A Framework for Feature Selection to Exploit Feature Group Structures

The 24th Pacific-Asia Conference on
Knowledge Discovery and Data Mining (PAKDD 2020)

Authors:

Kushani Perera **(University of Melbourne)**

Jeffrey Chan **(RMIT University)**

Shanika Karunasekera **(University of Melbourne)**

Outline

- Introduction
- Motivation
- Our Approach
- Experimental Results
- Conclusion & Future Work

Introduction

- Feature selection in high dimensional datasets:
  - Improves classification accuracy
  - Reduces computation costs
  - Produces simple learning models
- Problem: How to achieve high feature selection accuracy with low computational costs?
- External sources of correlations within feature groups
  - Pixels in images - Spatial locality
  - Words in text data - Semantics
  - Genes in genomic data - Gene Ontology terms

Motivation

- Feature selection methods
  - Filter methods
  - Wrapper methods
  - Embedded methods
- Filter feature selection methods
  - Computationally efficient
  - Classifier independent
  - Interpretable results
  - Simple and easy to implement
- Feature group information is rarely used to improve filter feature selection accuracy

Filter Methods for Feature Selection

Features (F_i) | Class (C)

Instances

	Cat	Computer	Tree	Dog	Document Type
Document 1	**1**	0	**1**	**1**	Biology
Document 2	**1**	0	**1**	**1**	Biology
Document 3	**1**	0	**1**	**1**	Biology
Document 4	0	**1**	**1**	0	Technical
Document 5	0	**1**	**1**	0	Technical
Document 6	0	**1**	**1**	0	Technical

Tree → Irrelevant ✘

Dog, Computer → Redundant ✘

Select Cat ✓

mRMR Algorithm

minimum Redundancy Maximum Relevancy (mRMR)

Objective:

$$\max \sum_{F_i \in S} Rel(F_i, C) - \frac{1}{|S|} \sum_{F_i, F_j \in S} Red(F_i, F_j)$$

Greedy method:

Selects a feature (F_i) at a time to

$$\max Rel(F_i, C) - \frac{1}{|S|} \sum_{F_j \in S} Red(F_i, F_j)$$

	Apple	Rice	Cow	Sheep	Document Type
Document 1	1	0	0	0	Botany
Document 2	1	1	0	0	
Document 3	1	1	0	0	
Document 4	0	1	0	0	
Document 5	0	0	1	0	Zoology
Document 6	0	0	1	0	
Document 7	0	0	0	1	
Document 8	0	0	0	1	
Document 9	0	0	0	0	Physics
Document 10	0	0	0	0	
Document 11	0	0	0	0	
Document 12	0	0	0	0	
Document 13	1	0	0	1	Agriculture
Document 14	1	0	0	1	
Document 15	1	1	1	0	
Document 16	0	1	1	0	

Features Selected by mRMR Algorithm

(1,1)

(1,0)

(0,1)

(0,0)

	Apple	Rice	Document Type
Document 1	**1**	0	**Botany**
Document 2	**1**	**1**	
Document 3	**1**	**1**	
Document 4	0	**1**	
Document 5	0	0	**Zoology**
Document 6	0	0	
Document 7	0	0	
Document 8	0	0	
Document 9	0	0	**Physics**
Document 10	0	0	
Document 11	0	0	
Document 12	0	0	
Document 13	**1**	0	**Agriculture**
Document 14	**1**	0	
Document 15	**1**	**1**	
Document 16	0	**1**	

Features Selected from Different Groups

(1,1)

(1,0)

(0,1)

(0,0)

	Apple	Sheep	Document Type
Document 1	**1**	0	**Botany**
Document 2	**1**	0	
Document 3	**1**	0	
Document 4	0	0	
Document 5	0	0	**Zoology**
Document 6	0	0	
Document 7	0	**1**	
Document 8	0	**1**	
Document 9	0	0	**Physics**
Document 10	0	0	
Document 11	0	0	
Document 12	0	0	
Document 13	**1**	**1**	**Agriculture**
Document 14	**1**	**1**	
Document 15	**1**	0	
Document 16	0	0	

	Apple	Rice	Cow	Sheep	Document Type
Document 1	1	0	0	0	Botany
Document 2	1	1	0	0	
Document 3	1	1	0	0	
Document 4	0	1	0	0	
Document 5	0	0	1	0	Zoology
Document 6	0	0	1	0	
Document 7	0	0	0	1	
Document 8	0	0	0	1	
Document 9	0	0	0	0	Physics
Document 10	0	0	0	0	
Document 11	0	0	0	0	
Document 12	0	0	0	0	
Document 13	1	0	0	1	Agriculture
Document 14	1	0	0	1	
Document 15	1	1	1	0	
Document 16	0	1	1	0	

Our Approach

- Propose a generic framework for filter feature selection methods to exploit feature group information from external sources of knowledge
- Incorporate feature group information in to mRMR objective
- *GroupMRMR:* A greedy feature selection algorithm
- Same computational complexity as mRMR

Modelling Feature Group Information

$$\min ||W||^2_{0,2} = \min \sum_{i=1}^{g} \frac{n_i^2}{\alpha_i}$$

features

W =

groups

	F_1	F_2	…	F_m
G_1	1	0	…	0
G_2	0	0	…	1
…	…	…	…	0
G_g	0	1	…	0

$n_i = |S \cap G_i| =$ No. of features selected from G_i

α_i = Weight of G_i , g = No. of groups

Modelling Feature Group Information

S = {Apple, Rice}

W =

	Apple	Rice	Cow	Sheep
Plant	1	1	0	0
Animal	0	0	0	0

$$||W||^2_{0,2} = (1+1+0+0)^2 + (0+0+0+0)^2 = 4$$

S = {Apple, Sheep}

W =

	Apple	Rice	Cow	Sheep
Plant	1	0	0	0
Animal	0	0	0	1

$$||W||^2_{0,2} = (1+0+0+0)^2 + (1+0+0+0)^2 = 2$$

GroupMRMR

Feature Selection Objective:

$$\max g(W) - \lambda \,.\, ||W||^2_{0,2}$$

Replace with mRMR objective:

$$g(W) = \sum_{F_i \in S} Rel(F_i, C) - \frac{1}{|S|} \sum_{F_i, F_j \in S} Red(F_i, F_j)$$

Greedy algorithm: selects one feature (F_i) at a time to

$$\max \left[Rel(F_i, C) - \frac{1}{|S|} \sum_{F_j \in S} Red(F_i, F_j) \right] - \lambda \,.\, \frac{2ni+1}{\alpha_i}$$

λ = User defined parameter, $n_i = |S \cap G_i|$

Experiment Datasets

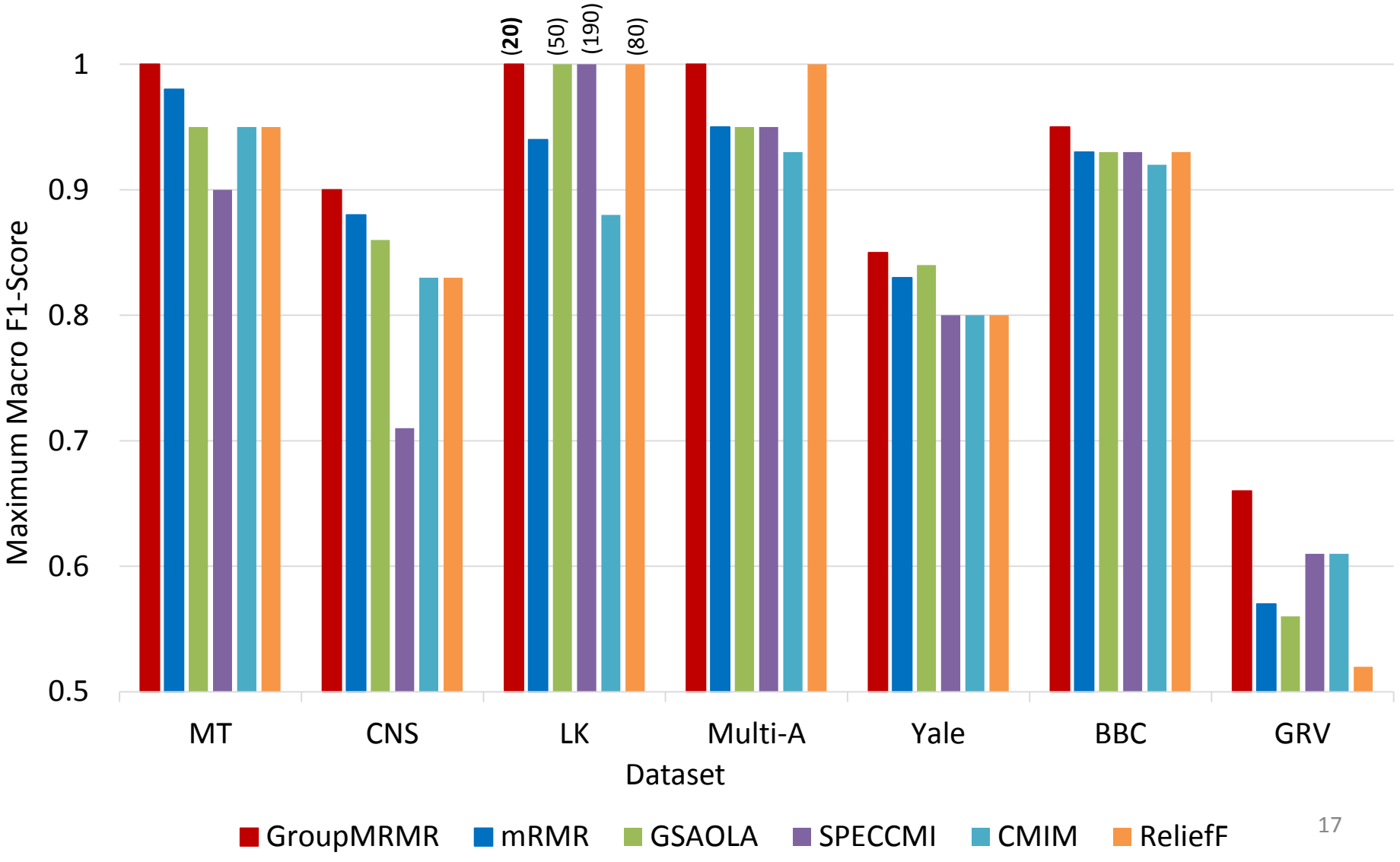
Dataset	Type	m	n	c	Feature Grouping
Multi-tissue (MT)	Genomic	1,000	103	4	Gene ontology
CNS	Genomic	989	42	5	Gene ontology
Leukemia (LK)	Genomic	999	38	3	Gene ontology
Multi-A	Genomic	5,565	103	4	Gene ontology
Yale	Image	1,024	165	15	Spatial locality (Group nearby pixels)
BBC	Text	9,635	2,225	5	Semantics (Word2Vec)
Groovy (GRV)	Software Defects	65	757	2	Code granularity

m: No. of features, **n**: No. of instances, **c**: No. of classes

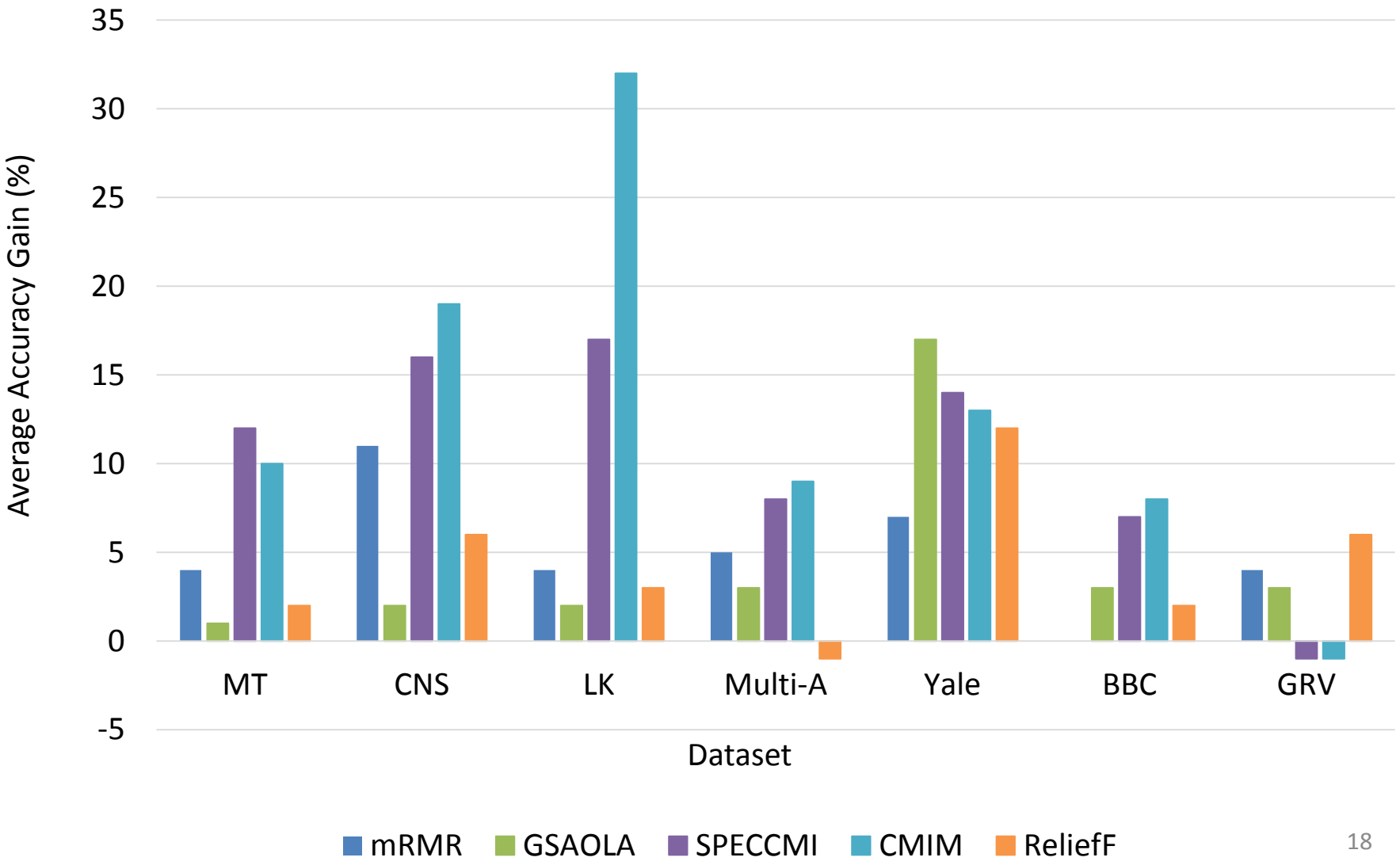
Baselines

- Minimum Redundancy Maximum Relevancy (mRMR)
- Group-SAOLA (GSAOLA)
- SPECCMI
- Conditional Mutual Information (CMIM)
- ReliefF

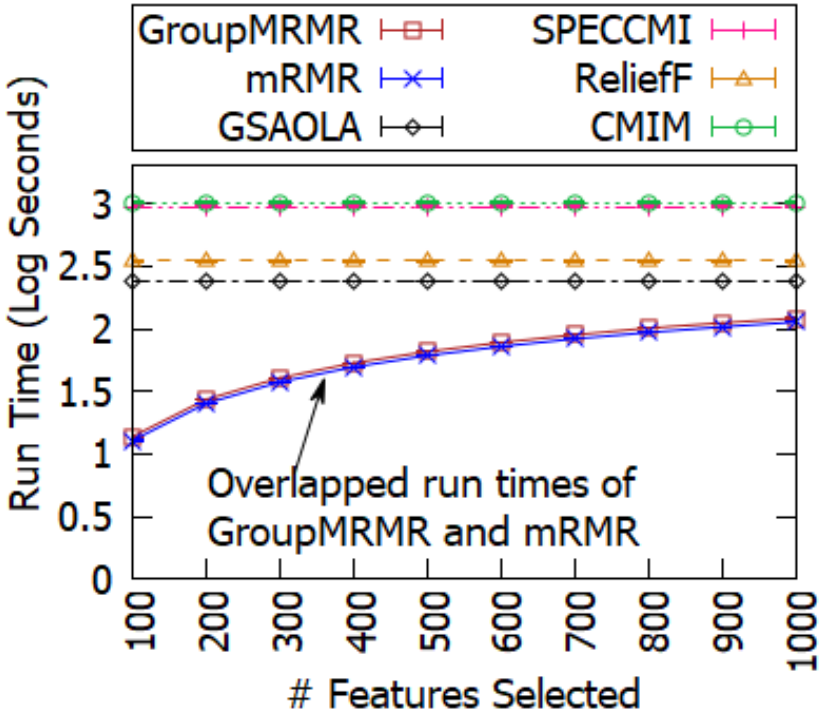
Maximum Classification Accuracy



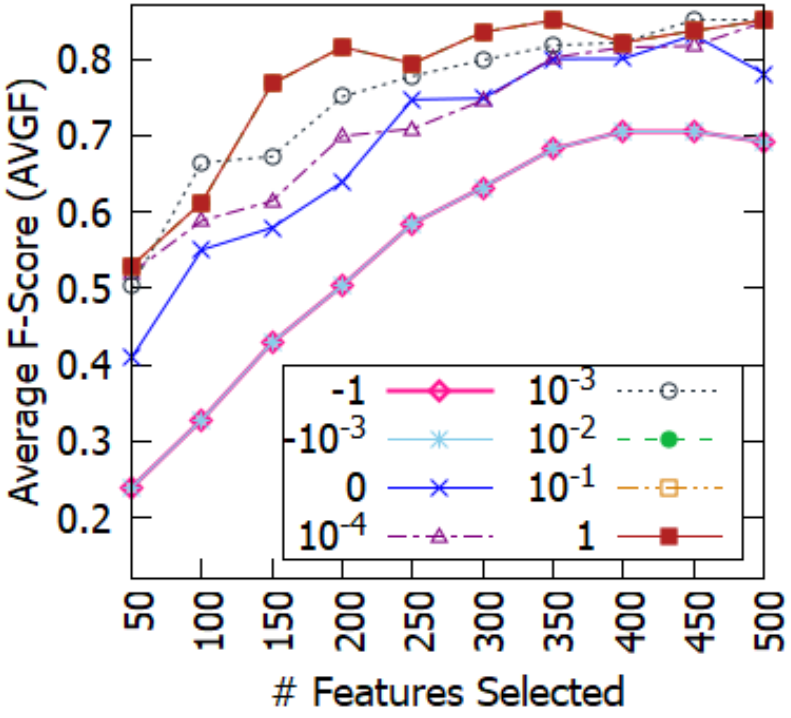
Average Accuracy Gain



Runtime and Parameter Sensitivity Results



Run time variation (BBC)



Accuracy variation with λ (Yale)

Conclusion and Future Work

Conclusion

- A framework which facilitates filter feature selection methods to exploit feature group information as an external source of information
- High feature selection accuracy with low computational costs

Future Work

- Experimenting the proposed framework for other filter methods

Thank you!