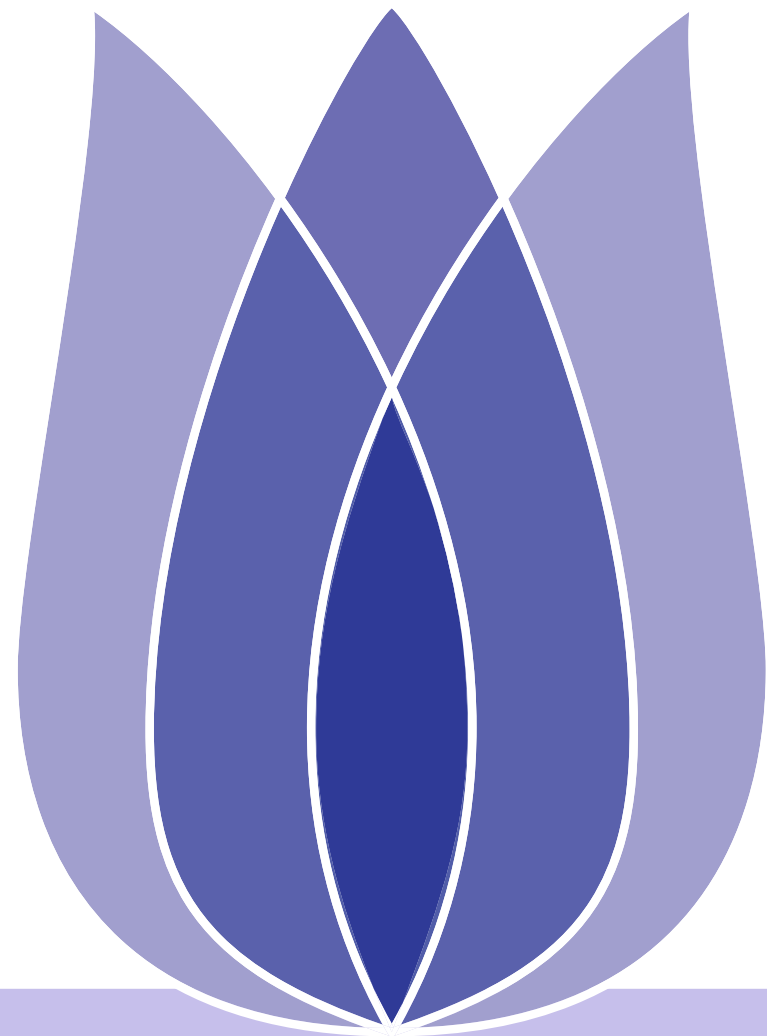


Group Outlying Aspects Mining

Dong Zhu

Deakin University
Chinese Academy of Sciences

(None)





Overview

- [Problem Definition](#)
- [Related Work and Challenges](#)
- [GOAM Algorithm](#)
- [Evaluation Results](#)
- [Conclusion](#)

Problem Definition

Outlying Aspects Mining

Group Outlying Aspects Mining

Related Work and Challenges

Related Work - Outlying Aspects Mining

Challenges (1)

GOAM Algorithm

Step One - Group Feature Extraction

Step Two - Outlying Degree Scoring

Step Three - Outlying Aspects Identification

Evaluation Results

Synthetic Dataset

NBA Dataset

Conclusion



Problem Definition

Outlying Aspects Mining
Group Outlying Aspects
Mining

Related Work and Challenges

GOAM Algorithm

Evaluation Results

Conclusion

Problem Definition



- Problem Definition
- Outlying Aspects Mining**
 - Group Outlying Aspects Mining
- Related Work and Challenges
- GOAM Algorithm
- Evaluation Results
- Conclusion

Defn

Outlying Aspects Mining aims to identify the outstanding features of the query object.

- A teacher may be interested in the **characteristics** that make **one student distinctive** from others.
- NBA coaches would prefer to find out the strengths and weaknesses of the player (a query object).

Player	3PT%	FTA	FT%	To



Outlying Aspects Mining vs Outlier Detection

Problem Definition

Outlying Aspects Mining
Group Outlying Aspects Mining

Related Work and Challenges

GOAM Algorithm

Evaluation Results

Conclusion

Player	3PT%	FTA	FT%	To

Outlying Aspects Mining

- Explain the distinctive **aspects** of the query object.
- The query object may (or may not) be an outlier.

Outlier Detection

- Find out **all** unusual **objects** in the whole dataset.
- **No** explanation on how they are different.



Group Outlying Aspects Mining

[Problem Definition](#)

[Outlying Aspects Mining](#)

[Group Outlying Aspects Mining](#)

[Related Work and Challenges](#)

[GOAM Algorithm](#)

[Evaluation Results](#)

[Conclusion](#)

Defn

Group outlying aspects mining aims to identify the outstanding features of the group of query object.

- Doctors desire to identify the merits & demerits between a group of cancer patients and normal people.
- NBA coaches are passionate about exploring the obvious advantages & disadvantages of the team.



Figure 1: Medical



Figure 2: NBA-Team



TULIP

Team for Universal Learning and Intelligent Processing



Problem Formalization

[Problem Definition](#)

[Outlying Aspects Mining](#)

[Group Outlying Aspects Mining](#)

[Related Work and Challenges](#)

[GOAM Algorithm](#)

[Evaluation Results](#)

[Conclusion](#)

Defn

Group outlying aspects mining aims to identify the top-k group outlying subspace in which the query group is distinctive with other groups.

- a set of groups.
- the query group.
- Other groups comparison groups.
- Each object in the group has features .



TULIP

Team for Universal Learning and Intelligent Processing

Term Definition

[Problem Definition](#)

[Outlying Aspects Mining](#)

[Group Outlying Aspects Mining](#)

[Related Work and Challenges](#)

[GOAM Algorithm](#)

[Evaluation Results](#)

[Conclusion](#)

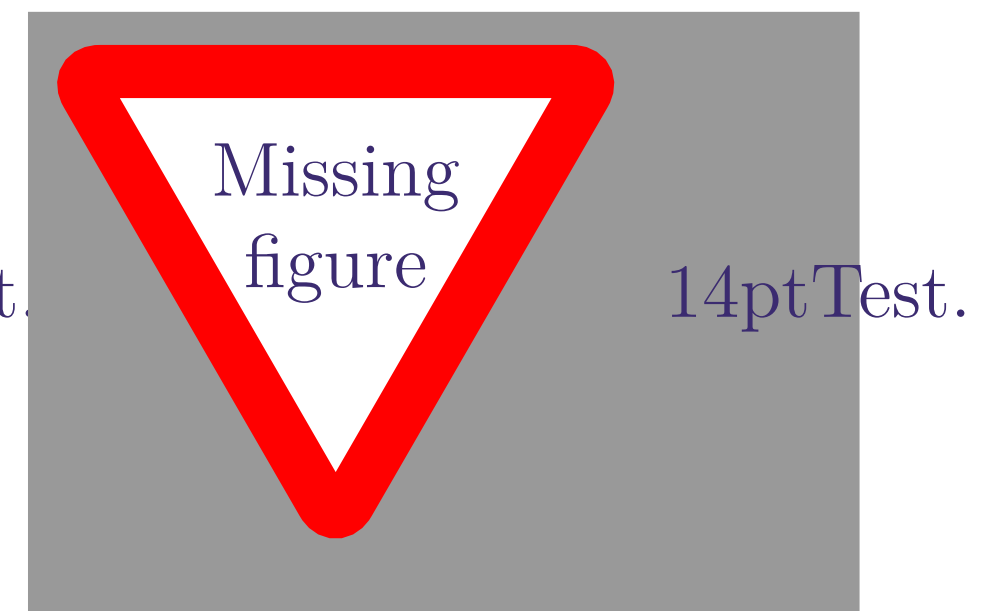
- Top-k group outlying subspaces
 - ◆ outlying scoring function.
 - ◆ quantifies the outlying degree of the query group in the subspace .
 - ◆ Order by DESC using scoring function to identify top K group outlying subspaces.



(a) Original Feature Spaces



(b) Group Outlying Spaces



(c) Another Subspaces



TULIP

Team for Universal Learning and Intelligent Processing



- Trivial Outlying Features
 - ◆ One-dimension subspaces.
 - ◆ f_i 's outlying degree $OD(f_i)$.

Table 1:

Feature	Outlying Degree
$\{f_i\}$	
$\{f_j\}$	
$\{f_k\}$	
$\{f_l\}$	
$\{f_m\}$	



- Non-Trivial Outlying Subspaces
 - ◆ Multi-dimension subspaces.
 - ◆ S 's outlying degree $OD(S)$.

Table 2:

Feature	Outlying Degree
$\{f_1\}$	
$\{f_2\}$	
$\{f_3\}$	
$\{f_4\}$	
$\{f_5\}$	



[Problem Definition](#)

[Related Work and Challenges](#)

[Related Work - Outlying](#)

[Aspects Mining](#)

[Challenges \(1\)](#)

[GOAM Algorithm](#)

[Evaluation Results](#)

[Conclusion](#)

Related Work and Challenges



Related Work - Outlying Aspects Mining

[Problem Definition](#)

[Related Work and Challenges](#)

[Related Work - Outlying Aspects Mining](#)

[Challenges \(1\)](#)

[GOAM Algorithm](#)

[Evaluation Results](#)

[Conclusion](#)

■ Existing Methods - Feature selection

- ◆ To distinguish two classes: the query point (positive) & rest of data (negative)

Disadvantages

- ◆ Positive and negative classes are **Not** balanced.
- ◆ **Not** quantify the outlying degree accurately.
- ◆ **Not** identify group outlying aspects.

Advantages

- ◆ Easy to operate.
- ◆ Resolve dimensionality bias.



TULIP

Team for Universal Learning and Intelligent Processing



Related Work - Outlying Aspects Mining

[Problem Definition](#)

[Related Work and Challenges](#)

[Related Work - Outlying Aspects Mining](#)

[Challenges \(1\)](#)

[GOAM Algorithm](#)

[Evaluation Results](#)

[Conclusion](#)

- Existing Methods - **Score-and-search**
 - ◆ Define an outlying score function.
 - ◆ Search subspaces.

Disadvantages

- ◆ Dimensionality bias.
- ◆ Search efficiency is **Not** high (dataset is large).
- ◆ **Not** identify group outlying aspects.

Advantages

- ◆ Quantify the outlying degree correctly.
- ◆ High Comprehensibility.



TULIP

Team for Universal Learning and Intelligent Processing



[Problem Definition](#)

[Related Work and Challenges](#)

[Related Work - Outlying Aspects Mining](#)

[Challenges \(1\)](#)

[GOAM Algorithm](#)

[Evaluation Results](#)

[Conclusion](#)

Group Outlying Aspects Mining

- Focus on differences between **groups**.
- **Multiple** points.

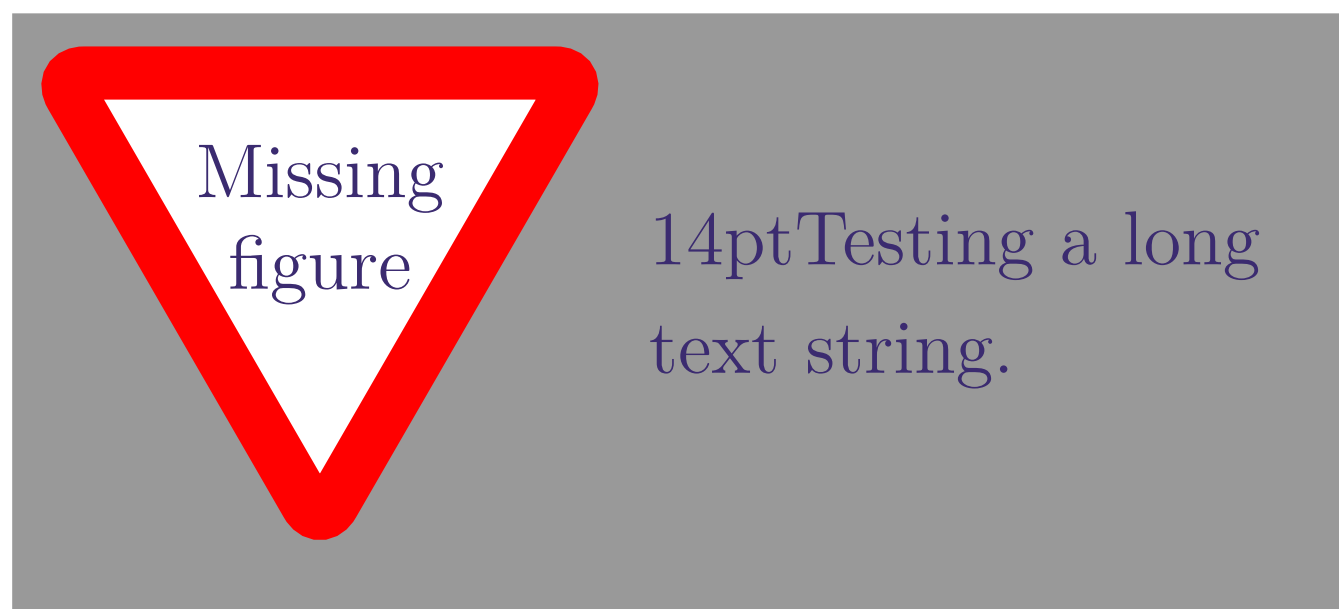


Figure 3: Group Outlying Aspects Target

Outlying Aspects Mining

- Concentrates on differences between **objects**.
- **One** point.

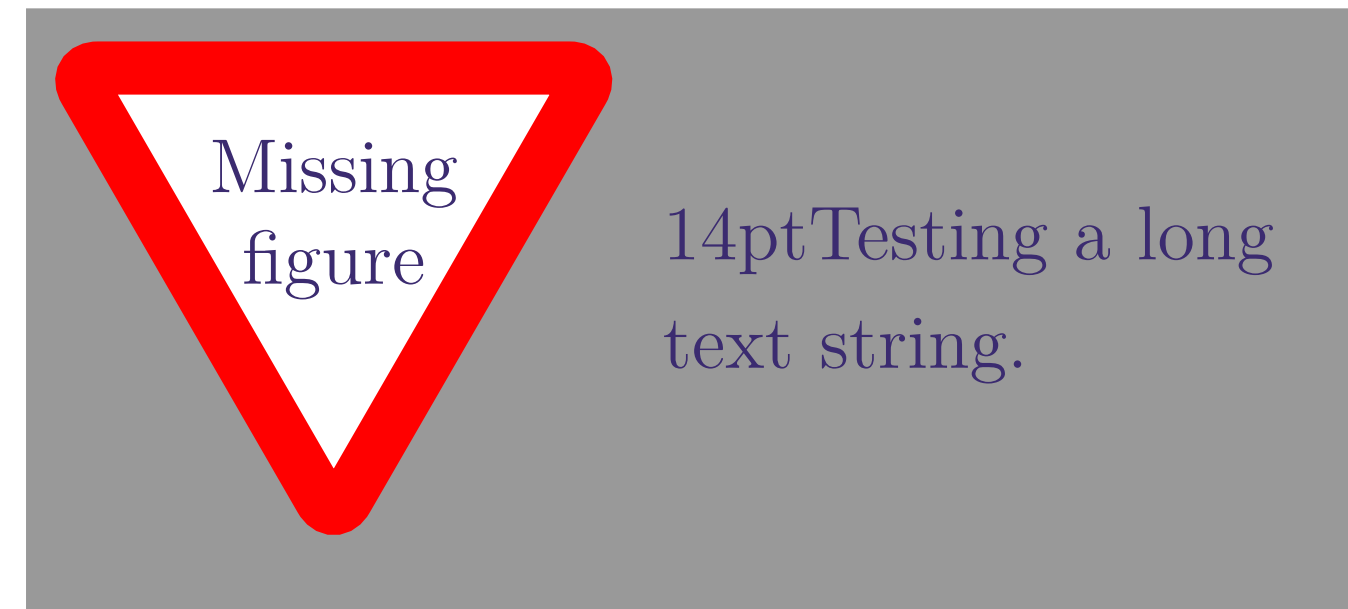


Figure 4: Outlying Aspects Target



TULIP

Team for Universal Learning and Intelligent Processing



Challenges (1)

[Problem Definition](#)

[Related Work and Challenges](#)

[Related Work - Outlying](#)

[Aspects Mining](#)

[Challenges \(1\)](#)

[GOAM Algorithm](#)

[Evaluation Results](#)

[Conclusion](#)

- How to **represent** the group features.
 - ◆ Can be affected by outlier values.
 - ◆ Can **Not** reflect the overall distribution of group features.



TULIP

Team for Universal Learning and Intelligent Processing



Challenges (2)

[Problem Definition](#)

[Related Work and Challenges](#)

[Related Work - Outlying](#)

[Aspects Mining](#)

[Challenges \(1\)](#)

[GOAM Algorithm](#)

[Evaluation Results](#)

[Conclusion](#)

- How to **evaluate** the outlying degree in different aspects.
 - ◆ Need design a scoring function when necessary.
 - ◆ Adopting an appropriate scoring function (without dimension bias) remains a problem.



TULIP

Team for Universal Learning and Intelligent Processing



Challenges (3)

[Problem Definition](#)

[Related Work and Challenges](#)

[Related Work - Outlying](#)

[Aspects Mining](#)

[Challenges \(1\)](#)

[GOAM Algorithm](#)

[Evaluation Results](#)

[Conclusion](#)

- How to **improve** the efficiency.
 - ◆ When the dimension of the **data is high**, the candidate subspace grows exponentially.
 - ◆ It will easily go beyond the limits of the computation resources.



TULIP

Team for Universal Learning and Intelligent Processing



[Problem Definition](#)

[Related Work and Challenges](#)

[GOAM Algorithm](#)

Step One - Group Feature

Extraction

Step Two - Outlying Degree

Scoring

Step Three - Outlying Aspects

Identification

[Evaluation Results](#)

[Conclusion](#)

GOAM Algorithm



- [Problem Definition](#)
- [Related Work and Challenges](#)
- [GOAM Algorithm](#)
 - [Step One - Group Feature Extraction](#)
 - [Step Two - Outlying Degree Scoring](#)
 - [Step Three - Outlying Aspects Identification](#)
- [Evaluation Results](#)
- [Conclusion](#)

Framework of GOAM algorithm:

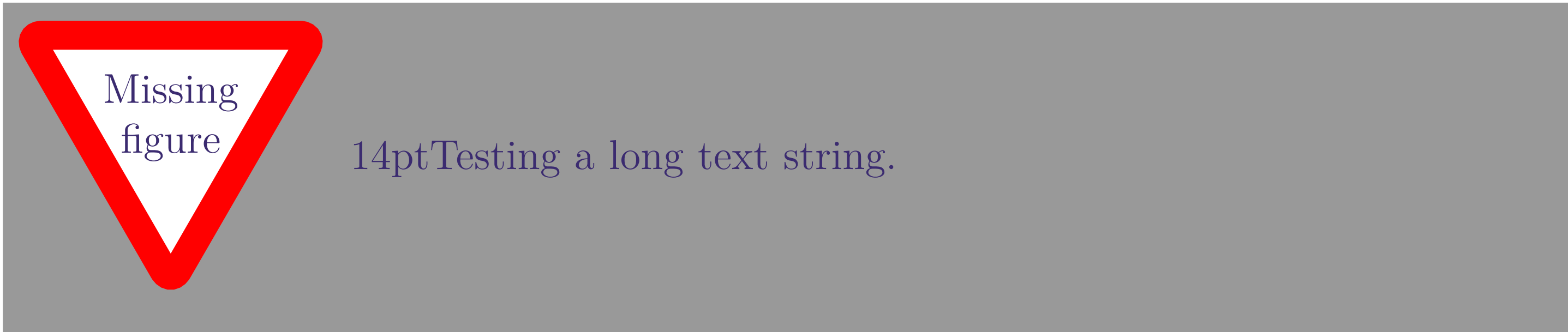


Figure 5: Framework of GOAM Algorithm



Step One - Group Feature Extraction

- Problem Definition
- Related Work and Challenges
- GOAM Algorithm
- Step One - Group Feature Extraction**
- Step Two - Outlying Degree Scoring
- Step Three - Outlying Aspects Identification
- Evaluation Results
- Conclusion

■ Suppose , , are three features of .

$$\begin{aligned} &: \{ \} \\ &: \{ \} \\ &: \{ \} \end{aligned}$$

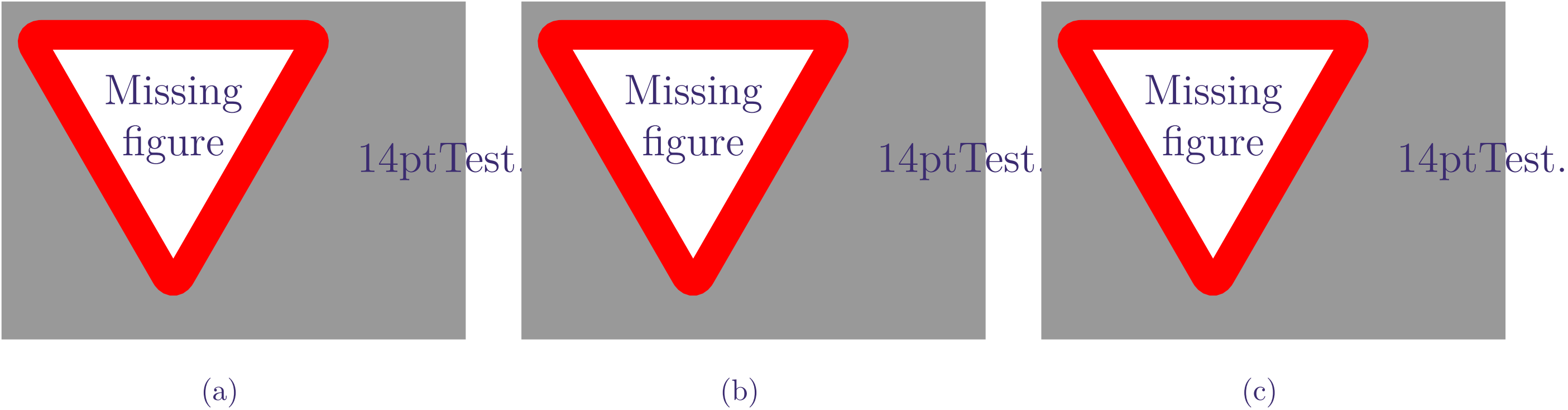


Figure 6: Histogram of on three features

Step Two - Outlying Degree Scoring

[Problem Definition](#)

[Related Work and Challenges](#)

[GOAM Algorithm](#)

[Step One - Group Feature](#)

[Extraction](#)

[Step Two - Outlying Degree
Scoring](#)

[Step Three - Outlying Aspects
Identification](#)

[Evaluation Results](#)

[Conclusion](#)

- Calculate Earth Mover Distance
 - ◆ Represent one feature among different groups
 - ◆ Purpose: calculate the minimum mean distance

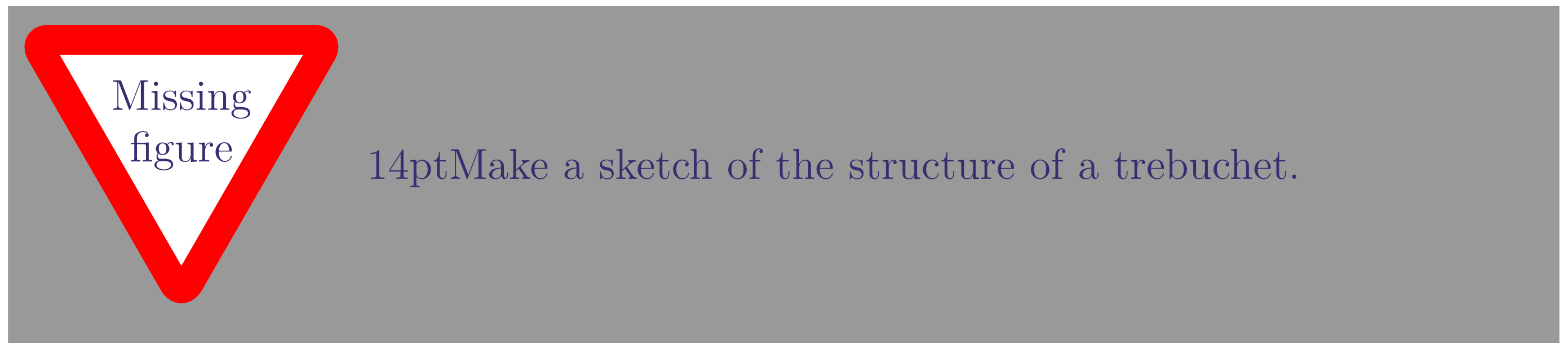


Figure 7: EMD of one feature



TULIP

Team for Universal Learning and Intelligent Processing



Step Two - Outlying Degree Scoring

- [Problem Definition](#)
- [Related Work and Challenges](#)
- [GOAM Algorithm](#)
- [Step One - Group Feature Extraction](#)
- [Step Two - Outlying Degree Scoring](#)**
- [Step Three - Outlying Aspects Identification](#)
- [Evaluation Results](#)
- [Conclusion](#)

- Calculate the outlying degree
 - ◆ n the number of contrast groups.
 - ◆ the histogram representation of in the subspace s .





Step Three - Outlying Aspects Identification

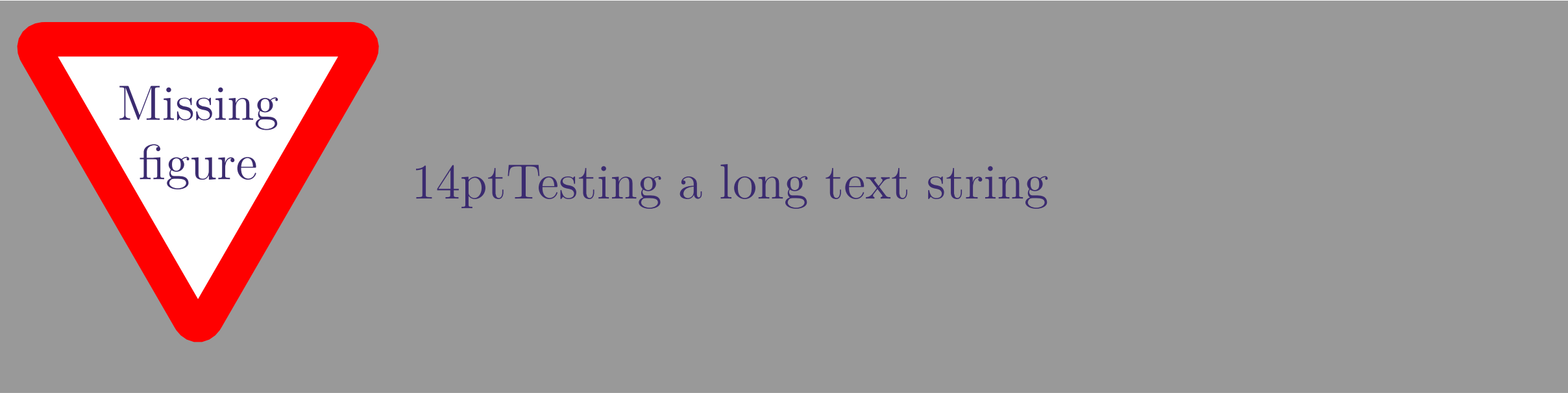
- [Problem Definition](#)
- [Related Work and Challenges](#)
- [GOAM Algorithm](#)
 - [Step One - Group Feature Extraction](#)
 - [Step Two - Outlying Degree Scoring](#)
 - [Step Three - Outlying Aspects Identification](#)
- [Evaluation Results](#)
- [Conclusion](#)

- Identify group outlying aspects mining based on the value of outlying degree.
- The greater the outlying degree is, the more likely it is group outlying aspect.



- [Problem Definition](#)
- [Related Work and Challenges](#)
- [GOAM Algorithm](#)
 - [Step One - Group Feature Extraction](#)
 - [Step Two - Outlying Degree Scoring](#)
 - [Step Three - Outlying Aspects Identification](#)
- [Evaluation Results](#)
- [Conclusion](#)

■ Pseudo code of GOAM algorithm





Illustration

Problem Definition

Related Work and Challenges

GOAM Algorithm

Step One - Group Feature

Extraction

Step Two - Outlying Degree

Scoring

Step Three - Outlying Aspects
Identification

Evaluation Results

Conclusion

Table 3: Original Dataset

10	8	9	8	7	7	6	6
9	9	7	9	8	9	9	8
8	10	8	8	6	7	8	9
8	8	6	7	7	7	7	8
9	9	9	8	8	6	6	7
8	10	8	8	9	8	8	8
9	9	7	9	7	7	7	9
10	9	10	7	8	6	6	8
9	10	8	6	9	8	8	7
9	9	7	9	8	7	9	8



Table 4: outlying degree of each possible subspaces

Feature	Outlying Degree	Feature	Outlying Degree
{ }	4.351	{ }	4.023
{ }	2.012	{ }	4.324
{ }	1.392	{ }	2.018
{ }	2.207	{ }	2.012

■ Search process:

 , save to .

 , save to .

 , save to .

 , save to .

 , save to .

 , save to .

 , remove.

 , remove.



Strengths of GOAM Algorithm

- [Problem Definition](#)
- [Related Work and Challenges](#)
- [GOAM Algorithm](#)
 - [Step One - Group Feature Extraction](#)
 - [Step Two - Outlying Degree Scoring](#)
 - [Step Three - Outlying Aspects Identification](#)
- [Evaluation Results](#)
- [Conclusion](#)

- Reduction of Complexity
 - ◆ Bottom-up search strategy.
 - ◆ Reduce the size of candidate subspaces.
- Efficiency
 - ◆ Before:
Now:



[Problem Definition](#)

[Related Work and Challenges](#)

[GOAM Algorithm](#)

[Evaluation Results](#)

[Synthetic Dataset](#)

[NBA Dataset](#)

[Conclusion](#)

Evaluation Results



Evaluation

- [Problem Definition](#)
- [Related Work and Challenges](#)
- [GOAM Algorithm](#)
- [Evaluation Results](#)
 - [Synthetic Dataset](#)
 - [NBA Dataset](#)
- [Conclusion](#)



P: Identified outlying aspects

T: Real outlying aspects



- [Problem Definition](#)
- [Related Work and Challenges](#)
- [GOAM Algorithm](#)
- [Evaluation Results](#)
- [Synthetic Dataset](#)**
- [NBA Dataset](#)
- [Conclusion](#)

■ Synthetic Dataset and Ground Truth

Table 5: Synthetic Dataset and Ground Truth

Query group	F_1	F_2		F_4				
	10	8	9	7	7	6	6	8
	9	9	7	8	9	9	8	9
	8	10	8	9	6	8	7	8
	8	8	6	7	8	8	6	7
	9	9	9	7	7	7	8	8
	8	10	8	8	6	6	8	7
	9	9	7	9	8	8	8	7
	10	9	10	7	7	7	7	7
	9	10	8	8	7	6	7	7
	9	9	7	7	7	8	8	8



Synthetic Dataset Results

[Problem Definition](#)

[Related Work and Challenges](#)

[GOAM Algorithm](#)

[Evaluation Results](#)

[Synthetic Dataset](#)

[NBA Dataset](#)

[Conclusion](#)

Table 6: The experiment result on synthetic dataset

Method	Truth Outlying Aspects	Identified Aspects	Accuracy
GOAM	,	,	100%
Arithmetic Mean based OAM	,	,	0%
Median based OAM	,	,	0%





NBA Dataset

[Problem Definition](#)

[Related Work and Challenges](#)

[GOAM Algorithm](#)

[Evaluation Results](#)

[Synthetic Dataset](#)

[NBA Dataset](#)

[Conclusion](#)

Data Collection

Source

Yahoo Sports website (<http://H.Bsports.H.Byahoo.H.Bcom.H.Bcn/H.BnbaH.B>)

Data

- Extract NBA teams' data until March 30, 2018;
- 6 divisions;
- 12 features (eg: *Point Scored*).



TULIP

Team for Universal Learning and Intelligent Processing



- [Problem Definition](#)
- [Related Work and Challenges](#)
- [GOAM Algorithm](#)
- [Evaluation Results](#)
- [Synthetic Dataset](#)
- [NBA Dataset](#)
- [Conclusion](#)

The detail features are as follows:

Table 7: Collected data of Brooklyn Nets Team

Pts	FGA	FG%	3FA	3PT%	FTA	FT%	Reb	Ass	To	Stl	Blk
18	12	42	2.00	50	7.00	100	0	4	3	0	0
15.7	14.07	41	5.45	32	3.05	75	3.98	5.1	2.98	0.69	0.36
14.5	11.1	47	0.82	26	4.87	78	6.82	2.4	1.74	0.92	0.66
13.5	10.8	42	5.37	37	3.38	77	6.66	2	1.38	0.83	0.42
12.7	10.59	39	5.36	33	3.37	82	3.24	6.6	1.56	0.89	0.31
12.6	10.93	40	6.94	37	1.70	84	4.27	1.5	1.06	0.61	0.44
12.2	10.39	44	3.42	35	2.70	72	3.79	4.1	2.15	1.12	0.32
10.6	7.85	49	4.51	41	1.35	83	3.34	1.6	1.15	0.45	0.24



- [Problem Definition](#)
- [Related Work and Challenges](#)
- [GOAM Algorithm](#)
- [Evaluation Results](#)
- [Synthetic Dataset](#)
- [NBA Dataset](#)
- [Conclusion](#)

■ Data Preprocess

Table 8: The bins that used to discrete data of each feature

Labels	Pts	FGA	FG%	3FA	3PT%	FTA
low	[0,5]	[0,4]	[0,0.35]	[0,1.0]	[0,0.2]	[0,1.0]
medium	(5,10]	(4,7]	(0.35,0.45]	(1.0,2.5]	(0.2,0.3]	(1.0,1.5]
high	(10,15]	(7,10]	(0.45,0.5]	(2.5,3.5]	(0.3,0.35]	(1.5,2.5]
very high	(15,]	(10,]	(0.5,1]	(3.5,]	(0.35,1]	(2.5,]
Labels	FT%	Reb	Ass	To	Stl	Blk
low	[0,0.6]	[0,2.0]	[0,1.0]	[0,0.6]	[0,0.2]	[0,0.25]
medium	(0.6,0.65]	(2,5]	(1,2]	(0.6,0.9]	(0.2,0.5]	(0.25,0.5]
high	(0.65,0.75]	(5,6]	(2,4]	(0.9,1.7]	(0.6,0.75]	(0.5,0.7]
very high	(0.75,1]	(6,]	(4,]	(1.7,]	(0.75,]	(0.7,]



Table 9: The identified outlying aspects of groups

Teams	Trivial Outlying Aspects	NonTrivial Outlying Aspects
Cleveland Cavaliers	{3FA}	{FGA, FT%}, {FGA, FG%}
Orlando Magic	{Stl}	None
Milwaukee Bucks	{To}, {FTA}	{FGA, FTA}, {3FA, FTA}
Golden State Warriors	{FG%}	{FT%, Blk}, {FGA, 3PT%, FTA}
Utah Jazz	{Blk}	{3FA, 3PT%}
New Orleans Pelicans	{FT%}, {FTA}	{FTA, Stl}, {FTA, To}



[Problem Definition](#)

[Related Work and Challenges](#)

[GOAM Algorithm](#)

[Evaluation Results](#)

[Conclusion](#)

Conclusion



Conclusion

[Problem Definition](#)

[Related Work and Challenges](#)

[GOAM Algorithm](#)

[Evaluation Results](#)

[Conclusion](#)

- Formalize the problem of *Group Outlying Aspects Mining* by extending outlying aspects mining;
- Propose a novel method **GOAM algorithm** to solve the *Group Outlying Aspects Mining* problem;
- Utilize the pruning strategies to reduce time complexity.



TULIP

Team for Universal Learning and Intelligent Processing



Questions?

- [Problem Definition](#)
- [Related Work and Challenges](#)
- [GOAM Algorithm](#)
- [Evaluation Results](#)
- [Conclusion](#)



Contact Information

Associate Professor Gang Li
School of Information Technology
Deakin University, Australia



GANGLI@TULIP.ORG.AU



TEAM FOR UNIVERSAL LEARNING AND INTELLIGENT PROCESSING

