BEAT: An Open-Science Web Platform

André Anjos, Laurent El-Shafey and Sébastien Marcel
Idiap Research Institute
rue Marconi, 19 Centre du Parc
1920, Martigny, VS, Switzerland
andre.anjos,laurent.el-shafey,sebastien.marcel@idiap.ch

Abstract

With the increased interest in computational sciences, machine learning (ML), pattern recognition (PR) and big data, governmental agencies, academia and manufacturers are overwhelmed by the constant influx of new algorithms and techniques promising improved performance, generalization and robustness. Sadly, result reproducibility is often an overlooked feature accompanying original research publications, competitions and benchmark evaluations. The main reasons behind such a gap arise from natural complications in research and development in this area: the distribution of data may be a sensitive issue; software frameworks are difficult to install and maintain; Test protocols may involve a potentially large set of intricate steps which are difficult to handle.

To bridge this gap, we built an open platform for research in computational sciences related to pattern recognition and machine learning, to help on the development, reproducibility and certification of results obtained in the field. By making use of such a system, academic, governmental or industrial organizations enable users to easily and socially develop processing toolchains, re-use data, algorithms, workflows and compare results from distinct algorithms and/or parameterizations with minimal effort. This article presents such a platform and discusses some of its key features, uses and limitations. We overview a currently operational prototype and provide design insights.

1 Introduction

One of the key aspects of modern computer science research lies in the use of personal computers (PCs) either for the simulation of known phenomena or for the evaluation of data collected from natural observations. Mashups of these data, organized in tables and figures are attached to textual descriptions leading to scientific publications. Frequently, data sets, code and actionable software leading to results are excluded upon recording and preservation of articles. This situation slows down potential scientific development in at least two major aspects: (1) re-using ideas from different sources normally implies the re-development of software leading to original results and (2) the reviewing process of candidate ideas is based on trust rather than on hard, verifiable evidence [12].

The need and benefits for reproducibility in computational science was already recognized by academia [5, 8, 13] and industry [1], though concrete actions to overcome inherent difficulties are yet to appear into *de facto* standards in this field. For example, it has been shown by MIT researchers [4] that the reviewing process that determines article acceptance in some conferences may by tricked by publications with machine generated content.

While scientific articles normally incorporate a stage of certification referred as peer-reviewing, the very same software frameworks and data leading to the stated conclusions, *when available*, are considered as a bonus and dismissed unreviewed. Even if knowledgable reviewers can predict when

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

written material is insufficiently discussed or poorly argued, one must also consider the hypothesis of rich arguments being coupled to poorly executed software implementations and data quality ending up in misleading conclusions, which do not translate in scientific development. It is a fact that correct and repeatable execution of software over data does not guarantee a good trend either, but less so does just an article. Only by coupling scientific reports, software and data to the extensibility and reuse which is required to verify evidence, can society achieve a faster and steady pace of development. Reproducibility and certification, in this context, should become a must for the future of articifial intelligence rather than a mere bonus.

Publication of research results is not the sole place where advances are needed. The conception of ideas, their embodiment in the form of computer routines and experiments, as well as the actual reviewing process could also benefit from technological advances in computer cloud infrastructures, programming tools (e.g. Web 2.0) and social networking. Albeit limiting, technological challenges do not hold exclusivity in irreproducibility. Many research domains such as those related to medical, biometrics and forensics applications also face legal barriers. Data used in these domains should be handled according to stringent law requirements related to human rights for privacy, which poses obstacles to reproducibility, but also knowledge sharing.

The idea behind a platform for the evaluation of reproducible machine learning and pattern recognition algorithms is not new. Software-based frameworks currently exist in different implementation languages and to attend different purposes. The current main trend seems to be biased towards the creation of web services that ease the management of challenges in machine learning and pattern recognition [10, 2] instead of run-yourself software solutions which were very popular in the last decade. Web services can offer convenient access through different types of devices (computers, tablets and mobile phones), while requiring only a compatible web-browser to be installed on the user machine. With the advent of modern web programming techniques and useful libraries, there is virtually nothing one cannot do through a browser window.

One of the key issues with leading platforms on this market is that of data sharing and privacy. At the same time reproducibility in computational science calls for open data access, certain research domains must respect privacy considerations when sharing data. With new EU privacy law requirements well on the way [3] and matching agreements being reached with other leading countries, personal data transfer must respect formal conditioning and safe keeping - for example, biometric data may be accessible only via end-user license agreements which, typically, disallow copying outside institutional premises. Such a trend will directly influence how research is able to access and share data which, in turn, will impact the adoption of existing solutions. In practice, platforms tackling research reproducibility must incorporate privacy by design[1] (PbD) on their blue prints. PbD can be beneficial to key players in academic domains in which data is not easily transferable, but also to improve the relationship between those, industry and governamental agencies which are also key players in research.

When one talks about research in machine learning and pattern recognition, they must not forget difficulties related to the implementation of its software building blocks, required for the needed repetitive testing, evaluation and performance tuning leading to discovery and reading material. Each of those blocks is implemented over and over using fashionable languages and paradigms through time, existing in a format which becomes outdated as new fashions and paradigms appear. At a point in time, the FORTRAN language was considered the *de facto* scientific programming tool. After that Matlab and nowadays a myriad of options exist to encode knowledge in this domain. You must have asked yourself many times *"Which to pick?"*. Because no right answer is on the horizon, solutions must also take into consideration the hybrid nature in this research domain. Building workflows require tools from a variety of languages to co-exist through time to build the perfect re-usable machinery.

In the remainder of this article, we introduce the BEAT platform,[2] a PbD-built architecture to take on these issues: social development, hybrid algorithm re-use, open-sourcing and confidentiality, providing a new paradigm for the development and evaluation of pattern recognition tools. We present the platform design in Section 2, outlining its main components and core technology. In Section 3 we examplify how it can be used to address data-driven problems in computational science

[1] `https://en.wikipedia.org/wiki/Privacy_by_design`

[2] Operational at `https://www.beat-eu.org/platform/`.

through different use-cases in education, challenge preparation and industry-academia relationship. Finally, at Section 4, we conclude the article with a summary of platform limitations and a look into its future.

2 The BEAT Platform

BEAT is a pan-european project composed of both academic and industrial partners in which one of the goals was the design and development of a free, open-source,[3] online web-based platform for the development and certification of reproducible software-based machine learning (ML) and pattern recognition (PR) experiments. The main intent behind the platform is to establish a framework for the certification and performance analysis of such systems while still respecting privacy and confidentiality of built-in data and user contributions. The framework, as per definition, is task-independent, being adaptable to different problem domains and evaluation scenarios. At the conceptual phase, the platform was bound to support a number of use-cases which we try to summarize:

- Benchmarking of ML and PR systems and components: users should be able to program and execute full systems so as to identify performance and computing requirements for complete toolchains or individual components;
- Comparative evaluation: it should be possible to run challenges and competitions on the platform as it is the case in similar systems such as Kaggle [2];
- Certification of ML and PR systems: the platform should be able to attest on the operation and performance of experiments so as to support the work of certification agencies or publication claims;
- Educational resource: the platform shall be usable as an educational resource for transmitting know-how about ML and PR applications. It should be possible to set-up interest groups that share work assignments such as in a teacher-student relationship.

2.1 Application Breakdown

The BEAT Platform is composed of three main applications: the web, the scheduler and one or more worker nodes. The main function of the web application is to handle authentication and authorization, while the main function of the back-end (scheduler and worker nodes) is to handle the execution of the experiments.

The web application is the main entry point of the platform. It consists of two different major components: a set of **Web Application Programming Interfaces (API)** and a **User Interface**. The BEAT Web API provides a set of entry points that can be used by external applications that wish to communicate with the platform, e.g. to get the status of a given experiment. This Web API is implemented as a RESTful API using the popular Django web framework [7]. On top of this API and also written using Django, a browseable user interface has been built, allowing a user to interact with the platform via a conventional web browser and to make use of the different functionalities provided by BEAT, such as implementing algorithms, starting an experiment or comparing a set of results (see Figure 1). The RESTful API also allows third-party applications to be developed in order to complement the user experience, for example, using smart phones or tablets.

The execution of experiments triggered via the web application is performed on one or more **workers**, which form the computation back-end. Job assignment is intermediated by a central **scheduler** process that assigns computing jobs to different nodes available respecting user quotas and hardware requirements. To achieve this flexiblity, the BEAT back-end closely resembles typical batch-queue submission systems, such as the Oracle Grid Engine™ or TORQUE and organizes its workers into processing queues. Each BEAT user can then assign whole experiments to such queues or individually define which resources each bit of an experiment must use. This technique allows for experimental toolchains which are composed of an heterogeneous pool processing environments composed of software libraries and hardware resources.

2.2 Object Model

All interactions between the web and the backend are done using an abstract object model representing an experiment and associated components, that was specifically crafted to represent machine

[3] Source-code: `https://gitlab.idiap.ch/beat/`

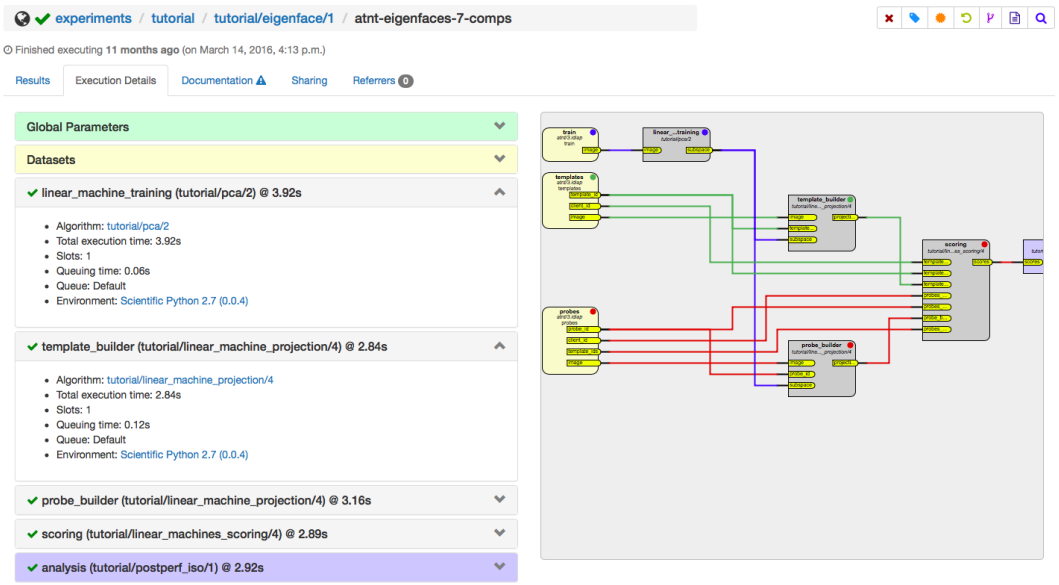

Figure 1: BEAT web user interface. Experiment view: after experiment execution, the user can check experimental details such as the results, but equally computing performance indicators. Each experiment may be certified by the platform guaranteeing its reproducibility, effectively making it read-only; Experimental results from different experiments may be combined into powerful reports that can be exported into publications. This experiment may be reviewed online at `https://www.beat-eu.org/platform/experiments/tutorial/tutorial/eigenface/1/atnt-eigenfaces-7-comps/`.

learning and pattern recognition problems. By configuring an experiment, a BEAT platform user puts together a **toolchain** (see drawing at Figure 1), **databases** and **algorithms** that produce the desired test setup.[4]

A toolchain (or workflow), defines a sequence of interconnected blocks that can perform a certain task (e.g. face recognition using eigen-faces). Each connection in a toolchain defines a distinct strongly typed data flow path, and determines the overall execution order for the blocks. What is not determined by the toolchain is which algorithms execute in each block or what is the input database. Once a toolchain is defined, the BEAT platform provides an easy to use web-based experiment configurator that allows the user to hand-pick algorithms and databases that fit together respecting the block configuration, input and output data format exchange between the block being configured and its surrounding siblings.

To implement this feature, the core object model defines **data formats**, which are user defined data structures implementing the atomic data elements that are exchanged in block connections (see more details in Figure 2). When the user hand-picks a particular database for the input of an experiment, such a database will yield data elements of a certain type through its fanouts which limits the choices for the input blocks containing user code. In the same way, each user algorithm defines input and output data types, that impose requirements on downstream blocks. The experiment configurator takes advantage of this feature, coupled with the algorithms and databases structure (number of fan-ins and fan-outs) to only provide possible combinations during experiment setup, improving user experience with the platform.

Each input data block in a toolchain (blocks on the left in Figure 1) outputs data through one or more channels that are synchronized with each other. This technique makes it possible for a database designer to determine which data points must be served together respecting data restrictions (e.g. associating labels to samples). Each block also determines what is the synchronization channel it will work on. Data is then fed into the algorithm via the block fan-ins in a paced manner, respecting the

[4]User guide: `https://www.beat-eu.org/platform/static/guide/index.html`

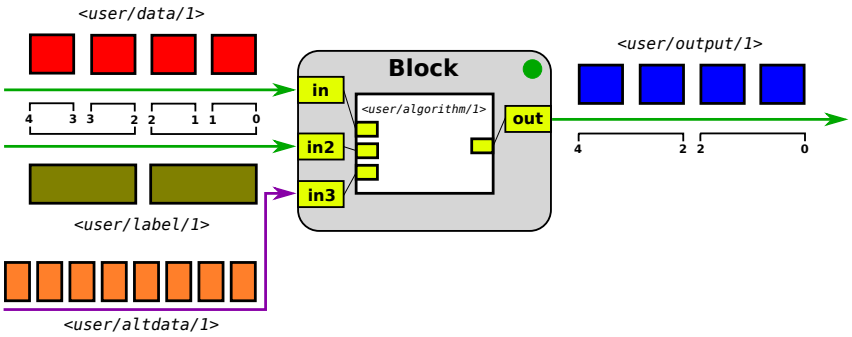

Figure 2: Schematic diagram showing the input of a block, composed of different types of objects which are defined by 3 different *data formats*, called `user/data/1`, `user/label/1` and `user/altadata/1`. Inside the block, an algorithm may be installed for as long as it can input those data types. The block fanout data type is determined either by the output type of the algorithm which, in turn, imposes a restriction on the downstream block algorithms that can be installed. Some inputs and the output of a block are automatically synchronized together so the BEAT plaform is aware of the relationship between original data and the outcome of each block in an experiment.

original database designer usage protocols. This technique allows the user code to be programmed loop-free, since iteration is carried out by the platform itself.

A less obvious yet powerful advantage of this approach is automatic parallelization. Because the platform controls input data iteration, it is possible to split processing in N-folds potentially speeding-up data processing without user intervention.

3 Social Research

Development of new techniques and ideas in academia is very often the result of team work. Ideas are exchanged via e-mails and virtual meetings, a common software framework is selected as a method for sharing implementations and analysis is carried out by the exchange of simple scalars, tables and, frequently, graphics comparing results of experimental setups. Go/no-go decisions are taken in group as part of the analysis process, which typically involves an even greater number of persons (e.g. research supervisors). The BEAT platform can be used in this context to ensure that the same software environment and analysis is executed for all experiments leading to a scientific report, guaranteeing homogenity and reproducibility at all times.

After databases are installed into the platform, one of the parties in the academic team creates (or copies) algorithms, toolchains and experiments that represent the state-of-the-art baseline systems one wishes to improve upon. At this point a clear set of metrics is defined leading to scalars and figures which should be generated on a per-experiment basis. These elements are then shared via the BEAT team feature so that only people in the academic pool of interest have access to core elements of the study and end results.

A search term is then created by a team member (and shared) such that it is possible to keep track of advances when comparing various experiments together. Figure 3 shows an example search filter setup on the BEAT platorm, examplifying a possible setup. The saved search query can be completed with a rich text description. Users have the ability to control which analysis figures to display in the aggregation table and plots. An unlimited number search terms may be stored to express different analysis points of views. For example, it would be possible to setup a search query that would compare the user current work against other algorithmic approaches or, in another instance, how the new user setup improves across databases or different parameter sets.

Once a conclusion has been reached and a scientific report needs to be generated, it is possible to attest all experiments together through BEAT platform *reports*. A report can be thought as a *macro-attestation*, certifying multiple experiments together in a single place. Reports can also be decorated with documentation, can contain tables and figures and be locked for external peer-reviewing, giving an assurance of reproducibility to external reviewers and the research team. A few articles have already been published using the platform [6, 9, 11], demonstrating its capabilities to serve as a reproducible research tool and a portal for extensible and re-usable experimental code and documentation.

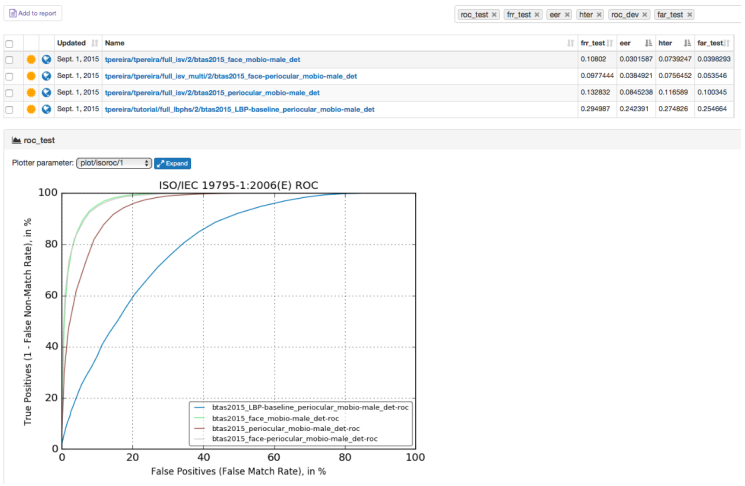

Figure 3: Example of tables automatically generated by the BEAT platform via its *search* feature. It allows you to store specific search terms that result in scalars and figures from the selected experiments to be compared. The user can select how to display the figures associated with their analysis metrics, how to sort and which columns to show on the comparison table. Available online via `https://www.beat-eu.org/platform/search/tpereira/btas2015_mobio_male/`. The resulting report can be found at `https://www.beat-eu.org/platform/reports/751803513/` and provides assurance experiments in the paper are reproducible.

4 Conclusions

The BEAT platform is an open-source software platform for data researchers and data owners, which allows executing and evaluating algorithms on image, audio, video, or multi-dimensional data sets. It can host data that cannot be distributed by conventional means, either because of their large size or because of confidentiality constraints (or both). It offers data-owners the possibility (in agreement with the researchers) to select the processing pipelines most appropriate for their needs, while offering researchers access to big data while minimizing the legal hassles, risks, and cost that accompany conventional data sharing and that currently hamper the research community to fully contribute to solving challenges associated to big data.

Despite the large number of features, the BEAT platform also presents limitations and therefore a lot of potential for improvements. Work leading to a more useful platform still needs to address multi-site scalability, increase the number of programming backends and improve development and debugging tools for new experiments.

- Scalability beyond a single site may allow BEAT platforms operating on different institutions to communicate algorithms, setups and results providing an uniform view to world-wide reproducible experiments;
- As of today, only a Python processing backend is implemented and available on the platform, allowing user algorithms to be programmed on;
- Development and debugging of experiments: Finally, current provisions for developing and debugging code for the BEAT platform could be improved. The migration from a desktop typical research environment into the web-based system may prove difficult for inexperienced users.

We continue to develop the platform towards these goals and welcome new groups and partnerships to enlarge the domains of application of our currently operational prototype.

5 Acknowledgements

The research and development required by the platform and leading to this article has received funding from the European Community's FP7 under the grant agreement 284989 (BEAT) and from the Swiss Center for Biometrics Research and Testing (`www.biometrics-center.ch`). We would like to thank also all the talented engineers from the Idiap research institute: Philip Abbet, Samuel Gaist, Flavio Tarsetti, Frank Formaz and Cédric Dufour. Without their help, an operational platform would not have been possible.

References

[1] Executable paper grand challenge. http://www.executablepapers.com/. Accessed: 2016-06-10.

[2] Kaggle: The home of data science. https://www.kaggle.com/. Accessed: 2015-12-14.

[3] Reform of EU data protection rules. http://ec.europa.eu/justice/data-protection/reform/index_en.htm. Accessed: 2016-06-10.

[4] Scigen. https://pdos.csail.mit.edu/archive/scigen/. Accessed: 2016-02-11.

[5] Monya Baker. 1,500 scientists lift the lid on reproducibility. *Nature: International Weekly Journal of Science*, 533(7604), May 2016.

[6] Tiago de Freitas Pereira and Sébastien Marcel. Periocular biometrics in mobile environment. In *IEEE Seventh International Conference on Biometrics: Theory, Applications and Systems*, pages 1–7. IEEE, September 2015.

[7] D. Greenfeld and A. Greenfeld. *Two Scoops of Django: Best Practices for Django 1.8*. 3rd. edition, Sep 2015.

[8] Barbara R. Jasny, Gilbert Chin, Lisa Chong, and Sacha Vignieri. Again, and again, and again... *Science*, 334(6060):1225, 2011.

[9] Pavel Korshunov and Sébastien Marcel. Joint operation of voice biometrics and presentation attack detection. In *IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–6, Niagara Falls, NY, USA, September 2016.

[10] P. Liang, I. Guyon, S. Escalera, and Viegas E. Codalab: Accelerating reproducible computational science. http://codalab.org/, 2015. Accessed: 2015-12-14.

[11] A. Morales, J. Fierrez, R. Tolosana, J. Ortega-Garcia, J. Galbally, M. Gomez-Barrero, A. Anjos, and S. Marcel. Keystroke biometrics ongoing competition. *IEEE Access*, 4:7736–7746, November 2016.

[12] Keith Price. Anything you can do, I can do better (No you can't)... *Computer Vision, Graphics, and Image Processing*, 36:387–391, 1986.

[13] Patrick Vandewalle. Code sharing is associated with research impact in image processing. *IEEE Computing in Science and Engineering*, 14(4):42–47, July 2012.