Accenture – Graph-Based Strain Optimization for Precision Fermentation

TIN-ACC-021

About the Organisation
Accenture Labs BioInnovation is Accenture’s dedicated arm of research and development on Bioinformatics and Artificial Intelligence for healthcare and life sciences. We offer a blend of industry and academic research activities, including an open publication policy and contribution to the open-source community. All supported by our Accenture Labs network of over 200 applied R&D specialists at seven sites worldwide.

Analysis Methods
Deep Learning, Knowledge Graphs, Graph Machine Learning

Data Source
Publicly available datasets on multi-omics data (genomic, biological pathways, etc) about bacteria, fungi, yeast and mammalian cells

Broader objectives
To design graph machine learning models to optimize strains towards particular objectives.

Role Description and Responsibilities
Recent advances such as Graph Neural Networks and Knowledge Graph Embeddings have been used successfully to predict missing, unseen edges in large graph databases [1,2,3]. In the healthcare context Accenture Labs adopts them in client and research projects for drug discovery, oncology decision-support, and clinical research hypothesis verification. Precision fermentation is currently an active focus of research for a new applicative context to these machine learning techniques.

Precision Fermentation is a protein farming approach finding its roots in the healthcare industry and more specifically pharmaceutics with, for example, the production of insulin [5]. Research is now ongoing to extend its application to the farming of food, for instance with fermented casein, and the replacement of petrochemical products in the cosmetics industry. The optimisation of different fermenting agent strains towards target goals is one of the key aspects of this expansion. Despite the years of experience available in the field and because of new technologies such as CRISPR gene editing, this optimisation remains challenging. There is a potential for graph machine learning to shed a new light on this inherently multi-
omics problem, in the same way it was successfully applied to the precision healthcare context so far. A particular area of interest will be the prediction of the outcome of different strain optimisation approaches.

The research intern will join a precision medicine project and will be in charge of designing, implementing, and evaluating novel principled ways to tackle one or more research problems listed above, with the ultimate goal of inferring knowledge from incomplete clinical and genomic knowledge graphs. The intern is expected to explore and experiment with a range of techniques from prior art, propose original research, and implement ideas that will be validated with the research team in Accenture Labs Dublin.

**Expected Outcomes**

Software prototype, technical report, submission to major AI academic conference (open publication policy), opportunity to contribute to our open-source machine learning library [4].

[4] https://github.com/Accenture/AmpliGraph

**Supervision and Mentorship**

On-site supervision will be by Christophe Gueret and other members of the research team in the Accenture Labs, Dublin.

**Person Specification**

**Essential Requirements**

- Being enrolled in a PhD programme in Computer Science or Bioinformatics
- Familiarity with Machine Learning foundations and Deep Learning architectures
- Familiarity with genomics and precision medicine foundations
- Strong scientific Python programming skills (e.g. NumPy)
- Hands-on experience with at least one machine learning framework (e.g. TensorFlow, PyTorch, JAX)
- Ability to work creatively and analytically in a problem-solving environment
- Interest in solving real-world scientific problems and in acquiring commercial experience
- Demonstrable experience of performing well both when working part of a team and individually
- Excellent verbal and written communication in English
- Ability to pause your PhD for the duration of the internship and return subsequently
Optional Requirements

- Previous exposure to *at least one* of these areas: Machine Learning for Knowledge Graphs (e.g. Knowledge Graph Embeddings), Graph Neural Networks, GNN architectures, Graph ML for Natural Language Processing, Self-supervised Learning on Graphs, Interpretable and Trustworthy Graph ML, Multi-omics data processing
- Previous exposure to Healthcare and Precision Medicine projects

Internship Logistics

This internship will be based in Accenture, Dublin - The Dock, 7 Hanover Quay. Please note that it is **not** possible for this internship to be performed remotely and it requires relocation to Dublin, Ireland.

The start date is expected to be September 2023, although this date is negotiable.

The duration will be 6 months.

The remuneration will be pro-rated, based on an equivalent annual salary of €40k.

*This is a full-time position, and we regret that we are unable to consider part-time applications.*

When applying for this role, please include your resume. Your resume should show any relevant links to illustrate programming experience (e.g. GitHub handle) and scientific accomplishments (e.g. Google Scholar, dblp, arXiv links, personal homepage) if not already captured in the application form.