Project Results

PAPUD

A unified approach to heterogenous data

EXECUTIVE SUMMARY

The aim of the ITEA project PAPUD is to help companies to exploit their large amounts of heterogenous data that currently go to waste. Within a platform based around a shared architecture, the project has developed models and algorithms for data analysis with a particular focus on Deep Learning. By tailoring the analyses of both structured and unstructured data to different domains, PAPUD will provide a competitive advantage for companies in the targeted markets.

PROJECT ORIGINS

From social media to Internet of Things, businesses are currently faced with a huge variety of autonomous, heterogeneous data sources. The corresponding 'data deluge' is too much for most to handle, yet almost every industry could benefit from the competitive insights that Deep Learning-based data analysis can unlock. Opportunities can be commercial, such as sentiment analytics, but could also be societal, as with radicalisation detection.

The PAPUD (Profiling and Analysis Platform Using Deep learning) project recognises that the value of Deep Learning lies not in independent analytics processes but rather a unified approach to different types of heterogenous data. It has therefore created a unique software platform (able to scale on a parallel architecture) and new Deep Learning algorithms to optimise the processing of this data. Five use-cases demonstrate the project's success: e-Commerce; Call Centre Operations; Recommendation System for Human Resources; Behaviour Analysis for Reverse Efficient Modelling (BAREM); and Prescriptive Maintenance for High Performance Computing (HPC).

TECHNOLOGY APPLIED

PAPUD's technological innovations are divided between the hardware of the platform, the Deep Learning software and the domain-specific use-case tools. The process begins with the acquisition of data from diverse sources such as surveys,

reviews or calls; this is fed into the platform via application programming interfaces and adaptors. Data can then be defined and characterised through HPC infrastructure and AI tools like TensorFlow, PyTorch and other open-source software libraries. In order to provide a complete application, PAPUD also accounts for privacy: Docker is used to bundle data into separate containers that allow partners to protect and control only their own information.

Following the pre-processing, data analysis is carried out within the Atos-hosted PAPUD platform. A combination of Deep Learning, Machine Learning and Data Mining tools, libraries and resources

(such as TensorFlow and Theano) produce models which are stored and visualised for end-users via a dashboard. The result is a series of recommendations which businesses can use to optimise or improve their processes and services. A concrete example is Call Centre Operations, for which KU Leuven has developed Deep Learning text models. 4C Consulting has integrated these into their AI platform for call centre data analysis, TellMi. The resulting cross-lingual, aspect-based model detects the customer's sentiment during the call and is the first step towards an automated response model. A selection of other tools developed by PAPUD have been made available in an open-source manner on GitHub.

ITEA3
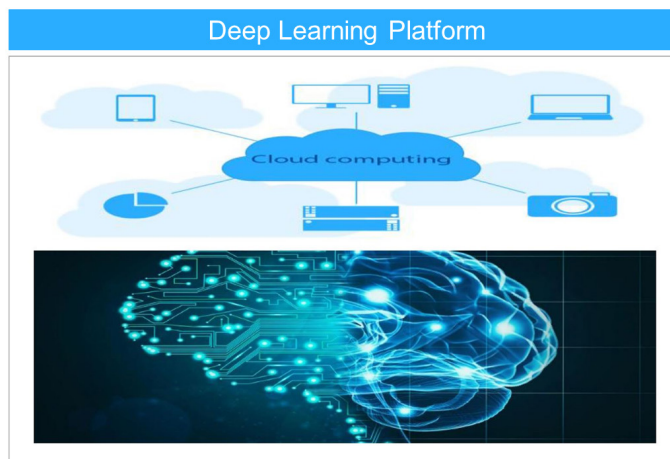
MAKING THE DIFFERENCE

In tangible terms, PAPUD's main contribution to businesses is greater efficiency achieved through sizeable improvements to the previous State-of-the-Art in Deep Learning. For example, the Area Under the Curve (AUC) – the ability of a classifier to distinguish between classes – stood at 0% for keyword extraction at the start of the project but is now 93.7%. Similarly, the accuracy of Deep Learning-based models for HPC prescriptive maintenance was increased from 50% to 95%. For HI Iberia Ingenieria y Proyectos, these kinds of improvements have cut the time taken to find a perfect match through CV processing from five days to three.

Greater efficiency translates into cost savings and increased sales, which the PAPUD project demonstrates as an opportunity for businesses of all sizes. Pertimm, for example, has developed an AI-based recommendation module which will serve as an added value module for their existing e-commerce platform; their expected sales boost (as well as possible collaboration with KoçSistem) has allowed them to hire two new engineers. As for a larger company, Atos will integrate part of

the results into their Codex AI Suite to tackle the most resource and performance demanding use-cases. One aspect is overheating: with PAPUD, 70% of overheating events can be predicted and preventive actions can reduce the costs of this by 65%. As Atos' HPC turnaround was EUR 600 million in 2019, large savings can be expected.

Less tangibly, PAPUD also represents a message to industry as a whole: it is becoming significantly easier for companies to benefit from data analytics. Use of the PAPUD platform and the creation of domain-specific Deep Learning tools allow businesses to bypass huge organisations which have dominated the field, saving them time and money while also allowing them to tailor resources more specifically to their own internal or commercial needs. This should then provide a competitive advantage over companies which fail to make the most of Deep Learning. PAPUD is currently working to disseminate this message, having so far released 29 publications and attended nine events. As awareness grows, so too will the number of markets that the project's results can benefit.

MAJOR PROJECT OUTCOMES

Dissemination
- 29 publications accepted (e.g. IEEE ISCC, ICANN, interspeech, …).
- 5 publications under review.
- 9 events (e.g. CeBIT, IMWorld, …).

Exploitation (so far)
New products:
- Deep learning recommendation for e-commerce.
- Tools for HPC prescriptive maintenance.
- Deep Learning product for HR recruiting process.
- Open Source contribution with PAPUD tools (Logflow, anomaly detection, …).

New services:
- Improvement of call center operations with cross lingual customer care.
- E-services continuous improvement with anomaly detection from logs and implicit feedback to the service owner.

PAPUD
16037

Partners

Belgium
4C Consulting
Katholieke Universiteit Leuven

France
Bull (Atos)
ContentSide
Institut Mines-Télécom
Université de Lille
Pertimm
Softeam
Université de Lorraine

Romania
Beia Consult International

Spain
HI Iberia

Turkey
Ericsson Ar-Ge
KoçSistem Information Communications Services
Setur
Türk Telekomunikasyon
Turkgen

Project start
January 2018

Project end
December 2020

Project leader
Cornel Crisan, Bull (Atos)

Project email
cornel.crisan@atos.net

Project website
https://papud.wp.telecom-sudparis.eu/