

Research Idea

Supervisors:

Dr Hisain Elshaafi

Dr Lei Shi

Big Data Trustworthiness

Big data refers to the extremely large amounts of mainly unstructured data captured and collected by individual organisations from internal business processes and external sources for the purpose of performing analytics and revealing patterns and insights. Big data brings a range of research challenges such as in the areas of storage scalability, data heterogeneity, security and visualisation. Big data is often characterised by 4Vs which are Volume, Variety, Velocity and Veracity. Veracity refers to the fact that big datasets have varying levels of accuracy, certainty, completeness, provenance, privacy and security. Big data is intrinsically heterogeneous with regard to data types, formats, representation and semantics. Erroneous data, conflicting data and missing data are commonplace in Big datasets. Those datasets also frequently contain hidden relationships and insufficient metadata. These characteristics are compounded by the big data's inherent large volumes, wide variety, speed and diverse applications. Some characteristics of big data such as redundancy can, however, compensate for some of those problems and issues. On the other hand, machine learning applications require uniform data structure and abundant metadata. This researcher will investigate the use of those and other wide range of characteristics to help establish the trustworthiness of large volumes of data collected from a variety of sources and their lifecycle. The multifaceted evaluation of data trustworthiness will help organisations improve the utilization of human expert preferences and making proper interpretations of the data.

In recent years, there has been a significant amount of research in the area of trust and reputation in computer science areas such as security, Web services, e-commerce and game theory. The ability to analyse past actions and to monitor current environmental metadata allows us make predictions about the likely behaviour in the future. These predictions are based on trust models and algorithms that seek to replicate social concepts of trustworthiness and that are designed to compute trust ratings and share trust-related information with other actors. The researcher will analyse existing trust models from different application areas and their relevance to the big data lifecycle with the aim of building modified or brand-new models.

Keywords

Big data, trustworthiness, trust model, reputation, machine learning, security.

References

- A. L'Heureux, K. Grolinger, H. F. Elyamany and M. A. M. Capretz, "Machine Learning With Big Data: Challenges and Approaches", in *IEEE Access*, vol. 5, pp. 7776-7797, 2017.
- H. V. Jagadish et al., "Big data and its technical challenges", *Commun. ACM*, vol. 57, no. 7, pp. 86-94, 2014.
- S. Yin and O. Kaynak, "Big Data for Modern Industry: Challenges and Trends", in *Proceedings of the IEEE*, vol. 103, no. 2, pp. 143-146, Feb. 2015.