test

by SHAFAAT SIDDHI

- **Introduction:** The motivation and goals of the topic are to:
  - Explore how AI can enhance language learning and teaching by simulating human conversation and providing feedback and guidance.

  - Understand the challenges and opportunities of identifying and using colloquial words in natural language processing (NLP), such as variability, sparsity, dynamics, accuracy, and robustness.

  - Teach AI to use colloquial language appropriately and effectively in different contexts and domains, such as e-commerce, healthcare, education, and entertainment.

- **Background:** The background information and concepts related to the topic are:

  - Conversational AI is a type of AI that can simulate human conversation by using natural language processing (NLP), a field of AI that allows computers to understand and process human language, and foundation models, such as BERT or GPT-3, that power new generative AI capabilities.

  - Colloquial words are informal words or phrases that are used in everyday speech, often depending on the region, culture, or context of the speaker. They can make the language more natural and engaging, but they can also be confusing or unfamiliar to some listeners or readers.

  - Some examples of colloquial words in different languages and domains are lol, wanna, or what's up in English, jajaja, vale, or qué tal in Spanish, or pop, biscuit, or swag in American or British English.

  - Some existing methods and tools for identifying and generating colloquial words in NLP are rule-based methods, which use hand-crafted rules or patterns to match colloquial expressions in text, such as regular expressions or lexicons, statistical methods, which use probabilistic models or features to learn colloquial expressions from data, such as n-grams, word embeddings, or topic models, and neural methods, which use deep neural

networks to automatically learn representations and classifiers for colloquial expressions, such as CNNs, RNNs, or transformers.

- **Methods:** Some methods and techniques for teaching AI general conversation and colloquial language are:

  - Collecting and annotating data for general conversation and colloquial language by using online sources, such as social media, blogs, or forums, or by creating corpora, such as the Switchboard corpus or the British National Corpus, that contain natural and spontaneous dialogues.

  - Designing and evaluating conversational AI systems that can handle colloquial words and expressions by using metrics, such as accuracy, fluency, coherence, diversity, and engagement, or by using user feedback, such as ratings, comments, or surveys.

  - Using pre-trained language models, such as BERT or GPT-3, to improve general conversation and colloquial language generation by fine-tuning them for specific tasks or domains, such as chatbots, captions, or summaries, or by using generative methods, such as ChatGPT.

  - Incorporating user feedback and adaptation mechanisms to enhance conversational AI performance and user satisfaction by using reinforcement learning, which allows the AI system to learn from its own actions and rewards, or by using personalization, which allows the AI system to tailor its responses to the user's preferences, needs, or goals.

- **Results:** Some results and findings from previous or ongoing research projects or experiments on the topic are:

  - The benefits and limitations of using colloquial language in conversational AI systems are that it can increase the naturalness and engagement of the

dialogue, but it can also decrease the accuracy and robustness of the system, especially when dealing with noisy, ambiguous, or rare expressions.

- The users' perception and interaction with conversational AI systems that use colloquial language are that they tend to have positive attitudes and satisfaction, but they also have moderate behavioral intention and expectations, and they may prefer different levels of formality and politeness depending on the context and domain.

- The effect of colloquial words on the accuracy and robustness of conversational AI systems are that they can improve the performance of some tasks, such as sentiment analysis or text normalization, but they can also degrade the performance of other tasks, such as speech recognition or machine translation.

- The variation of colloquial words across different languages, domains, and contexts are that they can have different forms, meanings, and usages depending on the region, culture, or situation of the speaker, and they can also change over time, creating new expressions or modifying existing ones.

- **Conclusion:** The main points and implications of the topic are:
  - The key challenges and opportunities for teaching AI general conversation and colloquial language are to balance the trade-off between naturalness and accuracy, to cope with the variability and dynamics of colloquial words, and to adapt to the user's needs and preferences.

  - The best practices and recommendations for developing and deploying conversational AI systems that use colloquial language are to use a combination of rule-based, statistical, and neural methods, to use pre-trained language models and generative methods, to use user feedback and adaptation mechanisms, and to use appropriate metrics and evaluation methods.

- The future directions and open questions for research and innovation on the topic are to explore the ethical and social implications of using colloquial language in conversational AI, to develop more robust and reliable methods for identifying and generating colloquial words, and to create more diverse and inclusive data sets and models for colloquial language.

Motivation:

- Explore people's suppressed and concealed emotions, as micro-expressions are often difficult to perceive and interpret by humans.

- Improve the naturalness and engagement of text generation systems, such as chatbots, captions, or summaries, by incorporating colloquial expressions.

- Enhance the performance of sentiment analysis systems, by considering the colloquial expressions used by the author, which may convey subtle or nuanced emotions.

A literature review for the proposal on idiomatic expressions and natural language processing:

- **Idiomatic expressions and their challenges for NLP**: Idiomatic expressions (IEs) are phrases that have a figurative or nonliteral meaning that cannot be easily understood based on the literal definitions of the words in the phrase. They are an important part of natural language and constantly being added to a language. However, they pose a classical challenge for NLP systems, as they require the ability to handle semantic non-compositionality, contextual ambiguity, and cultural specificity. Prior work has identified the impact of IEs on various NLP applications, such as sentiment analysis, paraphrase generation, natural language inference, and dialog modeling, and proposed methods to mitigate them .
- **Models and methods for idiom processing**: Several models and methods have been developed to address the problem of idiom processing, which can be broadly categorized into two main tasks: idiom type classification and idiom token classification. Idiom type classification aims to decide if a set of multi-word expressions (MWEs) can be used as IEs without considering additional context, while idiom token classification determines if a given potentially idiomatic expression (PIE) is used in a literal or figurative sense in a sentence. Most of the existing models and methods rely on linguistic features, such as syntactic structure, word embeddings, and external knowledge sources, to capture the idiosyncratic properties of IEs and to distinguish them from literal expressions . However, these models and methods often suffer from data sparsity, limited generalization, and insufficient semantic representation of IEs.
- **Recent advances and future directions**: Recent advances in NLP, such as pre-trained language models (PTLMs) and transformer-based architectures, have opened new possibilities for improving idiom processing. Some studies have probed the ability of PTLMs to handle IEs and found that they can capture some aspects of idiomaticity, but they still struggle with the contextual and semantic ambiguity of IEs . Other studies have proposed novel approaches to enhance the representation and comprehension of IEs, such as using adapters, denoising auto-encoders, similarity forcing, and contrastive learning . However, there are still many open challenges and opportunities for future research, such as generalizing to unseen IEs, incorporating more external knowledge, and applying idiom-aware PTLMs to downstream applications.

test

PRIMARY SOURCES

1	Ziheng Zeng, Suma Bhat. "Getting BART to Ride the Idiomatic Train: Learning to Represent Idiomatic Expressions", Transactions of the Association for Computational Linguistics, 2022 Publication	4%
2	Ziheng Zeng, Suma Bhat. "Idiomatic Expression Identification using Semantic Compatibility", Transactions of the Association for Computational Linguistics, 2021 Publication	1%
3	Li Yunhan. "Aviation Safety Ensemble Classification Model Based on Imbalanced Data", 2023 XX Technical Scientific Conference on Aviation Dedicated to the Memory of N.E. Zhukovsky (TSCZh), 2023 Publication	1%
4	cloud.google.com Internet Source	1%

5 Monika Henzinger, Steve Lawrence. "Extracting knowledge from the World Wide Web", Proceedings of the National Academy of Sciences, 2004
Publication

1 %

Exclude quotes   On   Exclude matches   Off
Exclude bibliography On

test

Article Error You may need to use an article before this word.

Tone This language may not be appropriate when you write an essay.

Nonstandard You have used an incorrect verb form.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Article Error You may need to use an article before this word. Consider using the article **a**.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Article Error You may need to use an article before this word.

Article Error You may need to use an article before this word.

Article Error You may need to use an article before this word. Consider using the article **the**.

Article Error You may need to remove this article.

Missing "," Review the rules for using punctuation marks.

Possessive

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Wrong Form You may have used the wrong form of this word.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.