HUNTING PHENOTYPE-ASSOCIATED GENES USING KERNEL DATA FUSION BASED ON ORDERED WEIGHTING AVERAGING

- **Pooya Zakeri, KU Leuven, Belgium**
- Yves Moreau, KU Leuven, Belgium

Abstract: Combining biological data sources at the decision level is viewed as a natural way to manage heterogeneous data sources. We enumerate the advantage of modeling multi-heterogeneous biological data fusion in the gene prioritization task using statistic order and the ordered weighting averaging (OWA) operators.

As a result, we develop several kernel-based gene prioritization frameworks that combine multiple genomic data sources through the late integration. After transforming genomic data matrices into kernel matrices, for each kernel matrix, a one-class support vector machine (SVM) is exploited to perform gene prioritization for each phenotype under investigation. Then, the fusion methods defined by the use of OWA operator weights and ordered statistics weights are applied to aggregate performance rankings.

We use several genomic data sources including STRING, literature, gene expression and several annotation-based data sources, as well as sequence-based protein features such as pairwise local sequence alignment and information extracted directly from position-specific frequency matrices (PSFM).

Finally, we assess our models by applying them with the disease data sets on which Endeavour has been benchmarked. We also evaluate our models using an unbiased and prospective benchmark developed on the basis of the OMIM associations. Experimental results on both Endeavour benchmark and the unbiased benchmark demonstrate that our model can improve the accuracy of the state-of-the-art gene prioritization model. Moreover, we already used this model for the second Critical Assessment of Functional Annotation (CAFA2) to predict Human Phenotype Ontology (HPO). We succeeded in obtaining the promising results among the participant groups in that challenge.