Media and Communication
2025 • Volume 13 • Article 9623
https://doi.org/10.17645/mac.9623

ARTICLE

Open Access Journal

Unmasking Machine Learning With Tensor Decomposition: An Illustrative Example for Media and Communication Researchers

Yu Won Oh [1] and Chong Hyun Park [2]

[1] School of Digital Media, Myongji University, Republic of Korea
[2] School of Business, Sungkyunkwan University, Republic of Korea

Correspondence: Chong Hyun Park (chypark@skku.edu)

Submitted: 15 November 2024 **Accepted:** 5 February 2025 **Published:** 24 April 2025

Issue: This article is part of the issue "AI, Media, and People: The Changing Landscape of User Experiences and Behaviors" edited by Jeong-Nam Kim (University of Oklahoma) and Jaemin Jung (Korea Advanced Institute of Science and Technology), fully open access at https://doi.org/10.17645/mac.i475

Abstract

As online communication data continues to grow, manual content analysis, which is frequently employed in media studies within the social sciences, faces challenges in terms of scalability, efficiency, and coding scope. Automated machine learning can address these issues, but it often functions as a black box, offering little insight into the features driving its predictions. This lack of interpretability limits its application in advancing social science communication research and fostering practical outcomes. Here, explainable AI offers a solution that balances high prediction accuracy with interpretability. However, its adoption in social science communication studies remains limited. This study illustrates tensor decomposition—specifically, PARAFAC2—for media scholars as an interpretable machine learning method for analyzing high-dimensional communication data. By transforming complex datasets into simpler components, tensor decomposition reveals the nuanced relationships among linguistic features. Using a labeled spam review dataset as an illustrative example, this study demonstrates how the proposed approach uncovers patterns overlooked by traditional methods and enhances insights into language use. This framework bridges the gap between accuracy and explainability, offering a robust tool for future social science communication research.

Keywords

automated content analysis; explainable AI; machine learning; PARAFAC2; tensor decomposition

1. Introduction

As vast amounts of communication data accumulate online, manual content analysis, which is widely employed in media studies within the social sciences, faces limitations in terms of coding scope, effort, and efficiency

(Kroon et al., 2024). This explains why automated approaches—despite suspicions of being "*incorrect* models of language" (Grimmer & Stewart, 2013, p. 268, emphasis in original)—are increasingly being tried in digital corpus analysis. Among these, automated methods combined with machine learning draw on the utility of the technique in classification and prediction and have been applied to detect, for instance, opinion spam (e.g., Oh & Park, 2021), incivility (e.g., Burnap & Williams, 2015), and misinformation (e.g., Tanvir et al., 2019).

Typically, machine learning approaches operate on training data—whether labeled or unlabeled—to predict the outcomes of the test data. The algorithms developed through this process show great promise for efficient analysis of large-scale online media and communication data. In the advancement of machine-learning algorithms that demonstrate satisfactory performance in the field of communication, one unsettling aspect is that the underlying mechanism behind the final prediction remains a *black box*. Most machine-learning models are designed with a primary focus on achieving high accuracy in decisions, and the development of such models is often a significant accomplishment in itself. At the same time, however, it also remains true that one cannot understand which features or variables—or combinations of them—drive the predictions. In other words, they are uninterpretable and unexplainable (Rudin & Radin, 2019).

This black-box nature is particularly unfortunate in communication research within social science disciplines. For example, although a study proposed an algorithm capable of distinguishing deceptive comments written by paid commenters from genuine ones with nearly 81% accuracy (Oh & Park, 2021), it did not indicate which features of the comments should raise suspicion. A kind of dilemma—models can be accurate but cannot be understood—makes it challenging to apply research findings from state-of-the-art methods to media literacy education or guidelines and, more fundamentally, to deepen our understanding of human communicative acts.

In this regard, explainable AI, which has been actively explored in other fields, has attracted attention. Explainable AI aims to communicate the meaning from resulting models without significantly compromising the performance advantages of machine learning in solving complex problems (Ali et al., 2023). It offers a way to improve the trustworthiness and transparency of models that ensure high prediction accuracy (Rai, 2020). However, despite the clear potential benefits its application could bring to the analysis of large online media and communication datasets, to date, few related attempts have been made in social science communication research (cf. Dobbrick et al., 2022).

As a proactive and forward-looking response, this study provides media scholars with a guide for digital content analysis using interpretable machine learning methods. We focused on tensor decomposition—specifically, PARAFAC2—among the several techniques worth considering. In this study, we illustrate this method and demonstrate how media and communication researchers can employ it to analyze online corpora and interpret the results. To explain it, we rely on a review dataset constructed by Ott et al. (2011, 2013).

2. Literature Review

2.1. Text Analysis Method

The evolution of text analysis methods began with the basic bag of words (BoW) approach, which progressively developed into more sophisticated techniques. Grimmer and Stewart (2013) highlighted the limitations of BoW models, noting that they analyze solely based on word frequency while ignoring word order and contextual information, thus failing to capture the structural meaning within texts. Despite these limitations, BoW remains widely used because of its computational efficiency and straightforward structure. Boumans and Trilling (2016) emphasized the efficiency of dictionary-based approaches such as BoW, explaining that the advancement of automated methodologies is essential for handling the increasing demand for data processing in text analysis.

Among dictionary-based methods, Linguistic Inquiry and Word Count (LIWC) provides an in-depth linguistic analysis by categorizing words into psychological, social, and linguistic domains. Van Atteveldt et al. (2021) noted that LIWC extends beyond BoW's simple analysis to assess psychological and emotional elements, although it still struggles to capture subtle contextual nuances. Recently, large language models such as BERT and GPT-x have demonstrated impressive contextual understanding, and are increasingly being applied to content analysis. Rogers et al. (2020) analyzed the internal mechanisms of BERT, underscoring the model's complex and difficult-to-interpret learning process. Similarly, Zini and Awad (2022) discussed how large language models such as GPT-x exhibit substantial generative capabilities from extensive data training but retain a black-box quality, posing challenges in predictability and explainability. While efforts to improve the explainability of such models are ongoing, the complexity of large language models remains a key concern.

In this context, our study adopted a dictionary-based LIWC-supported BoWs approach, which is better suited for social science analysis, where interpretability and explainability are paramount. LIWC allows for a clear interpretation of analysis results and provides explanations based on specific linguistic characteristics, aiding researchers in understanding psychological and linguistic patterns in communication data. Additionally, our analysis utilized only the linguistic dimensions of LIWC, not to exclude its psychological and social features, but rather to clarify the study's focus on presenting a methodological approach to corpus analysis.

2.2. Machine Learning Challenges in High Dimensions

Supervised machine learning models focus on prediction and classification tasks using labeled data and employ algorithms such as linear regression, logistic regression, support vector machine, and neural networks to learn the correct output for given inputs. These models are intuitive and predictive due to the presence of clear answers. However, they struggle to effectively handle complex nonlinear relationships or multidimensional interactions in high-dimensional data. Bishop (2006) highlighted these limitations, stressing the need for more sophisticated models to address the complexity of high-dimensional relationships. As data complexity increases, classic machine-learning techniques face difficulties in adequately capturing intricate connections and learning nonlinear patterns (Hastie et al., 2009). Addressing these challenges requires advanced analytical methods, emphasizing the growing importance of techniques capable of high-dimensional data analysis.

Meanwhile, unsupervised machine learning models have evolved to identify patterns and structures in unlabeled data. Algorithms such as K-means, density-based spatial clustering of applications with noise, principal component analysis, and autoencoders excel at extracting features and exploring patterns, particularly in clustering and dimensionality reduction. However, the results are often difficult to interpret. To overcome this problem, tensor decomposition methods have gained attention as powerful tools for analyzing multidimensional interactions in high-dimensional data. Shin and Woo (2022) emphasized the effectiveness of tensor decomposition for extracting significant patterns from complex datasets and uncovering hidden structures. Shi et al. (2018) also demonstrated its utility in capturing crucial features from multidimensional data for better understanding.

Our study applies the PARAFAC2 algorithm, which is typically utilized in unsupervised learning to detect essential patterns in multidimensional data, in the unique context of analyzing labeled data. Although PARAFAC2 is known for its effectiveness in identifying interactions within high-dimensional datasets (Sidiropoulos et al., 2017), our approach extends its conventional usage. By leveraging tensor decomposition in this manner, we aim to make the complex relationships within the labeled data more comprehensible. Kolda and Bader (2009) emphasized the broader implications of tensor methods in uncovering intricate data structures, supporting the application of this technique to provide interpretable insights.

2.3. Applications of Tensor Decomposition in Communication-Related Topics

Tensor decomposition methods have been in use for several decades (Carroll & Chang, 1970; Harshman, 1970; Tucker, 1966) and have seen widespread application across various fields, including chemometrics (Smilde et al., 2004), signal processing (Sidiropoulos et al., 2000), computer vision (Vasilescu & Terzopoulos, 2002), numerical analysis (Beylkin & Mohlenkamp, 2005), graph analysis (Kolda et al., 2005), and web search personalization, where query terms or anchor text serve as the third dimension (Kolda et al., 2005; Sun et al., 2005). Building on these diverse applications, another research direction is centered on improving the performance of tensor decomposition techniques. For example, Kolda and Sun (2008) explored tensor decompositions in multi-aspect data mining and optimized these methods for high-dimensional and sparse data. For a comprehensive overview of tensor decompositions, see Kolda and Bader (2009), which provides an in-depth discussion of the mathematical foundations, various decomposition models, and their applications across multiple fields.

Although the potential of tensor decomposition in social science communication research is gradually becoming more apparent, it unfortunately remains largely unfamiliar to social science media and communication researchers. Most studies applying tensor decomposition to communication data have been conducted in the fields of science and engineering, the so-called STEM. In the context of text analysis, tensor decomposition methods have proven particularly valuable for improving the interpretability of machine learning models because they allow for the extraction of underlying patterns and structures from high-dimensional text data. For instance, Acar et al. (2005) explored the use of tensor decomposition techniques across various types of data, including texts. Specifically, they used tensor decompositions of (user × keyword × time) data to distinguish conversation threads in chatroom data. This approach is highly beneficial for handling the complexity of text data compared with traditional, simpler analytical methods. Similarly, PARAFAC has been applied to email communications, as in Bader et al. (2008), where it was used to track discussions in the Enron email corpus. Papalexakis et al. (2016) further underscored the significance of tensor decomposition in data-mining applications, such as topic modeling and sentiment analysis, providing an efficient framework for managing high-dimensional data. Saha and Sindhwani (2012) introduced a method using dynamic tensor decomposition to analyze temporal topic evolution and user interaction patterns on social media, offering valuable insights for addressing time-series text data. Subsequent research continued to adopt tensor decomposition to analyze the structural intricacies of text data and social interactions, progressively broadening its scope of application.

Nevertheless, challenges remain in the broader adoption of high-dimensional data analysis techniques such as tensor decomposition in communication research. Stoll et al. (2023) identified the limitations of existing methods in their study on detecting incivility in German online discussions, highlighting the difficulties in analyzing the multidimensional nature of text data. To address these challenges, recent studies, such as Schuld et al. (2023), explored advanced techniques using machine learning to deepen the analysis of opinion discourse and enrich the field of text data analysis. Our research demonstrates how tensor decomposition methods, specifically PARAFAC2, can be used to enhance the analysis of text data in social science communication research. By leveraging this technique, researchers can extract complex multidimensional relationships, identify critical patterns, advance the precision of text analysis, and address the unresolved complexities in existing methodologies.

3. Methods

This section describes the data analysis method using tensor decomposition. Tensor decomposition is a technique that represents data as a multidimensional array and divides it into various components to extract hidden patterns. To illustrate the application of tensor decomposition methods in social science communication research, we selected an online spam reviews dataset as an illustrative example. We first used the LIWC tool to extract linguistic features from online reviews, then transformed the data into a tensor form, and applied a tensor decomposition algorithm to identify significant patterns. Furthermore, to demonstrate the effectiveness of tensor decomposition in extracting valuable information from high-dimensional data, we applied it to an analysis of deceptive reviews.

3.1. Overview of Tensor Decomposition

Here, for social science communication researchers, we explain the fundamental concepts necessary for understanding tensor decomposition. We introduce the definition and structure of tensors and provide a simple example to illustrate the key principles and methods of tensor decomposition. This content serves as a foundation for understanding the tensor decomposition algorithms discussed in subsequent sections.

3.1.1. Tensor

A tensor is a mathematical concept that represents data as a multidimensional array, allowing for a structured representation across multiple dimensions. For example, a scalar is a 0th-order tensor, a vector is a 1st-order tensor, and a matrix is a 2nd-order tensor. Arrays with dimensions higher than these are referred to as 3rd-order, 4th-order tensors, and so on, depending on the number of dimensions. A 3rd-order tensor X can be approximated as the outer product of three vectors. The outer-product operation combines two or more vectors to create a higher-dimensional object. For example, a 3rd-order tensor $X \in \mathbb{R}^{I \times J \times K}$ can be approximately expressed using three vectors $a \in \mathbb{R}^{I}, b \in \mathbb{R}^{J}$, and $c \in \mathbb{R}^{K}$. Mathematically, this is represented as:

$$X = a \circ b \circ c$$

Here, $\circ$ denotes the outer product. The outer-product operation combines the elements of the two vectors to form a matrix. For instance, the outer product of vectors $a \in \mathbb{R}^m$ and $b \in \mathbb{R}^n$ is defined as:

$$a \circ b = \begin{pmatrix} a_1 b_1 & a_1 b_2 & \cdots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \ldots & a_2 b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_m b_1 & a_m b_2 & \cdots & a_m b_n \end{pmatrix}$$

This results in an $m \times n$ matrix. The operation multiplies each element of a with each element of b, combining the results into a matrix. Consider a 3rd-order tensor $\mathcal{X} \in \mathbb{R}^{2\times2\times2}$ as an example. This tensor consists of two matrices stacked along the third dimension:

$$\mathcal{X} = \left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \right)$$

This structure can be approximated using three vectors, a, b, and c, which are defined as $a^T = \begin{pmatrix} 1 & 2 \end{pmatrix}$, $b^T = \begin{pmatrix} 1 & 3 \end{pmatrix}$, and $c^T = \begin{pmatrix} 1 & 4 \end{pmatrix}$. The outer product of these vectors approximates $\mathcal{X}$, computed as:

$$a \circ b \circ c = \left(\begin{bmatrix} 1\times1\times1 & 1\times1\times4 \\ 1\times3\times1 & 1\times3\times4 \end{bmatrix}, \begin{bmatrix} 2\times1\times1 & 2\times1\times4 \\ 2\times3\times1 & 2\times3\times4 \end{bmatrix} \right)$$

This yields:

$$a \circ b \circ c = \left(\begin{bmatrix} 1 & 4 \\ 3 & 12 \end{bmatrix}, \begin{bmatrix} 2 & 8 \\ 6 & 24 \end{bmatrix} \right)$$

Although this approximation generates a structure similar to $\mathcal{X}$, it may not match exactly. Techniques such as matrix factorization and tensor decomposition have been used to achieve more accurate approximations.

3.1.2. Tensor Decomposition

Tensor decomposition approximates a tensor as the sum of multiple rank-1 tensors. The more rank-1 tensors are used, the more accurately $\mathcal{X}$ can be approximated. CANDECOMP/PARAFAC (CP) decomposition is one such method that approximates a 3rd-order tensor as the sum of the outer products of vectors for each dimension. Figure 1 illustrates that a given 3rd-order tensor $X \in \mathbb{R}^{I\times J\times K}$ can be approximated as a sum of these outer products. The CP decomposition is mathematically expressed as:

$$X \approx \sum_{i=1}^{r} a_i \circ b_i \circ c_i$$

Here, $a_i \in \mathbb{R}^I$, $b_i \in \mathbb{R}^J$, and $c_i \in \mathbb{R}^K$ are the i-th component vectors corresponding to each dimension, and r represents the number of rank-1 tensors needed for the approximation.

Tensor decomposition is a highly effective unsupervised learning method used to extract features and patterns from high-dimensional data. It is well-known for its capability to classify data even in the absence of labeled training examples (Kolda & Bader, 2009). To illustrate the efficiency of tensor decomposition in identifying important features from multidimensional data, we analyzed a small example of social media user interaction data. In this example, suppose we aim to classify user groups based on their interaction times with various content types such as images and videos (Acar et al., 2005). The data can be represented as a three-dimensional tensor based on the user, content type, and time of day. Suppose that tensor $\mathcal{X}$ comprises

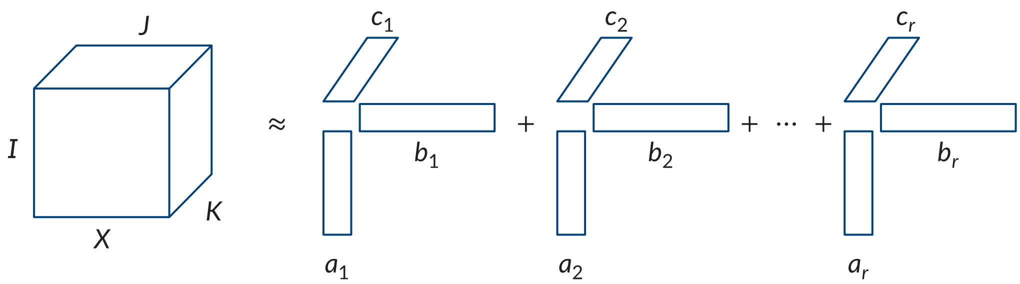

Figure 1. Tensor decomposition method.

interactions among three users, two content types (image and video), and two time periods (day and night). Each value in $\mathcal{X}$ represents the number of times a user interacted with a given content type during a specific time period. For instance, the first user interacted with image content five times during the day and twice at night and with video content once during the day and three times at night. The interaction data can be organized as follows:

$$X = \left(\begin{bmatrix} 5 & 2 \\ 1 & 3 \end{bmatrix}, \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix} \right)$$

Here, the first matrix represents interactions of user 1, the second matrix those of user 2, and the third matrix those of user 3. Using CP decomposition, we approximate $\mathcal{X}$ as the sum of two rank-1 tensors:

$$X \approx \lambda_1 a_1 \circ b_1 \circ c_1 + \lambda_2 a_2 \circ b_2 \circ c_2$$

Here λ_1 and λ_2 are scalar values, a_1, a_2 are user feature vectors, b_1, b_2 are content-type feature vectors, and c_1, c_2 are time-period feature vectors. Suppose that the first rank-1 tensor is given by:

$$\lambda_1 = 5, \;\; a_1 = \begin{pmatrix} 1.0 \\ 0.8 \\ 0.2 \end{pmatrix}, \;\; b_1 = \begin{pmatrix} 0.95 \\ 0.4 \end{pmatrix}, \;\; c_1 = \begin{pmatrix} 0.7 \\ 0.6 \end{pmatrix}$$

The outer product of these vectors generates the following three-dimensional array:

$$\lambda_1 a_1 \circ b_1 \circ c_1 = \left(\begin{bmatrix} 3.325 & 2.85 \\ 1.4 & 1.2 \end{bmatrix}, \begin{bmatrix} 6.65 & 5.7 \\ 2.8 & 2.4 \end{bmatrix}, \begin{bmatrix} 0.95 & 0.9 \\ 0.4 & 0.35 \end{bmatrix} \right)$$

Now, assume the second rank-1 tensor is defined as:

$$\lambda_2 = 3, \;\; a_2 = \begin{pmatrix} 0.3 \\ 0.6 \\ 0.9 \end{pmatrix}, \;\; b_1 = \begin{pmatrix} 0.1 \\ 0.8 \end{pmatrix}, \;\; c_1 = \begin{pmatrix} 0.5 \\ 0.9 \end{pmatrix}$$

The outer product of these vectors is computed similarly, and the sum of both rank-1 tensors approximates X:

$$\lambda_2 a_2 \circ b_2 \circ c_2 = \left(\begin{bmatrix} 0.045 & 0.081 \\ 0.36 & 0.648 \end{bmatrix}, \begin{bmatrix} 0.09 & 0.162 \\ 0.72 & 1.296 \end{bmatrix}, \begin{bmatrix} 0.135 & 0.243 \\ 1.08 & 1.944 \end{bmatrix} \right)$$

The final approximation is obtained by adding the two rank-1 tensors:

$$X \approx \left(\begin{bmatrix} 3.37 & 2.931 \\ 1.76 & 1.848 \end{bmatrix}, \begin{bmatrix} 6.74 & 5.862 \\ 3.52 & 3.696 \end{bmatrix}, \begin{bmatrix} 1.085 & 1.143 \\ 1.48 & 2.294 \end{bmatrix} \right)$$

Through this method, tensor decomposition serves as a tool for analyzing high-dimensional data. After CP decomposition, the feature matrices for users, content types, and time periods can be extracted. For instance, the user feature matrix A is:

$$A = \left[a_1\ ;\ a_2\right] = \begin{pmatrix} 1.0 & 0.3 \\ 0.8 & 0.6 \\ 0.2 & 0.9 \end{pmatrix}$$

The users can be grouped using these feature matrices. For example, by applying k-means clustering, we might group users with similar interaction patterns. Users 1 and 2 could belong to a group interacting mainly with content during the day, whereas user 3 might be grouped based on nighttime interactions with images. This analysis demonstrates that tensor decomposition effectively summarizes the key patterns in user behavior, enabling the grouping of similar users and the provision of personalized services (Wang et al., 2023).

3.2. Tensor Decomposition-Based Method

This section describes the method for analyzing the linguistic features of opinion spam using a tensor decomposition-based algorithm. It explains the process of transforming text data into a multidimensional tensor to extract linguistic patterns, which are then used to differentiate between fake and genuine reviews.

3.2.1. Dataset

For illustration and demonstration, we used the Deceptive Opinion Spam Corpus v1.4, developed by Ott et al. (2011, 2013). This reliable labeled opinion dataset consists of 1,600 reviews about 20 hotels located in Chicago, with 800 genuine and 800 fake reviews:

- Genuine reviews: A total of 400 positive reviews were collected from TripAdvisor. These reviews were based on actual lodging experiences and were sampled by Ott et al. (2011), excluding non-English and short reviews to ensure a matching length distribution among five-star reviews for the 20 hotels. Four hundred negative reviews written by travelers with genuinely negative experiences were collected from various travel websites, such as Expedia, Hotels.com, Orbitz, Priceline, TripAdvisor, and Yelp (Ott et al., 2013).
- Fake reviews: The 400 positive fake reviews were written by Amazon Mechanical Turk workers who were instructed to create positive reviews promoting specific hotels. The reviews had to be realistic and persuasive. The workers were US-based with a past approval rate of over 90%. All reviews were manually screened to ensure quality (Ott et al., 2011). A total of 400 negative fake reviews were also generated by Amazon Mechanical Turk workers who were tasked with writing reviews that portrayed competing hotels negatively (Ott et al., 2013).

3.2.2. Linguistic Feature Generation

Extracting linguistic features from review texts in our dataset using the LIWC program may hint at the characteristics of spam opinions. LIWC is a dictionary-based text analysis tool that connects word usage to various linguistic categories (Boyd et al., 2022). The LIWC analysis process involves several steps. First, LIWC examines each word in the text to determine whether it belongs to predefined linguistic categories, such as pronouns, conjunctions, or interjections, which reflect the structural characteristics of the text. LIWC then calculates the total frequency of words in each category and converts it into a percentage relative to the total word count. For example, if pronouns constitute 10% of the words in a given text, LIWC assigns a value of 10 to that category. This approach enables objective measurement of the linguistic features present in the text.

From the opinion spam reviews dataset, we extracted over 18 features, focusing on the linguistic dimensions categories of LIWC. The extracted features were represented using a BoW model based on the LIWC dictionary approach. The BoW model analyzes word frequency while ignoring word order, thus offering simplicity and high explainability (Kroon et al., 2024). Although this model may have limitations such as the loss of contextual information, it is widely used in tasks such as spam opinion analysis due to its effectiveness. Through a linguistic feature analysis, the linguistic pattern characteristics of spam opinions were explored.

Before constructing a tensor to analyze the linguistic features of spam opinions, we preprocessed the data. The 18 features utilized in this study had values distributed across different ranges and scales. These differences arise because the frequency of words belonging to various categories in LIWC analyses varies significantly. For instance, some categories may have high frequencies, resulting in large values, while others may have low frequencies, leading to lower values. To enhance data consistency and analytical accuracy, we applied min-max normalization. This technique transforms the values of each feature to fall within the range of 0 to 1, adjusting the minimum value to 0 and the maximum value to 1.

The formula for min-max normalization is as follows:

$$Z = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Here, Z represents the normalized value, x is the original value, and $\min(x)$ and $\max(x)$ denote the minimum and maximum values of the feature, respectively. By applying this normalization process, we ensured that all features were on the same scale, thus preventing any single feature from disproportionately affecting the analysis. This transformation uniformly distributes the data, allowing for effective comparisons between different values and minimizing the risk of skewed results during tensor construction. Such preprocessing forms a reliable foundation for accurately exploring the linguistic patterns of spam opinions.

The data generated in the previous preprocessing step are represented in the form of a tensor. Specifically, the preprocessed datasets for genuine and fake reviews consisted of 800 reviews and 18 features. Thus, both the genuine and fake review groups have dimensions of 800×18, where 18 represents the number of linguistic features extracted for each review group. Using these data, we constructed a three-dimensional tensor. The tensor structure is illustrated on the left in Figure 2. The final tensor size was defined as $I[2] \times J \times K$, which translated to $(800, 800) \times 18 \times 2$. Here, each dimension represented reviews ($I[2]$), linguistic features ($J = 18$), and review types ($K = 2$). The review dimension encompassed both genuine and fake review groups, the feature dimension consisted of the 18 linguistic features extracted using LIWC, and the review type dimension distinguished between genuine and fake categories. This tensor structure was utilized to analyze the complex interactions between the linguistic features of genuine and fake reviews and to uncover hidden patterns and relationships using tensor decomposition techniques.

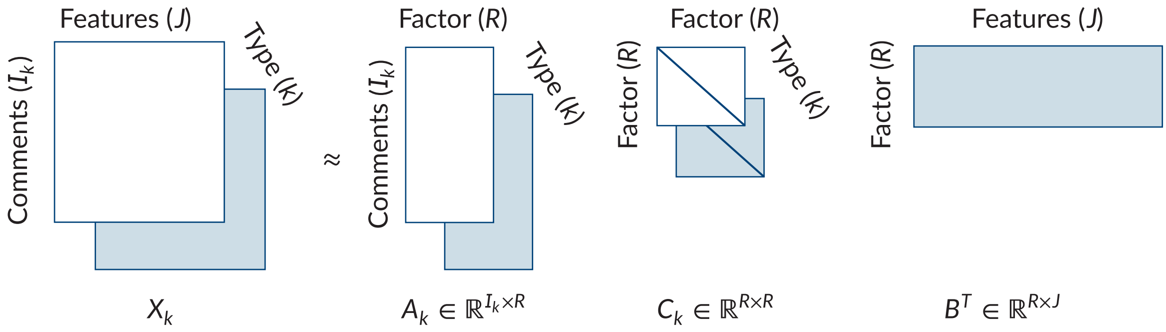

Figure 2. PARAFAC2 decomposition method.

We used the PARAFAC2 algorithm to decompose the constructed three-dimensional review dataset into component matrices A, B, and C. PARAFAC2 is a variant of CP that can be applied to a collection of matrices with the same number of columns but different numbers of rows, that is, when the matrices do not form a true tensor. One of the key advantages of PARAFAC2 is its ability to handle matrices with varying sizes in one mode, such as datasets with the same column dimensions but different row sizes. This flexibility makes PARAFAC2 well-suited for cases where the dimensions vary across slices, for example, in datasets with different numbers of genuine and fake reviews. The PARAFAC2 algorithm decomposes the tensor into multiple rank-1 components, thereby enabling the effective discovery of hidden patterns. Specifically, the three-dimensional tensor is divided into frontal slices, where each slice is represented as a two-dimensional matrix. For the k-th frontal slice $X_k \in \mathbb{R}^{I_k \times J \times K}$, the PARAFAC2 decomposition is mathematically expressed as:

$$X_k \approx A_k \bullet C_k \bullet B^T$$

Here, $k = 1, 2, \ldots, K$, $A \in \mathbb{R}^{I_k \times R}$, $B \in \mathbb{R}^{J \times R}$, $C_k \in \mathbb{R}^{R \times R}$ denote the component matrices. C_k is the k-th diagonal matrix, whose diagonal elements explain the relationships among factors. Through this decomposition, we obtain the component matrices A, B, and C. Matrix A_K represents the factor dependencies for each review group, and is composed of submatrices ($800 \times r$) that explain how each review is influenced by specific factors.

The value of r, which represents the number of such factors in the tensor decomposition, needs to be determined during the analysis. In tensor decomposition, selecting an appropriate rank r is crucial to ensure the quality and interpretability of the decomposition. The rank r represents the number of components or factors used to decompose the tensor. Choosing an optimal r helps balance the complexity and accuracy of the model. However, there is no straightforward algorithm to select the optimal r, and determining the best rank requires a balance between approximation accuracy and model complexity (Bader et al., 2008). If the rank is excessively small, the decomposition may fail to capture the underlying patterns in the data, leading to a poor approximation. On the other hand, if the rank is too large, the model may overfit the data and capture noise rather than meaningful features.

A widely used method for selecting the optimal rank is by minimizing the reconstruction error R_{error}, which is calculated using the Euclidean norm (also known as the Frobenius norm) between the original tensor $X \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ and the approximated tensor $\hat{X}$. The Euclidean norm of the error is expressed as:

$$R_{\text{error}} = \left\| X - \hat{X} \right\|_2$$

Here, $\|\cdot\|_2$ denotes the Euclidean norm of the tensor. For a tensor X, the Euclidean norm is computed as:

$$\|X\|_2 = \sqrt{\sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}\sum_{i_3=1}^{I_3} X_{i_1 i_2 i_3}^2}$$

In practice, as r increases, the reconstruction error typically decreases because a larger rank allows the model to approximate the original data more accurately. However, continuing to increase r beyond a certain value results in diminishing improvements, and at some point, the error reduction becomes negligible. Therefore, selecting the optimal rank involves finding a balance between the reduction in the reconstruction error and the model's complexity.

In addition, the variance of the reconstruction ratio R^2_{ratio} is computed as the ratio of the squared norm of the approximated tensor $\hat{X}$ to the squared norm of the original tensor X:

$$R^2_{\text{ratio}} = \frac{\|\hat{X}\|_2}{\|X\|_2}$$

This ratio indicates how well the decomposition approximates the original tensor. The optimal rank can often be determined by observing both the reconstruction error and the variance of the reconstruction ratio for different values of r and identifying the point where further increases in r lead to marginal or no improvement. This balance is crucial to avoid overfitting while still capturing meaningful data patterns. In practice, the elbow method is commonly used to identify this optimal point. This method involves plotting both the reconstruction error and variance of the reconstruction ratio against rank r and selecting the rank where further increases result in diminishing returns in terms of reconstruction error reduction. In doing so, we ensure that the selected rank achieves a good trade-off between model complexity and the ability to explain the underlying data structure. In this study, we determined the optimal rank r by running PARAFAC2 for each rank, starting with ten random initializations, and selecting the rank with the lowest R_{error}. The corresponding R^2_{ratio} for the selected R_{error} was then used to identify the optimal rank.

In addition, B is the feature-factor matrix ($18 \times r$) that describes the influence of linguistic features on the factors. Finally, C_k is the diagonal matrix ($r \times r$) that shows the relationship between review types and factors, indicating the effect of each factor on genuine and fake reviews. Our algorithms were written in Python, using the PARAFAC2 function from the tensorly.decomposition module. All tests were performed on a computer with an 11th Gen Intel(R) Core(TM) i7–11700@2.50 GHz processor and 8.00 GB of RAM.

In the next section, we demonstrate how the results of this decomposition can be used to analyze hidden linguistic patterns in the text and gain deeper insights into the characteristics of spam opinions.

3.2.3. Strength Distance Matrix

Based on the component matrices A, B, and C obtained through tensor decomposition, we calculated a distance matrix to better understand the relationships between linguistic and psychological features and review types. The main objective of this approach was to identify how strongly each feature influenced genuine or fake reviews. Component matrix B described the impact of each factor on specific linguistic and psychological features, while C_k explained how each factor contributed to different review types (genuine or fake). To calculate these distances, we used the Euclidean distance, which previous studies

(Shin & Woo, 2019, 2022) have proposed as an effective method for identifying the relationships between components derived from tensor decomposition and target categories, aiding in the interpretability of high-dimensional data.

We generated a distance vector by computing the Euclidean distance between each feature vector and the review-type vector. Specifically, the feature vector V_{f_n} consisted of the r factor values from matrix B, and the review-type vector V_{c_k} comprised the diagonal component values from matrix C. Mathematically, these vectors are defined as:

$$V_{f_n} = \{b_{1,n}, b_{2,n}, \dots, b_{r,n}\}^T$$

$$V_{c_k} = \{c_{k,1,1}, c_{k,2,2}, \dots, c_{k,r,r}\}$$

Here, n denotes the number of features (18), k represents the review type (genuine or fake), and r is the rank selected during tensor decomposition. We then computed the Euclidean distance between each feature vector and the review-type vector to form the distance vector D_{f_n}:

$$D_{f_n} = \left\{ \left\| V_{c_1} - V_{f_n} \right\|^2, \ \left\| V_{c_2} - V_{f_n} \right\|^2 \right\}$$

This distance vector helped determine the review type in which each feature had a stronger influence. A smaller distance indicates that a feature has a stronger influence on the corresponding review type. The distance matrix is constructed by vertically combining all the distance vectors, where smaller distances highlight features with a greater influence on a particular review type. This analysis provided a clearer understanding of the linguistic patterns that distinguished genuine from fake reviews and quantitatively evaluated the importance of each feature.

Figure 3 visually explains the process of calculating the distance matrix, showing how the relationships between features and review types are determined using the component matrices from tensor decomposition. In this tensor decomposition, we set the rank to 4 ($r = 4$), explaining the data structure using four factors. Matrix C is shown at the top left of Figure 3, where the original matrix C has been simplified to a two-dimensional form by isolating the diagonal elements. Matrix C captures the influence of each factor on review strength. As shown, the genuine review group is influenced by Factor 1 with a value of 3.6436 and by Factor 2 with a value of 1.4957, whereas the fake review group is influenced by Factors 1 (3.9033) and 2 (1.4894). These values quantitatively describe the relationship between review strength groups and each factor.

The bottom left of Figure 3 presents matrix B^T, originally of dimensions $J \times R$, where only 10 of the 18 features are displayed for clarity. Matrix B explains the relationship between the four factors and review features. In tensor decomposition, understanding the relative relationships among features is often more important than interpreting the exact values of the factors (Acar et al., 2005). This approach enables a comparison of the influences of different factors on each feature.

We then use the Euclidean distance to calculate the strength distance matrix and analyze how close each feature is to the genuine and fake review groups. The right side of Figure 3 displays the strength distance matrix for 10 features, with the red dashed lines indicating the vectors from C and B $\left(V_{c_1}, \ V_{f_1}\right)$ used to calculate the proximity of the first feature to the genuine review group. This distance matrix measures how close each

Component matrix $C \in \mathbb{R}^{K \times R}$

	factor 1	factor 2	factor 3	factor 4
Truthful	3.6436	1.4957	2.6930	3.6222
Fake	3.9033	1.4894	3.4366	3.2447

$$V_{c_1} = [3.6436 \;\; 1.4957 \;\; 2.6930 \;\; 3.6222]$$

Component matrix $B^T \in \mathbb{R}^{J \times R}$

	factor 1	factor 2	factor 3	factor 4
pronoun	2.7041	0.8279	2.0113	−1.0221
differ	−0.4659	1.9422	−0.3788	1.5774
socbehav	1.6414	−0.1212	1.1834	−0.8539
article	−.04667	0.8454	0.0467	3.0121
ipron	0.3727	.08765	0,6490	0.6392
prep	1.2257	1.2258	1.0889	1.0977
i	1.5283	0.3813	2.8846	−1.6464
discrep	0.2072	0.8404	0.4146	0.5277
cogproc	−0.0210	2.1250	0.1319	1.5693
prosocial	0.5022	−0.7226	0.3351	0.3390

$$V_{f_1} = [2.7041 \;\; 0.8279 \;\; 2.0113 \;\; -1.0221]$$

$$V_{f_1,c_1} = \| V_{c_1} - V_{f_1} \|^2 = 4.8335$$

	Truthful	Fake
pronoun	4.8335	**4.7024**
differ	5.5411	6.0524
socbehav	5.3794	5.4393
article	4.9691	5.5729
ipron	4,9151	5.2345
prep	3.8555	4.1666
i	5.7889	**5.5764**
discrep	5.1967	5.5315
cogproc	4.9597	5.4344
prosocial	5.5792	5.8756

Strength distance matrix $D \in \mathbb{R}^{J \times R}$

Figure 3. Calculation of the strength distance matrix.

feature vector V_f is to the vectors of the two review strength groups, demonstrating that smaller distances indicate a stronger influence of that feature on the review strength group.

3.2.4. Interpreting the Results

In the previous section, we presented a subset of the distance matrix results for illustration. Here, we provide complete results for all 18 linguistic features (Table 1). Table 1 quantitatively displays the Euclidean distances between each feature and the two review groups (truthful and fake). Each row contains the distance values for a specific feature. The final column of Table 1 presents the differences in distances for each feature between the genuine (truthful) and fake (fake) review groups. This allows for a quick assessment of the review group on which each feature has the strongest influence. The values highlighted with an asterisk indicate the minimum distance, signifying that the feature had the strongest influence on the corresponding review group.

To interpret, the analysis revealed that features such as pronouns (9.5034), personal pronouns (Personal pronouns; 9.5862), and first-person singular pronouns (First person singular; 9.7629) are prominent indicators of fake reviews. This suggests that fake reviews often emphasize personal and self-centered language, possibly reflecting an attempt to narrow the psychological distance with readers and appear more credible (c.f., Hancock et al., 2010; Newman et al., 2003).

Table 1. Distance matrix between 18 features and review type.

Features	Truthful	Fake	Difference
Impersonal pronouns	9.7766*	9.9235	−0.1469
Common adjectives	9.4235*	9.7097	−0.2862
Conjunctions	9.3392*	9.5170	−0.1778
Determiners	8.8372*	9.0319	−0.1947
Personal pronouns	9.6539	9.5862*	0.0677
Total pronouns	9.5272	9.5034*	0.0238
Adverbs	9.7837*	9.9726	−0.1889
First person singular (I)	9.8060	9.7629*	0.0431
First person plural (we)	10.2994*	10.4288	−0.1294
Third person plural (they)	10.0790*	10.1654	−0.0864
Articles	8.9768*	9.2555	−0.2787
Auxiliary verbs	9.7301*	9.9725	−0.2424
Quantities	9.2445*	9.4659	−0.2214
Verbs	9.5149*	9.6614	−0.1465
Second person (you)	10.0196*	10.1890	−0.1694
Negations	10.0051*	10.1517	−0.1466
Prepositions	8.9532*	9.0083	−0.0552
Total function words	9.1715*	9.3350	−0.1635

Note: * = The shorter distance between each feature and either truthful or fake.

4. Discussion

This study aims to offer a guide for social science communication researchers on how to apply a tensor-decomposition-based machine learning approach to the analysis of high-dimensional data, providing interpretable results. For illustration purposes, we systematically examined the linguistic features of fake reviews using a large-scale reviews dataset. Initially, we extracted the linguistic features using the LIWC tool. Following data normalization to ensure consistency, the tensor was decomposed using the PARAFAC2 algorithm. We then compared the influence of each feature on genuine and fake reviews using the Euclidean distance analysis. This comprehensive approach allowed us to quantify and understand the complex relationships between reviews, revealing that linguistic features such as pronouns, personal pronouns, and first-person singular pronouns were prominent in fake reviews. These features might be strategically used to foster intimacy with the reader or enhance emotional appeal. The insights gained from such interpretable results can be used to understand persuasive strategies, communication patterns, and other related aspects.

The versatility of the PARAFAC2 model should be particularly emphasized. This model offers significant flexibility in two ways. First, tensor-based representations can be applied to diverse domains. For instance, Acar et al. (2005) analyzed online chatroom data using a three-dimensional tensor (user–keyword–time) to track the evolution of social groups. Such tensor-based approaches are valuable for analyzing complex relationships in social media, news recommendation systems, and real-time communication networks, and have broad applicability in communication and data science (Bader & Kolda, 2006). Second, PARAFAC2 can handle tensors with varying dimensions along one mode, making it suitable for datasets of different sizes.

We used this model to conduct a nuanced analysis of the linguistic features in reviews, which can be applied to multidimensional studies, such as evaluating review credibility and analyzing advertising effectiveness in multi-label contexts.

This study focused on exploring the application of tensor decomposition to improve the interpretability of machine-learning models, particularly in the analysis of linguistic features in text data. By demonstrating the effectiveness of this approach, we highlighted its potential to improve the interpretability of detection algorithms and provide insights into complex data patterns. Our methodology emphasizes the importance of explainability in machine learning, offering a framework that can be adapted to various applications that require transparent and comprehensible analysis. Future research can further refine this approach by exploring how tensor decomposition models can be optimized to balance accuracy with interpretability across diverse data environments.

From a methodological perspective, determining the optimal rank for tensor decomposition is crucial, as it significantly affects the performance and accuracy of the algorithm. Although Section 3 in this article does not delve deeply into rank optimization, selecting an appropriate rank is essential to accurately represent data patterns and prevent overfitting. The Frobenius norm is commonly used to minimize reconstruction error (Kolda & Bader, 2009), and in our research, we adopted the elbow method to identify the optimal rank, which provides an intuitive and straightforward guideline by pinpointing inflection points on residual plots. For more detailed techniques of rank selection, readers may refer to Kolda and Bader (2009) and Cichocki et al. (2015).

Based on this study, there are several promising avenues for future research. While the Euclidean distance was chosen for its interpretability, further studies may investigate alternative measures for distance matrix computation, such as the dot product between component matrices B and C, which could provide additional insights. In addition, we implemented the standard PARAFAC2 algorithm, but some aspects can be applied more precisely. Specifically, the Python implementation that we used does not guarantee a unique solution, and there are various other ways to address this issue. For example, methods such as those proposed by Kiers et al. (1999) could be explored to improve the uniqueness of the results. Furthermore, to enhance both predictive power and interpretability, future studies could investigate the interactions between linguistic features particularly in the context of addressing the curse of dimensionality, which arises due to the exponential growth of parameters (e.g., Govindarajan et al., 2022; Novikov et al., 2017). Another promising direction could involve exploring the use of tensors to assess the linguistic similarity between fake and real texts without relying on decomposition, and utilizing faster and more numerically stable algorithms (Van Eeghem & De Lathauwer, 2020).

Conflict of Interests

The authors declare no conflicts of interests.

Data Availability

The codes and complete experimental results are available upon request from interested researchers. This ensures transparency and facilitates reproducibility while maintaining ethical considerations regarding data sharing and implementation.

References

Acar, E., Çamtepe, S. A., Krishnamoorthy, M. S., & Yener, B. (2005). Modeling and multiway analysis of chatroom tensors. In P. Kantor, G. Muresan, F. Roberts, D. D. Zeng, F.-Y. Wang, H. Chen, & R. C. Merkle (Eds.), *Intelligence and security informatics* (pp. 256–268). Springer.

Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Guidotti, R., Ser, J. D., Diaz-Rodriguez, N., & Herrera, F. (2023). Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence. *Information Fusion*, *99*, Article 101805. https://doi.org/10.1016/j.inffus.2023.101805

Bader, B. W., Berry, M. W., & Browne, M. (2008). Discussion tracking in Enron email using PARAFAC. In M. W. Berry & M. Castellanos (Eds.), *Survey of text mining II: Clustering, classification, and retrieval* (pp. 147–163). Springer.

Bader, B. W., & Kolda, T. G. (2006). Algorithm 862: MATLAB tensor classes for fast algorithm prototyping. *ACM Transactions on Mathematical Software (TOMS)*, *32*(4), 635–653.

Beylkin, G., & Mohlenkamp, M. J. (2005). Algorithms for numerical analysis in high dimensions. *SIAM Journal on Scientific Computing*, *26*(6), 2133–2159.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Boumans, J. W., & Trilling, D. (2016). Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars. *Digital Journalism*, *4*(1), 8–23.

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *The development and psychometric properties of LIWC-22*. University of Texas at Austin. https://www.liwc.app/static/documents/LIWC-22%20Manual%20-%20Development%20and%20Psychometrics.pdf

Burnap, P., & Williams, M. L. (2015). Cyber hate speech on Twitter: An application of machine classification and statistical modeling for policy and decision making. *Policy and Internet*, *7*(2), 223–242. https://doi.org/10.1002/poi3.85

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition. *Psychometrika*, *35*(3), 283–319.

Cichocki, A., Lee, N., Oseledets, I., Phan, A. H., Zhao, Q., & Mandic, D. P. (2015). Tensor networks for dimensionality reduction and large-scale optimization: Part 2 applications and future perspectives. *Foundations and Trends in Machine Learning*, *9*(6), 431–673.

Dobbrick, T., Jakob, J., Chan, C. H., & Wessler, H. (2022). Enhancing theory-informed dictionary approaches with "glass-box" machine learning: The case of integrative complexity in social media comments. *Communication Methods and Measures*, *16*(4), 303–320.

Govindarajan, N., Vervliet, N., & De Lathauwer, L. (2022). Regression and classification with spline-based separable expansions. *Frontiers in big Data*, *5*, Article 688496.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. (2010). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, *45*(1), 1–23.

Harshman, R. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics*, *16*, 1–84.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Kiers, H. A., Ten Berge, J. M., & Bro, R. (1999). PARAFAC2—Part I. A direct fitting algorithm for the PARAFAC2 model. *Journal of Chemometrics*, *13*(3/4), 275–294.

Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, *51*(3), 455–500.

Kolda, T. G., Bader, B. W., & Kenny, J. P. (2005). Higher-order web link analysis using multilinear algebra. In J. Han, B. W. Wah, V. Raghavan, X. Wu, & R. Rastogi (Eds.), *Proceedings of the 5th IEEE International Conference on Data Mining* (pp. 242–249). IEEE.

Kolda, T. G., & Sun, J. (2008). Scalable tensor decompositions for multi-aspect data mining. In F. Giannotti, D. Gunopulos, F. Turini, C. Zaniolo, N. Ramakrishnan, & X. Wu (Eds.), *2008 Eighth IEEE International Conference on Data Mining* (pp. 363–372). IEEE.

Kroon, A., Welbers, K., Trilling, D., & van Atteveldt, W. (2024). Advancing automated content analysis for a new era of media effects research: The key role of transfer learning. *Communication Methods and Measures*, *18*(2), 142–162. https://doi.org/10.1080/19312458.2023.2261372

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, *29*(5), 665–675.

Novikov, A., Trofimov, M., & Oseledets, I. (2017). *Exponential machines*. arXiv. https://doi.org/10.48550/arXiv.1605.03795

Oh, Y. W., & Park, C. H. (2021). Machine cleaning of online opinion spam: Developing a machine-learning algorithm for detecting deceptive comments. *American Behavioral Scientist*, *65*(2), 389–403. https://doi.org/10.1177/0002764219878238

Ott, M., Cardie, C., & Hancock, J. T. (2013). Negative deceptive opinion spam. In L. Vanderwende (Ed.), *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 497–501). Association for Computational Linguistics.

Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. In D. Lin (Ed.), *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* (pp. 309–319). Association for Computational Linguistics.

Papalexakis, E. E., Faloutsos, C., & Sidiropoulos, N. D. (2016). Tensors for data mining and data fusion: Models, applications, and scalable algorithms. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *8*(2), 1–44.

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*, 137–141. https://doi.org/10.1007/s11747-019-00710-5

Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). A primer in BERTology: What we know about how BERT works. *Transactions of the Association for Computational Linguistics*, *8*, 842–866.

Rudin, C., & Radin, J. (2019). Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. *Harvard Data Science Review*, *1*(2). https://doi.org/10.1162/99608f92.5a8a3a3d

Saha, A., & Sindhwani, V. (2012). Learning evolving and emerging topics in social media: A dynamic NMF approach with temporal regularization. In E. Adar & J. Teevan (Eds.), *WSDM '12: Proceedings of the fifth ACM international conference on Web search and data mining* (pp. 693–702). ACM.

Schuld, M., Durrheim, K., & Mafunda, M. (2023). Speaker landscapes: Machine learning opens a window on the everyday language of opinion. *Communication Methods and Measures*, 18(4), 315–331.

Shi, Q., Cheung, Y. M., Zhao, Q., & Lu, H. (2018). Feature extraction for incomplete data via low-rank tensor decomposition with feature regularization. *IEEE transactions on neural networks and learning systems*, *30*(6), 1803–1817.

Shin, Y., & Woo, S. S. (2019). What is in your password? Analyzing memorable and secure passwords using a tensor decomposition. In L. Liu & R. White (Eds.), *WWW '19: The World Wide Web Conference* (pp. 3230–3236). ACM.

Shin, Y., & Woo, S. S. (2022). PasswordTensor: Analyzing and explaining password strength using tensor decomposition. *Computers & Security*, *116*, Article 102634.

Sidiropoulos, N. D., De Lathauwer, L., Fu, X., Huang, K., Papalexakis, E. E., & Faloutsos, C. (2017). Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing*, *65*(13), 3551–3582.

Sidiropoulos, N. D., Giannakis, G. B., & Bro, R. (2000). Blind PARAFAC receivers for DS-CDMA systems. *IEEE Transactions on Signal Processing*, *48*(3), 810–823.

Smilde, A. K., Bro, R., & Geladi, P. (2004). *Multi-way analysis: Applications in the chemical sciences*. Wiley.

Stoll, A., Wilms, L., & Ziegele, M. (2023). Developing an incivility dictionary for German online discussions—A semi-automated approach combining human and artificial knowledge. *Communication Methods and Measures*, *17*(2), 131–149.

Sun, J.-T., Zeng, H.-J., Liu, H., Lu, Y., & Chen, Z. (2005). CubeSVD: A novel approach to personalized Web search. In A. Ellis & T. Hagino (Eds.), *WWW '05: Proceedings of the 14th international conference on World Wide Web* (pp. 382–390). ACM.

Tanvir, A. A., Mahir, E. M., Akhter, S., & Huq, M. R. (2019). Detecting fake news using machine learning and deep learning algorithms. In L. Gopal, D. Pradhan, & A. Singh (Eds.), *2019 7th International Conference on Smart Computing & Communications (ICSCC)* (pp. 1–5). IEEE. https://doi.org/10.1109/ICSCC.2019.8843612

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, *31*(3), 279–311.

van Atteveldt, W., van der Velden, M. A. C. G., & Boukes, M. (2021). The validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms. *Communication Methods and Measures*, *15*(2), 121–140.

Van Eeghem, F., & De Lathauwer, L. (2020). Tensor similarity in chemometrics. In S. Brown, R. Tauler, & B. Walczak (Eds.), *Comprehensive chemometrics: Chemical and biochemical data analysis* (pp. 337–354). Elsevier.

Vasilescu, M. A. O., & Terzopoulos, D. (2002). Multilinear analysis of image ensembles: Tensorfaces. In A. Heyden, G. Sparr, M. Nielsen, & P. Johansen (Eds.), *Computer Vision—ECCV 2002: 7th European Conference on Computer Vision, Part I* (pp. 447–460). Springer.

Wang, J., Zhang, S., Li, H., & Chen, Y. (2023). A tensor factorization approach for personalized recommendation with contextual information. *Journal of Information Science*, *49*(1), 85–98.

Zini, J. E., & Awad, M. (2022). On the explainability of natural language processing deep models. *ACM Computing Surveys*, *55*(5), 1–31.

About the Authors

Yu Won Oh (PhD, University of Michigan, 2015) is an associate professor in the School of Digital Media at Myongji University in Korea and the associate director of the Debiasing and Lay Informatics Lab in Oklahoma, USA. Her current research interests include the intersection of new media and civic communication with an emphasis on opinion expression, misinformation, issue development, and big data analytics.

Chong Hyun Park (PhD, Purdue University, 2016) is an associate professor in the School of Business at Sungkyunkwan University. His research interests include mathematical programming and machine learning modeling. He has been conducting interdisciplinary research that attempts to solve various social problems. He recently developed a machine learning algorithm that can detect manipulated opinion spam in the comment section. He has published research articles in *Production and Operations Management*, *European Journal of Operations Research*, *American Behavioral Scientists*, *Journalism and Mass Communication Quarterly*, etc.