سلطنة عمان

The Use of Social Robots Technology in the Educational Delivery Style for Academic Students

By

Reem Rashid Mohammed Al-Maawali

Asila Sarahan Abdullah Al-Rayasi

Dr. Falah hamode

1.1 Introduction and Statement of the Problem / Project

The availability of large textual information stored in free accessible resources, social media, for example, has created the potential for a vast amount of misinformation buried in these resources. This in turn has raised the need for computer-based techniques for utilizing relevant and useful information (i.e., facts and true claims). Misinformation discovery can generally be defined as the procedure of detecting low-quality information from texts, where "low quality" refers to some combination of false, unreliable, rumors, and misleading information. In other words, the goal of the misinformation discovery process can be summarized in finding out intentionally false information (i.e., false claims) from some textual collections [Hotho et al., 2005; Chan, L. et al. 2016; Verónica, et al. 2017; Zhou and Zafarani, 2019; Abdalgader and Al Shibli, 2020]. Despite the advantages provided by available textual information resources (e.g., social media), the quality of information on these resources is lower than traditional peer-reviewed and authentic resources. Due to its feasibility to provide cheap and fast information online that can be easily disseminated through these kinds of sources, large amounts of misinformation, i.e., that information with intentionally false claims, are created online for a variety of purposes, such as business, personal and governmental gains.

The widespread use of misinformation can have a serious negative impact on people, society, and organizations, where it can break the authenticity balance of information, creditability, and knowledge ecosystem. For example, it is found that the most popular misinformation was even more widely spread on social media than the most popular peer-reviewed textual resources. Furthermore, misinformation intentionally encourages people to accept biased or false claims, and it is usually processed by propagandists to convey the desired message. For example, some reports show that many people have created unreal/fake accounts on social media to spread false stories. Generating misinformation also changes the way people interpret and respond to true information. Some misinformation, for example, was just generated to cause disbelief in people and make them confused, obstructing their ability to distinguish between real information and unreliable information [Bin et al., 2016]. To help address the negative effects caused by the extensive spread of misinformation, it is critical that we need to develop linguistic techniques that automatically detect the misinformation on the available textual resources.

Tasks usually executed in misinformation discovery/detection processes include text classification (i.e., classifying text fragments as belonging to one or more predefined classes) [Namburu et al., 2005; Pitigala et al., 2011]; text clustering (i.e., grouping similar text-fragments) [Hatzivassiloglou et al., 2001; Verma, K. et al., 2013], and summary extraction (i.e., producing a summary which captures the main body of relevant content in some textual collections) [Chen et al., 2008; Archana and Sunitha, 2013]. These tasks are not independent, and activity focused on fact-checking or fake news detection, for example, may involve subtasks involving classification or clustering [Vidhya and Aghila, 2010; Srijan et al., 2018].

For example, consider the problem of fact-checking [Xuezhi et al., 2018; Zhou and Zafarani, 2019]. One method to factchecking is to detect the main themes or topics which describe a target textual collection, and then generate a summary of the unreliable information by appending, in a coherent manner, a description of each of those themes.

Presumably, fragments of text that are similar to each other (i.e., true claims) are more likely to relate/vote to the same theme than fragments that are less similar (i.e., false claims). Thus, clustering, using both an appropriate similarity measure and a suitable level of text fragmentation should provide a useful tool in allowing us to identify those themes. The research described in this proposal is motivated by the belief that being able to successfully capture and utilize such lexical and semantic relationships within a target text will lead to an increase in the breadth and scope of problems to which text clustering, text classification, language-features extraction, and other misinformation discovery processes can be developed. However, the performance of any of these linguistic techniques will be limited by the quality of the input data, and in the case of misinformation detection, the performance will depend fundamentally on the quality of the lexical and semantic similarity measure that is used. Thus, while the proposal will make contributions in a number of areas, the issue of developing a linguistic (lexical and semantic similarity measurement) technique holds a central position. Thus, the main specific questions that the research addresses are:

1. Can text similarity measurement for misinformation detection be improved through incorporating supervised or unsupervised approaches?
2. Can text-meaning identification performance be improved by better utilizing the context provided by the surrounding sentences/words?
3. Can relational clustering and classification technique be devised that is better able to capture the complex and subtle interrelationships between text objects at the sentence level?
4. Can the computational linguistic techniques developed for misinformation detection be deployed on a social-robot platform that supports natural language understanding?
5. Can the result of this research serve important sectors of industry/tourism in Oman (real application)?

1.2 Literature Review and Analysis of Related Work

Researchers in the area of natural language processing have been working for many years to define and develop approaches that have the ability to discover buried and useful information from a collection of texts (i.e., documents). Most of these attempts involve text summary, misinformation detection, sentiment analysis, and text mining in general. In the following section, we discuss some of the wellknown approaches to text information processing (i.e., general text mining approaches).

Many text mining techniques (i.e., discovering new information) have been proposed in literature [Srijan et al., 2018; Zhou and Zafarani, 2019; Valsamidis et al. 2013; Vinodhini and Chandrasekaran 2012; Aggarwal 2011; Ringel et al. 2010]. Valsamidis et al. (2010) for example, restrict the analysis to techniques that are specifically associated with largesized text classification. Another contribution is by Brucher et al. (2002), who stated several clustering-based algorithms for text collection retrieval and compared different clustering techniques for pattern extraction from documents. Furthermore, Durga and Govardhan (2011) proposed a new model for textual categorization to capture the relations between words by using lexical resources such as WordNet. This proposed approach maps the words that comprise the same concepts into one dimension and present better efficiently for text classification. Xu et al. (2008) indicated a best practice in the information extraction process based on lexical semantics capabilities and they highlighted many benefits in terms of efficient extraction of the information. Tekiner et al. (2009), introduced a general text mining method to discover relevant summaries from a large number of academic articles. However, the proposed approach ignored the semantic relations between words in the computed text fragments.

The main goal of information extraction methods is taking out specific information from natural language texts for a particular purpose [Wimalasuriya and Dau, 2010]. Usually, these texts contain much information that is not directly ready for processing (tokenization, sentence segmentation, and entity identification). However, machines can analyze a huge amount of text, extract useful information, and assign important attributes to it [Forman and Kirshenbaum, 2008]. Thus, information extraction can be considered as a limited form of full natural language understanding approach, where we already know what type of information we are searching for. As an example, the text source such as “Peter S. Richard, the executive manager of the Information Technology Center, is going to leave from August 1st to November 30th. He will be replaced by Suter J. Philip, the head of the computer science department”, we have to identify the language features, and then take out the information such as people's names, positions, organizations, and date. At this point, Hearst (1999) argues that by saying it cannot be real text mining, which discovers new pieces of knowledge; rather it only finds overall trends in textual sources. Furthermore, investigating new knowledge is like the detective following clues to find the criminal, rather than looking at crime statistics to obtain overall trends in some events.

Information retrieval (IR) is the finding of documents that contain relevant pieces of information a user requires. At the same time, the emphasis is also on the possibility of retrieving as few irrelevant documents [Haiduc et al. 2013]. To achieve this process, statistical methods and measures are used for the automatic processing of text documents, and then contrasting to the given query. Generally, it is based on the idea of question-and-answer approaches that retrieve documents, which are normally ranked according to some criteria (popularly the cosine similarity score between the document vector and the query vector). Besides, IR deals with the entire range of information processing, from data retrieval level to knowledge retrieval level. Due to the gained increased attention with the rise of text documents on the Internet and the need for sophisticated tools to deal with, (e.g., search engines), information retrieval is a relatively old research area that was founded in the 1970s [Haiduc et al. 2013].

While modern information retrieval systems are excellent at retrieving text documents relevant to some queries [Haiduc et al. 2013], they are not able to provide a synthesis, useful, novel, or informative summary (misinformation detection) of these documents. Therefore, text mining or misinformation discovery is different from what we commonly known about these systems. In search engines, for example, the user normally seeks information that has been already written down and known. This leads to pushing aside all results that currently do not satisfy the user's needs to obtain the relevant and useful information.

Information extraction has been confused by the text mining process. It analyses unrestricted text to extract the specific type of information, facts, events, or relationships. In addition, it has been described as the creation of a structured representation of selected information drawn from a collection of texts. For example, tagging all cars or companies' names in a text source is considered an information extraction process. In this case, there is no sound of novelty performed here, and it only extracts the information that is already presented in the text. This information extraction cannot be considered as real text mining, but it has been involved in the text mining processes. For example, one approach to text mining is to use extraction techniques as a pre-process step to obtain specific or interesting text documents and then apply text mining techniques for discovery or investigation.

The main argument of this research proposal is strongly in line with Hearst's analysis that investigating novelty is a vital requirement in text mining. However, we propose a more flexible and appropriate definition than Hearst's view of what is meant by "novelty". Also, we assume a subjective measure of what is or is not considered novel, which would be extremely useful (misinformation detection). For example, there is a different view to recognize a summarization produced as a novel. Therefore, text mining makes intuitive sense to extract as much value as possible from information resources. It is a quite-recent research area that represents a vital step ahead of text retrieval and changing the emphasis on text-based information approaches from low level "retrieve and extraction" to a higher level "analysis and investigation" capabilities.

We have reviewed the most important literature and related works that focus on the view of computational linguistic techniques that can be used for finding useful information and detecting misinformation from a collection of texts. In this proposed research, the belief is that the task aims at detecting misinformation from texts, and this can be achieved through appropriate clustering and classification techniques, which ultimately depends on the effective text semantic similarity measurement, which may, in turn, be based on correctly identifying the actual meaning in which the words are being used in those texts. The meaning of a word needs to be identified in the context of the text in which the word appears, and this presents a difficulty since some text provides only a very limited context.

1.3 Objectives

The main objective of this project is to use artificial intelligence in the field of teaching and education. Which contributes to the development of education in the Sultanate of Oman by using modern information technology, creating a new and advanced educational environment, and enhancing the role of the Fourth Industrial Revolution in education. Where we used a social robot called "Furhat Robot", which was purchased from the Swedish company that manufactured it (Furhat Robotics). Therefore, the sub-objectives of this research are:

- To introduce a novel dataset covering several different domains for evaluating fact-checking fake news and emotion cause detection methods.
- To develop a new method for determining the importance and actual meaning of the words/terms based on the context in which they appear. This is an initial step in the misinformation detection process.
- To develop a new method for text semantic similarity measurement.
- To develop a new method for text clustering classification and extraction.
- To deploy/apply the developed computational linguistic techniques for misinformation and emotion detection on the social robot for demonstrating a real-world application to evaluate the potential benefits.
- To evaluate the effectiveness of the developed methods (techniques) through deploying them to different real world related scenarios in Oman.

1.4 Benefits to Oman

A huge amount of new textual information (i.e., documents) is generated every day through academic, economic, tourism, weather monitoring, and forecasting, social activities, etc. In Oman, many with significant potential societal, economic, and learning values. This in turn has created the need for computerized linguistic techniques for discovering misinformation buried within these textual resources. Academic research, for example, mining, finding out and analytics of big-datasets are delivering efficiency and novel knowledge (false claims detection) in areas as diverse as finance, business, science, and social and law studies; the pharmaceutical and agricultural industries mine patents and research documents (articles) to improve medicine and planet-drug discovery respectively. Industries also use knowledge discovery techniques to analyze customer and competitor textual information to enhance competitiveness.

This leads to the fact that discovering or mining and analyzing of this scholarly literature and other available electronic textual information affords a real opportunity to support the innovation and development of new useful knowledge (i.e., misinformation discovery) in the Sultanate of Oman. To date, however, there has been no automatic linguistic technique that analyses the benefits and value of knowledge discovery or misinformation discovery in Oman. We gathered this fact from consultations with key stakeholders and a set of quick surveys. Oman also will benefit from this research by establishing the first linguistic computing laboratory/group at Sohar University as one of the secondary objectives of this research. This lab/group will provide an opportunity to deal with further research related to natural language processing all around Oman. Finally, this research is motivated by the belief that there is a significant effect of applying misinformation discovery techniques to facilitate and advance research activities across all majors in Oman.

The direct benefit should come, however, from applying the new technique to support different areas of industry and businesses in Oman. The case study planned as part of this project will investigate the application of new techniques in business and tourism. It should evaluate the benefits for different sectors of business and tourism. This case study, once successful, would constitute the start of big potential improvement in a variety of life activities in Oman. It also would pave the way to plenty of benefits for areas like social, logistics, and manufacturing.

REFERENCES

Abdalgader, K. & Al Shibli, A. (2020). Experimental Results on Customer Reviews Using Lexicon-Based Word Polarity Identification Method. IEEE Access, vol. 8, (pp. 179955-179969).

Archana A. B., Sunitha. (2013). “An overview on document summarization technique”. International Journal on Advanced Computer Theory and Engineering (IJACTE), Volume-1, Issue-2.

Aggarwal, C. (2011). Text mining in social networks. In Social Network Data Analytics. 2nd edn. Springer, 353-374. Bates, M. J. (1986). Subject access in online catalogue: A design model. Journal of the American Society for Information Science. 37, 6, 357-376.

Brucher, H., Knolmayer, G. &Mittermayer, M. (2002). Document classification methods for organizing explicit knowledge. In Proceedings of 3rd European Conference on Organizational Knowledge, Learning and Capabilities, Athens, 1-25.

Bin Guo, Yasan Ding, Lina Yao, Yunji Liang, and Zhiwen Yu. (2016). Future of Misinformation Detection: New Perspectives and Trends. Journal of the ACM, Vol. 1, No. 1, Article 1., 23 pages.

Chan Won Lee, Sherlock A. Licorish, Bastin Tony Roy Savarimuthu, Stephen G. MacDonell. (2016). "Augmenting Text Mining Approaches with Social Network Analysis to Understand the Complex Relationships among Users' Requests: A Case Study of the Android Operating System", 2016 49th Hawaii International Conference on system Sciences (HICSS), vol. 00, no. , pp. 1144-1153.

Chen, F., Han, K. and Chen, G. (2008). An approach to sentence selection-based text summarization. In Proceedings of IEEE TENCON02. 489-493.

Durga, A. K. &Govardhan, A. (2011). Ontology based text categorization-telugu document. International Journal of Scientific and Engineering Research, 2(9), 1-4.

Hatzivassiloglou, V., Klavans, J. L., Holcombe, M. L., Barzilay, R., Kan, M.-Y. and McKeown, K. R. (2001). SIMFINDER: A flexible clustering tool for summarization. In NAACL Workshop on Automatic Summarization. Association for Computational Linguistics. 41-49.

Haiduc, S., Bavota, G., Marcus, A., Oliveto, R., De Lucia, A. and Menzies, T., (2013). Automatic query reformulations for text retrieval in software engineering. In Proceedings of the 2013 International Conference on Software Engineering (pp. 842-851). IEEE Press.

Hearst, M. (1999). Untangling text data mining, In Proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics.

Hotho, A., Nürnberger, A. and Paaß, G. (2005). A brief survey of text mining. GLDV-Journal for Computational Linguistics and Language Technology. 20, 19-62.

Ringel, M. M., Teevan, J. & Panovich, K. (2010). What do people ask their social networks, and why: A survey study of status message question &answar behavior. In Proceedings of International Conference on Human Factors in Computing Systems (CHI 10), Atlanta, GA, USA, 56-62.

Salton, G. (1989). Automatic text processing: The transformation, analysis, and retrieval of information by computer. Addison-Wesley, Reading, Mass.

Srijan Kumar and Neil Shah. (2018). False information on web and social media: A survey. arXiv preprint arXiv:1804.08559.

Tekiner, F., Aanaiadou, S.,Tsuruoka, Y. & Tsuji, J. (2009). Highly scalable text mining parallel tagging application. In Proceedings of IEEE 5th International Conference on Soft Computing, Computing with Words and Perception in System Analysis, Decision and Control (ICSCCW), China, 1-4.

Forman, G. and Kirshenbaum, E. (2008). Extremely fast text feature extraction for classification and indexing. In Proceedings of 17th ACM Conference on Information and Knowledge Management, California, USA, 26-30.

Namburu, S.M., Tu, H., Luo, J., and Pattipati, K.R. (2005). Experiments on supervised learning algorithms for text categorization. IEEE Aerospace Conference. Big Sky, MT.

Pitigala, S., Li C., Seo, S. (2011). "A Comparative Study of Text Classification Approaches for Personalized Retrieval in PubMed," IEEE International Conference on Bioinformatics and Biomedicine Workshops.

Valsamidis, S., Theodosiou, T., Kazanidis, I. and Nikolaidis, M. (2013). A Framework for Opinion Mining in Blogs for Agriculture. Procedia Technology. Vol. 8: 264-274.

Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. (2017). Automatic detection of fake news. arXiv preprint arXiv:1708.07104.

Verma, K., M. K. Jadon, and A. K. Pujari. (2013). “Clustering Short-Text Using Non-negative Matrix Factorization of Hadamard Product of Similarities,” Information Retrieval Technology Lecture Notes in Computer Science, Volume 8281, pages 145–155.

Vidhya, K. A. and G. Aghila, G. (2010). Text Mining Process, Techniques and Tools: an Overview. International Journal of Information Technology and Knowledge Management. 2, 2, 613-622.

Vinodhini, G. and Chandrasekaran, R.M., (2012). Sentiment analysis and opinion mining: a survey. International Journal, 2(6).

Xu, X., Zhang, F. &Niu, Z. (2008). An ontology-based query system for digital libraries. In Proceedings of IEEE, Pacific-Asia Workshop on Computational Intelligence and Industrial Application, Wuhan, 222-226.

Wimalasuriya, D. C. and Dou, D. (2010). Ontology-based information extraction: An introduction and a survey of current approach. Journal of Information Science, 36(5), 306-323.

Xuezhi Wang, Cong Yu, Simon Baumgartner, and Flip Korn. (2018). Relevant document discovery for fact-checking articles. In Companion of the Web Conference 2018 on: The Web Conference 2018, pages 525–533. International World Wide Web Conferences Steering Committee.

Zhou, X., Zafarani, R. (2019). Network-based fake news detection: A pattern-driven approach. arXiv preprint arXiv:1906.04210.