RT-203: Meshing Capability and Threat-based Science & Technology Resource Allocation

Dr. Carlo Lipizzi (PI), Dr. Dinesh Verma (Co-PI), Dr. George Korfiatis (Co-PI), Dr. Razieh Saremi, Dr. Zhongyuan Yu, Mr. Pedro Sá, Mr. Mohammed Khan, Mr. Dario Borrelli, Mr. Rohit Shankar, Dr. Hoong Yan See Tao, Ms. Megan Clifford, Mr. Prasad Desai, Ms. Kara Pepe, Mr. Steven Hespelt, Dr. Tom McDermott, Dr. Jose Ramirez-Marquez, Dr. Steven Hoffenson

SYSTEMS ENGINEERING RESEARCH CENTER

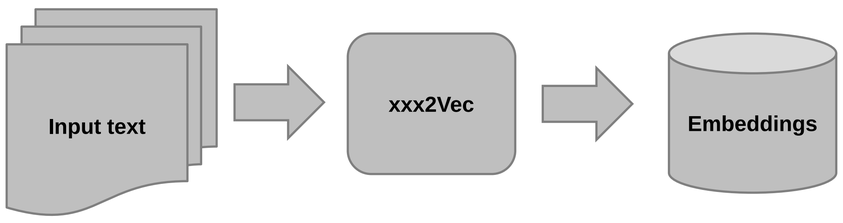
STEVENS INSTITUTE of TECHNOLOGY THE INNOVATION UNIVERSITY

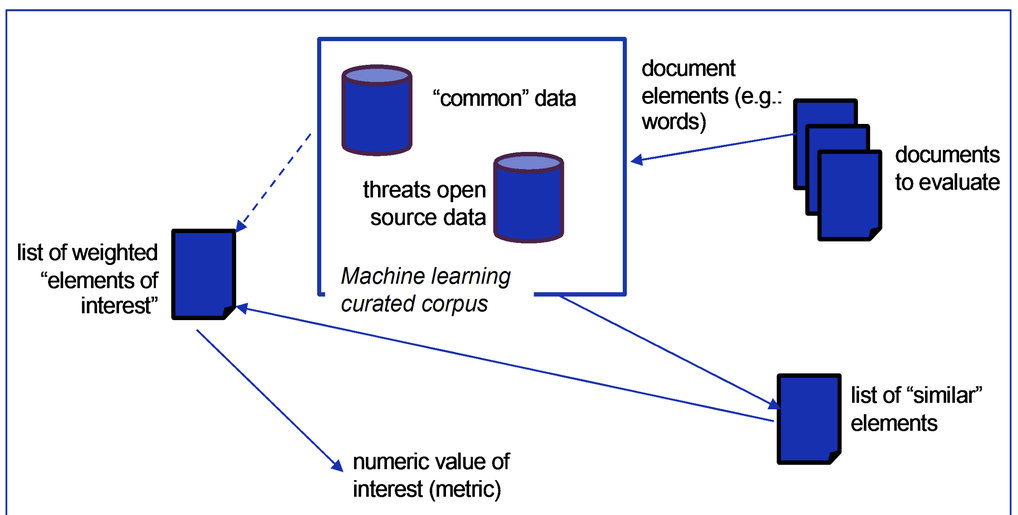
Research Task / Overview

- This research is focused on providing a computational model to support the planning cycle injecting relevant threat-based intelligence and operational scenarios into the more traditional capabilities-based planning
- This approach will better inform the technical communities charged with developing future weapons systems and has been piloted in late 2016 at the U.S. Army Armaments Research and Development Engineering Center (ARDEC) in the armament-systems domain
- This research project will be based on a tailorable approach to allow research iterations that continuously provide value

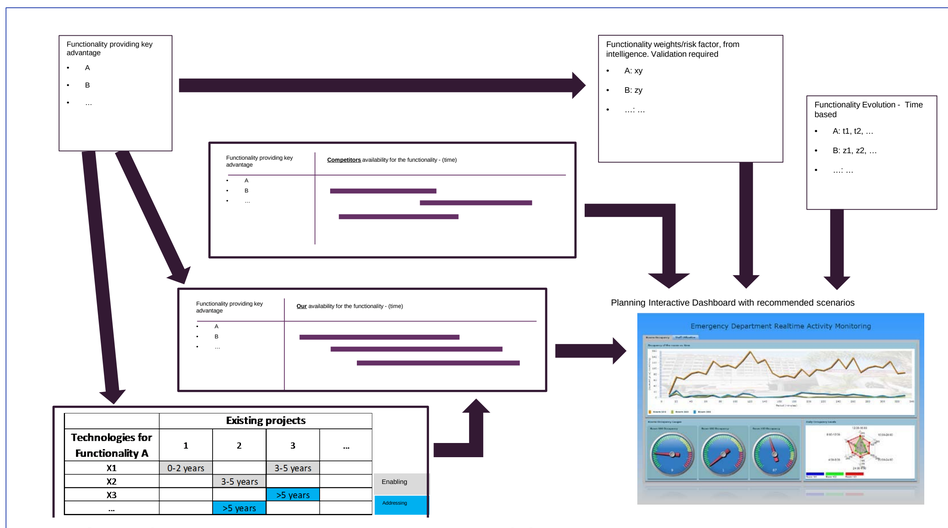
Data & Analysis

- This research uses a data/text-driven approach, initially focused on a proxy-domain to source the data
- Systems are developed as agile growing prototypes with modular components. Most of the components are developed separately for better reusability
- All our models are based on data/text, with human validation
- The data collected are texts related to a specific domain, and selected to be easily associated – for content and complexity – to the final target domain
- We use a combination of traditional Natural Language Processing (mainly for preprocessing) and embeddings. Embeddings are feature vectors for conversational elements in the text, calculated via Python libraries based on neural networks, such as Word2Vec
- From the embeddings, specific metrics are extracted to be used for risk evaluation and for visualization
- Dedicated GPUs are used to create the embeddings from the corpora, driving down the time to obtain the embeddings from days to hours





Component example: Document evaluator



Planning Support System: Architecture

Goals & Objectives

- **Replicate the process developed at ARDEC in 2016** to validate this notional computational architecture
- **Enhance the visualization and analytic capability** to allow rapid, high fidelity decision making
- **Introduce additional parameters and variables** to refine the decision making framework. Real-world scenarios will be modeled to project evolving threats, doctrine, partner force interoperability, and other operational environmental conditions
- Deliver the results as a planning support system with an agile approach, **developing modular prototypes/proofs of concepts with increasing capabilities, with the support of automatic learning components**

Methodology

- The research will be based on grounded theory, a research methodology that allows theories to emerge from collected data. Data will be primarily from text extracted from topic related sources and from interviews, existing planning and budgeting documents. The research will follow a systematic, yet flexible process to collect data, extracting metrics and represent results with interactive visualizations. An additional layer of Machine Learning will be added to provide suggestions
- The research will use standard methodologies for data mining - CRISP-DM (Cross Industry Standard for Data Mining) - adapted and detailed for the specific need to accommodate the decision-support components
- The Systems under development include:
  - **Risk panel – Planning Support System**. It provides an interactive panel that can be used for all the what-is analyses, with a future layer of Machine Learning trained by the user interactions and suggesting the "best" scenarios
  - **Technology Monitoring System**. It will scan for elements outside the defined domain, analyzing a broader set of sources looking for emerging technologies. It will provide the Sponsor with a way to be prepared for future technologies that may have an impact on their activities
- Systems will use Components that are developed separately. Components are:
  - **Basic :** Text Cleaning, Topic extraction, Chunking/named entity recognition, News scanner, Word2Vec embeddings matrix, Word/chunk semantic similarity, Opinion polarity evaluator
  - **Complex :** Weighted "critical topic"- list generator, Document evaluator, Social validator, Social Media collection and analysis

Future Research

- Systems are in an early prototyping phase, with the first working proof of concept ready by Mid of November
- All the basic components are ready to be used and a 1st release of the complex components will be available by Mid of November
- The proxy domain has not been selected yet. The team is working on a "reasonable option", that is ML/AI in connected environments
- Working on a 2 years project extension with the following goals:
  - 1st year of extension will bring the system from the proof of concept state to a working prototype. A preliminary scenario recommendation layer based on reinforced learning will be provided. The prototype will be tuned-up to work on the actual domain
  - 2nd year of extension will evolve the prototype and expand the reach of the scenario recommendation system. A handover to the sponsor will be also performed during this stage

Contacts/References

Dr. Carlo Lipizzi (clipizzi@stevens.edu)