


Incorporating Machine Learning Into Mascot Server 3.0 Using A Customizable Adapter

Richard Jacob¹, Patrick Emery², Ville Koskinen²

¹Matrix Science Inc, Boston, MA, USA

²Matrix Science Ltd, London, UK

richardj@matrixscience.com  @matrixscience.bsky.social

How it works: Mascot calls the adapter script, which reads the search results and extracts peptide sequence match information. It passes this to the Machine Learning (ML) pipeline and runs it. The adapter inserts the predicted features from the ML into the Percolator input file (PIP) along with the core features from Mascot. Mascot runs Percolator for multidimensional analysis and rescoring creating a Percolator output file (POP). Mascot combines the posterior error probability (PEP) results with the initial search results and runs protein inference. Mascot also generates a machine learning quality report.

Implementation: The initial implementation of the adapter script is with the MS²Rescore pipeline from the Martens Lab at the University of Ghent. Two ML components are included: DeepLC – which provides retention time prediction and includes 20 models covering different column types, gradient lengths and peptide properties; and MS²PIP - which provides spectral prediction and includes 13 models covering a range of fragmentation, instruments and peptide properties. All the ML analysis is run locally on the Mascot Server and does not require a GPU.

Results: By using the MS²Rescore pipeline better sensitivity is achieved at the same FDR. Depending on the data set increases of 30% more significant sequences is not unusual.

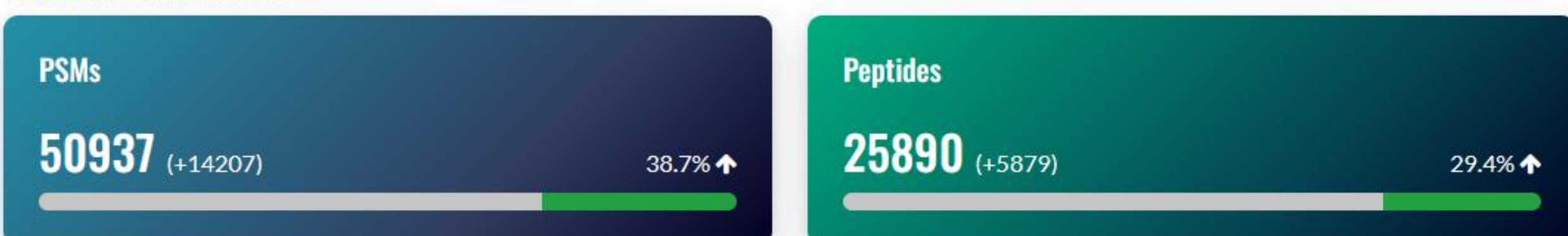
Developing Custom Adapters: Custom ML adapters for other pipelines can be written in any language supported by Mascot Parser (C#, C++, Java, Python, Perl)

Configuration: ML adapters are defined in a TOML file. This file specifies the program, interpreter (if required), whether it's visible in the user interface, command-line arguments, description, etc.

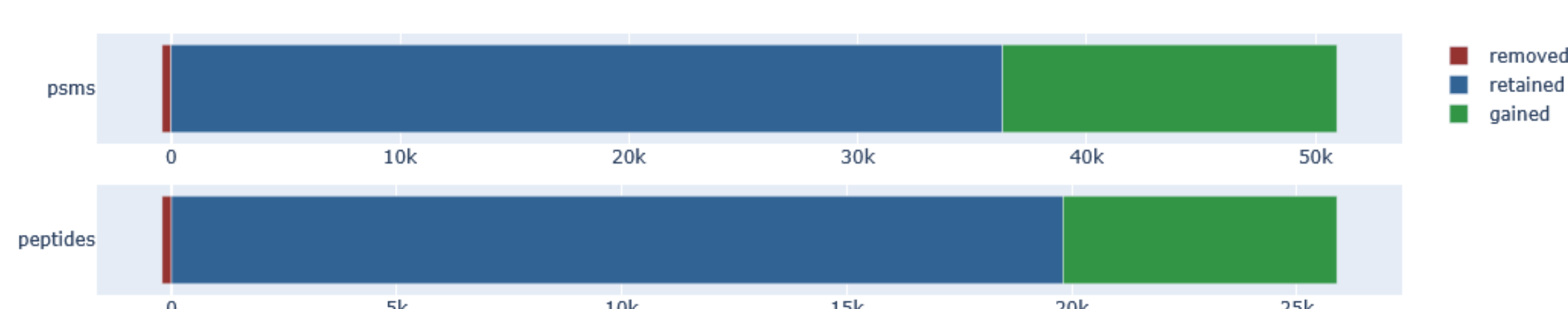
MS²Rescore Customization: It is possible to add custom DeepLC models but currently not custom MS²PIP models.

Example results: Thermo Orbitrap QE HF-X DDA run of human, yeast, E. coli mixture (PXD028735) – single data file with MS²PIP HCD2021

General statistics



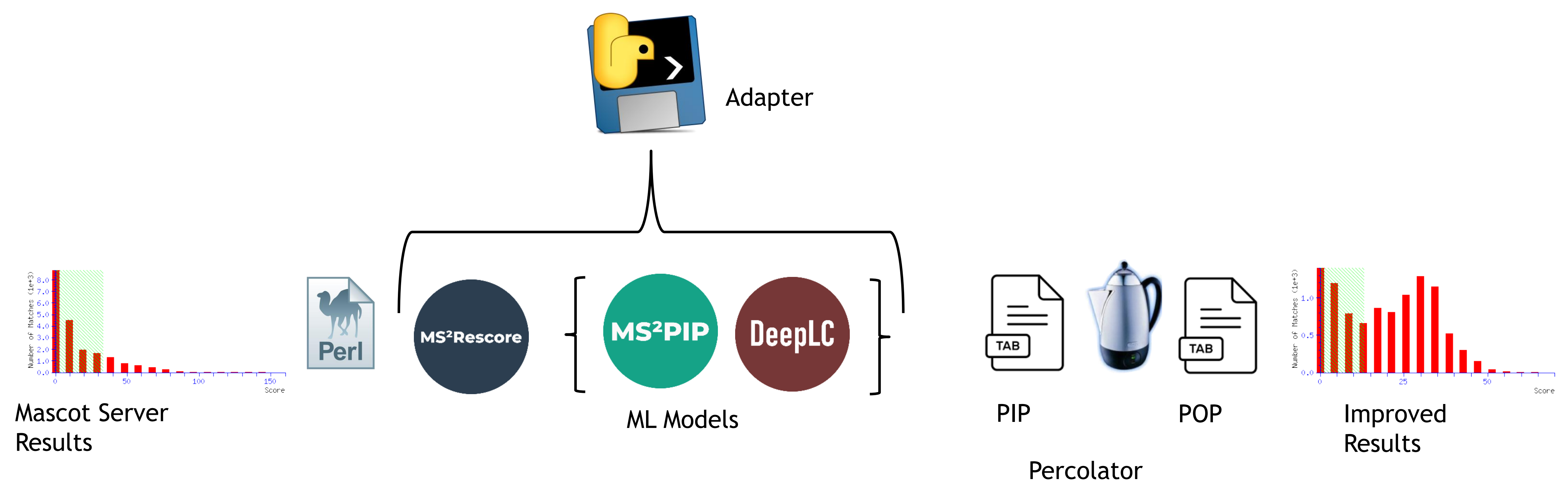
Removed, retained and gained by rescoring



Mascot Server now has a **customizable adapter framework** allowing integration of third-party machine learning pipelines.

The initial implementation is the **MS²Rescore pipeline**, which leverages **DeepLC** retention time and **MS²PIP** spectral prediction models to refine search results.

There are **significant improvements** in PSM and sequence matches after refining.

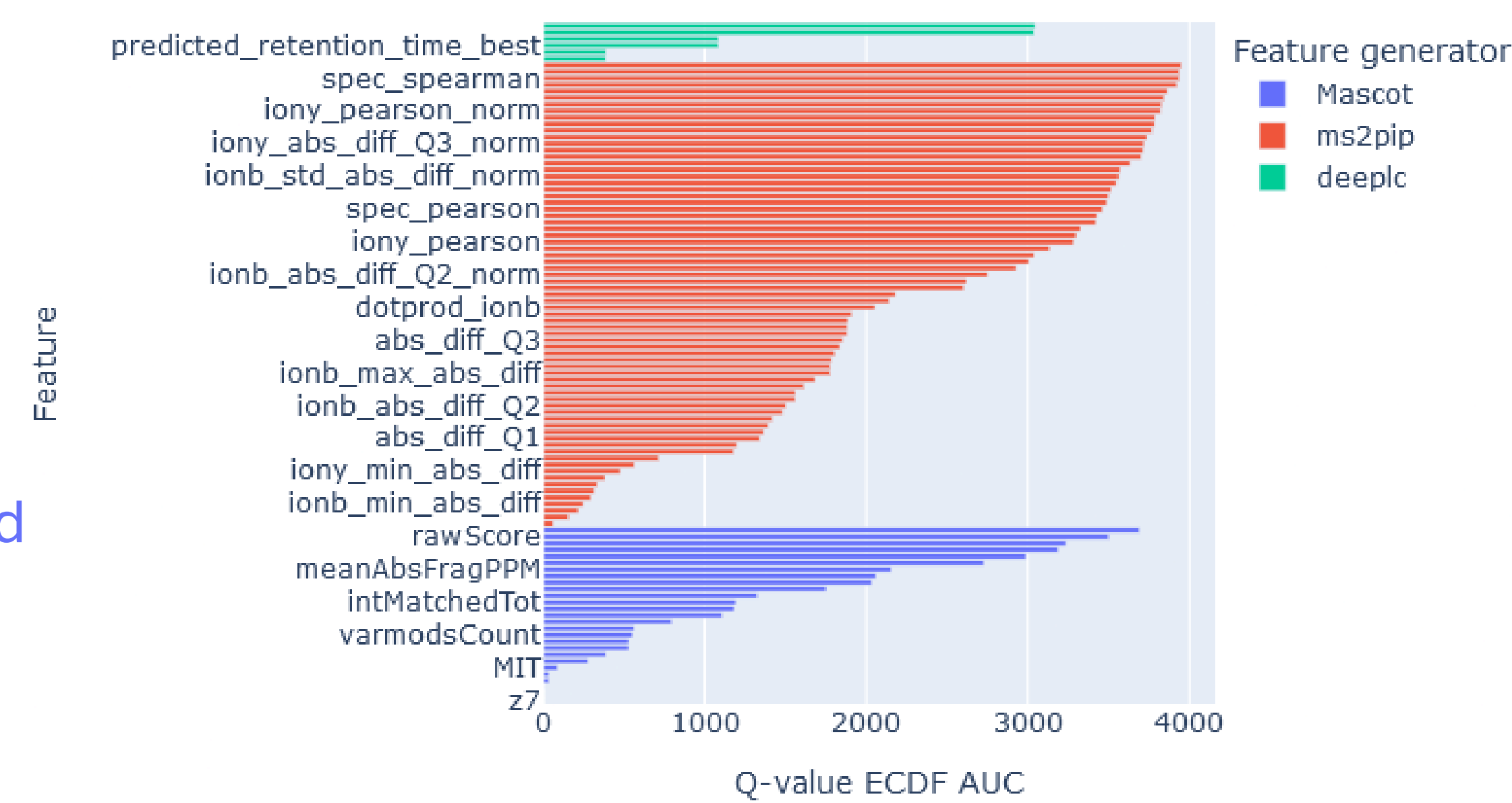


Machine learning report

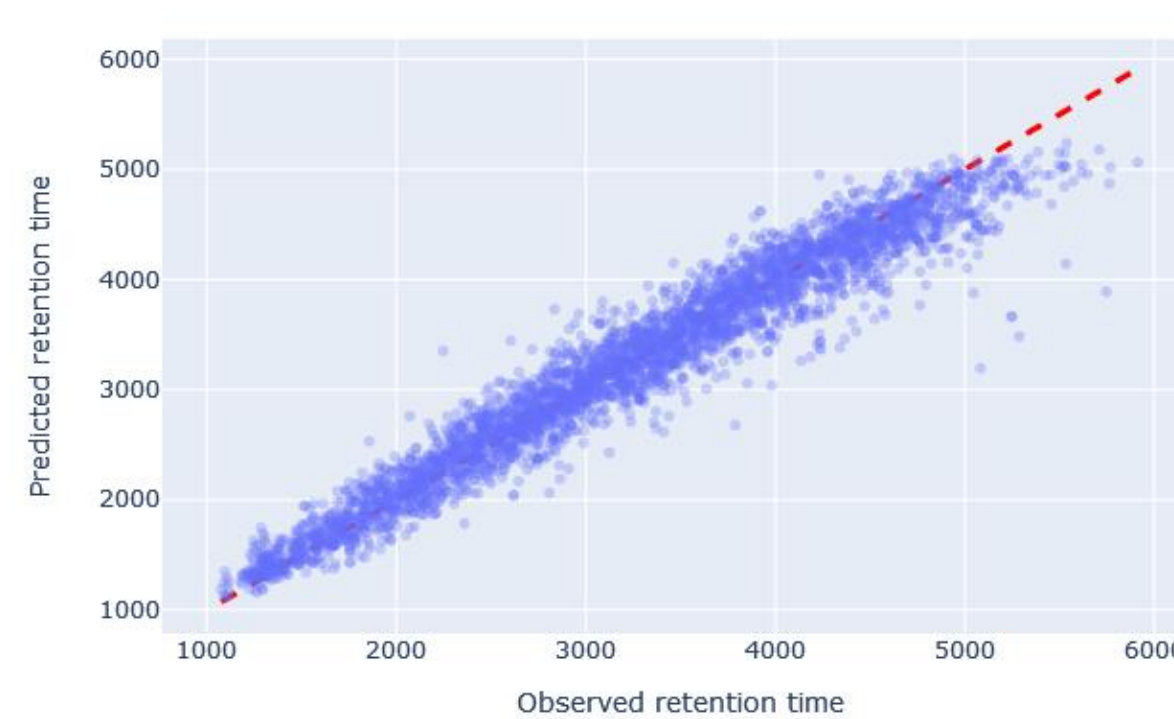
DeepLC retention time features

MS²PIP spectral correlation metrics

Core features calculated by Mascot

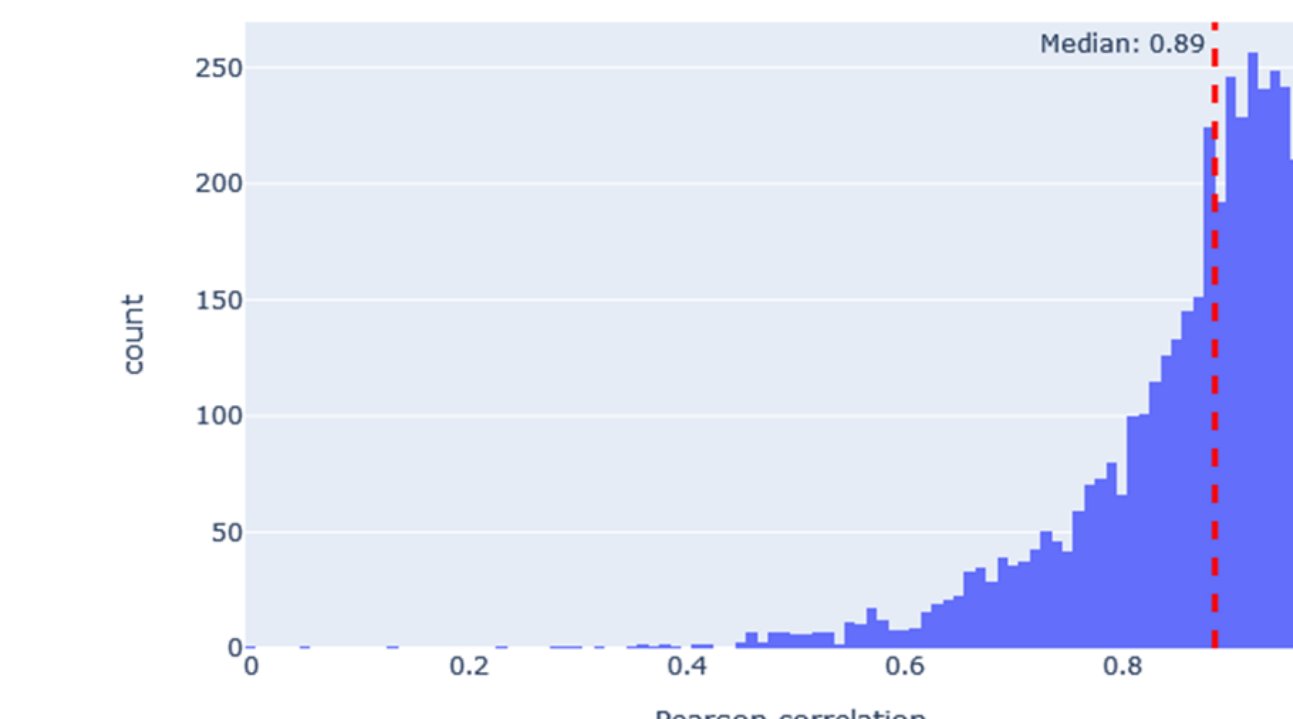


Predicted vs. observed retention times

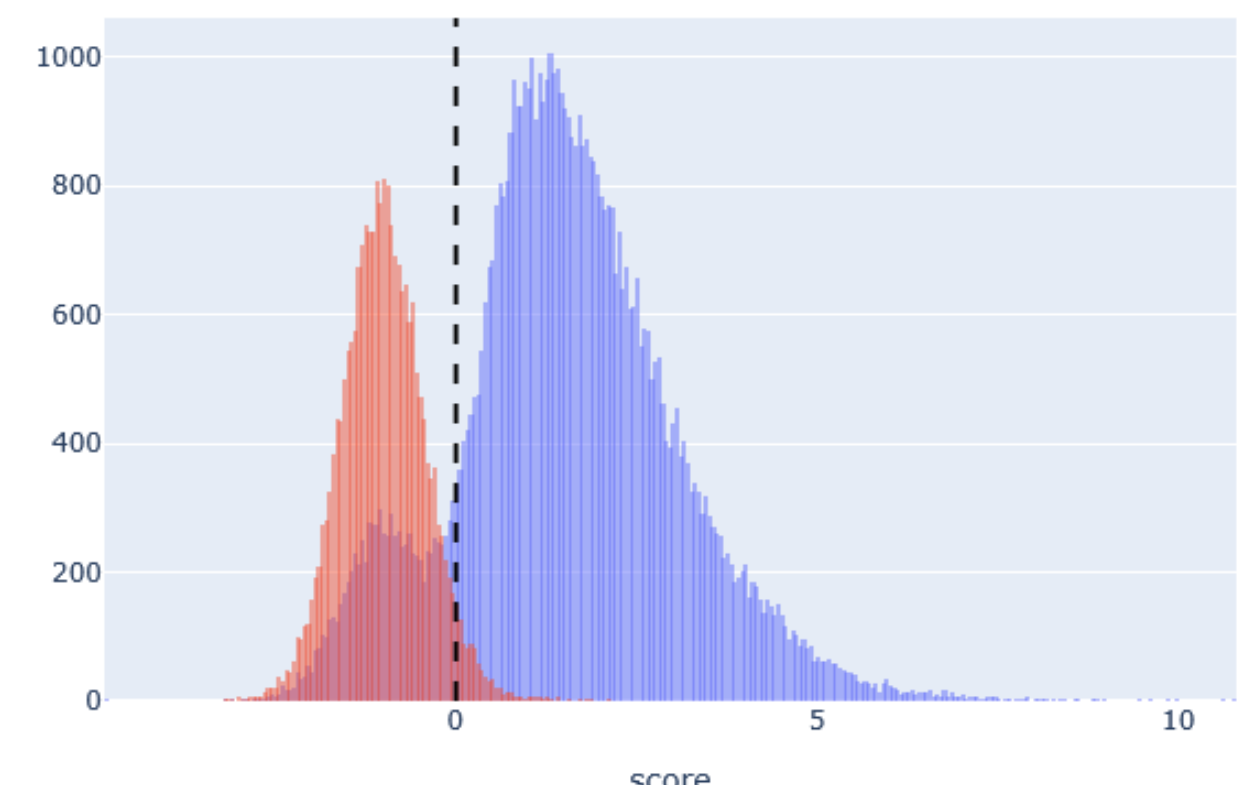


DeepLC

Median: 0.89



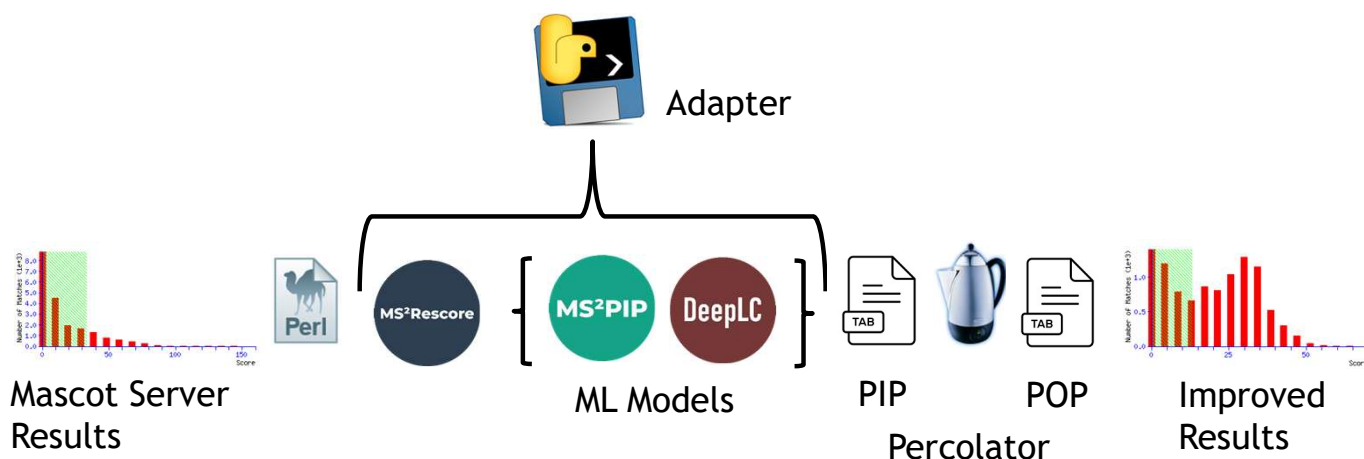
MS²PIP Pearson correlation coefficient



Score histogram for both target PSMs (blue) and decoy PSMs (red)



Incorporating Machine Learning Into Mascot Server 3.1 Using A Customizable Adapter



MASCOT

P05.11 Richard Jacob Poster session 1

© 2025 Matrix Science



Mascot Server has been using machine learning and percolator for about 10 years now. Mascot identifies peptides in the original search and a Perl script extracts those matches creating a PIP file that's passed to percolator, refining the results, outputting a pop output file, with the rescored results, that are presented to the end user.

In Mascot Server 3.0 and up we've incorporated additional machine learning tools and we're using a customizable adapter. Initially we have incorporated MS²Rescore but it's possible to add other machine learning pipelines. The MS²Rescore currently contains 2 components MS²PIP a spectra fragmentation prediction model and DeepLC a retention time prediction model. MS²Rescore is also customizable and you can extend it by adding additional models.

And you pip-pop, and you don't stop 'Till you refine the results and come to the poster for more details.

Incorporating Machine Learning Into Mascot Server 3.0 Using A Customizable Adapter

The latest release of Mascot Server has extended the machine learning options. Mascot Server uses Percolator, a semi-supervised machine learning algorithm, to rescore and refine results based on core features calculated from the Mascot matches. Mascot now provides a framework for the incorporation of additional third-party pipelines using a machine learning adapter. An adapter is a custom program which accepts a search result as input, manages a machine learning pipeline, and returns a tab delimited file with the machine learning results ready to be inserted into the Percolator input file. Using this new framework, Mascot Server 3.0 includes a machine learning adapter for the MS²Rescore pipeline, that uses two models to compare queries to predicted physico-chemical properties.

The integration of the MS²Rescore pipeline from Ghent University allows a researcher to select a DeepLC retention time prediction model and/or a MS²PIP spectral prediction model along with the other search parameters or to refine the results post search. Once the Mascot search is complete the machine learning models are used to compare the identified peptide against the predicted retention time and predicted spectra. The results of the comparison are evaluated with multiple features for each model being calculated. The features are combined and passed onto Percolator for multidimensional analysis. After rescoring, Mascot reads back the posterior error probability (PEP) values and finds a PEP threshold that yields the target false discovery rate. This provides a new score threshold that condenses the multidimensionality into a single number. A machine learning quality report provides an overview of the Percolator results. We show how the use of these models improves the results and identifies more peptides compared to older versions of Mascot Server.

Useful links:

https://www.matrixscience.com/help/mascot_server_3_0_release.html

<https://www.matrixscience.com/help.html#REFINING>

The custom adapter guide is in the set up and configuration manual. Please contact support@matrixscience.com if you would like to implement a ML pipeline.