Performance Complementarity in Topic Modeling: Strategic Algorithm Selection for Business Intelligence in African Markets

Darren Craig Roos
University of South Africa
Department of Decision Sciences
21001154@mylife.unisa.ac.za

Katherine Mary Malan
University of South Africa
Department of Decision Sciences
malankm@unisa.ac.za

Abstract

Topic modeling algorithms are increasingly vital for business intelligence in African markets, where understanding diverse textual data from multiple languages and contexts is crucial for informed decision-making. However, practitioners face the persistent question: which algorithm should be used for their specific business application? Through a comprehensive evaluation of eleven contextual topic modeling algorithms across ten diverse datasets and four performance metrics, we demonstrate that performance complementarity—rather than algorithmic superiority—characterizes this domain. Our findings reveal that in 84% of evaluation scenarios, all algorithms are Pareto optimal, each offering unique strengths that cannot be dominated by others. This evidence challenges the common practice of seeking a single "best" algorithm and instead advocates for strategic algorithm selection based on specific business requirements and data characteristics. For African businesses navigating complex multilingual markets and diverse data sources, understanding these performance trade-offs is essential for deploying effective AI-driven topic modeling solutions.

1. Introduction

African businesses increasingly rely on textual data analysis to understand market dynamics, customer sentiment, and emerging trends across diverse linguistic and cultural contexts. Topic modeling algorithms offer powerful unsupervised learning approaches to extract meaningful themes from large document collections, enabling businesses to process customer feedback, market research, financial reports, and social media content at scale [7, 9, 14, 16].

The business applications are particularly compelling in African contexts: analyzing multilingual customer reviews to understand product preferences, processing financial news to identify market trends, mining social media for brand sentiment across diverse communities, and extracting insights from regulatory documents in multiple languages. However, practitioners face a fundamental challenge: which topic modeling algorithm should they deploy for their specific business application?

Recent advances in Contextual Topic Modeling (CTM) algorithms [1, 8] leverage Large Language Model embeddings [6, 17, 22, 23] to capture semantic relationships beyond simple word co-occurrence. These models promise improved performance for business intelligence applications, particularly in multilingual African markets where traditional bag-of-words approaches may miss crucial cultural and linguistic nuances.

The literature consistently presents new algorithms as superior to their predecessors, typically demonstrating improved performance across selected datasets and metrics. This pattern creates confusion for business practitioners who must choose between numerous apparently "best" algorithms. Our research addresses this challenge by investigating whether successive improvements represent genuine algorithmic superiority or reflect a more complex performance landscape.

We present a comprehensive evaluation of eleven topic modeling algorithms across ten diverse datasets representing various business domains, evaluated using four performance metrics. Our findings reveal clear evidence of performance complementarity [11]: different algorithms excel under different conditions, with no single algorithm dominating across all scenarios. For African businesses, this implies that algorithm selection should be strategic, matching specific algorithms to particular business requirements and data characteristics rather than relying on blanket recommendations.

2. Methodology

Our experimental framework evaluates eleven topic modeling algorithms across fifty problem instances (ten datasets × five topic numbers: 10, 20, 50, 100, 200) with ten independent runs each, resulting in 5,500 model evaluations and 22,000 metric assessments.

2.1. Algorithms

We evaluated both traditional and contextual approaches: **Traditional:** LDA (Latent Dirichlet Allocation) [3, 4, 18], NMF (Non-negative Matrix Factorization) [13, 15] **Contextual:** ZeroShotTM [2], CombinedTM [1], BERTopic [8] (with both OpenAI and SBERT [19] embeddings), Gaussian Mixture Model, KeyNMF [12], and S³ [10]

2.2. Datasets

Ten diverse corpora represent various business domains: 20 Newsgroups (community discussions), IMDB Reviews (customer sentiment), Wikipedia Sample (general knowledge), TREC Questions (query understanding), Twitter Financial News (market sentiment), PubMed Abstracts (scientific literature), Patent Classification (innovation analysis), Goodreads Genres (content categorization), Battery Research Abstracts (technical documents), and ConvFinQA (financial Q&A).

2.3. Performance Metrics

Four complementary metrics assess different aspects of topic quality: **NPMI** (Normalized Pointwise Mutual Information) [5]: Traditional coherence based on word co-occurrence; **WEPS** (Word Embedding-based Pairwise Similarity) [20, 21]: Semantic coherence using embedding similarities; **WECS** (Word Embedding-based Centroid Similarity) [20]: Topic diversity through centroid comparisons; **NISH** (Negative Intruder Shift) [21]: Robustness to topic contamination.

3. Results and Business Implications

3.1. Performance Complementarity Evidence

Table 1 shows aggregate performance across all evaluations. While this provides a convenient ranking, it fundamentally obscures the nuanced performance patterns crucial for business decision-making. Different algorithms excel at different metrics: NMF leads in traditional coherence (NPMI), S³ dominates embedding-based coherence (WEPS), ZeroShotTM excels in topic diversity (WECS), and GMM achieves highest robustness (NISH).

3.2. Multi-Objective Analysis

When considering multiple metrics simultaneously using Pareto optimality analysis, we find that in 84% of problem instances, all eleven algorithms are Pareto optimal. This remarkable finding indicates that each algorithm offers unique strengths that cannot be dominated by others when evaluated comprehensively.

Table 2 shows the number of dominated algorithms per problem instance. Most instances (42 out of 50) have zero dominated algorithms, meaning all algorithms contribute meaningfully to the Pareto frontier.

3.3. Strategic Implications for African Businesses

These findings have profound implications for AI deployment in African business contexts:

Customer Sentiment Analysis: Businesses analyzing multilingual customer reviews should prioritize semantic coherence (WEPS) for accurate sentiment capture across languages. S³ and KeyNMF excel here, making them suitable for African markets with linguistic diversity.

Market Intelligence: Financial institutions processing market news and reports may prioritize topic diversity (WECS) to capture comprehensive market themes. ZeroShotTM variants perform strongly in this dimension.

Regulatory Compliance: Organizations analyzing regulatory documents may emphasize traditional coherence (NPMI) for interpretable topics that align with human understanding. NMF and GMM show superior performance here.

Brand Monitoring: Companies tracking brand mentions across social media platforms may prioritize robustness (NISH) to handle noisy, informal text. GMM and traditional approaches like LDA excel in this area.

4. Conclusion

Our comprehensive evaluation reveals that contextual topic modeling exhibits clear performance complementarity rather than algorithmic hierarchy. The finding that 84% of evaluation scenarios result in all algorithms being Pareto optimal fundamentally challenges the notion of a single "best" algorithm.

For African businesses deploying AI-driven topic modeling solutions, this research provides crucial guidance: algorithm selection should be strategic, matching specific algorithms to particular business requirements and data characteristics. Rather than seeking universally superior solutions, practitioners should understand the performance trade-offs and select algorithms that align with their specific business priorities.

This approach is particularly relevant in African markets, where businesses must navigate diverse linguistic contexts, varying data quality, and specific cultural considerations. Understanding performance complementarity enables more informed AI deployment decisions, potentially leading to more effective business intelligence applications across the continent.

Future research should develop algorithm selection frameworks that can automatically match algorithms to African business contexts, considering factors such as language diversity, data availability, and specific industry re-

Table 1. Overall performance summary. **Bold** indicates best performer(s), underlined indicates second-best (Mann-Whitney U test, $p < 0.05$).

Algorithm	NPMI	WEPS	WECS	NISH
LDA	-0.047 ± 0.255	-0.119 ± 0.026	-0.121 ± 0.024	<u>0.146 ± 0.028</u>
NMF	**0.136 ± 0.191**	-0.099 ± 0.018	-0.122 ± 0.023	<u>0.145 ± 0.026</u>
ZeroShotTM	-0.011 ± 0.310	-0.093 ± 0.012	**-0.105 ± 0.013**	0.127 ± 0.016
CombinedTM	0.019 ± 0.290	-0.097 ± 0.015	<u>-0.111 ± 0.017</u>	0.134 ± 0.020
BERTopic	0.080 ± 0.223	-0.100 ± 0.015	-0.118 ± 0.019	<u>0.141 ± 0.021</u>
ZeroShotTM-sbert	0.018 ± 0.302	-0.094 ± 0.011	<u>-0.109 ± 0.016</u>	0.131 ± 0.017
CombinedTM-sbert	0.037 ± 0.274	-0.097 ± 0.015	-0.113 ± 0.018	0.136 ± 0.021
BERTopic-sbert	0.075 ± 0.224	-0.100 ± 0.014	-0.119 ± 0.018	<u>0.142 ± 0.021</u>
GMM	<u>0.107 ± 0.242</u>	-0.103 ± 0.017	-0.129 ± 0.024	**0.152 ± 0.027**
KeyNMF	0.083 ± 0.243	<u>-0.088 ± 0.012</u>	-0.116 ± 0.021	0.132 ± 0.021
S³	-0.173 ± 0.347	**-0.084 ± 0.012**	-0.113 ± 0.015	0.130 ± 0.017

Table 2. Number of non-Pareto optimal algorithms per dataset-topic combination.

Dataset	10	20	50	100	200
battery-abstracts	1	1	0	2	0
goodreads-bookgenres	0	0	0	0	0
imdb-reviews	0	0	0	0	0
newsgroups	0	0	0	2	0
patent-classification	0	0	0	0	0
pubmed-multilabel	0	0	0	0	0
twitter-financial-news	1	0	1	0	0
others	0	0	0	0	0

quirements. Such frameworks could significantly enhance the practical deployment of topic modeling solutions in African markets.

References

[1] Federico Bianchi, Silvia Terragni, and Dirk Hovy. Pre-training is a Hot Topic: Contextualized Document Embeddings Improve Topic Coherence. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, June 2021.

[2] Federico Bianchi, Silvia Terragni, Dirk Hovy, Debora Nozza, and Elisabetta Fersini. Cross-lingual Contextualized Topic Models with Zero-shot Learning. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1676–1683, Online, 2021. Association for Computational Linguistics.

[3] David Blei, Andrew Ng, and Michael Jordan. Latent Dirichlet Allocation. *Advances in Neural Information Processing Systems*, 14, 2001.

[4] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022, 2003.

[5] G. Bouma. Normalized (pointwise) mutual information in collocation extraction. In *Proceedings of the Biennial GSCL Conference*, pages 31–40, 2009.

[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, 2019.

[7] Dat Tien Dieu and Dien Dinh. Using Topic in Summarization for Vietnamese Paragraph. *International Journal of Advanced Computer Science and Applications*, 14(10), 2023.

[8] Maarten Grootendorst. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. (arXiv:2203.05794), Mar. 2022.

[9] Bharathi Mohan Gurusamy, Prasanna Kumar Rengarajan, and Parthasarathy Srinivasan. A hybrid approach for text summarization using semantic latent Dirichlet allocation and sentence concept mapping with transformer. *International Journal of Electrical and Computer Engineering (IJECE)*, 13(6):6663, Dec. 2023.

[10] Márton Kardos, Jan Kostkan, Arnault-Quentin Vermillet, Kristoffer Nielbo, Kenneth Enevoldsen, and Roberta Rocca. Semantic Signal Separation. (arXiv:2406.09556), June 2024.

[11] Pascal Kerschke, Holger H. Hoos, Frank Neumann, and Heike Trautmann. Automated Algorithm Selection: Survey and Perspectives. *Evolutionary Computation*, 27(1):3–45, Mar. 2019.

[12] Ross Deans Kristensen-McLachlan, Rebecca M. M. Hicke, Márton Kardos, and Mette Thunø. Context is Key(NMF): Modelling Topical Information Dynamics in Chinese Diaspora Media. (arXiv:2410.12791), Oct. 2024.

[13] Daniel D. Lee and H. Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, Oct. 1999.

[14] Sheikh Sharfuddin Mim, Doina Logofatu, Gabriel Guerrero-Contreras, and Inmaculada Medina-Bulo. Leveraging Topic Modeling and Extractive Summarization for Unlocking Insights from NeurIPS Papers. In *2024 International Confer-*

ence on INnovations in Intelligent SysTems and Applications (INISTA), pages 1–6. IEEE, 2024.

[15] Fabian Pedregosa, Fabian Pedregosa, Gael Varoquaux, Gael Varoquaux, Normalesup Org, Alexandre Gramfort, Alexandre Gramfort, Vincent Michel, Vincent Michel, Logilab Fr, Bertrand Thirion, Bertrand Thirion, Olivier Grisel, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, Alexandre Tp, and David Cournapeau. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2011.

[16] Petr Pylov, Roman Maitak, and Andrey Protodyakonov. The Latent Dirichlet Allocation (LDA) generative model for automating process of rendering judicial decisions. *E3S Web of Conferences*, 431:05005, 2023.

[17] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving Language Understanding by Generative Pre-Training. https://openai.com/research/language-unsupervised, 2018.

[18] Radim Řehůřek and Petr Sojka. Gensim-python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, 3(2), 2011.

[19] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Conference on Empirical Methods in Natural Language Processing*, 2019.

[20] Silvia Terragni, Elisabetta Fersini, and Enza Messina. Word Embedding-Based Topic Similarity Measures. In Elisabeth Métais, Farid Meziane, Helmut Horacek, and Epaminondas Kapetanios, editors, *Natural Language Processing and Information Systems*, volume 12801, pages 33–45. Springer International Publishing, Cham, 2021.

[21] Anton Thielmann, Arik Reuter, Quentin Seifert, Elisabeth Bergherr, and Benjamin Säfken. Topics in the Haystack: Enhancing Topic Quality through Corpus Expansion. *Computational Linguistics*, pages 1–36, Jan. 2024.

[22] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. LLaMA: Open and Efficient Foundation Language Models. (arXiv:2302.13971), Feb. 2023.

[23] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is All you Need. *Conference and Workshop on Neural Information Processing Systems*, 2017.