Multi-View Learning in Biomedical Applications in the Big Data Era

Angela Serra, Paola Galdi and Roberto Tagliaferri
Introduction

- Multi-view learning is concerned with the problem of machine learning from data represented by multiple distinct feature sets.

- The recent emergence of this learning mechanism is largely motivated by the property of data from real applications where examples are described by different feature sets or different views.
  - Bioinformatics: microarray gene expression, RNASeq, PPI, gene ontology, etc.;
  - Neuroinformatics: Functional magnetic resonance imaging (fMRI), diffusion tensor imaging (DTI)


- Serra, Angela, Galdi, Paola, Tagliaferri, Roberto (2018). Multi-View Learning in Biomedical Applications. In Artificial Intelligence in the Age of Neural Networks and Brain Computing, Robert Kozma, Cesare Alippi, Yoonsuck Choe, Francesco Carlo Morabito Ed.s, in press.

Introduction

- How to put things together?

mRNA expression data

Protein-Protein Interaction

RNAseq

Medical Literature
Introduction

- Thanks to these multiple views, the learning task can be conducted with multi-view information.
In Bioinformatics multi-view approaches are useful since heterogeneous genome-wide data sources capture information on different aspects of complex biological systems.

Each source provides a distinct “view” of the same domain, but potentially encodes different biologically-relevant patterns.

Effective integration of such views can provide a richer model of an organism’s functional module than that produced by a single view alone.
Classification of Data Integration methodologies

Data Integration
- Type Of Analysis: Meta-analysis
  - Type of Data: Heterogeneous
    - Stage of Integration: Late
- Type Of Analysis: Integrative Analysis
  - Type of Data: Homogeneous or Heterogeneous
    - Stage Of Integration: Early/Intermediate/Late

Statistical Problem
- Supervised Learning
  - Embedding Methods
    - Feature Selection
    - Dimensionality Reduction
  - Subspace Learning
    - Clustering
    - Projective Methods
- Unsupervised Learning
  - Graph Integration
Meta-dimensional analysis can be divided into three categories.

a) Concatenation-based integration involves combining data sets from different data types at the raw or processed data level before modelling and analysis.

b) Transformation-based integration involves performing mapping or data transformation of the underlying data sets before analysis, and the modelling approach is applied at the level of transformed matrices.

c) Model-based integration is the process of performing analysis on each data type independently, followed by integration of the resultant models to generate knowledge about the trait of interest.

The analysis to be performed is somehow limited by the type of data involved in the experiment and by the desired level of integration. Analyses can be divided in two categories:

- **Meta-analysis** can be thought as an integrative study of previous results, usually performed aggregating the summary statistics from different studies. Due to its nature, meta-analysis can only be performed as a step of late integration involving Heterogeneous data.

- **Integrative analysis** considers the fusion of different data sources in order to get more stable and reliable estimates. Based on the type of data and the stage of integration, new methodologies have been developed spanning a landscape of techniques comprising graph theory, machine learning and statistics.
Data integration methodologies in systems biology can be divided into two categories based on the type of data: integration of homogeneous or heterogeneous data types.

- Usually biological data are thought to be homogeneous if they assay the same molecular level, for gene or protein expression, copy number variation, and so on.

- On the other hand if data is derived from two or more different molecular levels they are considered to be heterogeneous. Integration of this kind of data poses some issues: first, the data can have different structure, for example they can be sequences, graphs, continuous or discrete numerical values.
Integration Stage

Depending on the nature of the data and on the statistical problem to address, the integration of heterogeneous data can be performed at different levels:

- Early integration
- Intermediate Integration
- Late Integration
Early Integration

- Early integration consists in concatenating data from different views in a single feature space, without changing the general format and nature of data.
- Early integration is usually performed in order to create a bigger pool of features by multiple experiments.
- The main disadvantage of early integration methodologies is given by the need to search for a suitable distance function. In fact, by concatenating views, the data dimensionality considerably increases, consequently decreasing the performance of the classical similarity measures.
Intermediate integration consists in transforming all the data sources in a common feature space before combining them.

In classification problems, every view can be transformed in a similarity matrix that will be combined in order to obtain better results.
Late Integration

- In the late integration methodologies each view is analysed separately and the results are then combined.
- Late integration methodologies have some advantages over early integration techniques:
  - the user can choose the best algorithm to apply to each view based on the data;
  - the analysis on each view can be executed in parallel.
Supervised Learning

- In machine learning, supervised learning consists in inferring a function from labelled data.
- The input is a collection of samples defined as vectors on a set of features and a collection of labels, one for each sample.
<table>
<thead>
<tr>
<th>Data Type</th>
<th>Aim</th>
<th>Stage of Integration</th>
<th>Testing Data</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heterogeneous</td>
<td>Classification</td>
<td>Early - Intermediate - Late</td>
<td>Real Dataset from Stanford University</td>
<td>Gene functional classification from heterogeneous data. Pavlidis et al.</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>Classification</td>
<td>Early - Intermediate - Late</td>
<td>Genomic Cancer Datasets</td>
<td>Information content and analysis methods for Multi-Modal High-Throughput Biomedical Data. Bisakha et al.</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>Drugs classification and repositioning</td>
<td>Intermediate</td>
<td>CMAP Dataset</td>
<td>A multi layer drug repositioning approach. Napolitano et al.</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>Classification</td>
<td>Early</td>
<td>Webpage data and Advertisement data</td>
<td>Multi-view Fisher Discriminant Analysis (MFDA) which combines traditional FDA with multi-view learning. Chen et al.</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>Classification</td>
<td>Intermediate</td>
<td>PASCAL VOC (images)</td>
<td>Combines KCCA and SVM into a single optimisation termed SVM-2K. Larson et al.</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>Classification</td>
<td>Early - Late</td>
<td>CNN’s audio and video</td>
<td>AVIS: a connectionist-based framework for integrated auditory and visual information processing. Kasabov et al.</td>
</tr>
</tbody>
</table>
Brown et al. showed that SVM provides excellent classification performance on DNA microarray expression data.

Pavlidis et al. extend the methodology of Brown et al. to learn gene functional classifications from a heterogeneous data set consisting of microarray expression data and phylogenetic profiles.

SVMs are members of a larger class of algorithms, known as kernel methods, which can be non-linearly mapped to a higher-order feature space by replacing the dot product operation in the input space with a kernel function $K(\cdot, \cdot)$.

Gene functional classification from heterogeneous data

- The two types of data — gene expression and phylogenetic profiles — are combined in three different fashions, which we refer to as early, intermediate and late integration.

- In early integration, the two types of vectors are concatenated to form a single vector which serve as input for the SVM.

The two types of data — gene expression and phylogenetic profiles — are combined in three different fashions, which we refer to as early, intermediate and late integration.

In intermediate integration, the kernel values for each type of data are pre-computed separately, and the resulting values are added together. These summed kernel values are used in the training of the SVM.

Gene functional classification from heterogeneous data

- The two types of data — gene expression and phylogenetic profiles — are combined in three different fashions, which we refer to as early, intermediate and late integration.

- In late integration, one SVM is trained from each data type, and the resulting discriminant values are added together to produce a final discriminant for each gene.

Embedding Methods

- Dimensionality reduction of high dimensional multi-view data can be a non-trivial task because of the underlying connections between the features in the different views.
- A solution is to embed the multi-view patterns simultaneously into a low-dimensional space shared by all features.
Embedding Methods

- An example of embedding methods is Stochastic Neighbour Embedding (SNE) that constructs a low-dimensional manifold such that the density of low-dimensional data approximates the original density in the original high-dimensional space.

- Density is estimated as pairwise distances in the original feature space and the resulting embedding is obtained minimizing the Kullback-Leibler divergence among the high- and low-dimensional densities.

- Multi-view SNE is an extension of the original method that replaces the original estimated density with a combination of pairwise densities, each constructed from a different view. The corresponding objective includes $2$-norm regularization among the combination weights, plus a trade-off to balance the objective and the regularise.


Dimensionality Reduction: Feature Selection

- The goal of feature selection is to express high-dimensional data with a low number of features to reveal significant underlined information. It is mainly used as a pre-processing step for other computational methodologies.

- Three different approaches are proposed in literature:
  - The univariate filter methods
  - The multivariate wrapper
  - The multivariate embedded methods.

- They have the common goal of finding the smallest set of features useful to correctly classify objects. Accuracy and stability are the two main requirements for feature selection methodologies.
## Dimensionality Reduction: Feature Selection

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Aim</th>
<th>Stage of Integration</th>
<th>Testing Data</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heterogeneous</td>
<td>Feature Selection</td>
<td>-</td>
<td>Gene Expression</td>
<td>A Robust and Accurate Method for Feature Selection and Prioritization from Multi-Class OMICS Data. Fortino et al. [25]</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>Feature Selection</td>
<td>Late</td>
<td>Gene Expression Multiple Tissues</td>
<td>A sparse multi-view matrix factorization method for gene prioritization in gene expression datasets for multiple tissues. Larson et al. [31]</td>
</tr>
</tbody>
</table>
Dimensionality Reduction: Subspace Learning

- The aim of subspace learning approaches is to find a latent subspace shared by multiple views.
- One of the most cited approaches used to model the relationships between two (or more) views is Canonical Correlation Analysis (CCA).
- Consider two sets of variables $X$ and $Y$
- How to find the connection between the two sets of variables?
  - CCA: find a projection direction $w_x$ in the space of $X$ and $w_y$ in the space of $Y$, so that projected data onto $w_x$ and $w_y$ has max correlation.
  - Note: CCA simultaneously makes dimensional reduction for both the two feature spaces.
- It was defined for datasets with two views but it was later generalized to data with more than two representations in several ways (Kettenring, 1971 - Batch, 2002)
The problem with CCA is that most of the connections between objects in real datasets cannot be expressed with linear relations.

A solution is given by kernel methods that map data into a higher dimensional space and then apply linear methods in that space.

Kernel Canonical Correlation Analysis (KCCA) is the kernelized non linear version of CCA.
Dimensionality Reduction: Subspace Learning

- KCCA is widely used in genomics, in particular for the analysis of data from Genome-Wide Association Studies (GWAS).

- GWAS is used for the detection of genetic variants of complex diseases. So far, studies focused on the association of a Single Nucleotide Polymorphism (SNP) with a specific trait.

- Applying more sophisticated methods like KCCA, researchers can focus on more complex interactions between genes and specific traits of interest.
  - For example, Larson et al. developed a KCCA method able to identify associations between genes for complex phenotypes from a case-control study in genome-wide SNP data.
  - They applied the approach to find interaction between genes in an ovarian cancer dataset with 3869 cases and 3276 controls.
  - They were able to identify 13 gene pairs highly predictive of ovarian cancer risk.
Unsupervised Learning

- In machine learning, the unsupervised learning is defined as the problem of identifying hidden structures in unlabelled data.
- This means that the learner tries to group data by comparing the patterns based on their similarities.
- Here we focus in particular on multi-view clustering techniques.
Clustering is used when we want to extract information from data without any previous knowledge.

What does clustering mean?

Given a set of objects $X = \{x_1, \ldots, x_n\}$, clustering is a partition $P = \{P_1, \ldots, P_k\}$ of these objects such that

$$\bigcup_{i=1}^{k} P_i = X \quad \text{and} \quad P_i \cap P_j = \emptyset \quad \forall i \neq j$$

Each cluster contains similar objects and different objects are in different clusters.

The result depends on the (dis)similarity function.
Unsupervised Learning: Clustering differences between traditional and Multi View Clustering

- Traditional clustering methods take multiple views as a flat set of variables and ignore the differences among different views,
- Multiview clustering exploits the information from multiple views and take the differences among different views into consideration in order to produce a more accurate and robust data partitioning.
## Unsupervised Learning: Clustering

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Aim</th>
<th>Stage of Integration</th>
<th>Testing Data</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heterogeneous</td>
<td>Clustering</td>
<td>Early</td>
<td>Swissprot protein database and Image Dataset</td>
<td>Multi-view DBSCAN. Kailing et al</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>Clustering</td>
<td>Early</td>
<td>UCI Machine Learning Repository:</td>
<td>Multi-View weighted version of K-means. Chen et al.</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>Clustering</td>
<td>Late</td>
<td>Synthetic Dataset</td>
<td>A General Model for Multiple View Unsupervised Learning. Long et al.</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>Clustering</td>
<td>Late</td>
<td>Synthetic Dataset</td>
<td>Matrix Factorization. Greene, Derek.</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>Clustering</td>
<td>Late</td>
<td>Genomic Cancer datasets</td>
<td>A multi-view clustering integration methodology for cancer subtype. Serra et al.</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>Biclustering</td>
<td>Late</td>
<td>Ovaria Cancer</td>
<td>A non-negative matrix factorization method for detecting modules in heterogeneous omics multi-modal data Yang et al.</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>Clustering</td>
<td>Late</td>
<td>TCGA Dataset</td>
<td>Multi-omic integration approach that supports visual exploration of the data, and inspection of the contribution of the different genome-wide data-types. Taskesen et al.</td>
</tr>
</tbody>
</table>
Unsupervised Learning: Clustering
TW-Kmeans

- It is a two level variable weighting k-means clustering algorithm for multi-view data.
- The weights of views and individual variables are included into the distance function.
- It is an extension of the k-means algorithm with two more steps that should not require intensive computation so it should have the same computation complexity as k-means.

Unsupervised Learning: Clustering

**TW-Kmeans**

Let $X = \{X_1, X_2, \ldots, X_n\}$ be a set of $n$ objects represented by a set $A$ of $m$ variables.

---

Unsupervised Learning: Clustering TW-Kmeans

- Assume $A$ is divided into $T$ views $\{G_t\}_{t=1}^T$ where $G_t \cap G_s = \emptyset$ for $s \neq t$ and $\bigcup_{t=1}^T G_t = A$.

Unsupervised Learning: Clustering TW-Kmeans

- Let $W = \{w_1, w_2, \ldots, w_T\}$ be a set of $T$ weights, where $w_t$ indicates the weight that is assigned to the $t$th view and $\sum_{t=1}^{T} w_t = 1$.

Unsupervised Learning: Clustering TW-Kmeans

Let $V = \{ V_j \}$ be a set of $m$ variable weights, where $v_j$ indicates the weight that is assigned to the $j$th variable and

$$\sum_{j \in G_t} v_j = 1 \quad (1 \leq t \leq T), \quad \sum_{j=1}^{m} v_j = T$$

Unsupervised Learning: Clustering

TW-Kmeans

Assume that $X$ contains $k$ clusters. We want to identify:

- the set of $k$ clusters from $G$.
- the relevant views from the view weight matrix $W = [w_t]_T$
- the relevant variables from the variable weight matrix $V = [v_j]_m$

Late Integration

- Unification of patterns can also be seen as the next step of a data mining pipeline in which the preceding step is the clustering of objects on each single view. This distributed approach (as opposed to the centralized one) has some benefits as:
  - Clustering algorithms can be chosen with respect to the application domain.
  - Natural parallelization possibility.
  - Representation issues are avoided since clustering results are the inputs.
  - Suitable in privacy-preserving use cases.
Unsupervised Learning: Clustering

Notation and Formulation

- Given a set of views \( \{V_1, \ldots, V_v\} \) denoting \( n \) objects \( x_1, \ldots, x_n \), the goal is a consistent clustering between the views.

- The input is a set of clusterings \( C = \{C_1, \ldots, C_v\} \) where each \( C_h \) represents a clustering of the view \( V_h \). Clustering can be obtained by
  - Partitive clustering algorithms (k-means)
  - Probabilistic models (EM clustering)
  - Threshold based hierarchical clustering
  - Any other reasonable clustering method

Unsupervised Learning: Clustering

Notation and Formulation

- Each clustering is represented as a membership matrix.
- \( M_h \in \mathbb{R}^{n \times k_h} \) where \( k_h \) is the number of clusters on view \( V_h \). If an object is not present in \( V_h \), then the corresponding row is filled with zeros.

\[
\begin{align*}
C_1 &= \begin{bmatrix}
1 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
0 & 0 & 1 \\
\end{bmatrix} & \quad C_2 &= \begin{bmatrix}
0.5 & 0.5 \\
0.5 & 0.5 \\
0.7 & 0.3 \\
0.6 & 0.4 \\
0.2 & 0.8 \\
0.1 & 0.9 \\
\end{bmatrix}
\end{align*}
\]

Unsupervised Learning: Clustering

Matrix Factorization for Multi-View Clustering

- This algorithm combines information by factorizing the “matrix of clusters”.
- This factorization produces a projection of the original clusters into a new set of meta-clusters.
- These meta-clusters represent the additive combinations of clusters generated on one or more different views.

Matrix Factorization for Multi-View Clustering

- We start by transposing all the membership matrices and stacking them vertically obtaining the matrix of clusters $X \in \mathbb{R}^{l \times n}$ where $l$ is the total number of clusters in $C$. We want to find the best approximation of $X$ such that

$$X \approx PH \quad \text{and} \quad P \geq 0, \quad H \geq 0$$

Unsupervised Learning: Clustering

Matrix Factorization for Multi-View Clustering

- The rows of $P \in \mathbb{R}^{l \times k_f}$ project the clusters in a new set of $k_f$ meta-clusters.
- The columns of $H \in \mathbb{R}^{k_f \times n}$ can be viewed as the membership of the original objects in the new set of meta-clusters.

Unsupervised Learning: Projective Methods

- Projective methods are based on the concept of embedding the patterns into a new feature space learned by optimizing a criteria such as minimum reconstruction error from principal component analysis.
- Recently, this methodology has been applied in the context of multi-view data.
- For example Tyagi et al. proposed an intermediate integration approach for soft-hard clustering.

The method consists in mapping all the objects from the different views into a unit hypercube.

The projected views were concatenated and then clustered with k-means.

They tested the method on three different benchmark data sets: the first contains acoustic and seismic sensors for different types of vehicles, the second is the Handwritten Numeral dataset and the third is a multi-view image dataset.

The results were evaluated by using three performance measures: Clustering accuracy, Normalized Mutual Information (NMI) and Clustering purity.

They demonstrated that their methods have good performances and are not too sensitive to input parameters.

Multi-View Clustering on TCGA Dataset

- Taskesen et al. proposed a multi-omic integration approach (MEREDITH) that exploits the joint behaviour of the different molecular characteristics.
- It supports visual exploration of the data by a two-dimensional landscape.
- It is useful for inspecting the contribution of the different genome-wide data-types.
- Experiments were performed among 4,434 patients taken from The Cancer Genome Atlas (TCGA) across 19 cancer-types based on genome-wide measurements of four different molecular characteristics:
  - gene expression (GE; 18,882 features),
  - DNA-methylation (ME; 11,429 features),
  - copy-number variation (CN; 23,638 features)
  - microRNA expression (MIR; 467 features).

Multi-View Clustering on TCGA Dataset

Taskesen, Erdogan, et al. “Pan-cancer subtyping in a 2D-map shows substructures that are driven by specific combinations of molecular characteristics.” Scientific Reports 6 (2016).
Multi-View Clustering on TCGA Dataset

- Patient-sample projection in a two-dimensional map illustrating the cancer-landscape.
- The clustering is based on DBSCAN with the Davies-Bouldin index score for selecting the number of clusters.

Wang et al. proposed an **intermediate integration network fusion** methodology in order to integrate multiple genomic data and clustering patients.

Graph Integration: Similarity Network Fusion

- They first constructed a patients similarity network for each view.
- Then, they iteratively updated the network with the information coming from other networks in order to make them more similar at each step.
- At the end, this iterative process converged to a final fused network.

The authors tested the method to combine mRNA expression, microRNA expression and DNA methylation from five cancer data sets. They showed that the similarity networks of each view have different characteristics related to patients similarity while the fused network gives a more clear picture of the patients clusters.

They compared the proposed methodology with iClust and the clustering on concatenated views.

Results were evaluated with the silhouette score for clustering coherence, Cox log-rank test p-value for survival analysis for each subtype and the running time of the algorithms.

SNF outperformed single view data analysis and they were able to identify cancer subtypes validated by survival analysis.

MVDA: A Multi-view genomic data integration methodology

- We propose a multi-view approach in which the information from different data layers is integrated at the levels of the results of each single view clustering iterations by means of a matrix factorization approach.
- We performed experiment on six genomic datasets spanning on seven different views.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Response</th>
<th>N(0)</th>
<th>N(1)</th>
<th>N(2)</th>
<th>N(3)</th>
<th>Gene Expression</th>
<th>RNASeq</th>
<th>microRNA Expression</th>
<th>miRNAseq</th>
<th>Protein Expression</th>
<th>Copy Number</th>
<th>Clinical Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast Cancer patient from The Cancer genome Atlas (TCGA)</td>
<td>TCGA:BRC Pam50, (Her2, Basal, LumA, LumB)</td>
<td>24</td>
<td>13</td>
<td>55</td>
<td>59</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast Cancer patient from The Gene Expression Omnibus (GEO)</td>
<td>OXF:BRC1 Pam50, (Her2, Basal, LumA, LumB)</td>
<td>26</td>
<td>6</td>
<td>117</td>
<td>52</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OXF:BRC2 Clinical (Level1, Level2, Level3, Level4)</td>
<td>73</td>
<td>54</td>
<td>42</td>
<td>32</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prostate cancer patient from Memorial Sloan-Kettering Cancer Center (MSKCC), N=88</td>
<td>MSKCC:PRCA Tumor stages T1 vs. T2, T3, T4</td>
<td>53</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Ovarian cancer patient from The Cancer Genome Atlas (TCGA), N=93</td>
<td>TCGA:OVG Venus invasion present vs. absent</td>
<td>40</td>
<td>53</td>
<td>40</td>
<td>53</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glioblastoma Multiforme patient from The Cancer genome Atlas (TCGA), N = 167</td>
<td>TCGA:GBM (Classical, Mesenchymal, Neural, Proneural)</td>
<td>37</td>
<td>54</td>
<td>24</td>
<td>52</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**MVDA:** A Multi-view genomic data integration methodology

- **Goal:** input dimension reduction and relevant patterns discover.
- We tried different kinds of clustering algorithms using the Pearson coefficient as metric.
  - Pvclust
  - SOM
  - Hierarchical (Ward)
  - Pam
  - Kmeans

MVDA: A Multi-view genomic data integration methodology

- For each cluster a prototype element has been extracted

MVDA: A Multi-view genomic data integration methodology

- By selecting prototypes obtained at the previous step we find the most relevant features when working in the patients’ space.

- Feature selection is performed:
  - By computing the CAT t score.
    - The correlation-adjusted t-score (cat score) is a modification of the Student t-statistic to account for dependencies among variables.
    - Zuber and Strimmer have shown that the cat score improves ranking of genes to detect differential expression in the presence of correlation.
  - By computing the mean decrease accuracy index of the random forest classifier.

**MVDA**: A Multi-view genomic data integration methodology

- We select the top % relevant element for each view

**MVDA**: A Multi-view genomic data integration methodology

- The goal was to integrate the single view results in order to find patient clusters.
- We used a late integration approach.
- On each view we executed the same clustering algorithms of the first step to cluster patients.
- The algorithm used for multi-view data integration performed an iterative matrix factorization method.

Standard analysis aim at finding significant differences among groups defined \textit{a priori} based on clinical and expert knowledge.

We, instead, propose an approach in which we let the data group by themselves and then characterize \textit{a posteriori} significant differences emerged by this grouping with clinical information.
We consider each subject as an object represented in two different spaces, providing different kinds of information.

The features (or characteristics) of these spaces are the voxels of the rsfMRI and DTI data respectively.
Semi-supervised Subgroup discovery in ALS

- Dimensionality Reduction
- Single View Clustering
- Evaluation
- Multi View Integration
Dimensionality Reduction

- To overcome the issues deriving from HDLSS data we reduced the size of each dataset.
- Adjacent voxels are then aggregated with clustering. Each resulting area is then represented by a single value, derived by the clustered voxels.
- Voxel clustering can be seen as a data-driven parcelation.

How many clusters?

Fratello, Michele, et al. Submitted for publication
Semi-supervised Subgroup discovery in ALS

Clustered Voxels
Top: rsfMRI
Bottom: DTI

Fratello, Michele, et al. Submitted for publication
Single View clusterings are integrated together with side information on patient class labels, into 6 clusters.

With integration we can take into account simultaneously information from rsfMRI and DTI.
We looked for relations with clinical information.

We discovered that one of the clusters has an enriched group of subjects with lower limb onset and 2° clinical stage, with respect to the dataset.

The significance of the enriched group has been tested with a permutation test obtaining a p-value p=0.0033.
Conclusion

- And the future?
- Deep Learning for data fusion/integration

X-CNN: Cross-modal convolutional neural networks for sparse datasets

Petar Veličković¹,³, Duo Wang¹, Nicholas D. Lane²,³ and Pietro Liò¹

¹Computer Laboratory, University of Cambridge, UK
²Department of Computer Science, University College London, UK
³Nokia Bell Labs, UK

IEEE Symposium Series on Computational Intelligence 2016 (CIEL) 7 December 2016
Figure: X-CNN applied to imaging-based tumour detection. Three streams of X-CNN consume CT, MRI and PET images as input.
Thank You! Questions?

People who participated to this work:

ANGELA SERRA
PhD Student

MICHELE FRATELLO
PhD Student

VITTORIO FORTINO
Ph.D., Researcher

GIANCARLO RAICONI
Full Professor

FABRIZIO ESPOSITO
Associate Professor

DARIO GRECO
Ph.D., Assoc. Prof. Genetics

ROBERTO TAGLIAFERRI
Full Professor

Part of this project has been realized under the FP7 European project Nanosolutions (grant agreement FP7-309329), WP11


References


References


