International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

Argumentation Mining System for Corpus-based Discourse Analysis based on Structured Arguments

Shashi Prabha Anan[1], Vaishali Singh[2*]
[1]Research Scholar, Department of Computer Science and Engineering, Maharishi University of Information and Technology, Lucknow, Uttar Pradesh
[2]Assistant Professor, Department of Computer Science and Engineering, Maharishi University of Information and Technology, Lucknow, Uttar Pradesh*

Abstract: An Argumentation mining system can analyze a large volume of text data through a variety of sources. Nowadays it is highly useful in the areas of business, economics, and finance with digital marketing being the most promising field along with social media. It is the study of corpus-based discourse analysis that involves the automatic identification of argumentative structure in text. Initially, AM talks about extracting structured arguments from natural text, often unstructured or noisy text. Theoretical approaches of AM and pragmatic schemes that satisfy the needs of social media generated data, recognizing the need for adapting more flexible and expandable schemes, capable of adjusting to argumentation conditions that exist in social media. In this scenario it is a very challenging argumentation scheme able to identify the distinct sub-task and capture the needs of social media text, revealing the need for adopting a more flexible and extensible framework. Corpus-based Machine Learning of linguistic annotations has enabled researchers to identify repetitive linguistic patterns of language use and to uncover hidden meaning in all areas of Natural Language Processing.

Keywords: Argumentation mining system, Machine Learning, argumentative structure, pragmatic schemes, corpus-based discourse analysis, structure Arguments.

I. INTRODUCTION

In the era of social media heterogeneity of content and diversity of different types of jargon are very challenging tasks for annotating and automatically analyzing arguments. Arguments in social media and informal disclosure are sometimes the logical structures of an argument's components like premises, claims, and Warrants are not instantly distinguishable, after that analysis must take place to determine the distinctive components. In text derived from social media arguments frequently are missing, as it is common for a tweet or web post to contain just a stance on a specific topic without supporting it with evidence or reasoning associated with it. Opinion mining is the primary task in information retrieval research. The large volume of data originates from online shopping, online exams, online interviews, online digital elections, etc. With the great volume of opinionated data available on the Web, approaches must be derived to differentiate opinion from facts. Opinion mining involves two stages: relevance to the query and opinion detection. Traditionally topic-based retrieval takes place. After that, we start analyzing based on natural language processing using the neural network and support vector machines method for NLP. Natural language processing involves two techniques for machine learning that is Supervised Machine Learning and Unsupervised Machine Learning.

A. Supervised Machine Learning: The supervised learning technique is a popular technique that helps with training your neural networks on labeled data for a specific task.

Supervised Learning is trained using data that is well-labeled (or tagged). During training, those systems learn the best mapping function between known data input and the expected known output. Supervised NLP models then use the best approximating mapping learned during training to analyze never-seen-before input data to accurately predict the corresponding output. Usually, Supervised Learning models require extensive and iterative optimization cycles to adjust the input-output mapping until they converge to an expected and well-accepted level of performance. This type of learning keeps the word "supervised" because its way of learning from training data mimics the same process of a teacher supervising the end-to-end learning process. Supervised Learning models are typically capable of achieving excellent levels of performance but only when enough labeled data is available.

B. Unsupervised Machine Learning: Unsupervised machine learning deals with training a model without pre-tagging or annotating. Some of these techniques are surprisingly easy to understand.

Unsupervised Learning promises effective learning using unlabeled data (no labeled data is required for training) and no human supervision (no data scientist or high-technical expertise is required). This is an important advantage compared to Supervised Learning, as unlabeled text in digital form is in abundance, but labeled datasets are usually expensive to construct or acquire, especially for common NLP tasks like POS(Parts-of-Speech) tagging also called

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

grammatical tagging or Syntactic Parsing. Unsupervised Learning models are equipped with all the needed intelligence and automation to work on their own and automatically discover information, structure, and patterns from the data itself. This allows for the Unsupervised NLP to shine.

II. EXISTING RESEARCH STUDY

The first part deals with supervised learning from hand-annotated corpora, and presents a survey along three dimensions of classification. First, we outline different linguistic levels of analysis: Tokenisation, Part-of-Speech tagging, Parsing, Semantic analysis, and Discourse annotation. Secondly, we deal with Machine Learning applicable to linguistic annotation of corpora: N-gram and Markov models, Neural Networks, Transformation-Based Learning, Decision Tree learning, and Vector-based classification. Thirdly, we examine a range of Machine Learning systems for arguably the most challenging task to find the level of linguistic annotation and discourse analysis. The system architecture based on the data collection could depend on the areas in which you want to perform the analysis, then after a very important step to data pre-processing involves various criteria for removal of unwanted symbols, URL, username, etc. Then subjectivity detection takes place in this we remove the subjective statement and objective statements. After the data identification, we proceed with the data extraction for which different existing classification algorithms are used. Finally, the calculation of results takes place based on argumentation analysis patterns like opinion, sentiment, political, legal, essay, articles, etc.

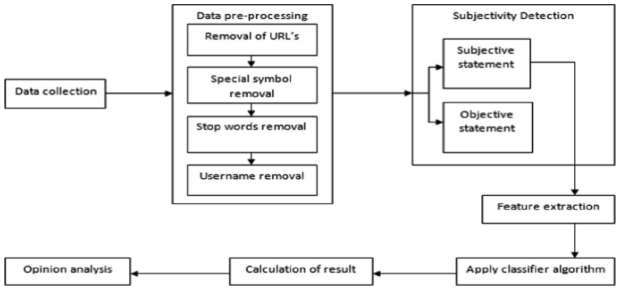

Fig. 1 The system architecture of the research study.

III. WEB BASED ARGUMENTATION MINING

Understanding arguments in social media would yield richer knowledge about the views of individuals and collectives. Extracting arguments from social media is difficult. Messages appear to lack indicators for argument, document structure, or inter-document relationships. In social media, lexical variety, alternative spellings, multiple languages, and alternative punctuation are common. Social media also encompasses numerous genres. These aspects can confound the extraction of well-formed knowledge bases of argument. We chart out the various aspects to isolate them for further analysis and processing. Argument mining (AM) has grown very effectively and focuses on the intersection of computational linguistics and computational argumentation. Given the increased usage of Twitter in political online discourse, investigating the extraction of argumentative text from tweets becomes especially important. In this paper, we provide the knowledge, first critical in-depth survey of the state of the art in social media-based AM. In particular, we have to follow two tasks:

I. Corpus Annotation, and
II. Argument Component and Relation Detection.

A. Corpus annotation: Corpus annotation refers to the practice of adding interpretative, linguistic information to an electronic corpus of spoken and/or written language data.AM model training usually depends on well-annotated data.

B. Argument component and relation detection: To detect all possible argumentative components, present in a text document and identify their relationship automatically. For relation detection solving this task is a necessary precondition for tasks like argument graph ranking method needed to draw based on claim and premises or attack and support.

IV. PROPOSED ARGUMENTATION MODELS

AM system has to perform many strictly interrelated tasks, the existing systems that have been developed till now implement a pipeline architecture (Fig 2) using which they take unstructured text documents as input and produce structured documents as output, where the relations detected in the argument are annotated to construct an argument graph. Each of the stages in the pipeline method corresponds to one subtask in the whole AM problem. The challenges in the AM field share many important similarities with the other subsets of AI fields like Natural Language Processing, discourse analysis, machine language, information extraction, knowledge representation, and computer linguistics.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

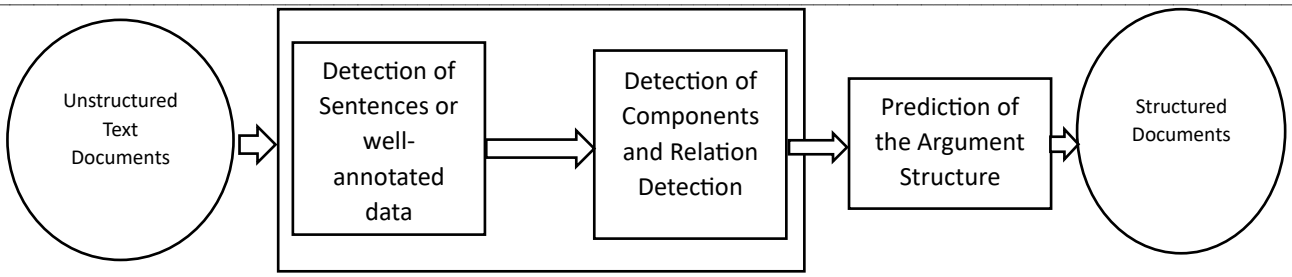

Fig 2: The pipeline architecture for the Argumentation Model

The proposed system consists of approximately 10% structured data in a specific format and the other 90% of unstructured data is still not formatted, which is a focal problem in our consideration. A format here has been regarded as a pure form or preliminary stipulation of this system. In other words, it is also the base state of the system whereas the unformatted data becomes a deviation from this base state. Such deviation can also be assumed as the noise for the associated system. Argument analysis seeks to address this problem by transforming unstructured text into organized argument data. This enables one to comprehend not just the individual indicates being made, but also the connection between them and how they interact to strengthen or weaken an argument as a whole. Despite the reality that there are proof claims that analysis may help people understand vast amounts of data. In the proposed method we tried to prove that if some data is structured in the corpus data analysis, then a huge difference is noticed in the result accuracy. If we can structure the maximum online data in a specific format then data extraction must be more efficient and accurate to enhance the vast impact of the result. Using existing filtration methods, we can apply a practical approach to achieve our target using various classification methods.

V. Conclusion & Future Scope

The two most well-known areas of AI research today are artificial intelligence (AI) and machine learning (ML), both of which hold considerable promise for customer profiling or market evaluation via the mining of internet and web data. After using opinion mining & and sentiment evaluation, which are both growing in popularity in the field of artificial intelligence [Habernal and colleagues, 2014] and are thereby offering reasoning engines for arguments originating on the web or social media, AM could represent the next step in AI. Opinion mining and AM vary in that opinion mining focuses on "what people believe about someone or something," whereas AM employs logic and causes to understand "why" individuals think the way they do to better understand why people have the mentality they do today. The purpose of AM is to investigate the 'human reasoning' process that humans use to rationally accept or reject a claim, view, or idea. The AM method may help in the development of advanced AI systems that can convert unorganized information into structured representations of knowledge in open areas for usage in the future. This concluded that if we can increase the structural arrangement then we would be able to control the result or accuracy to a large extent in our favor.

References

1) Ron Artstein and Massimo Poesio. Inter-coder agreement for computational linguistics. Comput. Linguist., 34(4):555–596, December 2008.
2) Tom Bosc, Elena Cabrio, and Serena Villata. Tweeties squabbling: Positive and negative results in applying argument mining on social media. In Computational Models of Argument – Proceedings of COMMA 2016, Potsdam, Germany, 12–16 September 2016, volume 287 of Frontiers in Artificial Intelligence and Applications, pages 21–32. IOS Press, 2016.
3) Muthuraman Chidambaram, Yinfei Yang, Daniel Cer, Steve Yuan, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil. Learning cross-lingual sentence representations via a multi-task dual-encoder model. CoRR, 1810.12836, 2018.
4) Jan Šnajder. Social media argumentation mining: The quest for deliberateness in raucousness, 2016.
5) Andreas Peldszus and Manfred Stede. From argument diagrams to argumentation mining in texts: A survey. Int. J. Cogn. Inform. Nat. Intell., 7(1):1–31, January 2013.
6) R. Yu, Y. Zhang, S. Gjessing, W. Xia, K. Yang, Toward cloud-based vehicular networks with efficient resource management, IEEE Netw. 27 (2013)
7) H.S. Narman, M.S. Hossain, M. Atiquzzaman, H. Shen, Scheduling internet of things applications in cloud computing, Ann. Telecommun. (2016).
8) C. Delgado, J.R. Gállego, M. Canales, J. Ortín, S. Bousnina, M. Cesana, On optimal resource allocation in virtual sensor networks, Ad Hoc Networks. 50 (2016) .
9) D. Zeng, L. Gu, S. Guo, Z. Cheng, S. Yu, Joint Optimization of Task Scheduling and Image Placement in Fog Computing Supported Software-Defined Embedded System, IEEE Trans. Comput. PP (2016).