EXECUTIVE SUMMARY

EXPLOITING AI CAPABILITIES TO FACILITATE AND ASSIST THE EVALUATION PROCESS OF IDEAS SUBMITTED IN THE OSIP PLATFORM - FEPOSI (FACILITATING THE EVALUATION PROCESS OF OPEN SPACE IDEAS)

Andres Garcia-Silva

01/06/2022
Expert.ai Research Lab

Executive Summary

The general goal of this project (Exploiting AI Capabilities to Facilitate and Assist the Evaluation Process of Ideas Submitted in the OSIP Platform, FEPOSI (Facilitating the Evaluation Process of Open Space Ideas) is to operationalize and deepen the findings and conclusions of the initial project "Exploring AI capabilities to support the evaluation process of ideas submitted in the OSIP platform." As a reminder, this initial solution was based on AI and aimed to use cognitive computing and natural language processing techniques to support the evaluation of ideas within ESA's Open Space Innovation Platform (OSIP). In fact, the initial project demonstrated that AI techniques and software (COGITO Discover) could be used positively to enhance and ease the evaluation of ideas, by providing some new indicators such as idea novelty (scoring 0-100) based on the notion of similarity between ideas and innovation projects in the space.

Among the criteria against which an idea is evaluated in OSIP, one target aspect to be evaluated is the novelty (innovation component) of the idea, especially how far this idea has been submitted in the past and if other ESA past activities addressing similar questions have taken place. The process to automatize the novelty assessment is as follows. First, the text content of ideas in evaluation status is extracted from OSIP platform (using HYPE API). Next the idea content is processed with Cogito Discover to extract the main lemmas/concepts of the idea and how they are interlinked together. These content metadata, i.e., lemmas/concepts, is used to compare ideas or documents describing projects in the space domain. The goal is to find documents/research papers/projects similar to the evaluated idea. Finally relying on the similar ideas or projects found we calculate on a 0-100 scale the novelty of the evaluated idea and automatically post back to the OSIP platform the summary of the AI evaluation. The intuition is that ideas with similar ideas or projects are less novel.

In FEPOSI we build upon the novelty score and the underlying similarity metric and generate the following outcomes:

New similarity metric for ideas

In the initial project, we define a similarity measure between ideas and projects based on the shared terminology obtained with the help of Cogito. Cogito reads the idea or the project text and generates relevant metadata such as the main concepts, lemmas, and multiword expressions. Ideas or projects with similar metadata is considered similar. Nevertheless, this approach fails when the concepts or lemmas are of high level (e.g., data, or PhD) since potential unrelated ideas are considered similar.

In FEPOSI we extract and assign descriptive and relevant pieces of metadata such as the main **challenges**, **solutions** proposed in an idea, and the **main topic** of the idea so that we can compare ideas in terms of such metadata. We manually annotate a dataset of ideas identifying the main challenges and solutions. Then, we use the dataset to fine-tune a language model in the task of classifying sentences as challenges and solutions. Regarding the topics we use the **NASA Subject and Scope categories taxonomy**. We gather a dataset of projects annotated with categories in the Subject and Scope categories extracted from NASA Technical Report Server. With this dataset we fine-tune a language model in the classification of ideas into the taxonomy. No changes are needed in the novelty score computation since the original formula was defined to work with any similarity metric.

Integration of new data sources

We extend the source of projects that we use in the novelty score computation to compare ideas with. The new source of project is NASA Innovative Advanced Concepts (NIAC) studies. We develop a web scrapping process to download the NIAC studies information from the web site and integrated in our database as source of related projects. The final list of sources of projects is Nebula Public Library, Horizon 2020 projects, and NIAC studies.

Graph visualization of similar ideas and shared vocabulary

This graph visualization available in OSIP for evaluators shows the idea of interest in the center and similar ideas and projects connected through edges where the thickness of the edge is proportional to the similarity score. Similar ideas and projects are also expanded with the shared

vocabulary with the reference idea. The shared vocabulary consists of the concepts added to the ideas by the semantic analysis process in Cogito Discover

Analysis of the Idea similarity graph

We upload the graph of similar ideas into a tool for network analysis and identify clusters of similar ideas sharing a common vocabulary. The clusters of ideas and their representative concepts can be used to identify common topics already covered in OSIP and thus foster the proposal of campaigns addressing less popular topics.

Custom Language model for OSIP: CosmicRoBERTa

The current state of the art in Natural Language Processing is based on deep learning approaches that learn language models that are then fine-tuned to carry out specific tasks. In FEPOSI we train a neural language model starting from an existing language model for the space (SpaceRoBERTa) and further train it on a text corpus extracted from OSIP, SCOPUS, and the NASA Technical Report Server (NTRS). We use this language model to train models able to extract challenges, solutions and to classify ideas into a taxonomy of topics.