Employability in Programme Development: Establishing a labour market to higher education feedback loop drawing on local labour market intelligence

ERASMUS+: Employability in Programme Development (EPD) Project

Erasmus+ Programme, Key Action 2: Strategic Partnerships (agreement number 2020-1-UK01-KA203-079171)

Intellectual Output: IO3

Intellectual Output title: Prototype Employability Dashboard

Document title: report setting out the future development pathways for employability dashboards

Contributing authors: [ǁ]Johan Loeckx, [ǁ]Corneliu Cofaru

Author affiliation: [ǁ]Vrije Universiteit Brussel (VUB)

Co-funded by the Erasmus+ Programme of the European Union

report setting out the future development pathways for employability dashboards.

Johan Loeckx, Corneliu Cofaru
Artificial Intelligence Laboratory
Vrije Universiteit Brussel

Introduction

A lot of hope is currently put into the hands of Artificial Intelligence technologies. When used correctly and fed with the right data, wonderful opportunities present themselves. In this document, we will review these opportunities, discuss risks, and suggest a path forward for the increased digitization and use of AI in the domain of employability.

Opportunities related to unstructured sources

1. Exploration of literature

Description
Information on employability is scattered throughout regions, domains and institutions. Information relevant for a researcher may be locked into a Higher Education Institution that is not easily discovered by an academic, or vice versa. AI techniques can help the different stakeholders to find information relevant to their specific context. This use case was elaborated in the first dashboard where a so-called "Insight Engine" was prototyped.

Relevant AI technologies
information retrieval (IR), text mining, Natural Language Processing, Learn-to-rank.

2. Prompt-based systems & summarization of information (LLMs)

Description
At a finer granularity, one could think of so-called "prompt-based" systems like ChatGPT that are trained on the corpus of relevant documents (the same as the use case above). One could then ask questions about the document set as a whole. The difference with the exploration use case is that the system now manipulates information present within a document, instead of processing at the "document"-level.
A related setting would be the *summarization* of documents that could be included in the Insight Engine, to quickly assess whether a document is relevant or not.

Relevant AI technologies
Large language models

Opportunities related to structured sources

The above use cases acted on *unstructured* data sources. Many interesting applications present themselves when we look at *structured* data sources, like surveys and other databases from institutions and governments like employer's surveys, graduate outcomes, employment information etc.

3. Mining of insight

Description
Data science can improve many processes by providing superior insight and draw inferences that go beyond human capabilities. In the picture below we show the main AI tasks. The possibilities are endless, so we will list just some of the options:

1. Predict future number of students or required degrees using *time series forecasting*

2. Use *topic modelling and tracking* to identify new, upcoming skills

3. Create “self-organised maps” to visualize with Neural Networks
4. Predict the likelihood of a student finishing a degree
5. Visualizing an aggregate view of student journeys
6. Assessing effectiveness of degrees by employability outcomes
7. Cluster degrees to find similar or overlapping ones
8. Analyse correlation / causalities between skills offered in degrees and career outcomes

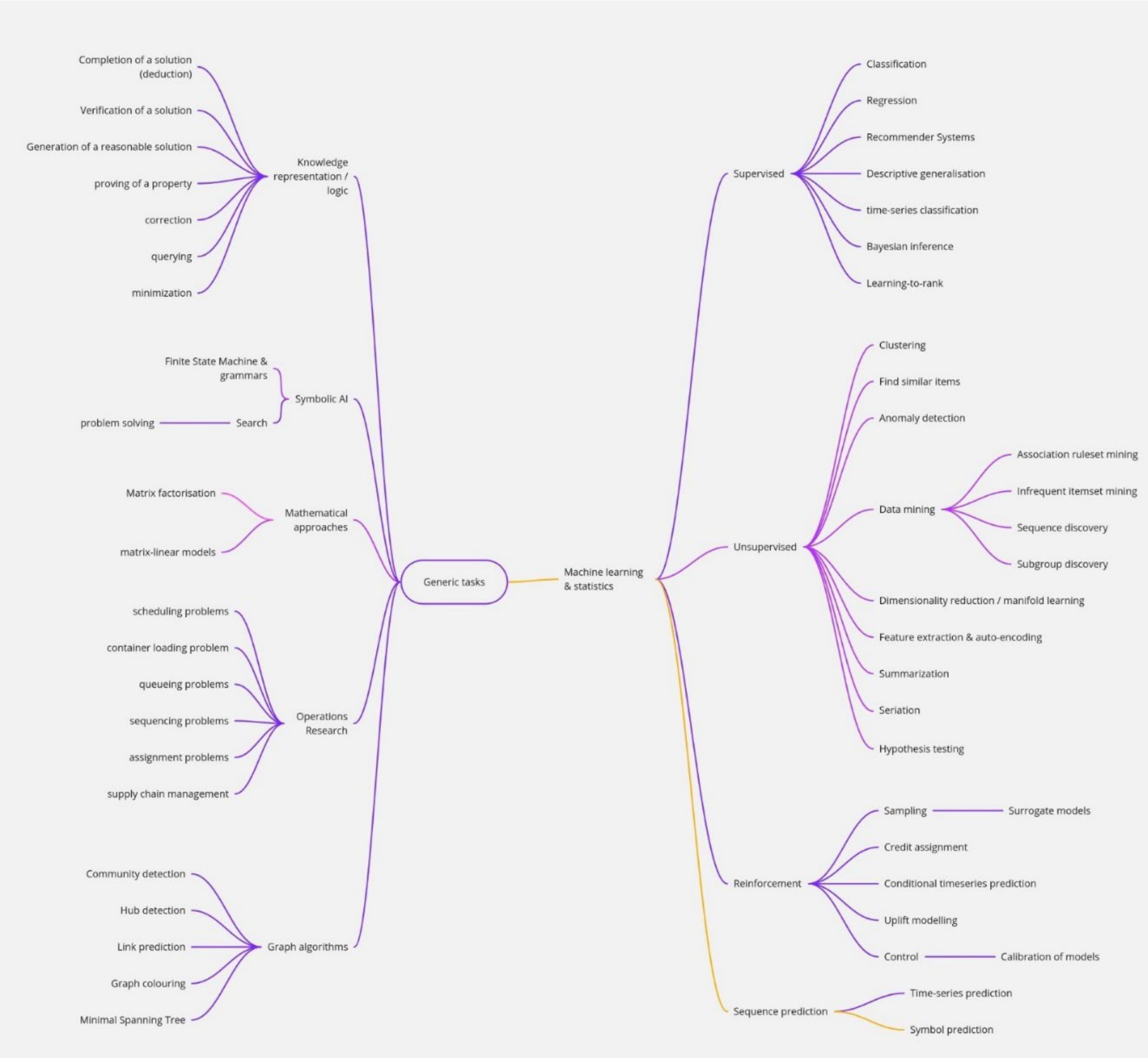
Relevant AI technologies
Data mining, Machine Learning, Knowledge Representation & Reasoning, Planning, Graph algorithms

Prompt-based analytics platform for researchers

One step further is to create a platform for researchers and technology-savvy social scientists where they can ask questions in natural language. These queries are then converted to computer code (in R, Python or another language) that performs the analysis. A prompt-based system, popularized by ChatGPT, is an ideal candidate user interface. as it allows analysts to express what they want in a natural manner.

This use case was elaborated in the second employability dashboard. Rather than relying on a mass of unstructured dataset, it builds on top of knowledge-based techniques that offer complete transparency and do not need excessive amounts of data.

Relevant AI technologies
(Knowledge-based) autoML, Natural Language Understanding, Fluid Construction Grammar



AI-assisted experimentation

Description

One of the potentially most interesting contributions of AI is the optimization of processes through experimentation. Randomized controlled trials (RCTs) can be made more efficient through advanced Bayesian and other techniques (including causal inference).

RCTs suffer from a high cost of "regret". Imagine a simple scenario with two interventions, one with an average pay-off of 5 and the other of 7. A classical approach would split participants into two groups and run 100 interventions for each group. As the first intervention clearly had a lower pay-off, we in fact lost 100 times (7-5). This is what we call regret. There are techniques, already used in medical settings, that can lower the regret once an intervention is performing less.

Another advantage of these techniques is their ability to learn to segment users and to detect the optimal intervention for each participant. These techniques (uplift modelling) are already used in marketing, for example to lower churn.

Relevant AI technologies

Multi-armed bandits, Reinforcement learning, Uplift modelling

Matchmaking

Description
The most obvious and often explored use case for AI in employability is matchmaking. The setting involves suppliers and people in need and could involve higher education institutions with degrees or employers with jobs as suppliers, or degrees that hone skills, and students or graduates as the ones in need.

The strength and main benefit of AI lies in the fact that the criteria for matchmaking should not necessarily be explicitly known but can – in principle – be learnt from examples of successful matches.

Relevant AI technologies
Matchmaking algorithms, Machine Learning

Personalized career paths & recommendations

Description
A similar use case as "matchmaking", but relying on different technologies, involves the recommendation of "items" to "users". It could be used to offer students suggestions on what to study or find jobs relevant to their experience. A step further presents not a single suggestion but creates a "plan" consisting of different steps to guide students toward their goals. It adapts according to the current offer and past success.

Relevant AI technologies
Recommender systems, planning algorithms

Creation of new curricula

Description
A final use case is situated at the "supra-curriculum" level. One could imagine an algorithm at the institution level to design a set of curricula that satisfies particular constraints, for example, to cover skills needs maximally. Another setting could be to adapt existing curricula minimally to suit the need of the labour market better.

Relevant AI technologies
Logic programming, Answer Set Programming

Roadmap

Above opportunities, unfortunately, do not come for free. A lot of preparatory work is needed before technology steps in. This section presents a high-level roadmap for empowering employability with the benefits of AI. Also, the risks will be briefly discussed.

<u>Phase I: Capacity building</u>

Despite the fast progress and spread of AI, there are still many misconceptions about what AI is, how it works, how it can be used and, most crucially, what is needed to implement it. Capacity building is needed:

- To understand what AI is and how it works.
- To identify good candidates and how they can be addressed with AI solutions.
- To identify challenges and manage expectations.

AI is different from other IT technologies in that *it does not exert a single function.* Describing a technology like relational databases, or Infrastructure-as-a-Service is quite straightforward. This is not the case with AI: it is a General-purpose Technology. Any problem that can be reformulated in a particular way (e.g., as a regression problem), can be then solved by a corresponding algorithm.

Capacity building is needed for all involved stakeholders: employers, policy makers, social scientists, career professionals, and students.

Unfortunately, a true and thorough understanding is difficult and takes time. Basic introductory training takes at least one week. Though typically skipped or minimized, it is a first and necessary step towards adopting AI in any field.

Phase II: Harmonization of data

The mantra “data is the new oil” stems from the fact that almost all AI algorithms need “data”, in the form of datasets or knowledge. Those who make their first steps into AI, may however be in for a bitter surprise: data needs a lot of preparation before it can be used! Furthermore, integrating different data sources can be hard to impossible.

Therefore, an important second step – that can run in parallel with the first – is setting up governance structures to create high-quality data that algorithms can exploit:

1. **Map out the existing data sources** related to employability, and characterize their content but also their format (data schema, granularity of information, document the process how it was collected, which surveys were used etc.)

2. **Improve the data quality** of the existing datasets by filling in missing data, standardizing encodings, and correcting errors.

3. **Create a repository** so that the data can be found. This requires agreement among partners and regions, on how to index, describe and search them.

4. **Define a strategy for semantic interoperability.** Different actors and regions will employ different approaches to collect, describe and store data. A strategy needs to be created on how to harmonize these sources. An ontology needs to be defined, as well as rules how to exchange and translate data from one source to another. One could look at efforts in other domains like health (SNOMED) for inspiration.

5. **Harmonize and integrate the existing data sources.**

Phase III: Technology

In a third phase, the technology can be implemented. Please bear in mind that AI systems are highly complex and hard to set up. Furthermore, there is a talent bottleneck. Implementing any project should thus be expected to take a considerable investment (money and time), even more than regular IT projects.

Phase IV: Governance & monitoring

Finally, once successful applications have been set up, good monitoring is needed. Due to the intrusive nature of AI technologies, they can have considerable consequences. People that rely on the technology, could lose skills due to overreliance. Or they could develop tunnel vision as algorithms have a limited focus. Or it could lead to alienation, as students or employees feel they are “but a number”. Finally, one should very closely watch the ethical consequences, as the use of algorithms may lead to bias and positive feedback loop that discriminate people or violate fundamental and other rights like privacy or freedom of expression.