I.J. Image, Graphics and Signal Processing, **2016, 5, 68-75**
Published Online May 2016 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2016.05.06

Author Based Rank Vector Coordinates (ARVC) Model for Authorship Attribution

N V Ganapathi Raju
Associate Professor, GRIET, Hyderabad, Research Scholar, JNTU Kakinada, India
E-mail: nvgraju@griet.ac.in

Dr. V Vijay Kumar
Dean, Professor of Computer Sciences, Anurag Institutions, Hyderabad, India
E-mail: drvvk_dean@cvsr.ac.in

Dr. O Srinivasa Rao
Associate Professor, Dept of CSE, JNTUK, Kakinada. Mail Id:
E-mail: osr_phd@yahoo.com

Abstract—Authorship attribution is one of the important problem, with many applications of practical use in the real-world. Authorship identification determines the likelihood of a piece of writing produced by a particular author by examining the other writings of that author. Most of the research in this field is carried out by using instance based model. One of the disadvantages of this model is that it treats the different documents of each author differently. It produces a matrix per each document of the author, thus creating a huge number of matrices per author, i.e. the dimensionality is very high. This paper presents authorship identification using Author based Rank Vector Coordinates (ARVC) model. The advantage of the proposed ARVC model is that it integrates all the author's profile documents into a single integrated profile document (IPD) and thus overcomes the above disadvantage. To overcome the ambiguity created by common words of authors ARVC model removes the common words based on a threshold. Singular value decomposition (SVD) is used on IPD after removing the common words. To reduce the overall dimension of the matrix, without affecting its semantic meaning a rank-based vector coordinates are derived. The eigenvector features are derived on ARVC model. The present paper used cosine similarity measure for author attribution and carries out authorship attribution on English poems and editorial documents

Index Terms—Threshold, Common words, Integration, ARVC model, SVD technique.

I. Introduction

Authorship attribution is the discipline of introducing and constituting characteristics of documents written by a particular author. It has many applications in day today life. Historically, human beings have shown a lot of interest and curiosity to identify the documents of unknown or disputed authors. This has created a great interest in deriving various methods of authorship attribution. History reveals a number of prominent cases where human beings have shown tremendous interest in authorship attribution as in the case of Hitler's Diaries published in 1983, the conflicts regarding its authorship, controversies concerning Shakespeare and his plays and the Telugu poet Vemana. Today, finding an anonymous author is not only the application of authorship attribution but it also finds a broad range of application, in areas such as, information retrieval (IR), computational linguistics, cybercrime, Natural Language Processing (NLP), and attribution of authors on the Internet etc.. The authorship attribution today deals with Internet anonymity as well [2].

In the olden days before computerization, researchers used to perform quantitative analysis of word usage, stylometric feature analysis, word frequencies, analysis of richness of vocabulary, etc. for authorship attribution. Analyzing the huge text of documents and considering a wide variety of documents written by the author, was a difficult task. The innovations of complex computerized techniques with artificial, neural intelligence and various classifiers, have become a great catalyst in deriving novel approaches for analyzing huge numbers of documents. The main problem involved is tractability of the data size and the number of potential authors considered. To achieve good performance, despite the above problems, one needs to derive novel approaches. If one considers all the terms of the authors, then it becomes a huge representation and the attribution results may take long time and may not be suitable for real time applications. The present paper overcomes the above obstacles, by deriving a novel approach for achieving good performance. In today's digital era, there is an unbelievable quantity of data available on the net, to the extent of 281 Billion Gigabytes. The indexed web contains at least 4.29 billion pages.

There is a common criteria between information, image retrieval and author attribution models. All these

models should derive significant features and apply measures to match the query content with the database of training contents [5, 8, 14, 15, 19, 20, 21, 22]. However, lexical matching methods can be inaccurate when used to match a user's query [5, 8, 17, 18] in author attribution models. There are many ways to express a given concept (synonymy). The literal terms in a user's query may not match those of a relevant document. These kinds of situations never occur in Information Retrieval. The prevalence of synonyms tends to decrease the "recall" performance of retrieval systems. In addition, most words have multiple meanings (polysemy), so terms in a user's query will literally match terms in irrelevant documents. Polysemy is one factor underlying poor "precision"[5, 8, 17, 18]. A better approach would allow users to retrieve information on the basis of a conceptual topic or meaning of a document [8]. Latent Semantic Analysis (LSA), is a method that relies on a mathematical technique, called singular value decomposition. It is used to identify patterns in the relationship between terms and concepts within an unstructured mass of text. LSA has a large number of practical applications in the field of information retrieval which includes information discovery, relationship discovery, authorship attribution, etc. [5, 8, 17, 18].

The present paper is organized as follows. The literature is presented in section two. The section 3 and 4 describes the methodology and results and discussion. The conclusions are presented in section 5.

II. Related Works

Each author has his own unique style of writing pattern, which is the signature (uniqueness) of that author. The stylistic choices of an author are far more difficult to capture and quantify compared to topic-related information. It is preferable to focus on a stylistic choices that are unconsciously made by the author and stable throughout the text. In natural language processing the stylistic features of the authors are extracted and quantified at lexical, syntactic, semantic and application specific levels. Most of the earlier authorship attribution studies were based on the following: gathering a suitable corpus, identifying significant features that capture writing patterns of the authors, extracting feature vectors from each corpus document axiomatically and building a supervised classification algorithms like Naïve Bayes, Support Vector Machines, Artificial neural nets, k-nearest neighbours and Decision Trees or unsupervised clustering algorithms like K-Means, hierarchical, Fuzzy C-Means to identify the author of an unknown document. Various researchers used statistical methods like multivariate analysis, principle component analysis and linear discriminant analysis. Various researchers used similarity/dissimilarity based methods for calculating pairwise similarity/dissimilarity between the unknown document and all the training documents [1, 23]. Other researchers used compression algorithms like Zip, PPM and various similarity measures for author identification [24].

The present model inherently uses latent semantic analysis (LSA) for authorship attribution. The LSA method is used by many researchers in various domains and applications [3, 4, 5, 6, 7]. Satyam et.al. [3] used LSA as a dimensionality reduction technique. Castillo et.al. [6] used LSA as a similarity measure for author identification task however the main disadvantage of this method is that preprocessing and identification of unique terms have not been done and that's why the results of the method is not good and it increased dimensionality. Victor Wennberg [7] used LSA for author attribution using dependency grammars. The main disadvantage of this method is the difficulty in deriving appropriate dependency grammars for the documents considered. Others [4, 5, 8] used LSA method for the purpose of indexing and ranking of the documents. Burrows [16] Delta method calculate the z distributions for a set of 150 function words on English Poems producing remarkable results. It has been demonstrated that it is a very effective attribution method for texts of at least 1,500 words. Hoover [15] suggested that the larger samples come from collected Early Poems: 1950-1970, while the smaller ones comes from Midnight Salvage: Poems, 1995-1998. This reminds us that some authors' styles change dramatically during their careers, and that some authors use very different styles in different texts. The study of Don Foster (1990) was to identify the authorship of Shakespeare's poem, "A Funeral Elegy". By applying a series of stylometric tests on this obscure poem, he found that the elegy was the work of William Shakespeare.

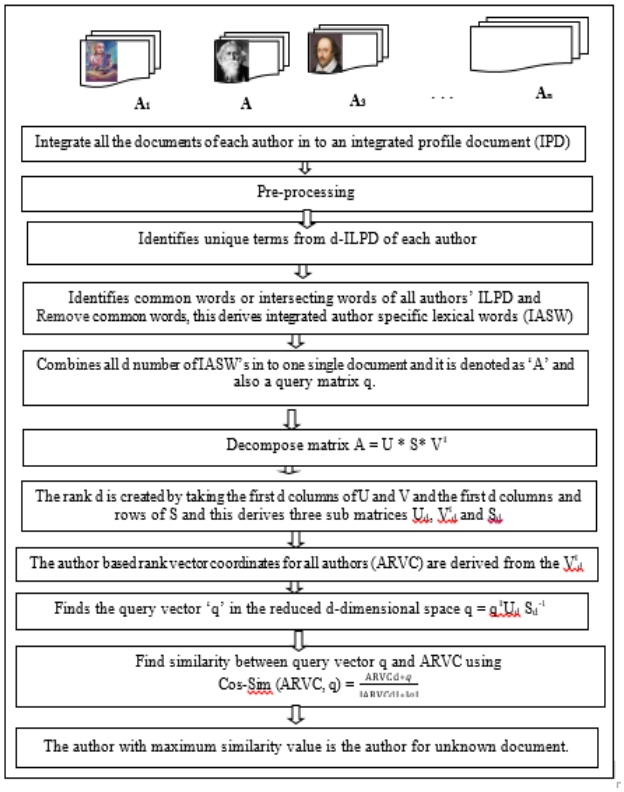

Fig.1. Block Diagram of ARVC Model

III. METHODOLOGY

The proposed ARVC model holds all the available documents of authors as one single document file. The proposed model builds d document files, where d represents the number of authors. These document files are used to derive the attributes of an author's style. The advantage of this representation is that it eliminates the separate representation of each text sample.

The proposed ARVC consists of 9 steps as given below.

Step 1: Integrate all the documents of each author into an integrated profile document (IPD). This leads to d-integrated profile documents (d-IPD), one per author, and d represents the number of authors. The size of the IPD of each author increases as the number of documents of the author increases. The number of IPD is directly proportional to the number of authors considered. The documents of an author can be of different sizes.

Step 2: A Preprocessing step is applied on all the d-IPD. The preprocessing is applied in two stages. In stage 1 the present paper removes spaces, numbers and special symbols or characters from each of the d-IPD. In stage 2 the proposed model filters stop words from each of the d-IPD of stage 1 and then stem them using a Porter stemming algorithm. This derives d-integrated lexical based profile documents (d-ILPD) one per author.

Step3. Identifies unique terms from d-ILPD of each author and evaluates the frequency of unique lexical terms of each author.

Step 4: Identifies common words or intersecting words of all authors' ILPD of step 3 and removes common words. This reduces the dimensionality. The common words of the authors create ambiguity in identifying the author. There are various approaches to remove common words from a document of each author. A few researchers identified a word as a common word if it appears in all author's profile documents. The present paper used a threshold 't' to identify common words and removed them from all ILPD. In our case t=d/2. This derives integrated author specific lexical words (IASW) by removing common words from each ILPD. This creates d number of IASLW.

Step 5: Combines all d numbers of IASW into one single document denoted as 'A' and also a query matrix q.

Step 6: Derives singular valued decomposition (SVD) on 'A'.

Step 6.1: Decompose 'A' into U, S and V matrices, which could be multiplied back together to give original matrix 'A'.

$$A = U * S * V^T \tag{1}$$

The present paper used JAMA package of Java to decompose matrix A into U, S and V matrices. JAMA is a basic linear algebra package for Java. It provides user-level classes for constructing and manipulating real, dense matrices.

Where U is an m * n matrix whose columns are the eigenvectors of the AA^T matrix, S is an n * n matrix whose diagonal elements are singular values of A and V is an n* n matrix whose columns are the eigenvectors of the A^TA matrix. The eigenvalues of U and V are always the same, so either of them can be used for S. The other property is of U and V are that $U^TU=I$ and $V^TV=I$, where I is the identity matrix.

Step 6.2: The resulting three matrices U, S and V are of high dimension. To reduce the dimensions of the above matrices without losing attributes, the present model used an approximation of rank d, where d is the number of authors. Rank d is created by taking the first d columns of matrices U and V and the first d columns and rows of S matrix and this derives three sub matrices U_d, V^t_d and S_d respectively.

Step 6.3: The author-based rank vector coordinates for all authors (ARVC) are derived in this paper from the V^t_d matrix of step 6.2. The rank is used as d, where d is the number of authors. Rows of V^t_d hold eigenvector values or coordinates of individual authors.

Step7. The present paper finds the query vector 'q' in the reduced d-dimensional space as shown in equation 2.

$$q = q^T * U_d * S_d^{-1} \tag{2}$$

Step 8: The similarity between query vector q and ARVC is found using Cosine similarity measure.

$$\text{Cos}-Sim(ARVC, q) = \frac{ARVC * q}{|ARVC| * |q|} \tag{3}$$

Step 9: The author with maximum similarity value is the author for unknown document.

IV. RESULTS AND DISCUSSION

The proposed ARVC is applied on a large set of poems, editorial documents and a combination of both of them of three cases.

Table 1. Attributes of IPD of Each Author

author	no of poems	before preprocessing	after preprocessing	
		size IPD in KB	size of IPD in KB	number of lexical words
JK	216	693	392	22952
RF	158	209	100	8122
RT	217	261	99.2	7721
SN	52	39	22.8	2592
WB	138	179	101	8399
WS	404	316	174	8544

Case1: The corpus of poems is collected from the website: www.poemshunter.com. The website consists 404,217,216,158,138 and 52 poems of William Shakespeare (WS), John Keats (JK), Rabindranath Tagore (RT), Robert Frost(RF), William Blake(WB) and Sarojini Naidu(SN) respectively. We have considered all

the poems of the above six poets for the sake of our experiment. The size of these poems ranges from 4 lines to four pages. Table 1 shows the basic attributes of the IPD of each author (combination of all poems of each author).

The proposed ARVC model is experimented with different combinations of training and randomly picked test poems (t, r) from the overall documents of each author. The t and r represent the percentage of poems picked randomly for training and test. The test (r) set is not a part of training data set (t). The present model treated all the training poems of an author as an IPD and applied pre-processing, and then removed common words, and constructed ILPD for all authors. On this SVD is applied. To reduce the dimensionality of the decomposed matrix, a rank is applied. The rank of ARVC in case 1 is 6 since we have considered 6 poets. The cosine similarity measure is applied to find the anonymous author.

The Table 2 and 3 shows the cosine similarity measure between the query vector derived from the test case poems and the ARVC of the training set. The tables are evaluated on 1 or 2 test samples per author. These test samples are not part of training database. The author attribution is carried out based on the highest similarity value, and it is shown in bold color in the tables.

Table 2. Cosine Similarity Values Between Query and ARVC for Poems

	Test samples of Poems of authors					
	JK 1	RT 1	RF 1	SN 1	WB 1	WS 1
JK	0.7957199	0.0601482	0.1221121	0.1005864	0.0921504	0.0234866
RF	0.7125939	0.7426384	0.7156782	0.0631342	0.2275352	0.1888486
RT	0.1614909	0.5177522	0.3757932	0.0742989	0.1539996	0.0070586
SN	0.4978918	0.1545169	0.0848349	0.9349703	0.4702816	0.0906435
WB	0.1939671	0.2345178	0.0606555	0.1697683	0.8330659	0.1391303
WS	0.3052226	0.312936	0.5663892	0.2781822	0.0292688	0.9675543

Table 3. Cosine Similarity Values Between Query and ARVC for Poems

	Test samples of Poems of authors					
	JK 2	RT 2	RF 2	SN 2	WB 2	WS 2
JK	0.635011	0.121805	0.188124	0.03772	0.077614	0.166362
RF	0.399827	0.840377	0.598648	0.448843	0.73598	0.263371
RT	0.312547	0.386737	0.681002	0.00382	0.483136	0.075663
SN	0.466525	0.21777	0.142518	0.849575	0.014918	0.419416
WB	0.305433	0.174783	0.411579	0.229075	0.450324	0.164159
WS	0.565054	0.226699	0.280996	0.151162	0.125994	0.833293

In table4 (t, r) represents percentage of poems picked randomly for training and test case of an author. For example (90, 10) of William Shakespeare represent 90% and 10% of (217, 24) poems chosen randomly for test and training cases, respectively, and the training set is not part of test data set.

Table 4. Percentage of Accuracy of Different Authors for Various (t, r) on Poems

	Different percentages of (t, r)							
	(95,5)	(90,10)	(85,15)	(80,20)	(75,25)	(70,30)	(60,40)	(50,50)
JK	100	85	90	88	84	88	81	80
RT	100	100	100	100	100	100	97	95
RF	50	40	50	50	25	30	40	40
SN	100	100	100	100	100	100	100	95
WB	50	40	40	50	40	40	60	40
WS	100	100	100	100	100	100	95	93
Verage accura	83.33	77.5	80	81.33	74.83	76.33	78.83	73.83

The accuracy results of the proposed ARVC with cosine similarity measure are shown in table 4. The accuracy is evaluated by the equation 4. The same is also plotted in Fig.3. in the form of a graph. The results have

shown that accuracy of author attribution slightly decreases by decreasing the training dataset 't' and increasing the number of test cases 'r'. We have also experimented the extreme cases with 50% of training test data set, where the test data set is not present in the training data set. The results are a little inconsistent among the authors and confirmed the general rationale of the proposed model.

$$\text{Accuracy} = \frac{\text{Number of poems whose author was correctly identified}}{\text{Total number of attempts}} * 100 \quad (4)$$

Additionally, we observed that poets are classified into two groups based on the graph of Fig.2. The first group of poets, John Keats, Rabindranath Tagore, Sarojini Naidu and William Shakespeare, have shown an approximate accuracy rate of above 85% with consistency. The other group of authors, Robert Frost and William Blake have shown an average of 50% with no consistency. Several factors may account for the discrepancy of performance between the two groups of authors, like different writing styles. The other major reason is that some of the poems of the above authors are of 4 to 8 lines. For choosing these poems (randomly) as a test case, the accuracy decreased drastically.

Since William Blake used a combination of varieties of styles in his poetry and since he is a mystic and visionary, it is difficult to identify the writer in his poetry. Robert Frost is an abstract poet. Since his religious and political beliefs which seen to be abstract for common readers, it is difficult to identify the author in his poetry.

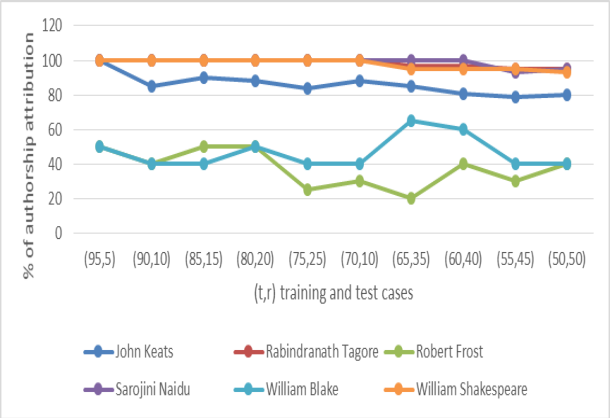

Fig.2. Accuracy of Various Authors with Different (t, r) on Poems

Case 2: The proposed ARVC framework is also implemented on 700 English editorial documents collected from various newspapers such as The Hindu, Times of India and Sunday Guardian of seven authors (100 documents each), namely, M.J.Akbar (MJA), C.P.Chandrasekhar(CPC), Chetan Bhagat (CB), C.R.L.Narasimhan(CRLN), A.S.Panneerselvan (ASP), C.Raja Mohan(CRM) and Tavleen Singh(TS).

The Table 5 and 6 shows the cosine similarity measure between the query vector derived from the test case poems and the ARVC of the training set. The tables are evaluated on 1 or 2 test samples per author. These test samples are not part of training database. The author attribution is carried out based on the highest similarity value, and it is shown in bold color in the tables.

Table 5. Cosine Similarity Values between Query and ARVC for Editorial Columns

	Test samples of editorial columns of authors						
	MJA 1	CB 1	CPC 1	CRLN 1	ASP 1	CRM 1	TS 1
MJA	0.610379533	0.21913768	0.054017697	0.084663057	0.193311036	0.141623217	0.038022572
CB	0.141874445	0.75363466	0.129180481	0.163370359	0.21506876	0.040351357	0.088651332
CPC	0.223722251	0.118708847	0.859080089	0.52228139	0.42057193	0.188532667	0.033374745
CRLN	0.397005073	0.272486799	0.474426614	0.818018611	0.493984733	0.094145214	0.178754833
ASP	0.188067266	0.273844669	0.075242983	0.04698008	0.631794251	0.110103186	0.053780312
CRM	0.490287521	0.254331325	0.106542535	0.138718011	0.241199456	0.950419793	0.063036706
TS	0.351983869	0.394964941	0.016806641	0.052537406	0.195277886	0.135964153	0.975070312

Table 6. Cosine Similarity Values between Query and ARVC for Editorial Columns

	Test samples of editorial columns of authors						
	MJA 2	CB 2	CPC 2	CRLN 2	ASP 2	CRM 2	TS 2
MJA	0.507159999	0.18133199	0.12536197	0.076469372	0.235887364	0.201246043	0.184510406
CB	0.099452298	0.626597096	0.132898773	0.165514661	0.231849175	0.03733075	0.193957188
CPC	-0.003200914	0.286312995	0.767569554	0.548852882	0.35563389	0.127237234	0.015427771
CRLN	0.189975417	0.477365122	0.484765925	0.807330729	0.364014979	0.165911484	0.10786377
ASP	0.195199858	0.225593287	0.190441765	0.055083146	0.710876338	0.05825898	0.197011286
CRM	0.708132921	0.286201043	0.272794908	0.050709814	0.262787575	0.943064958	0.268535735
(TS	0.396537328	0.363096172	0.178258641	0.090162645	0.239203398	0.14705403	0.897518257

The results are summarized in Table 7 and in Fig.3. for different percentages of (t, r). The proposed ARVC model achieved above 84% accuracy with different (t, r), where t ranges from 50% to 95% and r ranges from 50% to 5%. The results showed a consistence classification. We observed that the size of editorial document is large and not just like 4 lines of poems. That's why the editorial documents showed high consistency and accuracy of author attribution by the proposed ARVC. Since M.J.Akbar is a politician and a political commentator, is assessment of the achievements of different poets and their styles differed from those of other columnists.

Table 7. Percentage of Accuracy of Different Authors for Various (t, r) on Editorial Documents

	different percentages of (t,r)						
	95,5	90,10	80,20	75,25	70,30	60,40	50,50
M.J.Akbar	75	80	50	80	53	55	72
Chetan Bhagat	100	80	90	100	100	95	100
C.P.Chandrasekha	100	60	70	70	80	90	54
C.R.L.Narasimhan	100	100	100	100	86	90	81
A.S.Panneerselva	100	80	100	91	93	85	88
C.Raja Mohan	100	100	100	100	100	95	96
Tavleen Singh	100	100	100	100	93	95	92
	96.43	85.71	87.14	91.6	86.43	86.43	83.29

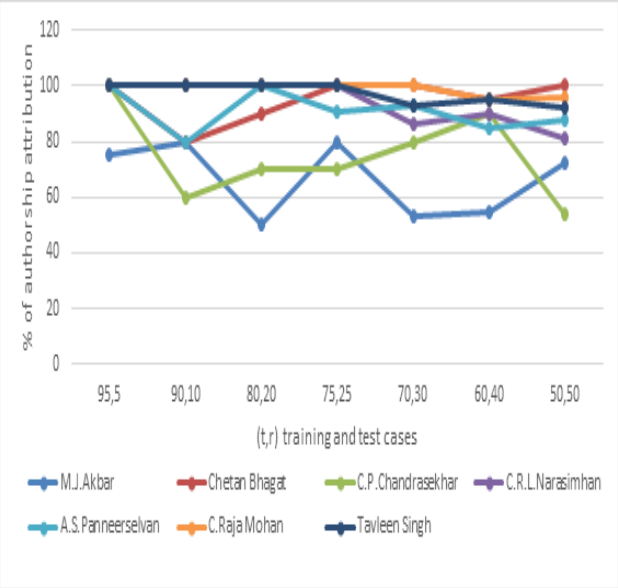

Fig.3. Accuracy of Various Authors with Different (t, r) on Editorial Documents

Case 3:- The proposed ARCV framework is experimented by combining poems and editorial documents with different (t, r) and the results are shown in the form of a graph in Fig.4. The author identification accuracy results are the same as the accuracy results of poems and editorial documents.

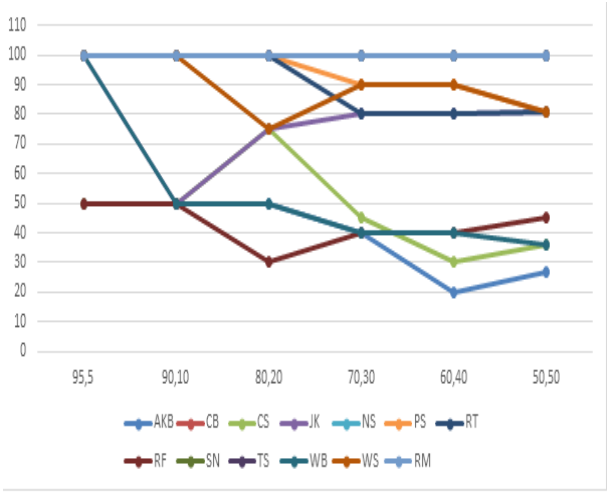

Fig.4. Accuracy of Various Authors with Different (t, r) on Poems and Editorial Documents

V. Conclusions And Future Work

In this paper we proposed a model for authorship attribution of English poems and editorial documents. To derive effectiveness of this model, we conducted experiments on different sets of English poems and editorial documents of well-known poets and editors. The experimental results showed the effectiveness of the proposed ARCV approach in identifying the poets and editors. The present paper addressed the issues related to tractability of the size of large documents by removing common words and by using a rank-based model. The proposed lexical word features after removing common words have shown high discriminatory capabilities for authorship identification on editorial documents rather than on poems. Thus the proposed ARCV model is also suitable for real time applications. One of the main reasons for this is that the size of the poems used in test cases is of 4 to 8 lines only. We believe the proposed model has the potential to assist in tracking identification in cyberspace, online messages, books, novels, etc. One of the future research directions on the proposed framework is that it can be applied on different languages.

Acknowledgment

The authors would like to express their gratitude to Management, Anurag Group of Institutions (AGOI), Hyderabad for providing necessary infrastructure for Centre for Advanced Computational Research (CACR) at AGOI, which is bringing various research scholars across the nation to work under one roof. The CACR is providing a research platform for exchanging and discussing various views on different research topics related to computer science. Authors extended their gratitude to Management, Gokaraju Rangaraju Institute of Engineering & Technology, Bachupally, Kukatpally, Hyderabad, India. This research has been supported by UGC under minor research project grant MRP-4590/14 (SERO/UGC) in March 2014.

REFERENCES

[1] Efstathios Stamatatos, "A Survey of Modern Authorship Attribution Methods", Journal of the American Society for Information Science and Technology, Volume 60 Issue 3, Pages 538-556, March 2009.

[2] P. Juola, "Authorship Attribution", Journal of Foundations and Trends in Information Retrieval, Vol 1, Issue 3, 2006, pp 233-334, **7** March 2008**.**

[3] Satyam, Anand, Arnav Kumar Dawn, and Sujan Kumar Saha, "A Statistical Analysis Approach to Author Identification using Latent Semantic Analysis", Notebook for PAN at CLEF 2014.

[4] Thomas K Landauer, Peter W. Foltz, Darrell Laham, "An Introduction to Latent Semantic Analysis", Discourse Processes, Volume 25, Pages 259-284, 1998.

[5] Scott Deerwester, Susan T. Dumais George W. Furnas Thomas K. Landauer, Richard Harshman, "Indexing by Latent Semantic Analysis", Journa of the American Society for Information Science 41 (6): Pages 391–407, 1990.

[6] Esteban Castillo, Ofelia Cervantes, DarnesVilari ño, David Pinto, "Unsupervised method for the authorship identification task Notebook for PAN at CLEF 2014".

[7] Victor Wennberg, "A Structural Approach to Authorship Attribution using Dependency Grammars", Bachelor of Science Thesis, Fall 2012.

[8] Barbara Rosario, "Latent Semantic Indexing: An overview", College of Engineering,, Michigan state university, Springer 2000.

[9] Edel Garcia, "Latent Semantic Indexing (LSI) A Fast Track Tutorial", 2006.

[10] Efstathios Stamatatos, Walter Daelemans, Ben Verhoeven, Martin Potthast, Benno Stein, Patrick Juola, Miguel A. Sanchez-Perez, "Overview of the Author Identification Task at PAN 2014", Proceedings of CLEF Conference on Authorship Identification, Sep 2014.

[11] Online edition (c) 2009 Cambridge UP" Matrix Decompositions and latent semantic indexing". Cambridge University Press, 2009.

[12] Latent semantic indexing—Wikipedia, the Free Encyclopedia, 2014.

[13] V.Vijay Kumar, N V Ganapathi Raju, O. Srinivasa Rao,"Histograms of Term Weight Feature (HTWF) model for Authorship Attribution", International Journal of applied Engineering Research, vol. 10, no. 16, pages 36622-36628, 2015.

[14] Moshe Koppel, Jonathan Schler, Shlomo Argamon. "Computational Methods in Authorship Attribution", Journal of the American Society for Information Science and Technology, Volume 60 Issue 1, Pages 9-26, January 2009.

[15] David L. Hoover, "Word Frequency, Statistical Stylistics, and Authorship Attribution", AHRC ICT Methods Network, Centre for Computing in the Humanities, 2006.

[16] Burrows, J.F.," 'Delta': A measure of stylistic difference and a guide to likely authorship", Literary and Linguistic Computing, 17(3), 267-287, 2002.

[17] Michael W. Berry, Susan T. Dumais, Todd A. Letsche**, "**Computational Methods for Intelligent Information Access", Proceedings of IEEE/ACM 1995 conference.

[18] Gavin W. O'Brien, Gavin W. O'Brien," Information Management Tools for Updating an SVD-Encoded Indexing Scheme", October 1994.

[19] T.V. Madhusudhana Rao,S.Pallam Setty, Y.Srinivas, "An Efficient System for Medical Image Retrieval using Generalized Gamma Distribution", I.J. Image, Graphics and Signal Processing,May 2015.

[20] Ibrahim S. I. Abuhaiba, Ruba A. A. Salamah, "Efficient Global and Region Content Based Image Retrieval", IJIGSP, vol.4, no.5, pp.38-46, 2012.

[21] Mohamed M. Fouad,"Content-based Search for Image Retrieval", IJIGSP, vol.5, no.11, pp.46-52, 2013.

[22] Hiroki Kobayashi, Masashi Toda, "Utilization of Textural Features in Video Retrieval System by Hand-writing Sketch", I.J. Image, Graphics and Signal Processing, August 2012.

[23] Moshe Koppel, Jonathan Schler, Shlomo Argamon,"Computational Methods in Authorship Attribution", Journal of the American Society for Information Science and Technology, Volume 60 Issue 1, Pages 9-26, January 2009.

[24] W. Oliveira Jr, E. Justino, L.S. Oliveira, Comparing compression models for authorship attribution", Forensic Science International, pages 100–104, 2013.

Authors' Profiles

N V Ganapathi Raju is working as an Associate Professor in CSE department, Gokaraju Rangaraju Institute of Engineering & Technology, Hyderabad. He is a Research Scholar under Dr.V.Vijaya Kumar Director - Centre for Advanced Computational Research (CACR) and Dr O.Srinivasa Rao, Associate Professor, College of Engineering JNTUK, Kakinada, A.P.,India. He received M.Tech (C.S.T.) from Andhra University. His research interests include Information Retrieval and Natural Language Processing. He got UGC minor project grant MRP-4590/14 (SERO/UGC) in March 2014.

Dr. Vakulabharanam Vijaya Kumar is working as Dean - Computer Sciences (CSE & IT) at Anurag Group of Institutions (AGOI), Hyderabad. He received integrated M.S.Engg, in CSE from USSR in 1989. He received his Ph.D. degree in Computer Science from Jawaharlal Nehru Technological University (JNTU), Hyderabad India in 1998. He has served JNT University for 13 years as Assistant Professor and Associate Professor. He is also the head for Centre for Advanced Computational Research (CACR) at AGOI, Hyderabad where research scholars across the state are working. He has received best researcher and best teacher award. His research interests include Image Processing, Pattern Recognition, Digital Water Marking, Cloud Computing and Image Retrieval Systems. He is the life member of CSI, ISCA, ISTE, IE (I), IETE, ACCS, CRSI, IRS and REDCROSS. He has published more than 100 research publications till now in various National, International conferences, and Journals and guided 28 research scholars for PhD. He has also established and also acted as a Head, Srinivasa Ramanujan Research Forum (SRRF) at GIET, Rajahmundry, India from May 2006 to April 2013 for promoting research and social activities.

Dr. O.Srinivasa Rao did his B.Tech in ECE, M.Tech in CS and Ph.D in Network Security and obtained the Ph.D from JNTU Kakinada. He published 16 international journal papers and guiding 2 Ph.D.'s. Presently he is working as Associate Professor in CSE department, University College of Engineering(A), Kakinada, JNTUK, Kakinada, Andhra Pradesh, India.

How to cite this paper: N V Ganapathi Raju, V Vijay Kumar, O Srinivasa Rao,"Author Based Rank Vector Coordinates (ARVC) Model for Authorship Attribution", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.8, No.5, pp.68-75, 2016.DOI: 10.5815/ijigsp.2016.05.06

