



Journal Website:
<https://theusajournals.com/index.php/ijll>

Copyright: Original content from this work may be used under the terms of the creative commons attributes 4.0 licence.

KARAKACROSS: A KARAKA-BASED APPROACH TO CROSS-LINGUAL SENTIMENT ANALYSIS

Submission Date: Aug 22, 2023, Accepted Date: Aug 27, 2023,

Published Date: Sep 01, 2023

Crossref doi: <https://doi.org/10.37547/ijll/Volume03Issue09-01>

Dipti Rai

Professor of Language Technologies at International Institute of Information Technology, Hyderabad (IIIT-H), India

ABSTRACT

Cross-lingual sentiment analysis is a challenging task in natural language processing due to the linguistic diversity across different languages. Existing approaches often struggle to accurately transfer sentiment knowledge between languages with distinct syntactic and semantic structures. In this research, we propose a novel approach called "Karakacross" for cross-lingual sentiment analysis. Inspired by the Karaka theory, which models the semantic roles of words in sentences, our method leverages semantic role labeling and cross-lingual transfer learning techniques. The Karakacross approach enables the alignment of sentiment-related semantic roles across languages, facilitating the transfer of sentiment knowledge. We conduct extensive experiments on multilingual datasets, demonstrating the effectiveness of Karakacross in achieving superior cross-lingual sentiment analysis performance compared to state-of-the-art methods. Our research contributes to advancing the field of cross-lingual sentiment analysis and offers new insights into leveraging semantic role information for better sentiment transfer between languages.

KEYWORDS

Cross-lingual sentiment analysis, Karakacross, semantic role labeling, sentiment transfer, cross-lingual transfer learning, natural language processing, linguistic diversity, multilingual datasets, semantic roles, sentiment knowledge, syntactic and semantic structures.

INTRODUCTION

Sentiment analysis, also known as opinion mining, is a fundamental task in natural language processing (NLP) that aims to automatically identify and categorize the sentiment expressed in textual data. With the increasing volume of multilingual content on the web, the demand for cross-lingual sentiment analysis has grown significantly. Cross-lingual sentiment analysis involves analyzing sentiments expressed in different languages and is essential for various applications, including social media monitoring, market research, and customer feedback analysis.

However, cross-lingual sentiment analysis presents several challenges due to the linguistic diversity across languages. Existing approaches often struggle to effectively transfer sentiment knowledge between languages with distinct syntactic and semantic structures. The variations in word order, word meanings, and sentiment expression across languages hinder the direct application of sentiment models trained on one language to another.

In this research, we propose a novel approach called "KarakaCross" to address the challenges of cross-lingual sentiment analysis. The KarakaCross approach draws inspiration from the Karaka theory, which is used in linguistics to model the semantic roles of words in sentences. We leverage semantic role labeling and cross-lingual transfer learning techniques to enable sentiment knowledge transfer between languages.

METHOD

Data Collection and Preprocessing:

We collected multilingual datasets comprising texts in different languages, each labeled with sentiment

labels (positive, negative, or neutral). The datasets were preprocessed to handle language-specific challenges, including tokenization, stemming, and stop-word removal.

Semantic Role Labeling (SRL):

Semantic role labeling is employed to extract the semantic roles played by words in sentences. SRL helps to capture the relationships between words and their roles in expressing sentiment, enabling a deeper understanding of sentiment expression in different languages.

Building a Cross-lingual Transfer Model:

We construct a cross-lingual transfer model that can learn to transfer sentiment knowledge from a source language to a target language. This model leverages the semantic role information extracted through SRL to facilitate sentiment knowledge alignment across languages.

Cross-Lingual Sentiment Knowledge Alignment:

The core aspect of the KarakaCross approach is to align sentiment-related semantic roles across languages. This alignment allows for the transfer of sentiment knowledge between languages, even when they have distinct linguistic structures.

Model Training and Evaluation:

The KarakaCross model is trained on the source language dataset and then fine-tuned using the target language dataset. We use standard evaluation metrics, such as accuracy, precision, recall, and F1 score, to

assess the performance of the KarakaCross approach in cross-lingual sentiment analysis.

Comparison with State-of-the-Art Methods:

We compare the performance of KarakaCross with state-of-the-art cross-lingual sentiment analysis methods to demonstrate its superiority in accurately transferring sentiment knowledge across languages.

Extensive Experimentation:

To ensure the robustness of the KarakaCross approach, we conduct extensive experiments on various multilingual datasets, covering different languages and domains. The experiments aim to validate the effectiveness of KarakaCross in achieving better cross-lingual sentiment analysis results.

The KarakaCross approach presents a novel contribution to cross-lingual sentiment analysis by leveraging semantic role information for sentiment knowledge alignment between languages. Through experimentation, we aim to demonstrate the effectiveness of KarakaCross and its potential to advance the field of cross-lingual sentiment analysis, addressing the challenges posed by linguistic diversity in multilingual data.

RESULTS

The performance of the KarakaCross approach was evaluated on various multilingual datasets, containing texts in different languages with sentiment labels. The experiments aimed to assess the effectiveness of KarakaCross in achieving cross-lingual sentiment analysis and compare its performance with state-of-the-art methods.

The results of the experiments demonstrated that KarakaCross outperformed existing cross-lingual sentiment analysis methods in terms of accuracy, precision, recall, and F1 score. The semantic role information provided by the Karaka theory proved to be valuable for sentiment knowledge alignment across languages, enabling more accurate sentiment analysis in multilingual contexts.

DISCUSSION

The superior performance of KarakaCross can be attributed to its ability to capture the relationships between words and their semantic roles in expressing sentiment. By aligning sentiment-related semantic roles across languages, the KarakaCross approach effectively transferred sentiment knowledge, even in the presence of linguistic diversity.

The use of semantic role labeling allowed KarakaCross to understand the underlying structure of sentiment expression in different languages. This understanding facilitated the alignment of sentiment features, enabling the model to generalize sentiment knowledge from the source language to the target language.

Furthermore, the KarakaCross approach demonstrated robustness across diverse multilingual datasets, covering various languages and domains. The consistent performance across different datasets suggests the generalizability and effectiveness of the KarakaCross method in real-world applications.

CONCLUSION

In conclusion, the KarakaCross approach offers a novel and effective solution to the challenges of cross-lingual sentiment analysis. By leveraging semantic role

information and aligning sentiment-related features across languages, KarakaCross successfully transfers sentiment knowledge and achieves superior sentiment analysis results in multilingual contexts.

The findings of this research have significant implications for various NLP applications, particularly in the analysis of sentiment in multilingual data. KarakaCross provides a valuable tool for organizations and researchers working with data from diverse linguistic backgrounds, enabling more accurate and meaningful sentiment analysis across different languages.

Future research directions may focus on further enhancing the KarakaCross approach, exploring its applicability to additional languages, and investigating its potential in other NLP tasks beyond sentiment analysis. Additionally, incorporating domain adaptation techniques and exploring the impact of domain differences on sentiment transfer could be areas for further investigation.

In summary, the KarakaCross approach advances the field of cross-lingual sentiment analysis, offering new insights into leveraging semantic role information for sentiment knowledge alignment. The successful performance of KarakaCross in sentiment analysis demonstrates its potential for facilitating more effective and robust sentiment analysis in multilingual environments, contributing to the advancement of natural language processing in cross-lingual settings.

REFERENCES

1. J. Aditya, A. R. Balamurali, and P. Bhattacharyya. "A fall-back strategy for sentiment analysis in hindi: A case study," in Proc. The 8th ICON, 2010.
2. B. Akshat, P. Arora, and V. Varma. "Hindi subjective lexicon: A lexical resource for Hindi polarity classification," in Proc. Eight International Conference on Language Resources and Evaluation (LREC), 2012.
3. M. Namita et al., "Sentiment analysis of Hindi review based on negation and discourse relation," in Proc. International Joint Conference on Natural Language Processing, 2013.
4. K. Amandeep and V. Gupta, "A survey on sentiment analysis and opinion mining techniques," Journal of Emerging Technologies in Web Intelligence, vol. 5, no. 4, 2013, pp. 367-371.
5. P. Pooja and S. Govilkar. "A framework for sentiment analysis in Hindi using HSWN," International Journal of Computer Applications, vol. 119, no. 19, 2015.
6. J. S. Philip, D. C. Dunphy, and M. S. Smith, "The general inquirer: A computer approach to content analysis," 1966.
7. H. Vasileios and K. R. McKeown. "Predicting the semantic orientation of adjectives," in Proc. Eighth Conference on European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, 1997.
8. W. Janyce et al., "Learning subjective language," Computational Linguistics, vol. 30, no. 3, 2004, pp. 277-308.
9. D. Amitava and S. Bandyopadhyay, "SentiWordNet for Indian languages," Asian Federation for Natural Language Processing, China, 2010, pp. 56-63.
10. P. Bo, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment classification using machine learning techniques," in Proc. the ACL-02 Conference on Empirical Methods in Natural Language Processing, vol. 10, 2002.