the handwriting samples

The participants were asked to copy the 5-page text on A4 paper in only one direction using a blue pen. The data was collected from the handwritings of the participants who copied the text they were given (Appendix-0). The text is not given in the article as it is a simple text in Turkish which was not considered important.

Second step: data collection survey

A survey with 10 questions was created to find the level to which the persons are able to recognize their own handwritings. It was filled by the participants from whom handwriting samples were collected (Appendix-1).

Third step: data analysis tool

The text that was written by the participants in the first step was re-examined by the researcher taking the survey into account and the results were recorded in the survey filled by the researcher (Appendix-2).

The following steps were used to detect the gender of the author from his/her handwriting within the framework of the study.

Firstly, the criteria for the character analysis were prioritized. According to research, the data analysis tool was decided to have eight subtitles as below;

- Pressure
- Borders and Spaces
- Dimensions and Baselines
- Slanting
- Handwriting Fields-Areas
- Logical, Intuitionistic, Solutioner
- Crown, Arch, Triangle, Fine Line
- Personal Image (T and D)

The descriptions of the data analysis tool subtitles are presented below:

Pressure: You need to apply more or less pressure on the paper in order to write. The amount of the pressure applied gives us clues about the individual and his/her state of mind. The observation of the pressure applied by the individual shows the amount of energy s/he has. The amount of energy used implies personal traits like being energetic, lively and tough. Although pressure differs from person to person, it can be divided into three categories, which are high, medium and low.

Borders: Borders not only defines the frame of the handwriting but also communication limits of the author. Some people live a full and busy life just as they leave no space between the lines. Others need time to concentrate. They need space in their lives just as some people who leave blank spaces on a page to make additions later on.

Spaces: Distent or intertangled lines symbolize our need for arrangement and design. Also, it implies our desire to become distanced and be together with others. Line blankings show the intimacy or even isolation from the surroundings. Inter-word spaces reflect the real desires in social relations with other people. The length between the words indicates the distance that the individual wishes to keep with others.

Dimensions and baselines: All of the handwriting must be seen and recognized. The lines going from left to right all over the page shows whether the person is determined to reach his/her goals. Type sizes and the links between the lines should be evaluated as a whole including all spaces in order to get a general picture of the organizing skills of a person (Mahony, 1997).

Slanting: In fact, slanting does not reveal one's feelings, yet it shows how quickly those emotions arise. Slanting is necessary to assess one's self-control (Mahony, 1997).

Handwriting fields-areas: Handwriting consists of three areas which show the physical, mental and social traits of a personality. Each letter has three areas which are called top, bottom and middle. Your handwriting and your body may have similar portions. The top area is similar to the head, which forms the intellectual part of the body. The middle part represents the body which provides social interaction. The bottom part stands for the biological drives, physical and sexual desires.

Logical, intuitionistic, solutioner: It is possible to observe the handwriting of a person just as we do for body language. The habits and individual behaviors redound on handwriting.

Crown, arch, triangle, fine line: The opposite way round the human behaviors that are tough and incomprehensible, handwriting can be quite clear and distinctive. You might think that handwriting is haphazard, but actually, it tells you something distinct and explicit. You may not know what to do in this case. These internal groupings in the handwriting are divided into four categories and they explain how we behave and communicate with others.

Personal image (T and D): Letters "t" and "d" represent the feelings that the author has for his/her self-image. The letters are the biggest indicators of the changes we make even if they are small. They help to reveal the emotional and physical self and assess self-respect. Letter "t" expresses self-esteem and self-sufficiency while "d" points out physical appearance.

7. Data processing

After the designation of the data collection survey and method, these scales were given to the participants. Data collection process is made up of three stages.

- Collection of Handwriting (Appendix-0)
- Data Collection Survey (Appendix-1) (in order to test whether the participants can recognize their own handwriting)
- Data Analysis Tool (Appendix-2) (in order to detect the gender of the participants from their handwritings with Weka program)

The participants were asked to write a text of five pages in handwriting. Data analysis tool consists of 133 attributes in total. These 133 attributes were transformed into numeric data and then into MS Excel spreadsheets. After that, the data collected from 80 participants were evaluated based on data analysis table and they were checked on the table.

Then, the following steps were carried out:

1. First, Appendix-1 list was transferred to Excel in order to obtain meaningful results. Each value in Appendix-1 were turned into a title. Then, all the data were transformed into the binary system and entered on the table. The data that the users marked and did not mark on the Appendix-1 list marked as 1 (true) and 0 (false), respectively.
2. Secondly, the handwriting of the participants in Appendix-0 was reevaluated on the same table and they were added to the table. The first column below each alternative in the table was marked by the participants while the markings the second column belonged to the researcher indicating the results after the analysis. After that, the same alternatives that both the participant and the researcher marked were found and they were compared to find out the percent which the participant can recognize his/her handwriting. Results for each participant were found. Although there were 10 questions in Appendix-1, the percentage results for the participants were obtained only for 9 questions because the participants were asked to use a blue pen while writing to make a better observation. Thus, the fifth question was eliminated from the evaluation.
3. After all the results were obtained from the participants, the percentages of the answers that were marked as 1 (true) both by the participants and the researcher were calculated. While percentages are calculated, correct answers given by a participant to nine questions from appendix-1 (except 5) were marked as 1, whereas incorrect answers were marked as 0. Then, correct answers were counted, multiplied by 100, and divided into 9 (Number of correct answers $*100/9$).
4. The threshold value was calculated by averaging the percentages of 80 participants. It was concluded with manual scanning that the ones above the threshold value can recognize their handwriting while the ones below the threshold value cannot do so. The threshold value was found 48.33%. 38 of 80 participants were above the threshold value. The data obtained so far forms the first part of the study and revealed that 47.5% of the sample was able to detect their own handwriting in manual scanning. This finding was retrieved by the researcher through investigating the appendix-1 form filled out by the sample. All the steps taken from this point forward are related to automated gender detection.
5. Appendix-2 was filled for each of the participants after the 5-page handwriting samples were examined by the researcher.
6. A table in Excel was created with the transfer of the Appendix-2 decision tree and list of rules to the digital environment to obtain results.
7. The table includes gender, whether the participant uses his/her left or right hand, Appendix-2 titles, subtitle items and item characteristics. The data was prepared by giving 1 (true) value to the items that were marked by the participants in Appendix-2.
8. During the transfer of this data to Weka, item characteristics were numbered from A1 to A133 and the gender column was transformed into "female" and "male" statements and Excel table was prepared.
9. Table input was transferred over to Weka software so that Appendix-2 data could detect gender through the system.

The algorithms that were applied in the study will be in accordance with Appendix-2. ID3 and J48 algorithms will be used in the study.

8. Experimental results

Results for manual handwriting detection

Firstly, results of the data retrieved from the questions in Appendix-1 are examined.

The mean of all the values	The number of the participants above the average	The percentage of the participants who were able to recognize their own handwriting
48.33	38	% 47.5

Appendix-1 shows to what extent a participant can recognize his/her own handwriting. According to the results, the mean for 80 participants was found 48.33% and 38 of 80 participants, or 47.5%

of the participants, in other words, were found to be able to recognize their own handwriting.

Results for gender detection through data mining

The aim of the Appendix-2 study was to detect the gender of the person by looking at his/her handwriting. The results were obtained from 134 attributes (133 + gender) in 80 participants. They were tested with ID3 and J48 algorithms on Weka program. Forming the rule set that was to be used for gender detection through handwriting correctly was the most critical step of the study. The percentage of the data that was trained to reach accurate results was altered to test its reliability. The amount of trained data was always intended to be maximized, so when decision-making, an optimal level was attained. Data selection procedure was executed by Weka sequentially.

8.1. J48 algorithm results

70% of the 80 data were used as training data and the rest 30% were used as test data in the decision tree created with J48 algorithm. The size of the tree was found 17 and it has 9 branches. The gender was observed to be estimated with a 70.83% success rate in the decision tree created by J48 algorithm test results.

The list of rules created with J48 is presented below:

List of Rules:

If A115 = No, A22 = Yes then Gender = Male
If A115 = No, A22 = No, A15 = No then Gender = Female
If A115 = No, A22 = No, A15 = Yes then Gender = Male
If A115 = Yes, A75 = No, A93 = Yes then Gender = Male
If A115 = Yes, A75 = No, A93 = No then Gender = Female
If A115 = Yes, A75 = Yes, A48 = No then Gender = Male
If A115 = Yes, A75 = Yes, A48 = Yes, A16 = No then Gender = Male
If A115 = Yes, A75 = Yes, A48 = Yes, A16 = Yes, A33 = No then Gender = Female
If A115 = Yes, A75 = Yes, A48 = Yes, A16 = Yes, A33 = Yes then Gender = Male

The attributes in the list of rules created by J48 algorithm are as follows:

A15 => (Borders and Spaces) => (Borders) => (Wide Borders)
A16 => (Borders and Spaces) => (Borders) => (Medium Borders)
A22 => (Borders and Spaces) => (Right and Left Borders) => (Wide Left Border)
A33 => (Borders and Spaces) => (Inter-word Spaces) => (Narrow Inter-Word Spaces)
A48 => (Slanting) => (Types of Slanting) => (A)
A75 => (Handwriting Fields-Areas) => (Bottom Area Length) => (Weak Connections)
A93 => (Crown, Arch, Triangle, Fine Line) => (Triangle-Angle) => (Normal Angularity)
A115 => (T and D) => (Cross Line Position and Shape) => (Umbrella Shaped Horizontal Line)

In the list of rules above, the first, second, third and fourth columns represent the attribute code, main heading, heading items and item attributes, respectively.

Based on the J48 algorithm results, kappa statistic, mean absolute error and root mean squared error were found 0.3913, 0.3056 and 0.485, respectively. The decision tree is shown in Fig. 1.

8.2. ID3 algorithm results

80% of the data set including 80 data and 133 attributes were used as training data and the rest 20% were used as test data. The gender was observed to be estimated with a 93.75% success rate with the ID3 algorithm and the data set. In other words, only one of the 16 data that made up the test data was misclassified. The list of rules created by ID3 algorithm is as follows:

List of rules:

If A115=No, A22=Yes then Gender=Male
If A115=No, A22=No, A75=No then Gender = Female
If A115=No, A22=No, A75=Yes, A90=No then Gender = Female
If A115=No, A22=No, A75= Yes, A90=Yes, A91=Yes then Gender = Male
If A115=No, A22=No, A75= Yes, A90=Yes, A91=No, A11=No then Gender = Female
If A115=No, A22=No, A75= Yes, A90=Yes, A91=No, A11=Yes then Gender = Male
If A115=Yes, A48=Yes, A117=No, A38=Yes then Gender = Female
If A115=Yes, A48=Yes, A117=No, A38=No, A67=Yes then Gender = Male
If A115=Yes, A48=Yes, A117=No, A38=No, A67=No then Gender = Female
If A115=Yes, A48=Yes, A117=Yes, A16=No then Gender = Male
If A115=Yes, A48=Yes, A117=Yes, A16=Yes then Gender = Female
If A115=Yes, A48=No, A102=No then Gender = Male
If A115=Yes, A48=No, A102=Yes, A4=Yes then Gender = Female
If A115=Yes, A48=No, A102=Yes, A4=No then Gender = Male

The attributes in the list of rules created by ID3algorithm are as follows:

A11 => (Pressure) => (Width versus Depth) => (Thin lines)
A16 => (Borders and Spaces) => (Borders) => (Medium borders)
A22 => (Borders and Spaces) => (Right and Left Borders) => (Wide Left Border)
A38 => (Borders and Spaces) => (Time spared for oneself) => (Moderate spaces within the words and inter-words)
A48 => (Slanting) => (Types of Slanting) => (A)
A67 => (Handwriting Fields-Areas) => (Middle area width) => (Moderate middle area)
A75 => (Handwriting Fields-Areas) => (Bottom area length) => (Weak connections)
A90 => (Crown, Arch, Triangle, Fine Line) => (Crown) => (Normal Crown)
A91 => (Crown, Arch, Triangle, Fine Line) => (Arch) => (Normal Arch)
A115 => (T and D) => (Cross Line Position and Shape) => (Umbrella shaped horizontal line)
A117 => (T and D) => (Cross Line Position and Shape) => (Horizontal line on the right side of the body)

In the list of rules above, the first, second, third and fourth columns represent the attribute code, main heading, heading items and item attributes, respectively.

Based on the ID3 algorithm results, kappa statistic, mean absolute error and root mean squared error were found 0.875, 0.0625 and 0.25, respectively. The decision tree is shown in Fig. 2.

9. Discussion

This study intended to detect writer's gender through their handwriting and achieved remarkable success. It was confirmed that gender detection through handwriting analysis was possible. To better understand its importance, this study should be compared with existing literature. In Hannad et al. (2016) study, texture-based descriptors were employed to identify writers. The method practiced mainly concentrated on a local approach that depended on texture analysis of fragments. Each font used was identified with Local Phase Quantization (LPQ) histogram, and used as

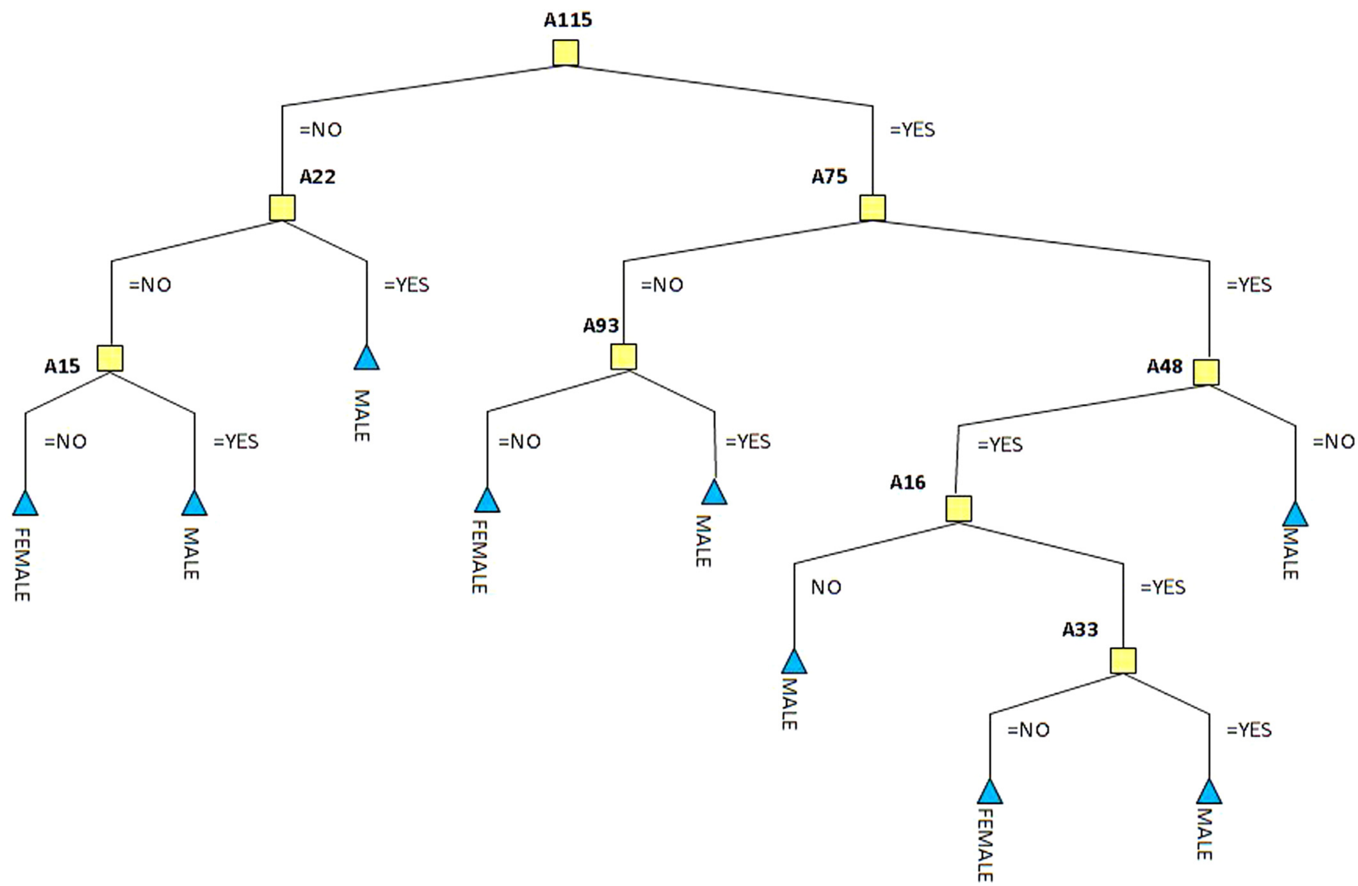

Fig. 1. J48 Decision Tree.

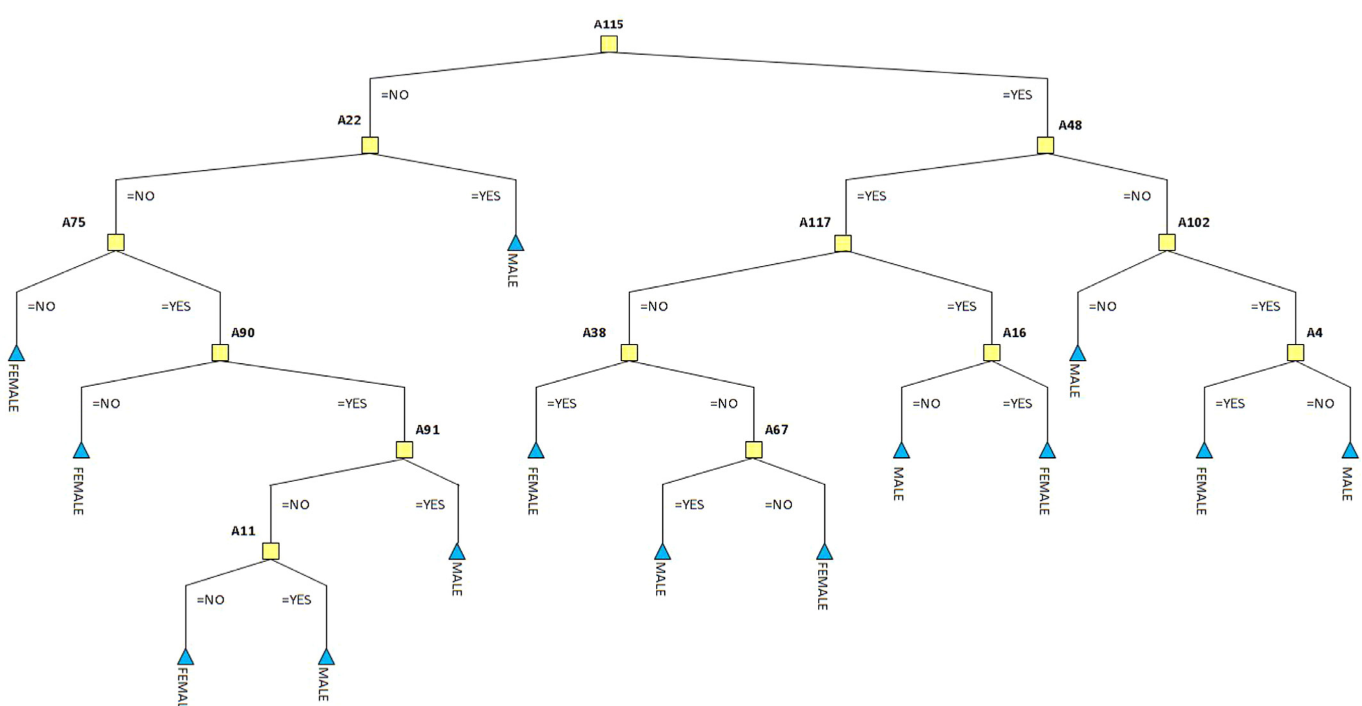

Fig. 2. ID3 Decision Tree.

an attribute during classification. In this study, writer identification achieved considerable success.

In another research design, Herrera-Luna et al. (2011) suggested a method rested on off-line graphicometric static attributes in writer identification. In this study, both line-level and word-level features were extracted from text images. The success rate of 92% was obtained through these graphometry features.

Newell and Griffin (2014) used a similar technique to that of Hannad et al. (2016) while identifying writers. oBIF (oriented Basic Image Feature) was used as a filtration approach to extract features out of texts. In a training set, an average oBIF histogram was calculated, and the difference between the histogram and the oBIF was used as a distinguishing descriptor. With IAM data set for 300 writers, the study reached 99% success rate.

Helli and Moghaddam (2010) used feature relation graph for each participant and developed Gabor and Extended Gabor features. This graph was formed through identifying relationships among extracted features with a fuzzy method. This showed the similarity of the features derived from different handwriting samples of each participant. The data was obtained from 100 participants' 5 pages of handwriting independent of the text. With 80% of trained data, identification accuracy reached 100%.

Mutalib et al. (2008), on the other hand, concentrated on personality and handwriting in their research, and questioned psychometric reliability and validity. It was confirmed that individuals tend to identify some certain personality features, especially emotional control, through their handwriting. It was also in this study that individuals' emotional control levels were identified through variable features and fuzzy rules. 30 handwriting samples were obtained for the experimental study. Writers agreed that the study was able to identify their emotional control through handwriting analysis.

The approach suggested in this study is different from the abovementioned literature as most studies focus on writer identification. Therefore, the literature review within this study focused on the closest research areas as there is no exact match in terms of research concentration.

Whether gender can be estimated through handwriting did not find a solid answer in previous research. Therefore, this study sheds light into future research. Handwriting changes by person and gender. Yet, with this research, changes have been transferred into a standardized system, so detecting gender through handwriting is much easier now. Furthermore, by having a machine to investigate handwriting, utmost precision can be achieved with the help of attributes mentions in the study. While trying to detect differences, the findings showed that detecting gender through handwriting was possible. Our study concentrated on gender detection and one's recognizing his/her own handwriting by using data mining techniques. The study achieved great success especially in verifying gender through handwriting analyses. In future research, a study design which investigates whether human resources departments of businesses can detect preferred personality traits through handwriting is planned. As a result, accuracy degree of labeling handwriting as brain-writing will be investigated. It is believed that this research will enlighten other research in the fields of human resources management, law, medicine, education, and etc.

10. Results

As mentioned earlier, the study has three stages. In the first stage, the participants were asked to copy a 5-page standard text in Turkish (Appendix-0). In the second stage, the participants were asked to fill a survey which asked questions about their handwriting characteristics (Appendix-1). In the third stage, all the data obtained from the participants was examined by the researcher and two algorithms were used to detect the gender from the handwriting.

The mean for the data obtained from 80 participants was found 48.33 for the Appendix-1. This numerical result showed that the participants were able to recognize their own handwriting with a 47.5%.

The second stage was the analysis of the data obtained from 80 participants for Appendix-2. A decision tree and two lists of rules were created in the study which included 133 attributes.

In the decision tree created with J48 algorithm, 70% of the data was used for training while the rest 30% was used for testing purposes and J48 algorithm was found 70.83% successful at detecting the gender of the author. Besides, 8 of the 133 attributes in the list of rules were found sufficient to detect the gender of the author.

Furthermore, 80% of the data was used for training purposes and 20% was used for testing purposes with the ID3 algorithm. The success rate of the algorithm was found 93.75% at detecting the gender of the author, which means that only one person's gender couldn't be detected among the 16 participants. 11 of 133 attributes were used to reach the list of rules for this algorithm. We can say that our hypothesis was verified with 93.75%.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.eswa.2017.03.001.

References

Altınköprü, A. (1999). Karakter Bilim ve İnsan Tanımada Testler. *Hayat Yayıncılık* ISBN: 9758243578.

Altınköprü, T. (2005). *Yazı ve Karakter*. İstanbul: Hayat Yayınevi ISBN: 9758243586.

Ataç, Y., Aydoğdu, E., & Bora, T. (2012). Adli Bilimlerde El Yazısının Kişiye Aidiyetinin Tespiti. *Polis Bilimleri Dergisi*, 113–132. http://www.acarindex.com/dosyalar/makale/acarindex-1423910970.pdf

Ayre, L. B. (2006). *Data mining for information professionals*. San Diego, California, USA. https://pdfs.semanticscholar.org/8f1c/ae76b3c3b3b1ade99b5afe9eb4e576652eef.pdf

Bangerter, A., König, C. J., Blatti, S., & Salvisberg, A. (2009). How widespread is graphology in personnel selection practice? A case study of a job market myth. *International Journal of Selection and Assessment, 17*(2), 219–230.

Berry, M. J., & Linoff, G. (2000). *Mastering data mining*. New York: John Wiley & Sons Inc.

Brewer, J. F. (1999). Graphology. *Complementary Therapies in Nursing and Midwifery, 5*(1), 6–14. doi:10.1016/S1353-6117(99)80065-1.

Dirks, H. (2000). Die Handschrift. *Deuten und Beurteilen*. München: Bassermann Inspiration. ISBN:9783572010141.

Garz, A., Würsch, M., Fischer, A., & Ingold, R. (2016). Simple and fast geometrical descriptors for writer identification. *Electronic Imaging, 17*, 1–12 2016.

Gökmener, A. (2009). *Graphology and employee selection* MSc thesis. İstanbul: Institute of Social Sciences, Yeditepe University.

Halder, C., Thakur, K., Phadikar, S., & Roy, K. (2015). Writer identification from handwritten Devanagari script. In *Information systems design and intelligent applications* (pp. 497–505). India: Springer.

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: concepts and techniques*. Elsevier ISBN:978-0-12-381479-1.

Hannad, Y., Siddiqi, I., & El Kettani, M. E. Y. (2016). Writer identification using texture descriptors of handwritten fragments. *Expert Systems with Applications, 47*, 14–22 2016.

Helli, B., & Moghaddam, M. E. (2010). A text-independent Persian writer identification based on feature relation graph (FRG). *Pattern Recognition, 43*(6), 2199–2209.

Herrera-Luna, E. C., Felipe-Riveron, E. M., & Godoy-Calderon, S. (2011). A supervised algorithm with a new differentiated-weighting scheme for identifying the author of a handwritten text. *Pattern Recognition Letters, 32*(8), 1139–1144. doi:10.1016/j.patrec.2011.03.002.

Kaya, H., & Köymen, K. (2008). Data mining concept and application areas. *Doğu Anadolu bölgesi araştırmaları, 6*(2), 159–164.

Kaymaz, K., Gürsakal, N., & Eroğlu, U. (2010). Analyzing the relationship between signature and personal traits: A research on managers. *Guc: The Journal of Industrial Relations & Human Resources, 12*(3), 29–40.

Ketizmen, A. (2013). Adli Belge İncelemede Güzel Sanatlar Akademisyenlerine Gereksinimler, Yeterlilikleri ve Uygulamadan Kaynaklanan Sorunlar. *Türk Eğitim Bilimleri Dergisi, 11*(3), 258–267.

Kusiak, A., Kernstine, K. H., Kern, J. A., McLaughlin, K. A., & Tseng, T. L. (2000). Data mining: Medical and engineering case studies. In *Industrial Engineering Research Conference* (pp. 1–7).

Mackinnon, M. J., & Glick, N. (1999). Applications: Data mining and knowledge discovery in databases–an overview. *Australian & New Zealand Journal of Statistics, 41*(3), 255–275.

Mahony, A. (1997). *El yazısı Örnekleriyle Karakter Analizi*. İstanbul: Say Yayınevi 1997. ISBN: 978-975-4682-22-9.

Munis, E. (2004). *Evrimi ile Yazı Sanatı*. Konya: Tablet Basım Yayın 1974. ISBN: 9759869101.

Mutalib, S., Ramli, R., Rahman, S. A., Yusoff, M., & Mohamed, A. (2008). Towards emotional control recognition through handwriting using fuzzy inference. In *International Symposium on Information Technology, 2008. ITSim 2008.: 2* (pp. 1–5). IEEE.

Newell, A. J., & Griffin, L. D. (2014). Writer identification using oriented basic image features and the delta encoding. *Pattern Recognition, 47*(6), 2255–2265.

Özekes, S. (2003). In *Veri madenciliği modelleri ve uygulama alanları: 2* (pp. 65–82). İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi.

Özkan, Y. (2013). *Veri madenciliği yöntemleri*. Papatya Yayıncılık ISBN: 978-975-6797-82-2.

Santana, O., Travieso, C. M., Alonso, J. B., & Ferrer, M. A. (2008). Writer identification based on graphology techniques. *IEEE Aerospace and Electronic Systems Magazine, 25*(6), 35–42.

Silahtaroğlu, G. (2013). *Veri Madenciliği: Kavram ve Algoritmaları*. İstanbul: Papatya Yayıncılık ISBN: 978-975-6797-81-5.

Souza, S., & Abe, J. M. (2014). Handwritten numerical character recognition based on paraconsistent artificial neural networks. In *Recent developments in computational collective intelligence* (pp. 93–102). Springer International Publishing.

Vasquez, J. L., Travieso, C. M., & Alonso, J. B. (2013). Using calligraphies features for off line writer identification. *2013 47th International Carnahan Conference on Security Technology (ICCST)*.

Yao, Y. Y., & Zhong, N. (1999). An analysis of quantitative measures associated with rules. In *Methodologies for knowledge discovery and data mining* (pp. 479–488). Berlin, Heidelberg: Springer.