

December 2017.

Everware - towards challenge-driven education through reproducibility

Andrey Ustyuzhanin

YSDA, HSE NRU

andrey.u@gmail.com

Tim Head

self-employed

thead@gmail.com

Igor Babuschkin

Manchester University

Abstract: Organization of a computational challenge requires a lot of efforts from organizers. Besides thorough testing of the challenge setup (dataset, metric, leakage) they have to provide participants with sensible problem, meaningful description and runnable/comprehensible results that represent current state-of-the-art approaches. Such baselines should be executable easily by students. Everware is a jupyterhub-based solution that integrates github, docker and jupyter-based solutions. It has been used for several challenge-driven educational events and has shown usefulness and easiness at the same time.

Challenge-driven education model gains significant popularity as diversity of educational models grows. Such model allows getting rapid results by focusing participants minds on specific problem those were not familiar before. thus such model is very attractive for organizers - problem providers and for participants. The latter learn quite a lot during short period of time if the challenge has been prepared properly. Examples of such educational model include challenges on kaggle.com, (data analysis) hackathons, summer schools.

Key aspects of the challenge-driven education models include:

- novelty of the problem to participants;

- time-boxed event;
- heavy computational part (optional).

Events following this model may face the complications like those:

- explanation of basic terms/conditions/assumptions;
- give participants hands-on experience with the problem;
- syncing of the solutions;
- ensuring organizers can reproduce the results;
- access to computational resources. accessibility of challenges are very similar to reproducibility of scientific research problems.

Proposed solution for computational part (basic assumptions) is:

- based on Jupyter;
- facilitates literate programming;
- supports variety of programming languages.

We present project Everware that seamlessly integrates github/gitlab, Docker and Jupyter helping with a) sharing results of real research and b) boosts education activities. With the help of everware one can not only share the final artifacts of research but all the depth of the research process. This been shown to be extremely helpful during organization of several data analysis hackathons and machine learning schools. Using everware participants could start from an existing solution instead of starting from scratch. They could start contributing immediately. Everware possibilities:

- allows recreation of an environment
- repetition of the results
- run on own laptop and on the cloud
- integration with github and hence CI for re-check of the results

Examples of everware use:

- summer schools on machine learning for HEP (MLHEP)
- goto hackathons organized by Higher School of Economics
- Yandex hackathons
- kaggle challenge starter-kits
- LHCb master class

Technologies behind everware include jupyterhub, docker swarm, github. Everware is supported by the Mozilla science lab and Yandex. It is being evaluated as an option for analysis preservation at LHCb. As an open-source project it welcomes contributions of all kinds at: <https://github.com/everware/everware>

REFERENCES

Everware source code repository - <https://github.com/everware/everware>

Jupyterhub source code repository - <https://github.com/jupyterhub/jupyterhub>