**LEVEL 0 SUMMARY**

* **Name of student:** Reckia Jiffard
* **Name of your Level 1:** Mohamed Anissat
* **Source (e.g. scholars.google.com):** Google scholar
* **Paper title:** Augmenting Modelers with Semantic
* **Keywords specific to the paper:** “business process modeling” “deep learning”

**Summary of the main contributions (Use text paragraphs, tables and if necessary, figures):**

This paper presents a new technique for assisting business process modelers during the design process. It is aimed to suggest automatically the next task or element that will be added when making a new process model. It is called "autocompletion" and it helps fasten the modeling. It also ensures consistency by recommending standard elements that are based on existing successful processes. The approach is "semantic" in that it considers the meaning and purpose of process fragments, not just their structure. This allows it to handle elements with similar but not identical labels. To express processes in the form of text, this way allowing analysis based on similarities, the technique uses methods from natural language processing (NLP). The objective is to explore current collections of processes and identify the most suitable choice for advancing an incomplete model.

This paper details the process, from the extraction of segments with specific length to processes within a data set. A slice is a series of tasks connected that you do in order. To convert a slice to text, you need to associate element labels and types. It then encode texts as numerical vectors with the help of a pre trained NLP model, this way, capturing semantic information. The top matches lead to recommendations - the elements following the matched slices in their original processes. For example, if a university admissions process matches patterns where students were invited for interviews, that recommendation would surface.

Several datasets from different domains were evaluated. Metrics included precision, recall, BLEU, METEOR and cosine similarity between predictions and true next tasks. A randomized baseline provided comparison. Experiments analyzed the impact of slice length and ability to generalize across domains. Key findings showed the technique achieves high accuracy for various processes and is applicable across domains. Precision averaged around 0.2-0.3, indicating one correct suggestion out of three on average since recommendations list three options. Recall reached nearly 0.9, showing most true next tasks were covered. Semantic similarity metrics like BLEU and METEOR correlated strongly, exceeding 0.85 in many cases. This validates they appropriately measure recommendation quality.

Compared to random selection, the approach improved precision up to 64 times and recall up to 63 times when focusing just on activity tasks. Slice length had a mild effect - most metrics peaked around 3 elements but did not change drastically with longer slices. This parameter can thus be tuned for new datasets during preprocessing. The proprietary dataset evaluation showed potential for real-world use. While only activity tasks were evaluated to isolate algorithm performance, inclusion of all nodes like gateways or end events would give a more realistic picture.

Future work could conduct user studies to directly assess usability or expand the approach to leverage additional data sources like process execution logs or documentation. Establishing this work as a baseline also enables comparison to future techniques. In summary, this research presented a novel method for semantically enhanced autocompletion of business process models during design. It translated the challenges of limited available data and element variability into an empirical solution leveraging well-established NLP techniques.

Evaluation on several real-world process collections demonstrated the technique can accurately recommend continuations across domain boundaries. This semantic recommendation approach has potential to significantly aid process modelers through an intelligent user experience much like code or email autocompletion systems. Further refinements and user validation could strengthen its practical application for accelerating process modeling.

* **AI model used (e.g. Neural network, etc.)**

The Universal Sentence Encoder (USE) is the AI model used in the document.

* **Introduce the AI models**

This paper introduces the Universal Sentence Encoder (USE) as the AI model used for the embeddings computation. This model has been trained on English sentences and is employed to encode process elements as vectors, enabling comparison through cosine similarity. The authors also mention the potential for future exploration of multi-language models , other encoders to enhance their autocompletion solution. The USE model comes into play to transform text into high-dimensional numerical vectors. The goal is to facilitate some tasks such as text classification, semantic similarity assessment, clustering, and other natural language processing (NLP) tasks. The embeddings generated by the USE model enable the comparison of pairs of sentences by computing a similarity score between the vectors representing the sentences. This approach allows for the transformation of sentences of arbitrary length into vectors of real numbers of the same length, thereby enabling the comparison of pairs of sentences. The paper put in light the benefits of incorporating semantic rooting. The goal is to capture similarities beyond exact label matches and leverages knowledge from a large input dataset of processes, rather than requiring massive training data. Some evaluations were made to demonstrate the accuracy as well as the diversity of recommendations provided by this approach to assist in process modeling.

* **How do they contribute the idea proposed by the paper?**

The AI model used contributes significantly to the idea proposed by the paper by enabling the semantic autocompletion of processes. By leveraging the USE model, the paper introduces a method to encode process elements such as vectors, for allowing the comparison of these vectors through cosine similarity. This approach is used to allow the identification of semantically similar process elements and should go beyond label matches. The USE model's ability to encode text as high-dimensional numerical vectors facilitates the assessment of semantic similarity, element crucial for identifying relevant process elements, may not be exact matches but share semantic similarities. This contributes to the core idea of the paper to provide modelers with a semantic autocompletion tool that leverages existing process knowledge to recommend the next steps in process design. The USE model's embeddings enable the comparison of process elements, and the identification of similar elements, ultimately enhancing the accuracy and diversity of recommendations for process modeling.

* **Supported by a software application? (If yes, provide more details)**

Yes, the idea proposed by the paper is supported by a software application. The paper mentions the development of a recommendation engine that provides accurate suggestions for data sets with different characteristics and from different domains. Additionally, the evaluation of the solution on a proprietary data set demonstrates its suitability for applications in commercial products. The authors also express their intention to conduct a user study to assess the efficiency of their approach in practical settings. This user study would involve process modelers rating the predictions made by the tool, providing further support for the practical application of the proposed idea. Therefore, the paper not only introduces the concept of semantic autocompletion of processes but also demonstrates its implementation through a software application, which has been evaluated and deemed suitable for real-world use.