

Evaluating Performances of Two Unsupervised Methods for Classification Crime Text

Natiq M. Abdali^{1*}, Asaad Noori Hashim², Salah Al-Obaidi³

¹ College of Arts, University of Babylon, Babel 51001, Iraq

² Faculty of Computer Science and Mathematics, University of Kufa, Najaf City, Al-Najaf 54001, Iraq

³ Department of Computer Science, College of Science for Women, University of Babylon, Babel 51001, Iraq

Corresponding Author Email: art.natiq.mutashar@uobabylon.edu.iq



Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijss.140116>

ABSTRACT

Received: 11 December 2023

Revised: 15 January 2024

Accepted: 23 January 2024

Available online: 29 February 2024

Keywords:

crime news, crime classification, machine learning, text mining, crime analytics, text classification

Crime is a pervasive societal issue that has a negative impact on both a community's economic growth and overall quality of life. The bulk of crimes committed in everyday times are documented online by the wherewithal of news reports, blogs, and social networking sites. To improve crime analytics and community protection in response to rising crime. Law enforcement agencies continue to promote effective electronic information systems and crime data mining. Consequently, the aim of this study is to design a system of crime that depends on unsupervised machine-learning techniques that categorize five types of crime text. Two famous unsupervised algorithms: Independent Component Analysis based on natural gradient (NG-ICA) and Fast Independent Component Analysis (Fast-ICA) were used, to recover the latent components from observations. In order to evaluate the proposed system, the BERNAMA dataset, which had been manually annotated was used. Two experiments were conducted, and the results showed that the approaches that were employed satisfied promising results. Where the NG-ICA achieved an average F-measure of 83.3% and the Fast-ICA achieved 87.1%. This outcome demonstrates the appropriateness of these techniques in the implementation of text mining.

1. INTRODUCTION

Criminal information refers to material, data, photographs, or facts related to the investigation or prosecution of criminal exercise. In online newspapers and electronic archives, for example, valuable criminal information can be found in human-readable text format. However, there are still few software programs that can extract and offer pertinent information. The interesting in analyzing the crime domain is driven by its societal significance. This leads to increase the effort to reduce the spread of crime news and their risks. Researchers in the field of information extraction have paid close attention to particular detail [1]. However, manually controlling on the spread of crime-related data accessible on the internet poses significant challenges such as retrieving, analyzing, and utilizing the valuable information contained in these texts [2]. With the increased amount of the crime news found on the internet, processing this information manually is prone to be inefficient. Therefore, presenting an automatic tool to extract the information efficiently and effectively from the crime broadcast is required to meet the challenges of manual extraction [3]. The primary objective of this paper is to develop a model for extracting crime-related information from text sources. An unexplored framework for unsupervised crime recognition tasks by using Independent Component Analysis (ICA) is presented. In this research focuses on the development of crime text clustering in English-language

crime documents. ICA is unsupervised machine learning model that plays an important role in several domains such as biomedical engineering [4], text mining [5, 6]. Also has been extensively utilized in other various fields such as mobile communications, signal processing, audio signal separation, multispectral image demixing, and seismic signal processing. Therefore, ICA can be exploited to categorize crime news documents to be useful in identifying crimes that have been identify in the past and that will take place in the present. The remaining parts of the paper are arranged as follows: Section 2 presents the related works. Section 3 describes the proposed methodology. The proposed experimental setting and solution approach for crime classification are presented in Section 4, together with the experimental findings. Finally, the conclusions are presented in Section 5.

2. RELATED WORK

Several investigations into crime data mining have been carried out. The outcomes are frequently employed in developing software tools that are designed to detect and examine data relating to criminal activity. In this field, we can recognize two themes: Traditional models and machine learning-based models.

A rule-based strategy to extract information was described by the Rahem and Omar [7] on the basis of a set of drug

criminal gazetteers as well as a set of grammatical and heuristic rules. Experiments were used to validate this research. Results indicated that the method created here has promising potential. The Named Entities Extracted were the drug name, nationality, crime location, drug quantity, drug price, and drug hiding way. The better value of the F-measure was 0.96% for Drug names. A machine learning technique was used to construct a crime-named entity recognition system based on the extraction of crime scenes, weapons, nationalities, and online crime reports [8]. In order to extract the nationalities of criminals, locations of crime, and types of crime weapons from online crime documents. This system was developed based on Naïve Bayes (NB) and the Support Vector Machine (SVM) algorithms, were produces a good results. The results of the SVM algorithm were 91.08% of weapons, 96. 25% of nationalities, and 89.28% of crime locations. While the findings of NB were 86.73% of weapons, 94. 02% of nationalities, and 87.66% of crime locations. The work of ToppiReddy et al. [9] adopts a variety of visualization approaches and machine learning algorithms to predict the distribution of crime in a given area. Finding spatial and temporal crime hotspots is the main goal of Almanie et al. [10], where demonstrates how to predict potential crime types using the Decision Tree classifier and Naïve Bayesian classifier. In paper of Soni et al. [11] illustrates an innovative approach to locating the safest way to have the lowest risk score. In this study, the average risk score of clusters and regions is calculated using updated crime and accident data from NYC Open Data. Machine learning techniques were used to calculate the danger grade of a path based on the average risk rate of neighboring clusters and regions. In the study of Mata et al. [12] employed a semantic processing-based hybrid strategy to recover pertinent data from sources of unstructured data and a classifier algorithm to gather pertinent crime data from reports of the official government via a mobile application. In order to find safe routes, government reports that were geocoded and related to crime events in Mexico City were kept in a geospatial repository. The information gathered was used to forecast potential crime events that would happen in a specific location. The data mining approaches were applied to crime data for predicting features that implicate the high crime rate [13]. Decision trees, Naïve Bayes and Regression were used as techniques in data mining and machine learning on gathered data and therefore used for predicting the features responsible for causing crime in a region or locality. The Police Department and Crimes Record Bureau can take the appropriate steps to reduce the likelihood that a crime will occur based on the rankings of the features. The reasoning behind employing an unsupervised Independent Component Analysis (ICA) algorithm is in its capacity to group unannotated data into clusters. This study will present a strategy for categorizing crime documents into five clusters.

3. METHODOLOGY

This section offers the schema of text crime analysis to categorize a crime text news corpus of the data used in this research were collected from the Malaysian National News Agency (BERNAMA) dataset [3]. The proposed model for text crime news classification is depicted in Figure 1. There are three main steps were involved in a text analysis framework: Text preprocessing, text representation, and information discovery.

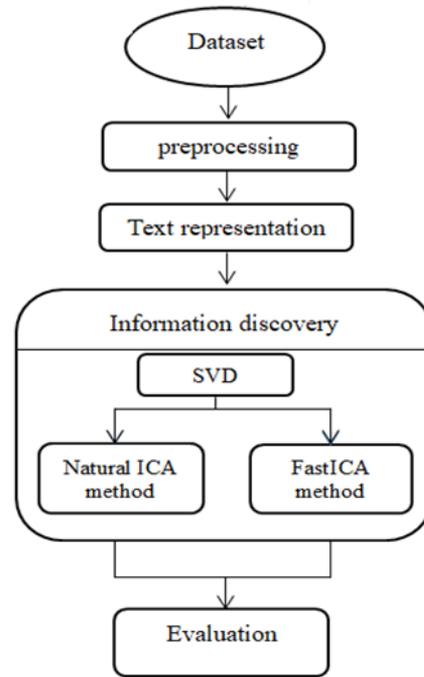


Figure 1. Architecture of the proposed model

3.1 Text preprocessing

Text is non-numerical unstructured data. As a result, computer applications cannot handle raw text easily. In order to index text, a document made up of sentences must be divided into specific phrases or single words [14]. The end result of using text indexing is a collection or list of terms. Three stages are used to outline the fundamental steps of indexing text. Tokenization is the initial stage in natural language processing and refers to the procedure of dividing text into smaller components known as tokens. Tokens encompass various linguistic units such as words, phrases, paragraphs, and so on. Tokenization is a technique used to divide text into meaningful pieces, which may then be efficiently processed by NLP models. The second stage involves stemming, which refers to the procedure of reducing words to their fundamental or foundational form by eliminating suffixes. As an illustration, the words "running", "runs", and "ran" can be reduced to the stem "run". Stemming aids in diminishing the word count in text and streamlining the lexicon [15]. The ultimate stage of text processing involves the removal of stopwords, which involves the elimination of frequently occurring words that contribute little meaning or information to the text. For instance, words such as "the", "a", and "and" can be used as examples. Stopword elimination aids in diminishing the superfluous content and magnitude of text, allowing for a more concentrated emphasis on significant terms [14].

3.2 Text representation

Encoding text is the process of transforming text into a format that is more organized. During the procedure of text encoding, numerous terms that originate from text indexing are selected as features, As the process of encoding each document progresses, numerical values are assigned to them. Text representation creates numerical vectors where use Vector Space Model (VSM) [16]. The VSM represent as in Eq.

(1). Variable in the VSM model has a different numerical weight [17].

$$D = \begin{bmatrix} T_{11} & T_{12} & \dots & T_{1N} \\ T_{21} & T_{22} & \dots & T_{2N} \\ \vdots & \vdots & \dots & \vdots \\ T_{m1} & T_{m2} & \dots & T_{mN} \end{bmatrix} \quad (1)$$

where, T is represent the terms, D is represent the text, m represent the number of words, and N represent the number of texts. The occurrences of terms corresponding to features in the given document are assigned as values of features in different schema, in this work, the formula for determining the weight of individual words within a document is denoted as Eq. (2):

$$w_{ij} = \frac{f_{ij}}{\sum_{i=1}^M f_{ij}} \quad (2)$$

The variable f_{ij} denotes the frequency of the occurrence of the word i within the text j . The aforementioned schema is utilized for the purpose of determining the level of importance of individual words within a given corpus of texts.

3.3 Singular value decomposition technique

A low-dimensional space which reflects semantic notions is created by words and documents using the automated indexing approach known as latent semantic indexing (LSI). In order to overwhelm the limitations of employing simply terms-based analysis, it analyzes text conceptually by projecting documents into a semantic space [18]. The fundamental goal of LSI is to generate a set of concepts connected with a collection of documents and the words they include in order to analyses the relationships between the documents and words. The LSI technique involves conducting a Principal Component Analysis (PCA) on the term-document matrix. Singular Value Decomposition (SVD) is a widely recognized method for performing this analysis, as noted in the study of Wang et al. [19]. The concept is that the SVD identifies a small number of "concepts" that connect the matrix's rows and columns.

3.4 Information discovery

ICA was initially designed to solve the problem of blind source separation. Its goal is to retrieve separate source signal from linear mixes of these signals as accurately as possible [20]. The process of independent component analysis involves linearly transforming multivariate data (observations) in order to minimize dependencies between variables, resulting in so-called independent components (ICs). As a result, the objective of ICA is to identify both the separating matrix and sources (ICs), and the problem is not identifiable if more than one component has a Gaussian distribution [21]. Negentropy maximization, a measurement of the non-Gaussianity of components, is obtained using ICA by decreasing the mutual information among component estimations [6]. Eq. (3) can be utilized to establish the ICA model in the following manner.

$$Z = AS \quad (3)$$

When the linear mixture is Z , the symbols " S " and " A " respectively represent the source signals and the constant mixing matrix whose constituent mixture coefficients are unknown. S is estimated using ICA as in Eq. (4):

$$S = W^T Z \quad (4)$$

where S is the estimated components, and W is the demixing matrix. Two methods of ICA will be employed in this system: natural gradient-based ICA, and FastICA algorithms.

3.4.1 Natural gradient based ICA(NG-ICA) algorithm

The application of natural gradient learning in the Riemannian parametric space of nonsingular matrices represents a highly accurate steepest-descent method. The natural gradient has been utilized in simple gradient descent algorithms within the context of ICA [22]. There are equivariant characteristics in the Fisher-efficient natural gradient algorithm (NGA). Therefore, whitening observed signals during NGA preprocessing enables Stiefel manifolds to utilize the natural Riemannian gradient. The loss function is denoted by Eq. (5):

$$E = -\ln|\det W| - \sum_{i=1}^n \log p_i(y_i(t)) \quad (5)$$

where, $y = WX_{(t)}$, and $p_i(\bullet)$ stands for the probability density function (pdf) of the latent variable $s_i(t)$.

The NG-ICA algorithm employs an iterative approach to minimize the loss function specified in Eq. (5). The formula representing the natural gradient is as in Eq. (6):

$$W_{(t+1)} = W_{(t)} + \eta [I - g(y(t))y^T(t)]W_{(t)} \quad (6)$$

The given expression involves a positive step size denoted by $\eta > 0$, and a vector function $g(y) = (g_1(y_1), \dots, g_n(y_n))^T$, where each element of the vector corresponds to the negative score function [22]. When dealing with source signals exhibiting positive kurtosis, it is observed that the function $g(y) = \tanh(y)$ generates the super-Gaussian components.

3.4.2 FastICA algorithm

The considerably popular linear ICA model is FastICA, which uses a fixed-point iteration algorithm. The effectiveness of this approach relies on the implementation of crucial preprocessing steps, namely centering and whitening, performed on the mixed data [23]. The parallel learning principle FastICA finds a direction unit vector w and uses the $w^T x$ projection to maximize the non-Gaussianity value. Non-Gaussianity is assessed using the negentropy approximation [24]. The following constitutes the fundamental formula for the parallel FastICA:

1. Initialize the weight vector W in a random manner.
2. Iterate until reaching a stable solution:
 - 2.1: Suppose $W^+ = E\{Xg(W^T X)\} - E\{g'(W^T X)\}W$
 - 2.2: Suppose $W = \frac{W^+}{\|W^+\|}$

Eq. (4) utilize to get the source matrix after performing the estimated demixing matrix (W).

3.5 Softmax function

The class list, also known as the probability distribution of category candidates, is provided by Softmax. It is typically utilized in the classification task's last layer [25]. Also, it is

used in neural networks to assign the unnormalized outputs of the network to a probability distribution across the predicted output class. The present work involved the conversion of separated signals (i.e., sources) into "class probabilities" through the utilization of the Softmax function [26], followed by the restoration of the source signals through normalization. Consequently, the documents are allocated to autonomous components (ICs) utilizing the probability IC value as a criterion, as in Eq. (7):

$$P(S)_i = \frac{e^{S_i}}{\sum_{j=1}^k e^{S_j}} \quad (7)$$

where, $i=1, \dots, k$; and $S=(S_1, \dots, S_k)$. Each member S_i of the input vector S is subjected to the conventional exponential function. In order to obtain the probability value of the separated source, it is necessary to normalize these values by dividing them by the sum of all exponentials.

4. DATA SETUP AND EXPERIMENT RESULTS

On a crime corpus that has been manually labeled, the performance of the NG-ICA and FastICA methods was evaluated. 1279 documents from the Malaysian National News Agency (BERNAMA) were included in the corpus. Each file contained a journalistic piece covering one or more crimes. The data in the corpus has been divided into five types of crimes: Drugs, kidnapping, money laundering, murder, and sexual crimes. The drugs class contained 222 documents, the kidnapping class contained 200 documents, the money laundering contained 260 documents, the murder class contained 298 documents, and the sexual crimes class contained 299 documents.

Two experiments were carried out to evaluate the performance of the suggested approaches. First experiment was performed using the NG-ICA algorithm, and the second experiment employed the FastICA algorithm. After constructing the term-document matrix D from the crime corpus as shown in Eq. (1), it was sent into the LSI approach, which applied the SVD.

The SVD method was utilized to decompose the D matrix. The matrix $D= \Sigma VT$ was computed using the maximal eigenvalues of the diagonal matrix V . The proposed ICA algorithms utilized the first five largest principal components of D as its input to generate five independent components (ICs) that were subsequently employed as distinct classes. Finally, The Softmax function was utilized to separate the signals within the matrix S into distinct categories based on their respective probabilities. The predicted ICA category assigned to a particular document corresponded to the component S_k that exhibited the greatest likelihood. The conventional assessments were utilized to evaluate the efficiency of the system to classify the documents available in the presented techniques. The major criteria for evaluating the efficacy of the crime categorization system are recall (R), precision (P), F-measure, and macro-average (F1) [27].

First experiment: the first experiment applied the NG-ICA technique with step Size=0.01, the maximum number of iterations in natural gradient algorithm=100, and compute loss function according to Eq. (2). The algorithm's performance was evaluated using a crime corpus that was manually labeled. The classification of the crime dataset based on the five IC components is presented in Table 1, which compares the

results obtained through NG-ICA and manually labeled documents using a confusion matrix. It is structured such that each class is represented by a column and each IC component is represented by a row.

Table 1. NG-ICA classifier confusion matrix per document class

	Class1	Class2	Class3	Class4	Class5
IC1	190	6	5	10	4
IC2	29	160	16	22	37
IC3	0	0	233	0	0
IC4	1	14	5	251	26
IC5	2	20	1	15	232

Second Experiment: the second experiment employed the parallel FastICA method, whose maximum number of iterations was 150 and whose tolerance on update at each iteration was 0.3. This technique approximated negentropy using logcosh. The manually labeled crime corpus was used to gauge how well the FastICA algorithm performed. The five IC components, each of which represented the class of a group of documents, were used to carry out the experiment. Table 2 displays the confusion matrix pertaining to the categorization of the crime corpus, which is based on the five IC components. The confusion matrix for the crime corpus categorization based on the five IC components is shown in Table 2.

Table 2. FastICA classifier confusion matrix per document class

	Class1	Class2	Class3	Class4	Class5
IC1	192	5	0	8	5
IC2	20	172	11	17	25
IC3	6	4	246	2	0
IC4	2	12	1	261	17
IC5	2	17	2	10	252

Tables 3 and 4 exhibit the classifier's performance in each class where measured by the precision, recall, and F-measure (expressed as percentages), for two classifiers: NG-ICA and FastICA respectively.

Table 3. Average F-measure of the NG-ICA classifier in each document class

	Precision (%)	Recall (%)	F-measure (%)
Class1	0.884	0.856	0.87
Class2	0.606	0.8	0.703
Class3	1.0	0.896	0.948
Class4	0.845	0.842	0.844
Class5	0.85.9	0.776	0.818
Macro-F-Measure			0.833

Table 4. Average F-measure of the FastICA classifier in each document class

	Precision (%)	Recall (%)	F-measure (%)
Class1	0.914	0.865	0.89
Class2	0.702	0.819	0.761
Class3	0.953	0.946	0.95
Class4	0.891	0.876	0.883
Class5	0.89	0.843	0.867
Macro-F-Measure			0.871

These techniques produced promising results that particularly highlight the suitability of using the unsupervised

machine learning algorithms NG-ICA and FastICA for the categorization of crimes. The first experiment's outcomes indicate that the NG-ICA algorithm is appropriate for conducting crime classification. Where this is shown to an overall F-measure average of 83.3%, as seen in Table 3. The outcomes of the second experiment indicate that Fast-ICA algorithm outperforms NG-ICA algorithm, which yielded an F-measure of 87.1% for Crime classification, as presented in Table 4. The results of FastICA are better because it achieves better separation of components than NG-ICA, Due to The Natural algorithm relies on step size, while Fast uses the fixed point method, which is devoid of the local minimum caused by step size. The good results of NG-ICA and FastICA in crimes classification, give a view that it is suitable for further applications of text mining.

5. CONCLUSION

With the significant growth in crime throughout the world, there is a requirement for crime data analysis to help reduce crime rates. This research introduces two unsupervised methodologies that can be employed to cluster crime-related text and assist law enforcement authorities. There are plans to utilize additional classification algorithms in the future to analyze crime data and enhance the accuracy of predictions. Our objective is to construct a model that captures stream data and regularly updates the results based on this data. This will enhance our ability to predict trends in crime and provide the public with general information to increase awareness. Two unsupervised approaches, NG-ICA and FastICA, were used in this work to categorize five categories of crime. Where an overall F-measure average of 83.6% of the NG-ICA algorithm, and the outcomes of the Fast-ICA algorithm yielded an F-measure of 87.1% for Crime classification. The outputs of this system might be used to increase public awareness about risky locales and to assist law enforcement in forecasting future crimes in a given location at a certain time. This allows the police and community to take the appropriate steps and solve crimes more quickly. The results demonstrated that the methodologies used were appropriate for crime categorization and that the FastICA algorithm outperformed the NG-ICA methodology because NG-ICA may fall into the trap of local minima due to the use of the learning rate, whereas FastICA met the highest measurement values of accuracy due to its use of Newton's method, which is based on a fixed-point iteration algorithm. This shows that FastICA is considered more appropriate for addressing crime data as compared to NG-ICA. Also, the promising results of NG-ICA and FastICA in crime classification, give a view that it is suitable for further applications of text mining. For future, we can implement this system in other languages such as Arabic or Spanish and also made it adaptable to handle real-time data (stream data).

REFERENCES

[1] Rahem, K.R., Omar, N. (2015). Rule-based named entity recognition for drug-related crime news documents. *Journal of Theoretical & Applied Information Technology*, 77(2): 103.

[2] Shabat, H.A., Omar, N. (2015). Named entity recognition in crime news documents using classifiers combination. *Middle-East Journal of Scientific Research*, 23(6): 1215-

1221. <https://doi.org/10.5829/idosi.mejsr.2015.23.06.22271>

[3] Albera, L., Kachenoura, A., Comon, P., Karfoul, A., Wendling, F., Senhadji, L., Merlet, I. (2012). ICA-based EEG denoising: A comparative analysis of fifteen methods. *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 60(3): 407-418. <https://doi.org/10.2478/v10175-012-0052-3>

[4] Shabat, H.A., Abbas, N.A. (2020). Enhance the performance of independent component analysis for text classification by using particle swarm optimization. In *2020 International Conference on Advanced Science and Engineering (ICOASE)*, Duhok, Iraq, pp. 1-6. <https://doi.org/10.1109/icoase51841.2020.9436547>

[5] Shabat, H.A., Abbas, N.A. (2020). Independent component analysis based on natural gradient algorithm for text mining. In *2020 1st Information Technology to Enhance e-learning and Other Application (IT-ELA)*, Baghdad, Iraq, pp. 72-76. <https://doi.org/10.1109/IT-ELA50150.2020.9253072>

[6] Tharwat, A. (2021). Independent component analysis: An introduction. *Applied Computing and Informatics*, 17(2): 222-249. <https://doi.org/10.1016/j.aci.2018.08.006>

[7] Rahem, K.R., Omar, N. (2014). Drug-related crime information extraction and analysis. In *Proceedings of the 6th International Conference on Information Technology and Multimedia*, Baghdad, Iraq, pp. 250-254. <https://doi.org/10.1109/icimu.2014.7066639>

[8] Shabat, H., Omar, N., Rahem, K. (2014). Named entity recognition in crime using machine learning approach. In *Information Retrieval Technology: 10th Asia Information Retrieval Societies Conference, AIRS 2014*, Kuching, Malaysia, pp. 280-288. http://doi.org/10.1007/978-3-319-12844-3_24

[9] ToppiReddy, H.K.R., Saini, B., Mahajan, G. (2018). Crime prediction & monitoring framework based on spatial analysis. *Procedia Computer Science*, 132: 696-705. <https://doi.org/10.1016/j.procs.2018.05.075>

[10] Almanie, T., Mirza, R., Lor, E. (2015). Crime prediction based on crime types and using spatial and temporal criminal hotspots. *arXiv preprint arXiv:1508.02050*. <https://doi.org/10.5121/ijdkp.2015.5401>

[11] Soni, S., Shankar, V.G., Chaurasia, S. (2019). Route-the safe: A robust model for safest route prediction using crime and accidental data. *International Journal of Advanced Science and Technology*, 28(16): 1415-1428.

[12] Mata, F., Torres-Ruiz, M., Guzmán, G., Quintero, R., Zagal-Flores, R., Moreno-Ibarra, M., Loza, E. (2016). A mobile information system based on crowd-sensed and official crime data for finding safe routes: A case study of Mexico city. *Mobile Information Systems*, 2016: 8068209. <https://doi.org/10.1155/2016/8068209>

[13] Yerpude, P. (2020). Predictive modelling of crime data set using data mining. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, 7(4): 43-58. <https://doi.org/10.5121/ijdkp.2017.7404>

[14] Kowalski, G.J., Maybury, M.T. (2000). *Information Storage and Retrieval Systems: Theory and Implementation*, (Vol. 8). Springer Science & Business Media.

[15] Singh, J., Gupta, V. (2017). A systematic review of text stemming techniques. *Artificial Intelligence Review*, 48: 157-217. <https://doi.org/10.1007/s10462-016-9498-2>

- [16] Lappin, S., Fox, C. (2015). *The Handbook of Contemporary Semantic Theory*. John Wiley & Sons. <https://doi.org/10.1002/9781118882139>
- [17] Ko, Y. (2012). A study of term weighting schemes using class information for text classification. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Portland Oregon USA, pp. 1029-1030. <https://doi.org/10.1145/2348283.2348453>
- [18] Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6): 391-407. [http://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6%3C391::AID-ASII%3E3.0.CO;2-9](http://doi.org/10.1002/(SICI)1097-4571(199009)41:6%3C391::AID-ASII%3E3.0.CO;2-9)
- [19] Wang, F., Li, H., Li, R. (2006). Data mining with independent component analysis. In *2006 6th World Congress on Intelligent Control and Automation*, Dalian, China, pp. 6043-6047. <https://doi.org/10.1109/wcica.2006.1714240>
- [20] Journée, M., Absil, P.A., Sepulchre, R. (2007). Optimization on the orthogonal group for independent component analysis. In *Independent Component Analysis and Signal Separation: 7th International Conference, ICA 2007*, London, UK, pp. 57-64. https://doi.org/10.1007/978-3-540-74494-8_8
- [21] Comon, P. (1994). Independent component analysis, a new concept. *Signal Processing*, 36(3): 287-314. [https://doi.org/10.1016/0165-1684\(94\)90029-9](https://doi.org/10.1016/0165-1684(94)90029-9)
- [22] Amari, S.I., Cichocki, A., Yang, H. (1995). A new learning algorithm for blind signal separation. *Advances in Neural Information Processing Systems*, 8.
- [23] Hyvärinen, A., Hurri, J., Hoyer, P.O. (2009). *Independent Component Analysis*, Springer London. http://doi.org/10.1007/978-1-84882-491-1_7
- [24] Comon, P., Jutten, C. (2010). *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press.
- [25] Horiguchi, S., Ikami, D., Aizawa, K. (2016). Significance of Softmax-based features over metric learning-based features. <https://doi.org/10.1109/TPAMI.2019.2911075>
- [26] Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*. Oxford University, Press. <https://doi.org/10.1093/oso/9780198538493.001.0001>
- [27] Ceri, S., Bozzon, A., Brambilla, M., Della Valle, E., Fraternali, P., Quarteroni, S., Quarteroni, S. (2013). An Introduction to Information Retrieval. *Web Information Retrieval*, 3-11.