**LLM-based patent claim classification**

By Lingxi Tang & Camila Acevedo Carrillo

Github Repo: https://github.com/ascentedlife/patent-webapp

Motivation

Patents are crucial representations of technological innovation and are often analysed to study technological evolutions. Patent documents contain claims, where each claim refers to a novel aspect of the invention referred to by the claim. Within the research area of technological studies, there is value in identifying whether a patent claim is an innovation in the product itself or an innovation in the manufacturing process of the product. This is proven by the multiple studies where specific analysis on product and process innovation has been conducted [1][2][3].

The goal of this project is to enable the mass classification of patent claims using LLMs. Then, the classification tool can be used to compare the share of product-process innovation across technologies and draw other insights.

Project Overview

1. Available data: ~8.600 labelled patent claims & ~2 million unlabelled patent claims across 6 technologies
2. Hyperparameter optimization via grid search of XLNet [4] and BERT for Patents [5], using labelled patent claims
3. Using the model with the best performing combination of hyperparameters, perform inference on unlabelled patent claims
4. Developed webapp as data visualization for inference results

Difficulties faced and solutions

1. Issue: Inference of 2 million samples take a long time
   1. Solution: Utilised GPU on Google Colab to speed up data processing
   2. Solution: Inference code designed to be able to stop and continue running across multiple sessions

Contributions of the webapp

The webapp has 3 main features:

1. Patent claim classifier: Try it yourself! Enter a patent claim and it tells you if it’s innovation in product or process.
2. Word Cloud: Review the most common words in patent claims based on specific filters
3. Product/Process Innovation Ratio: See how the product-process innovation share differs across technologies

[1] WM Cohen, S Klepper, Firm size and the nature of innovation within industries: The case of process and product RD. Rev. Econ. Stat. 78, 232–243 (1996).

[2] J Bena, E Simintzi, Machines Could Not Compete with Chinese Labor: Evidence from U.S. Firms’ Innovation. SSRN Electron. J. (2019).

[3] N Banholzer, et al., Knowledge spillovers from product and process inventions in patents and their impact on firm performance. end report, (ZEW-Gutachten und Forschungsberichte), Technical report (2019).

[4] Z Yang, et al., XLNet: Generalized Autoregressive Pretraining for Language Understanding (2019). https://doi.org/10.48550/arXiv.1906.08237

[5] R. Srebrovic, J Yonamine, Leveraging the BERT algorithm for Patents with TensorFlow and BigQuery