

Recognizing Sentiments In Regional Language Text From Social Media Using Machine Learning Approach: the way ahead through cross domain applications

Sudarshan Sirsat¹ and Nitish Zulpe²

¹Assistant Professor, Department of Data Science and Technology, K.J. Somaiya Institute of Management, Somaiya Vidyavihar University, Mumbai, India

²Principal, Research Guide, College of computer science and information technology, Latur

Email: sudarshan@somaiya.edu¹ | nitishzulpe@gmail.com²

Manuscript Details

Available online on <https://www.irjse.in>
ISSN: 2322-0015

Editor: Dr. Arvind Chavhan

Cite this article as:

Sudarshan Sirsat and Dr. Nitish Zulpe. Recognizing Sentiments In Regional Language Text From Social Media Using Machine Learning Approach: the way ahead through cross domain applications, *Int. Res. Journal of Science & Engineering*, 2024, Special Issue A14: 89-96. <https://doi.org/10.5281/zenodo.12702214>

Article published in Special issue of National Conference on Machine Learning and Data Science (NCMLDS-2024) organized by College of Computer Science and Information Technology (COCSIT) Ambajogai Road, Latur, Maharashtra, India on date April 16th to 17th 2024.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

Abstract

India is a country of 120 languages and its enormous dialects with almost its 43% population living in the rural area, around 430 million people availing the internet services and many amongst the undigitized chunk are joining the mainstream internet users. Not everyone is well versed with English foreign language and willing to access all internet services, contents and devices with their regional language. Maharashtra states it amongst many exponentially developing states of India, and people are accessing internet contents in Marathi language as matter of choice. Researchers, scholars and enterprises are trying to analyze and merge these undigitized chunks of the population through their study, innovation and services. This marathi language sentiment analysis can widespread its applications like social media sentiment analysis, analyzing the reviews of the users like google reviews, youtube reviews or product or service reviews through microblogging websites. Contents which are available in English as global language and official language of many countries can be transliterated into the regional language contents to make it available to all of the respective states. Marathi language sentiment analysis as regional language sentiment analysis which can reflect the emotions and sentiments of the state public can help in understanding and analyzing the mass sentiments on common issues and trends like political, cultural, governance, education, business enterprise and many such state affairs. Devices like alexa, seri in regional language with sentiment analysis can do wonders, mobile applications and many such social media browser based extensions can help both the parties i.e state and citizens to understand each other in a better sense.

Keywords: Regional language, sentiment analysis, marathi text sentiment analysis, Natural Language Processing, Regional language sentiment analysis (RLSA), Regional Language sentiment analysis based Approach (RLSAA), Point of Dataset Generation (PoDSG)

Introduction

Approximately 7151 languages are spoken worldwide like English, Mandarin, Hindi and Spanish leading on the list. The Largest chunk of the population can speak only one language, mostly their first language; approximately 3.3 million people are bilingual. Marathi being 11th on the list of top first language used by people around the globe, currently 83.1 million speakers own it as first language, 16 million people prefer it as second language and 99.1 million total speakers can make use of it as mode of verbal communication [1]

72% Marathi speakers prefer regional language contents over the internet, 43% entertainment, 35% sports and 27% search for political content. This reflects strong inclination and interest of major regional language natives i.e. Marathi internet users are more interested in their mother tongue and geographical based news contents [2].

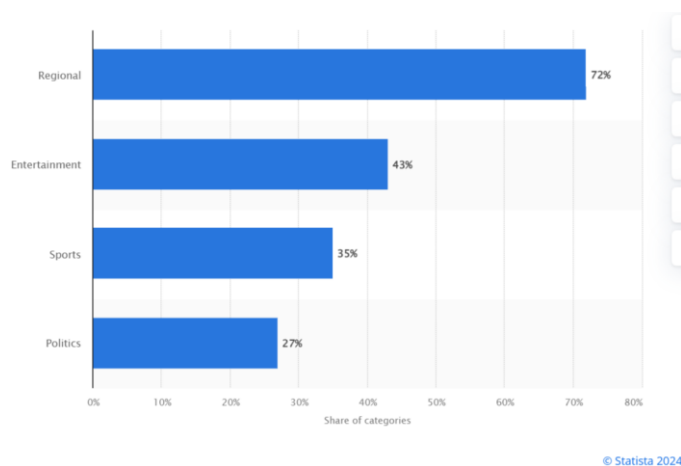


Figure: digital news category preferences among Marathi internet users across India, 2016*

Literature review

The excerpt highlights the significant impact of language diversity on management decisions within multinational corporations. It notes the challenges in empirically investigating and theoretically conceptualizing these effects. The authors systematically review 264 articles on language in international

business, analyzing geographic distributions, theoretical advancements, and core findings. They propose a future research agenda emphasizing interdisciplinary approaches, multi-level studies, cross-national collaborations, and exploration of new data sources and analysis methods to deepen understanding in this field. The study through this paper discusses the evolution of language's significance in international business over the past three decades. It highlights how language influences organizational communication, knowledge creation, and the construction of organizational realities. The excerpt reviews scholarly approaches to language in business, focusing on national languages in multinational corporations, corporate language policies, and English as a global business language. It also examines the interplay between national languages, corporate languages, and English, alongside economic research on language-related factors. The review identifies a growing sophistication in conceptualizing language in business studies, with an increasing emphasis on its complexity and dynamism. It advocates for interdisciplinary collaboration with linguistics and translation studies to deepen understanding and suggests a future research agenda. [6].

The research study discusses the significance of language barriers in intercultural business communication, emphasizing their detrimental effects on business transactions. It identifies various causes of language barriers, including differences in language, regional accents, unclear speech, use of jargon and slang, word choice, literacy, grammar, and spelling. Additionally, it suggests solutions such as repetition, code-switching, machine translation, external translators, language training, bilingual employees, and locally hired non-native personnel. The study concludes by recommending strategies to overcome language barriers and proposes future research directions focusing on challenges faced by businesses in non-English speaking countries and the impact of language differences in different cultural contexts [5].

the importance of Business Process Integration and Management (BPIM) in enterprise transformation, particularly for small and medium-sized businesses

(SMBs). It emphasizes the need for fast, efficient engagement methodologies, reusable frameworks to reduce costs, and lightweight platforms for easy revision and execution of solutions. Model-driven technologies, specifically Model Blue developed by IBM China Research Laboratory (CRL), are proposed as a solution to address these challenges. Model Blue offers model-driven BPIM methods, frameworks, tools, and a runtime environment. The effectiveness of Model Blue was demonstrated through its deployment with Bank SinoPac, a mid-sized bank in Taiwan. Despite challenges such as remote work and the outbreak of Severe Acute Respiratory Syndrome (SARS), the project was completed successfully within schedule and budget, resulting in up to 30% efficiency improvement compared to similar projects. The paper illustrates how each phase of business process integration and solution development was guided by business process modeling, highlighting key experiences gained. Technical aspects discussed include two-dimensional business process modeling, lightweight processing logic automation, and an end-to-end BPIM methodology integrating models and documents into existing methodologies and software engineering approaches [7].

The authors highlight key components of their approach, including optical character recognition (OCR) software, natural language processing (NLP) algorithms, and speech-to-text engines. These technologies form the foundation for translating spoken English in videos into text, which is then processed and translated into Indian regional languages. Additionally, the paper emphasizes the application of open innovation principles to drive the development and refinement of these machine translation methods. The proposed methodology involves several stages. Firstly, the English audio from videos is transcribed into text using speech-to-text engines. Next, optical character recognition software is employed to extract text from on-screen visuals or subtitles. Subsequently, natural language processing algorithms analyze and translate the extracted text into the desired regional languages. The authors likely discuss the technical intricacies involved in each stage, such as data preprocessing, language modeling, and translation algorithms. The

overarching goal of the study is to enhance accessibility to video content for speakers of Indian regional languages, thereby promoting linguistic inclusivity and cultural diversity. By leveraging machine translation technologies, the authors aim to bridge the language barrier that often hinders the dissemination of information and knowledge across linguistic boundaries. The paper likely presents insights into the challenges encountered during the machine translation process, such as linguistic variations, semantic nuances, and the accuracy of translation outputs. Additionally, it may discuss the effectiveness of the proposed approach through experimental results or case studies involving the translation of real-world English videos into Indian regional languages. Overall, the study contributes to the advancement of machine translation techniques for multimedia content and underscores the importance of open innovation in addressing linguistic diversity and accessibility challenges in the digital age [8].

In The article "On the Generation of E-Learning Resources Using Business Process, Natural Language Processing, and Web Services" The authors emphasize the importance of leveraging technology to streamline the development of e-learning resources, particularly in the context of rapidly evolving industries and educational paradigms. They propose a methodology that combines business process analysis, NLP techniques, and web services integration to automate the creation of educational content. Key components of the proposed approach include task analysis to identify learning objectives and content requirements, NLP algorithms to extract relevant information from textual sources, and web services to aggregate multimedia resources and deliver personalized learning experiences. By automating the generation of e-learning materials, the authors aim to improve efficiency, scalability, and customization in educational content development. The paper likely discusses the technical implementation of the proposed methodology, including the selection and integration of NLP tools, the design of web service interfaces, and the development of content generation algorithms. Furthermore, it may present case studies or experimental results demonstrating the effectiveness of the approach in

creating diverse types of e-learning resources, such as interactive modules, instructional videos, and assessment materials. Overall, the study contributes to the advancement of e-learning technology by introducing a systematic approach to resource generation that leverages business process analysis, NLP, and web services integration. By automating content creation processes, the proposed methodology has the potential to enhance educational outcomes, promote lifelong learning, and address the growing demand for accessible and engaging online learning experiences [9].

Business Application areas for Regional Language Sentiment Analysis:

Sentiment based Translations

Translation services are the future of the AI industry. This industry has revolutionized from scratch to cloud9 in just a few years. The majority of smartphone companies are trying to provide inbuilt translation services as their core product or service. The trend of this industry is shown dramatic and dynamic evolution, adapting to the technological advancements of the industry specific changing global needs. This industry

has seen a huge surge in transcreation and adaptation to new technologies and markets for new audiences and cultural considerations, this also includes increase in video translation services due to exponential growth in video content consumption by the internet users.

2020 - efficiently integrating machine translation with human post-encoding (MTPE) due to increased demand in non-english languages and collaboration of human translation experts and AI.

2021 - translation and localization leveraged the technology like never before with AI and NLP enhancing machine translation post editing (MTPE). Majorly catering the online translation services due to pandemic era, online meetings, video remote interpretation services, adapting contents to various cultural contexts.

2022 - MTPE gained popularity and higher accuracy due to advanced human and AI translation, advance in speech recognition technology were the trend setters. Media localization and increased business translation transactions changed global commerce.



Image: role of good translation services [4]

2023 - witnessed the blend of efficiency and human expertise for the B2B framework. Became the global problem solver and communication medium for diverse markets.

2024 - AI and Neural Machine Translation (NMT) technology offering never before accuracy and efficiency in the industry services along with the real time translation service solving the language barrier in global communication and businesses with the proficiency.

Efficient speech-to-speech and speech-to-text translation technologies with the advancements in technology have facilitated real-time and precise spoken language translations, significantly enhancing communication by overcoming barriers with unprecedented efficiency.

Proposed RLSAA solution:

Along with these domain specific technologies and their advancements sentiment-based translation can give all together different directions and business applications towards global business ranging from C2C and B2B areas.

The above image talks about the role of good professional translation services, translation methods for efficient translation for businesses.

Multilingual/Regional customer support

Currently AI ML based customer support is for efficient instance, error and ticket solving is done with few AL ML based services like IBM watson, watson conversation service, watson tone analyzer, Microsoft's Language Understanding Intelligent Service (LUIS), Google's dialogflow etc.

Proposed RLSAA solution:

We propose parallel regional language based multilingual support along with understanding the analysis of regional language service based sentiments of the customers or consumers through voice and text sentiment analysis.

Market Expansion

Leveraging regional language and regional language sentiment analysis taps into the previously untapped markets with tailored products and services for selective

customer segments. This opens the whole product and service lifecycle process adaptation to the regional language and local target customers and consumers.

The steps for the market expansion through regional language analysis are as follows

- Market research
- Localization strategy
- Product adaptation
- Local partnership
- Website and digital marketing and presence
- Promotional campaigns
- Feedback & Customer support
- Measure performance

Proposed RLSAA solution:

The whole product or service lifecycle for market expansion can be ideated and established around regional language and its sentiment analysis throughout the lifecycle

Regional Government compliances

Government compliances and regional language are tightly integrated and intertwined, specifically in the region with diverse linguistic demographics like India. It will allow both the parties to understand the sentiments associated with business and ease of doing business in the state and region. Areas and scope of government compliances and regional language includes

- Legislation and regulations
- official language policies
- Language accessibility
- Localization requirements
- Education and literacy
- Cultural preservation
- Public communication
- Legal documentation

Proposed RLSAA solution:

Regional language sentiment analysis with regional government compliances will enable governance objectives like "Make in India" initiatives of the government of India. And it will help achieve visions like ease of making business for the central and state governance authorities.

Localization Services and Regional Languages

Localization services entail the adjustment of a product or service to meet the linguistic, cultural and other specific needs of a particular target market or locale. This is done by not only translating texts into the target language but also observing various cultural variations, local regulations as well as preferences.

As far as localization services are concerned, regional languages have immense influence due to their reflection of diverse linguistic choices within a given geographical location. In fact, these languages serve as primary communication channels across different parts of a country.

In addition to software and mobile app development, website localization services, video game localization services and e-commerce platforms are some industries and sectors that require localization services. Consequently, this will mean that businesses can provide content in regional languages making it easy for users to engage with them thereby boosting accessibility and customer satisfaction in general.

Proposed RLSAA solution:

Why should businesses only adapt their products and services to local languages and not to the regional language sentiments? It will boost the popularity of these products and services amongst locals.

Academics and online Education Platforms:

Things are moving with lightning speed when it comes to digital platforms for education and the number of people getting associated with them. All the stakeholders' engagement nowadays is done through global language English, whereas a huge chunk of people are not able to join the mainstream because of language barriers. Government portals like Swayam and NPTEL are using regional resources for generating similar content in regional languages.

Recently the government of Maharashtra introduced a few medical courses in the regional language like homeopathy, nursing, dentistry and MBBS and introduced the text books in the regional language Marathi.

Also recently Dr APJ Abdul Kalam Technical University (AKTU) proposed and planned for BTech first year students to teach and make study material available in Hindi for the regional language students.

Tamil Nadu government official Lieutenant Governor proposed to start a medical college offering MBBS in Tamil medium.

Madhya Pradesh government implemented a decision to teach medical students subjects like anatomy, physiology, bio-chemistry etc in Hindi their regional language in all 13 medical colleges from 2022.

In 2022, the All India Council of Technical Education (AICTE) came up with an initiative to teach technical education and courses in non-English language and allowed a budget for second year course material and 12 regional Indian Languages.

Proposed RLSAA solution:

All these regional languages can be analyzed over regional language sentiment analysis for better understanding and stakeholder engagement throughout. All the teaching learning associated processes can be mapped and analyzed over regional language sentiment analysis.

Point of Dataset Generation (PoDSG) for businesses and academic organizations



Fig: point of dataset generation for business and academic organizations

A device can be set up at the entry exit point of the organization, highly accessible to the stakeholders to enter their feedback about the product or services in the regional language. All these feedbacks will be collected on a cloud-based data server. This regional language based feedbacks will be analyzed for regional language-based sentiment analysis and used for continuous improvement.

Future Scope

We will try to understand the business applications through proper surveys, interviews, and domain analysis in detail to know how exactly an area can be served through regional language sentiment analysis. These application areas are considered from the generic point of view of the research scholars, which can in future be explored technically as an consultation opportunity. The area demands real time sentiment analysis of the regional language contents to satisfy the ongoing business needs.

Conclusion

Regional language is changing the business scenarios starting from market expansion to adding regional language clusters to the main stream. Localization services, content translation services for web contents, audio video content analysis for multinational business organizations. Multilingual support for regional languages and regional language services are transforming the business from human to intelligent auto and high efficient optimization. In the future dynamic and ongoing regional language sentiment analysis for business applications can help organizations understand their stakeholders sentiment on products and services in real time, which can increase the satisfaction of the stakeholders exponentially.

References

1. Richard Rogers (September 15, 2023). 74+ Language Statistics For 2024 (Trends, Facts & Data) Language statistics, <https://myclasstracks.com/language-statistics/>

2. Tanushree Bsuroy (Mar 19, 2021). Category preference in digital news among Marathi in India 2016, <https://www.statista.com/statistics/719666/popular-categories-in-digital-news-among-marathi-users-india>
3. Anu Parthiban (October 29, 2022). MBBS in Marathi: Maharashtra to introduce medical textbooks in Marathi from next year, <https://news.careers360.com/mbbs-in-marathi-maharashtra-introduce-medical-textbooks-in-marathi-from-next-year>
4. Excel translations, The Role of Good Language Translation Services for Technical Products, <https://exceltranslations.com/role-good-language-translation-services-technical-products/>
5. Adanlawo, Eyitayo Francis & Reddy, Mike & Rugbeer, Hemduth. (2021). Intercultural Business Communication: The Implications of Language Barriers. *Journal of Education & Psychology*. 58. 6281-6290.
6. Tenzer, H., Terjesen, S. & Harzing, AW. Language in International Business: A Review and Agenda for Future Research. *Manag Int Rev* 57, 815-854 (2017). <https://doi.org/10.1007/s11575-017-0319-x>
7. Zhu J et al., "Model-driven business process integration and management: A case study with the Bank SinoPac regional service platform," in *IBM Journal of Research and Development*, vol. 48, no. 5.6, pp. 649-669, Sep. 2004, doi: 10.1147/rd.485.0649.
8. Pulipaka SK, Kasaraneni CK, Sandeep VN, Vemulapalli and S. S. Mourya Kosaraju. Machine Translation of English Videos to Indian Regional Languages using Open Innovation," 2019 IEEE International Symposium on Technology and Society (ISTAS), Medford, MA, USA, 2019, pp. 1-7, doi: 10.1109/ISTAS48451.2019.8937988.
9. Fragoso-Diaz OG, V. López-Caballero, J. C. Rojas-Pérez, R. Santaolaya-Salgado and J. G. González-Serna, "On the Generation of E-Learning Resources Using Business Process, Natural Language Processing, and Web Services," in *IT Professional*, vol. 23, no. 2, pp. 40-44, 1 March-April 2021, doi: 10.1109/MITP.2021.3054640.

© The Author(s) 2024

Acknowledgements

We acknowledge the Swami Ramanand Teerth Marathwada University (SRTMUN) college of computer science and information technology (COCSIT) research center for providing us all the facilities and infrastructure necessary to conduct the study.

Conflicts of interest: The authors stated that no conflicts of interest.

Publisher's Note

IJLSCI remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Correspondence and requests for materials should be addressed to Sudarshan Sirsat.

Peer review information

IRJSE thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <https://www.irjse.in/reprints>

Submit your manuscript to a IRJSE journal and benefit from:

- ✓ Convenient online submission
- ✓ Rigorous peer review
- ✓ Immediate publication on acceptance
- ✓ Open access: articles freely available online
- ✓ High visibility within the field

Submit your next manuscript to IRJSE through our manuscript management system uploading at the menu "**Make a Submission**" on journal website

<https://irjse.in/se/index.php/home/about/submissions>

For enquiry or any query email us: editor@irjse.in