

موضع ارائه:

Transformer-based Topic Modeling

فهرست مقالات مرور شده:

- **Semantic-Driven Topic Modeling Using Transformer-Based Embeddings and Clustering Algorithms**

Mersha, M. A., Yigezu, M. G., & Kalita, J. (2024). Semantic-Driven Topic Modeling Using Transformer-Based Embeddings and Clustering Algorithms. *Procedia Computer Science*, 244, 121–132- 2024
- **CWTM: Leveraging Contextualized Word Embeddings from BERT for Neural Topic Modeling**

Zheng Fang¹, Yulan He^{1,2,3} and Rob Procter^{1,3} 1Department of Computer Science, University of Warwick, UK, 2Department of Informatics, King's College London, UK, 3The Alan Turing Institute, UK -2021
- **FASTopic: Pretrained Transformer is a Fast, Adaptive, Stable, and Transferable Topic Model**

Xiaobao Wu¹ * Thong Nguyen² Delvin Ce Zhang³ William Yang Wang⁴ Anh Tuan Luu- 2024
- **BERTopic for Topic Modeling of Hindi Short Texts: A Comparative Study**

Atharva Mutsaddi, Anvi Jamkhande, Aryan Thakre, Yashodhara Haribhakta-2023
Department of Computer Science and Engineering, COEP Technological University {atharvaam21, jamkhandeaa21, aryanst21, ybl}.comp@coeptech.ac.in
- **A survey on neural topic models: methods, applications, and challenges**

Xiaobao Wu¹ · Thong Nguyen² · Anh Tuan Luu¹
Accepted: 19 December 2023 / Published online: 25 January 2024
© The Author(s) 2024

▪ Semantic-Driven Topic Modeling Using Transformer-Based Embeddings and Clustering Algorithms

▪ هدف اصلی مقاله:

این مقاله به محدودیت اصلی روش‌های کلاسیک Topic Modeling می‌پردازد: عدم درک معنای وابسته به Context روشنایی مانند LDA فقط بر اساس تکرار کلمات عمل می‌کند و معنای مفهومی را در نظر نمی‌گیرند. حتی برخی مدل‌های مبتنی بر Transformer نیز در مرحله نهایی استخراج موضوع، به طور کامل از قدرت درک زبانی این مدل‌ها استفاده نمی‌کنند.

▪ CWTM: Leveraging Contextualized Word Embeddings from BERT for Neural Topic Modeling

▪ هدف اصلی مقاله:

این مقاله یک مدل Topic Modeling جدید به نام CWTM را معرفی می‌کند که از رمزگذاری‌های Contextualized Word Embeddings از مدل زبانی BERT استفاده می‌کند تا محدودیت‌های Topic Modeling کلاسیک مبتنی بر Bag-of-Words (BOW) را برطرف کند. هدف اصلی، بهبود انسجام موضوعات، مدیریت واژه‌گان و استخراج موضوعات معنادارتر بدون اتكا به نمایش BOW است.

▪ FASTopic: Pretrained Transformer is a Fast, Adaptive, Stable, and Transferable Topic Model

▪ هدف اصلی مقاله:

ارائه یک مدل موضوعی جدید به نام FASTopic است که از دو جزء اصلی : Dual Semantic-relation و Embedding Transport Plan و Reconstruction تشکیل شده تا بجای استفاده از روش‌های کلاسیک، مستقیماً ارتباط معنایی بین اسناد و موضوعات و کلمات را با افزایش سرعت، سازگاری، پایداری، قابلیت انتقال و کارایی و با هدف کشف موضوعات پنهان در متون مدل کند.

▪ BERTopic for Topic Modeling of Hindi Short Texts: A Comparative Study

▪ هدف اصلی مقاله:

این مقاله به ارزیابی کارایی مدل BERTopic در مدل‌سازی موضوعی متون کوتاه زبان هندی می‌پردازد و آن را با روش‌های کلاسیک و مدرن دیگر مقایسه می‌کند تا برتری آن را در مواجهه با چالش‌های زبانی خاص هندی، مانند نبود حروف بزرگ، املای غیراستاندارد و عبارات تأکیدی تکراری نشان دهد.

▪ A survey on neural topic models: methods, applications, and challenges

▪ هدف اصلی مقاله:

مقاله یک مرور جامع و بروز بر مدل‌های عصبی (Neural Topic Models - NTMs) (ارائه می‌دهد و با مقایسه NTMs با مدل‌های کلاسیک مانند LDA شروع می‌کند و بر برتری‌های NTMs در انعطاف‌پذیری، مقیاس‌پذیری و عدم نیاز به استنتاج مدل‌محور تأکید می‌کند.