

IL14**Semantic Ion Vectors - deep learning applied to mass spectrometry**

Starčević Antonio^{1*}, Zućko Jurica¹, Damir Oros¹, Diminić Janko¹, Cindrić Mario²

¹*Faculty of Food Technology and Biotechnology, University of Zagreb, Pierottijeva 6, 10000 Zagreb, Croatia*

²*Ruđer Bošković Institute, Bijenička cesta 54, 10000 Zagreb, Croatia*

*E-mail: astar@pbf.hr

Protein mass spectrometry is the dominant method used for protein characterization. Peptide mass fingerprinting (PMF) is a phrase given to application of mass spectrometry (MS2) to protein identification. In case of PMF, peptides do not fragment sequentially and the process is not entirely random, with some fragmentations being more preferred over others. The resulting fragmentation spectrum captures fragment ions.

A novel approach which relies on Deep learning techniques to capture distributed representations of ions into Paragraph Vectors using unsupervised algorithm was employed to predict subset of informative b and y ions and distinguish them from the “noise” ions. Unlike Word embedding this approach has taken a different turn. Numerical data were turned into words (tokens), which were grouped into sentences and afterwards embedded using Paragraph Vectors. Several tokenization schemes have been implemented and performance has been tested under different parameters. Sole fitness criteria used was the performance of simple binary classification of into informative “ b-y ions” and “noise” (the rest). The best performing combination produced Semantic Ion Vectors, which were fed into classifier of choice and served as a model to make predictions.

Resulting Semantic Ion Vectors can be used in variety of classification tools and provide accurate predictions of informative ions. Tokenization method developed, efficiently reduced complexity of this task. Intensity patterns and other information, commonly used as a set of rules in a synthetic manner (more suited to organic brain) were replaced by simple tokens and embedded into Paragraph Vectors.