Multi-agent Feature Selection for Integrative Multi-omics Analysis

Presenter: Sina Tabakhi

Supervisor: Haiping Lu

Department of Computer Science, University of Sheffield

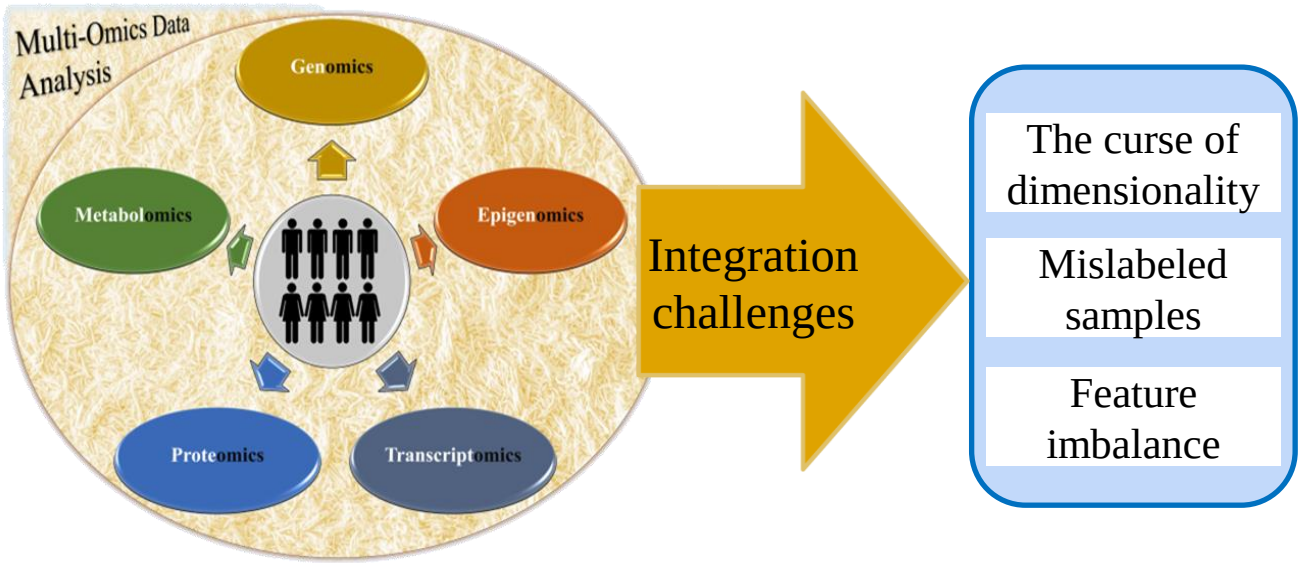
44th International Engineering in Medicine and Biology Conference (EMBC 2022)

The University Of Sheffield.

EMBC 2022 Glasgow

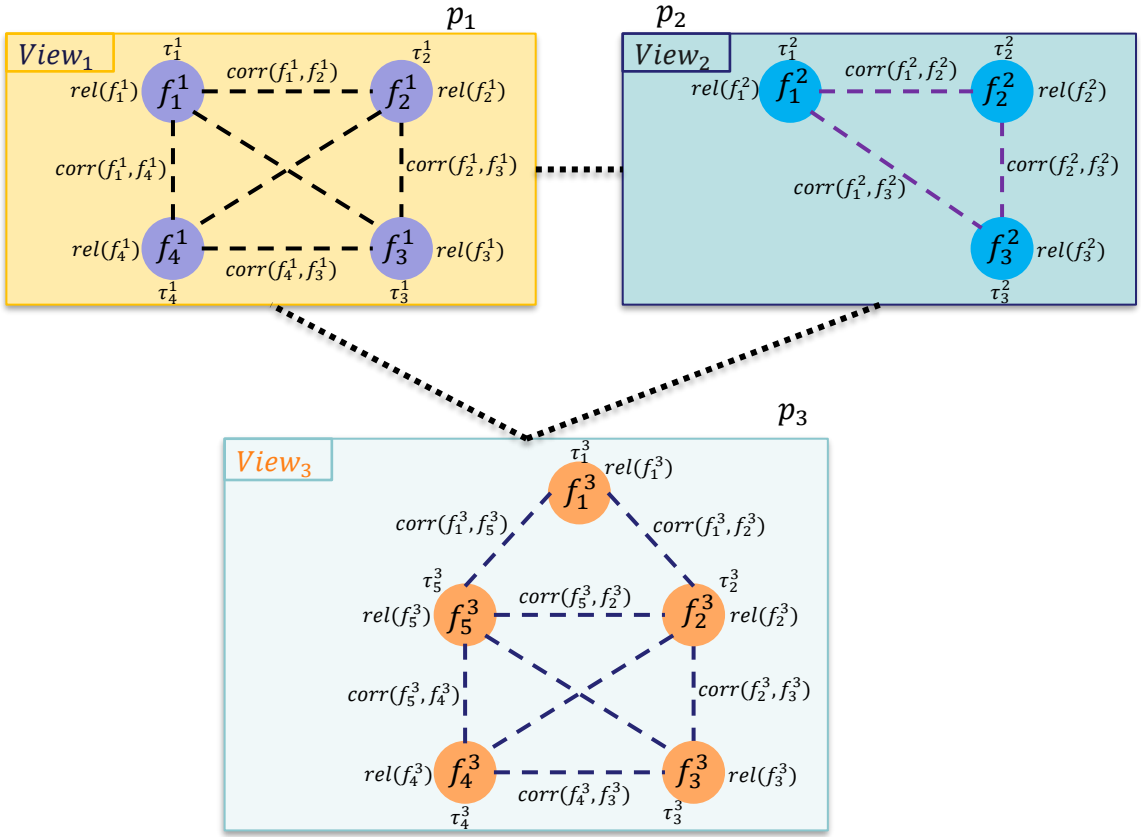
Motivation

- Multi-omics data
  - Deep understanding of complex molecular mechanisms
  - Diagnose, treat, and cure cancers through biological omics data



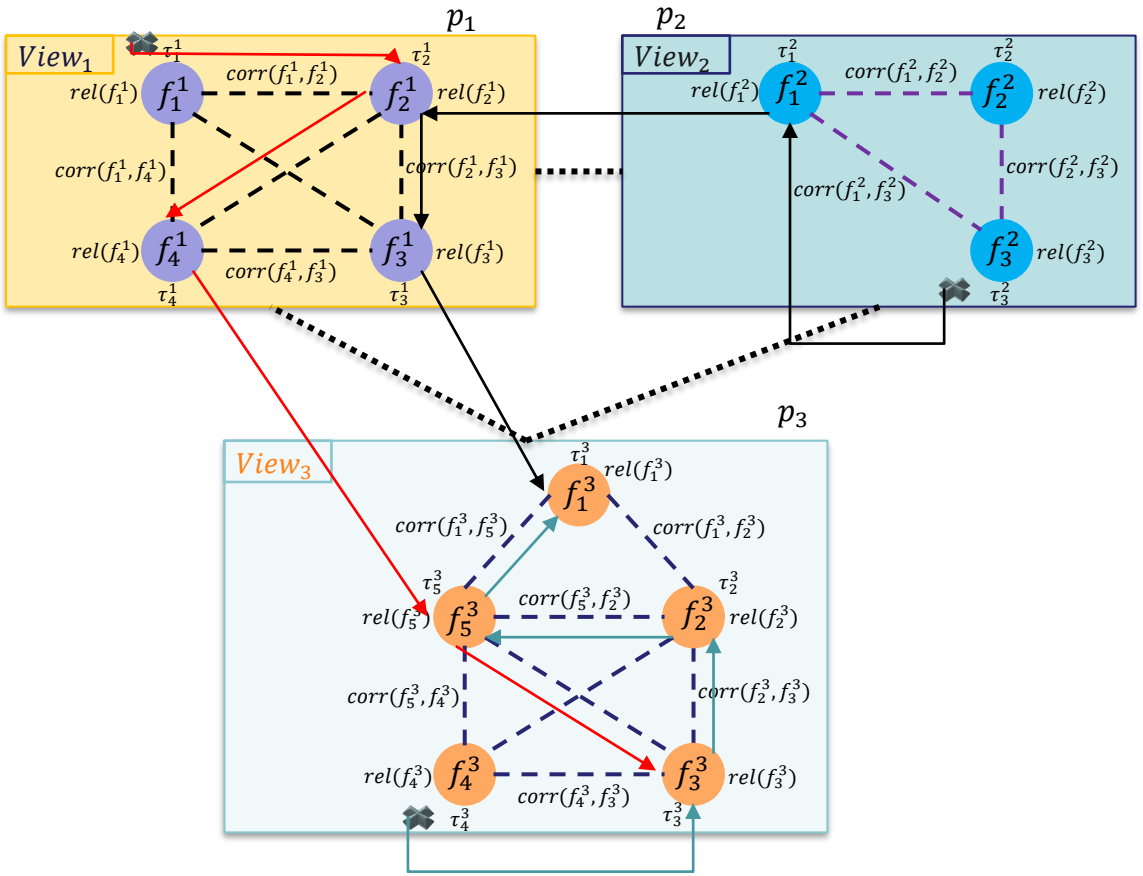
Representation for multi-agent feature selection

- A complete weighted graph
- corr(.,.): correlation function
- rel(.): relevance function
- τ_i^j: Pheromone level



Multi-agent feature selection algorithm

- Goal
  - Minimum redundancy
  - Maximum relevance



State transition rules in the framework

- The greedy rule

$$f_j^v = \arg\max_{j \in \Omega_{iv}^v} \left[\tau_j^v(t) \left[\eta_1 \left(f_j^v \right) \right]^\alpha \left[\eta_2 \left(f_i^v, f_j^v \right) \right]^\beta \right] \text{ if } q \leq q_0$$

Pheromone level

Relevance of feature

Inverse of correlation

- The probabilistic rule

$$v' = \text{choice}(V, \mathbb{P}) \quad \text{if} \quad q > q_0$$

$$P(f_j^{v'} | f_i^v) = \frac{\tau_j^{v'}(t) \left[\eta_1 \left(f_j^{v'} \right) \right]^\alpha \left[\eta_2 \left(f_i^v, f_j^{v'} \right) \right]^\beta}{\sum_{u \in \Omega_{iv}^{v'}} \tau_u^{v'}(t) \left[\eta_1 \left(f_u^{v'} \right) \right]^\alpha \left[\eta_2 \left(f_i^v, f_u^{v'} \right) \right]^\beta}$$

Fitness function to evaluate subsets

$$\text{fitness}(S_a) = \frac{\sum_{f_i^v \in S_a} \text{rel}(f_i^v)/|S_a|}{\sum_{f_i^v, f_j^{v'} \in S_a} \text{corr}(f_i^v, f_j^{v'}) / \sum_{i=1}^{|S_a|-1} i}$$

Maximal relevance

Minimal redundancy

Experiments

Dataset

- The ovarian cancer data from The Cancer Genome Atlas (TCGA)

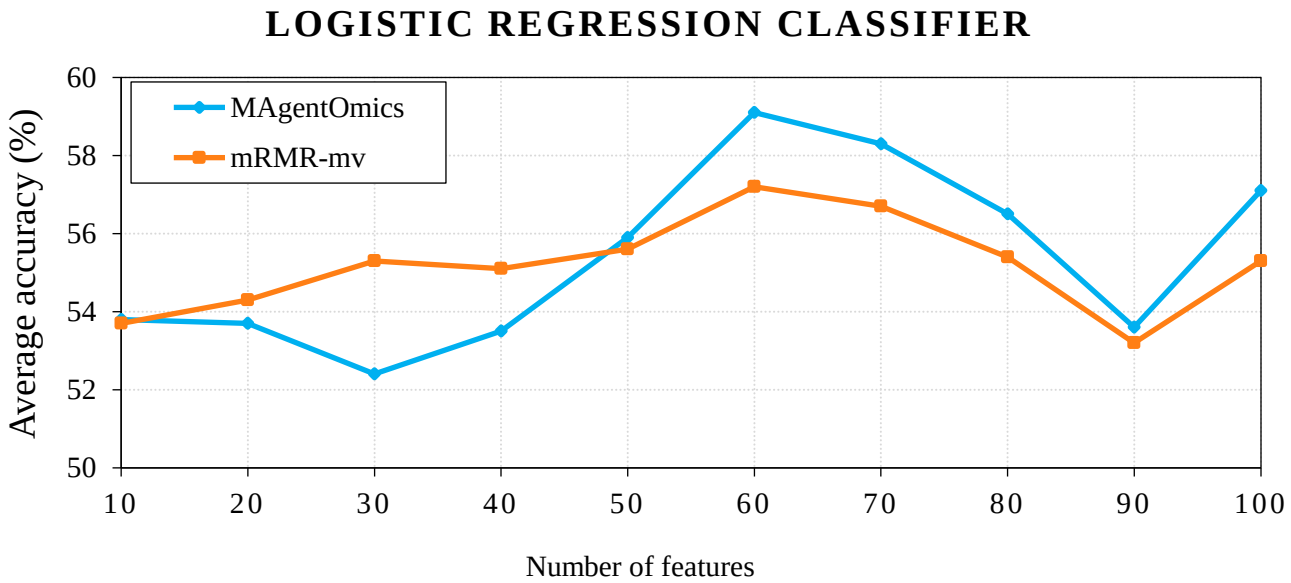
Omics Type	#Features	#Samples
DNA methylation	27,578	616
Gene-level copy number variation	24,776	579
Gene expression RNA-seq	20,530	308

Preprocessing

- Remove features with missing values
- Normalize values to the range of [0, 1]
- Remove features with variance lower than 0.05

Results

- Performance of the proposed method (MAgentOmics)



Conclusion

Summary

- Tackled the high-dimensionality challenge of integrative multi-omics
- Assessed the relative importance of each view
- Demonstrated the MAgentOmics outperforms the mRMR-mv

Future directions

- Improve efficiency by considering a sparse graph
- Design of new fitness functions
- Apply the proposed framework to other multi-omics data

Thank You!