

IDENTIFICAÇÃO DE AUTORIA EM MANUSCRITOS UTILIZANDO A CARACTERÍSTICA GRAFOMÉTRICA INCLINAÇÃO AXIAL

Nayane Batista Costa (PIC/UEM), Yuri Pires Alves (PIC/UEM), Aline Maria Malachini Miotto Amaral (Orientadora). E-mail: ra117189@uem.br.

Universidade Estadual de Maringá, Centro de Tecnologia, Maringá, PR.

Área e subárea do conhecimento: Ciências Exatas e da Terra / Metodologia e Técnicas da Computação.

Palavras-chave: computação forense; reconhecimento de padrões; documentoscopia.

RESUMO

A escrita como elemento biométrico tem sido alvo de muitas pesquisas e diferentes soluções computacionais para identificação de autoria em documentos manuscritos vêm sendo apresentadas na literatura. Nesse sentido, o suporte automatizado à extração e análise das características grafométricas, que são as utilizadas pelos peritos forenses, pode ser uma importante ferramenta de apoio ao trabalho desses profissionais. Uma característica importante e que normalmente apresenta resultados positivos é a Inclinação Axial, que representa o ângulo de inclinação da escrita. Sendo assim, este trabalho tem como objetivo desenvolver uma ferramenta que suporte a extração e a análise dessa característica grafométrica. Além disso, foram realizados experimentos para analisar o potencial da implementação em uma base de cartas forenses na língua portuguesa. Obtivemos resultados promissores em termos de taxa de identificação de autoria, destacando o uso das configurações do modelo SVM para classificação dos manuscritos, onde testamos valores para o parâmetro de regularização (-C) e diferentes tipos de *kernels*, incluindo *PolyKernel* com expoente (-E) variando de 1.0 a 3.0, e *RBFKernel* com gama (-G) variando de 0.01 a 1.0. Com o desenvolvimento dessa estrutura base, os índices de acerto nos testes feitos foram de até 100% para conjuntos menores de escritores, sendo esses de 5 e 10 autores, e 40,88% para 587 escritores.

INTRODUÇÃO

Na área da identificação de autoria, existem propostas de pesquisa com o objetivo de apresentar métodos que automatizam todo ou parte do processo de extração, análise das características e classificação da escrita humana. Uma das principais vantagens do uso de características grafométricas para a identificação de autoria em

manuscritos é sua aceitação no âmbito jurídico, uma vez que seus métodos de extração implicam na automatização das ações realizadas pelos peritos forenses (Amaral; Freitas; Bortolozzi, 2013).

Um dos elementos da grafia que deve ser observado, pois apresenta um alto poder discriminatório é a **Inclinação Axial**. Amaral (2014) explica que essa característica refere-se ao ângulo de inclinação da escrita, em relação ao eixo vertical de um sistema de eixos cartesianos, onde o eixo horizontal é representado por uma linha de base imaginária. Dentro do contexto apresentado, o objetivo deste trabalho é desenvolver uma ferramenta que suporte o processo de extração e análise de características grafométricas em documentos manuscritos. O processo de extração será composto pela implementação e avaliação do desempenho da **Inclinação Axial**.

MATERIAIS E MÉTODOS

Para a confecção deste trabalho, a base de dados trabalhada foi a *Brazilian Forensic Letter Database* (BFL), pois se trata de um acervo de cartas forenses constituído por amostras de manuscritos englobando todas as letras do alfabeto, números, acentos e pontuações. Atualmente, a base possui 600 escritores com 3 cartas atribuídas a cada um deles, escritas em folha A4 não pautada e digitalizadas em 300 *dpi*, com 256 escalas de cinzas (Freitas *et al.*, 2008).

Para a implementação da Inclinação Axial consideramos as coordenadas dos vértices dos contornos detectados em cada símbolo identificado pela biblioteca *OpenCV*. Após a extração e cálculo das inclinações axiais, os dados são armazenados e preparados para análise no *WEKA 3* em um arquivo de formato *.arff*. O processo adotado para a identificação de manuscritos segue fluxo geral de aplicação de reconhecimento de padrões: pré-processamento, extração da Inclinação Axial, definição do modelo, e posterior classificação. Nota-se que tanto as cartas conhecidas quanto às questionadas são submetidas ao pré-processamento e a extração da Inclinação Axial para possibilitar a criação do modelo e a atribuição de autoria.

RESULTADOS E DISCUSSÃO

Para realização de experimentos de validação com o *WEKA 3* e obtenção dos resultados, seguiu-se o protocolo experimental de Amaral (2014), onde o vetor de primitivas moldado durante a extração das características seria usado como entrada do algoritmo de classificação. Para consolidar os primeiros estágios de treinamento e teste, tal vetor de primitivas foi inserido ao classificador SVM no *WEKA 3* (Bouckaert *et al.*, 2009).

Nos experimentos, para o treinamento foram utilizadas as duas primeiras cartas de cada um dos 587 escritores processados com sucesso, pois algumas cartas forenses falharam nessa fase, totalizando 1.174 cartas distintas e, para os testes, utilizou-se a terceira carta de cada um, totalizando 587 cartas para esta etapa. Sumarizando, três cartas são randomicamente extraídas da Base de Cartas, as duas primeiras serão usadas para o treinamento do modelo e a última é usada para validar o teste. Apresentamos os resultados obtidos de forma tabular a seguir.

Tabela 1. Taxas de acerto

Número de Escritores	Taxa de Acerto (%)
5	100,00
10	100,00
20	85,00
49	67,34
99	63,26
199	48,24
397	42,82
587	40,88

Direcionando-se aos parâmetros e *kernels* utilizados para classificar manuscritos, foram testados diferentes valores para o parâmetro de regularização (-C) e diferentes tipos de *kernels*, incluindo *PolyKernel* com expoente (-E) variando de 1.0 a 3.0, e *RBFKernel* com gama (-G) variando de 0.01 a 1.0. Percebemos que, para um pequeno número de autores (5 e 10), a acurácia foi de 100%, indicando uma excelente performance do modelo. Todavia, conforme o número de autores aumenta, a precisão diminui gradativamente.

Para 20 autores, a acurácia variou entre 75% e 85%, enquanto para 49 autores, a porcentagem caiu para cerca de 61,2% a 67,3%. Com 99 autores, a taxa de acerto ficou entre 49,4% e 55,5%. Este padrão de declínio continua à medida do aumento do número de escritores, com 36,18% a 48,24% de acurácia para 199 autores, 36,02% a 42,82% para 397 autores, e finalmente 29,13% a 40,88% para 587 autores.

Essas métricas indicam que, embora o modelo SVM seja altamente eficaz para um quantitativo limitado de autores, sua precisão diminui significativamente à medida que o número cresce, sugerindo a necessidade de aprimoramentos e ajustes nos parâmetros ou no uso de técnicas adicionais para manter uma acurácia satisfatória em contextos mais complexos.

CONCLUSÕES

O processo de extração de características grafométricas e quais podem ser utilizadas perante um processo judicial, ainda se mostra uma área de exploração recente na Computação Forense. Este trabalho cumpriu o objetivo proposto diante do desenvolvimento realizado e os resultados retornados, obtendo um modelo capaz de identificar a autoria em manuscritos utilizando uma base de cartas através da característica de Inclinação Axial. Os testes feitos enunciaram descobertas promissoras, comparáveis à literatura, especialmente ao trabalho de Amaral (2014). Este estudo não só apresentou resultados semelhantes à tese citada, mas também explorou de forma mais abrangente novas maneiras de implementar as características grafométricas, com tecnologias atuais, apontando que há diversas abordagens que atendem ao mesmo problema. Para trabalhos futuros, sugere-se a implementação de novas funcionalidades, novas características grafométricas e testes adicionais com conjuntos e combinações diferentes de *features*, analisando as combinações feitas com métricas bem definidas.

REFERÊNCIAS

AMARAL, A. M. M. M. **Identificação de Autoria de Documentos Manuscritos Utilizando Características Grafométricas**, 2014. Tese (Doutorado em Informática) — Pontifícia Universidade Católica do Paraná, Curitiba, 2014.

AMARAL, A. M. M. M.; FREITAS, C. O. A.; BORTOLOZZI, F. *Combining multiple features based on graphometry for writer identification as part of Forensic Handwriting Analysis*. **Proceeding of International Document Image Processing**. Patras, Grécia: International Association for Pattern Recognition (IAPR), v. 1, p. 23-30, 2013.

BOUCKAERT, R. R.; FRANKE, E.; HALL, M.; KIRKBY, R.; REUTEMANN, P.; SCUSE, D.; SEEWALD, A. **WEKA manual for version 3-6-1**. Hamilton: University of Waikato, 2009. 212 p.

FREITAS, C. O. A.; SABOURIN, R.; BORTOLOZZI, F.; SUEN, C. Y. *A new database for writer identification and verification research*. **IEEE**. 19th International Conference on Pattern Recognition. [S.l.], 2008, p. 1–4.