

Guided Analytics Software for Smart Aggregation, Cognition and Interactive Visualisation

Aleksandar Karadimce¹, Natasa Paunkoska (Dimoska)¹, Dijana Capeska Bogatinoska¹, Ninoslav Marina¹, and Amita Nandal²

¹ University of Information Science and Technology "St. Paul the Apostle" – Ohrid
aleksandar.karadimce@uist.edu.mk, natasa.paunkoska@uist.edu.mk,
dijana.c.bogatinoska@uist.edu.mk, Rector@uist.edu.mk

² Dept. of Comp. and Comm. Eng., Manipal University Jaipur, India
amita.nandal@jaipur.manipal.edu

Abstract. The development of tools that improve efficiency and inject intelligent insights into social media businesses through guided analytics is crucial for consumers, prosumers, and business markets. These tools enable contextualised socially aware and spatial-temporal data aggregation, knowledge extraction, cognitive learning about users' behaviour, and risk quantification for business markets. The proposed Tools for Analytics and Cognition framework will provide a tool-set of guided analytics software for smart aggregation, cognition and interactive visualisation with a monitoring dashboard. The aggregation, monitoring, cognitive reasoning, and learning modules will analyse the behaviour and engagement of the social media actors, diagnose performance risks and provide guided analytics to consumers, prosumers and application providers to improve collaboration and revenues, using the established Pareto-trust model. This framework will provide a seamless coupling with distributed blockchain-based services for early alert, real-time tracking and updated data triggers for reach and engagement analysis of events. Moreover, this will allow users to analyse, control and track their Return on Investment to enhance monetary inclusion in collaborative social media.

Keywords: Guided Analytics, Data Aggregation, Geospatial, Temporal, Augmented Cognitive, Microservices, Social Media

1 Introduction

Social media platforms, such as Facebook, YouTube, WhatsApp, WeChat, Instagram, LinkedIn, and others, are very attractive these days. Besides the main aim of their existence, they have the potential to shape and mobilise communication patterns, practices of exchange and business, creation, learning and knowledge acquisition. One drawback of these social media platforms is their centralisation. Generally, they represent a centralised entity with a single proprietary organisation controlling the network. This feature brings critical trust and governance

issues for the content created and propagated in that environment. Especially concerns the fact that very often centralised intermediaries are involved in the regular data breaches. Therefore, there is a necessity for a new innovative solution that will address the centralised content issue and will facilitate global reach, improved trust, and decentralised control and ownership of the underlying social media environment. The idea is that the next generation social media ecosystem must bring together a range of diverse and fragmented social media actors, such as individuals, startups, Small and Medium Enterprises (SME), and providers, under one systemic umbrella with decentralised ownership, allowing them to participate in collaborative decision-making and the sharing economy.

Such an ongoing project funded by the European Commission under the Horizon 2020 programme is ARTICONF ('smART social media eCOsystem in a blockchaiN Federated environment')³. The main purpose of the project is to create a decentralized and federated social media ecosystem, supported by an underlying blockchain technology that will ensure portable intra-platform and cross-platform social media data, interpretable in a range of different contexts for preservation, analysis and visualization via four frameworks (i) Trust and Integration Controller (TIC), (ii) Co-located and Orchestrated Network Fabric (CONF), (iii) Semantic Model with self-adaptive and Autonomous Relevant Technology (SMART) and (iv) Tools for Analytics and Cognition (TAC).

The four frameworks are integrated part of the ARTICONF social media platform. TIC framework provides a mobile/Web app-enabled backend system for the end-users integrated to a decentralized blockchain platform; CONF consists primarily of a suite of microservices to customize virtual infrastructures; SMART framework comprises of a suite of software services for semantic linking and decision making in a collaborative network of social objects; and TAC provides guided analytics software toolset with smart aggregation, cognition and interactive visualization with a monitoring tool.

This paper focuses on the development of the TAC framework.

This framework is mostly important for the consumers, prosumers and application providers in the direction of enhancing collaboration and revenues. The tool includes guided analytics with social and predictive models for improving the efficiency and inject intelligent insights into operational and mission-critical social media businesses. Hence, the TAC tool deals with contextualised socially aware and spatial-temporal data aggregation, knowledge extraction, cognitive learning about users' behaviour, and risk quantification for business markets. The contributions of this papers are as follows:

1. We propose a TAC framework to enable the social media service owner to have a reach and engagement analysis of the performances of the particular service he/she offers.
2. We implemented a guided analytics dashboard as a proof of concept, using a car-sharing use case. The output of this tool is a visualization micro-service, that will (i) provide better user experience for the social media service users,

³ <https://articonf.eu/>

and (ii) optimize the business, reduce the costs, and increase the revenue of the social media service owners.

The remainder of the paper is organized as follows. Section 2 gives an overview of the related work. Section 3 explains in detail the TAC architecture. Section 4 provides TAC implementation. Finally, section 5 concludes the paper.

2 Related work

A social media network can be seen as an umbrella that covers all web-based applications that allow content generation by individuals [1]. Social media through active collaboration makes people active participants and engages them in exchange of information. Another aspect of looking at the social network is as an online group that brings together people with common interests [2]. Most social networking sites offers integrated and accessible tools to simplify and speed up the information publication process. Nowadays the companies already changed the collaboration approach of focused on individuals within a single company to more sophisticated that facilitate social sessions across organizations. Indeed, the business sector is the one that has effective use of these collaboration tools [3].

The existence of different social media services leads to the production of heterogeneous data. The data has various formats such as images, videos, maps, and geolocation data, hence, they come from various data sources and have inconsistent file formats. Thus, the process of aggregation becomes difficult for managing. One alternative for data management is using large-scale storage, and multidimensional data management in a single integrated system [4], [5]. The convergence of GIS and social media made progress in collecting spatial and temporal data from social media, which withdraw the emphasis on the dynamic process of time-critical or real-time monitoring and decision-making [6]. Further, the interpretation of spatial analysis actualizes concepts like proximity or access, isolation or exposure, neighbourhoods and boundaries, neighbourhood effects, and diffusion [7]. The social media industry provides computer-mediated tools, which enable people or companies to create, share or exchange information [8]. The social-media industry is trying to catch or target users at optimal times in ideal locations. The ultimate aim is to convey information or content that is in line with the consumer's mindset [9]. Social media change fundamentally the communication between businesses, organizations, communities, and individuals [9].

Measuring the impact of the social media integration and specific use case performances on the ARTICONF project is a way of providing guided analytics to consumers, prosumers and application providers, and improvement of collaboration and revenues. The impact assessment of using different tools of the members' collaboration and teams' performance usually is done by quantitative or qualitative measurements [10]. The Return on Investment (ROI) remains as the most common indicator that measures avoidance and reduction of costs, optimizing the business and giving faster business decisions.

3 TAC architecture

In this section, we give a description of the proposed TAC framework, as well as how this framework interacts with other frameworks described above (TIC, SMART, and CONF) within the ARTICONF project.

TAC as a fundamental part of the decentralised social media ecosystem enhances business productivity by tracking updated data triggers from diverse social media events. The TAC tool seamlessly interacts with TIC and provision socially contextual geospatial and temporal data aggregation to gain intelligent insights and prediction through augmented cognition and reasoning. SMART initiates the TAC configuration providing aggregation, monitoring, cognitive reasoning and learning modules that analyse the behaviour and engagement of the application and social media actors, diagnose performance risks, and provide guided analytics to consumer prosumers and application providers to improve collaboration and revenues. TAC interacts with CONF to intelligently provision services based on abstract social media application requirements, operational conditions at the infrastructure level, and time-critical event triggering. TAC initially collects and aggregates smart data in accordance with guidelines provided by the TIC, SMART and CONF recommendation engines. The extracted cross-contextual cognitive inferences together with the established Pareto-trust SLA preferences act as input to the TAC guided analytics engine, which provides visual results on the ARTICONF graphical user interface. The proposed TAC framework architecture is shown in Fig. 1.

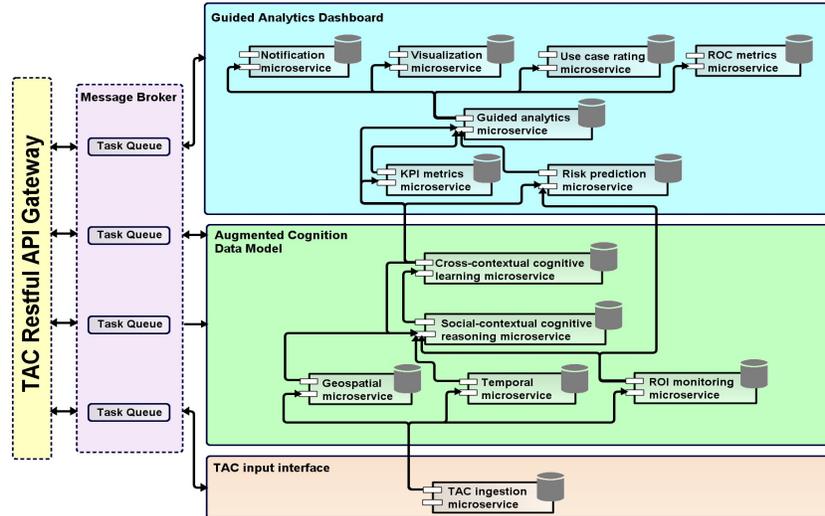


Fig. 1. TAC framework architecture

API gateway provides a RESTful interface to query different microservices provided by the module. **Message broker** facilitates communication between microservices.

Augmented cognition data model consists of four microservices:

- **Geospatial microservice** handles the gathering, display, and manipulation of Global Positioning System (GPS) data, satellite photography, geotagging and historical data, usually represented in terms of geographic coordinates, or implicitly in terms of a street address, postal code, or forest stand identifier.
- **Temporal microservice** offers support to analyse complex social networks providing users with actionable insights even against a large amount of data over a short time period, coupled with visualisation to uncover influential actors.
- **ROI (Return of Investment) monitoring microservice** provides business insights and measurable ROC (Return on Collaboration) metrics.
- **Social-contextual cognitive reasoning and cross-contextual cognitive learning microservices** reduces uncertainty, double-checks the validity of information and their sources in a hostile environment.

Guided analytics dashboard consists of four microservices:

- **KPI (Key Performance Indicators) metrics microservice** runs the testing, measurement and evaluation of the KPI success metrics of the ecosystem, calculated for every use case objective, further exploited according to the social media application provider needs and requirements.
- **Risk prediction microservice** analyses the data received from the ROI monitoring microservice.
- **Guided analytics microservice** provisions the process of guiding the social network users through the workflow and analyses the parameters of its interest through a use case recommendation, moving the analysis beyond reporting shallow summary data to acquire strong and actionable insights from users.
- **Use-case rating microservice** notifies and visualises results from the KPI metrics microservice and the ROC metric from the risk prediction microservice.

TAC input interface consists of a **TAC ingestion microservice** acting as an input interface for the TIC, SMART, and CONF tools for pushing all openly available and anonymised semantic data.

4 TAC implementation

This section will first provide a short overview of the open-source software product ELK stack, as an implementation solution for the proposed TAC framework. Thereafter, the implementation of the proof of concept will be explained using a car-sharing use case scenario.

The implementation of the TAC is done using a collection of three open-source products: Elasticsearch, Logstash, and Kibana, also known as ELK Stack. The main benefit of choosing the ELK Stack as an implementation solution for the TAC framework is a platform that can ingest and process data from different data sources, then store that data in one central data store that can vertically scale as data grows. Finally, TAC will provide a set of tools to analyze the data. The process of collection of the different types of data will be handled by edge hosts, installed with Beats lightweight agents into the ELK stack.

The TAC framework is running the data preprocessing and aggregation using the Logstash tool, which can collect data from various input sources the ARTICONF platform tools (TIC, CONF, SMART, TAC) and use case partners data-set traces. The Logstash can run different transformations and enhancements to the selected data and then ships it for parsing in the Elasticsearch database. The open-source Elasticsearch and analysis engine is based on the Apache Lucene search engine uses the document-oriented index entries which can be associated with a schema and combines “full text”-oriented search options for text fields with more precise search options for other types of fields, like date + time fields, geolocation fields, etc. Finally, Kibana is a web-based visualization layer that works on top of Elasticsearch mainly used for exploring and visualizing data. The modern browser-based interface (HTML5 + JavaScript) provides the end-users with the ability to analyze and visualize the data.

For the purposes of the project, two servers are installed and configured. The first one is an Ubuntu server, that contains the ELK stack installed using a simple docker installation, and the second one is an MS Windows server which contains the Open Source Search (OSS) versions, so we can compare features and have two different testing environments to decide upon.

To demonstrate and validate the ARTICONF ecosystem, four diverse and carefully selected complementary social media use cases are selected. They include (i) crowd journalism with news verification, (ii) car sharing, (iii) co-creation of financial value with video and (iv) smart energy. ARTICONF targets a broad and diverse set of potential customers not limited to its pilot case studies. For the purposed of our proof of concept, the car-sharing use case is selected.

Car-sharing use case, as an example of the sharing economy concept, allows the users to rent and share a car at any place and time through smart contracts based on blockchain. That way, a social network of users is created, so the users can interact with others and report issues directly to the company, vehicle owner or other users. This collaborative consumption will save money, reduce pollution and increase the quality of life in general. The platform will collect and store all the information generated by the anonymous user (e.g. geolocation, social network interaction, external events like a traffic jam). SMART framework within ARTICONF will aggregate and optimize the collected data and send them to the TAC framework. TAC analyses user data in order to classify users and set strategies to target them. Moreover, TAC will provide information about patterns, behaviours, and events for further predictive analytics. The ultimate

goal for this particular use case is to boost user experience and the economic benefits for users and vehicle providers involved in the model.

Using the car-sharing database schema provided by the ARTICONF use case partner, we created two demo datasets with randomly generated data, similar to the real data for this use case. In this scenario, using the simulated data, we have created a simplified dashboard, visualizing some pieces of information important for the car-sharing companies, private vehicle owners and/or users. Different types of information are visualized using different types of graphs (see Fig. 2). Some of the information that we have visualized are for instance the statistical information on the kilometres travelled other is related to the most popular rating score per travel amongst the consumers. The interesting analysis was to observe is the customer's preference for the car-sharing service, whether they prefer to rent the vehicles for kilometres or they have a time-based preference and they are actually passing in average more kilometres. The average price balance is changing per week during the year and this trend is visible with the horizontal bars on the TAC dashboard. Furthermore, we can show external influencing factors, such as weather conditions that affect the usage of vehicles and users habit to use the car-sharing service. Any other information can be visualized on demand.

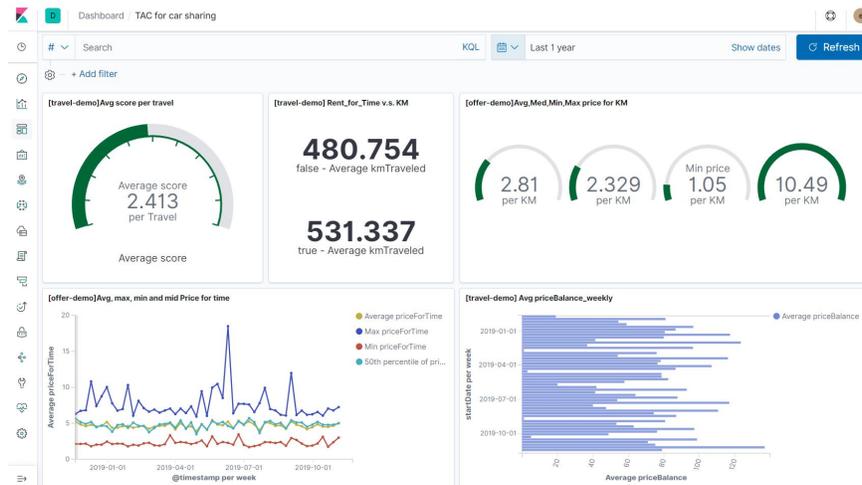


Fig. 2. TAC Dashboard car sharing implementation

The TAC car-sharing dashboard provides the following insights (see Fig. 2):

- A gauge showing the average satisfaction rating score for travel;
- Metrics for average km travelled compared between customers that prefer a charge of priceforKm (false) or charge of priceForTime (true);
- Four gauges showing the average, median, minimum and maximum price for KM for travels in the past year;

- Area plot showing the average, median, minimum and maximum price for a time in the past year;
- Horizontal bar charts showing the average travel prices balance per week in the past year.

If the stakeholders are satisfied with this proof of concept, the development will continue with the full implementation.

5 Conclusion and future work

The main aim of the Tools for Analytics and Cognition (TAC) framework is to uncover which strategy for data aggregation data and group recommendation generation is most appropriate when dealing with groups that are made up of users within a specific use case. Due to a large number of activities, which users carry out as part of a group rather than individually, the TAC framework will improve collaboration amongst intelligently defined communities elaborating over the shared knowledge acquisition and learning. The ARTICONF ecosystem will develop robust tools for monitoring and reasoning social and cognitive states, which will provide social media consumers with enhanced cognitive abilities, especially under complex collaborative participation scenarios using active and automated learning methods. In this way, the augmented cognition reasoning model will improve collaboration amongst intelligently defined communities in a network of social objects elaborating over the shared knowledge, acquisition and learning.

References

1. J. Waters and L. Lester, *The Everything Guide to Social Media: All you need to know about participating in today's most popular online communities*. Karen Cooper, 2010.
2. H. Harrin, *Social media for project managers*. Project Management Institute, 2010.
3. C. Weise, "The return on collaboration: Assessing the value of today's collaboration," San Jose, CA: Cis. Sys., 2010.
4. K. Boulil, S. Bimonte, F. Pinet, Conceptual model for spatial data cubes: A UML profile and its automatic implementation, *Comp. Stand. & Inter.* 38, 113-132, 2015.
5. M. Sarwat, *Interactive and Scalable Exploration of Big Spatial Data – A Data Management Perspective*, *Mobile Data Management (MDM)*, 2015 16th IEEE International Conference, Volume 1, pp. 263-270, Pittsburgh, 2015.
6. D. Sui and M. Goodchild, "The convergence of GIS and social media: challenges for GIScience," *Int. J. Geogr. Inf. Sci.*, vol. 25, no. 11, pp. 1737–1748, Nov. 2011.
7. J. R. Logan, W. Zhang, and H. Xu, "Applying spatial thinking in social science research," *GeoJournal*, vol. 75, no. 1, pp. 15–27, Feb. 2010.
8. Fliphodl, "Social Media Alternatives Series, What You NEED to Know," 2018.
9. K. Hwang and M. Chen, *Big-Data Analytics for Cloud, IoT and Cognitive Computing*, 1st ed. Wiley Publishing, 2017.
10. B. Gholami and S. Murugesan, "Global IT Project Management Using Web 2.0:," *Int. J. Inf. Technol. Proj. Manag.*, vol. 2, no. 3, pp. 30–52, Jul. 2011.