FLIP00 Interim Inspection

Cong Ma

(None)

Overview

Problem Definition

Background

Defn	Kobe Bryant marked his retirement from the NBA by scoring 60 points in his final game as a Los Angeles Laker on Wednesday, April 12, 2016. Drafted into the NBA at the age of 17, Kobe earned the sport's highest accolades throughout his long career.

Related Work and Challenges

Related Work - data collection

- Existing Methods - Download from kaggle
- Existing Methods - Configure the running environment and load the required packages

Challenges (1)

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

GOAM Algorithm

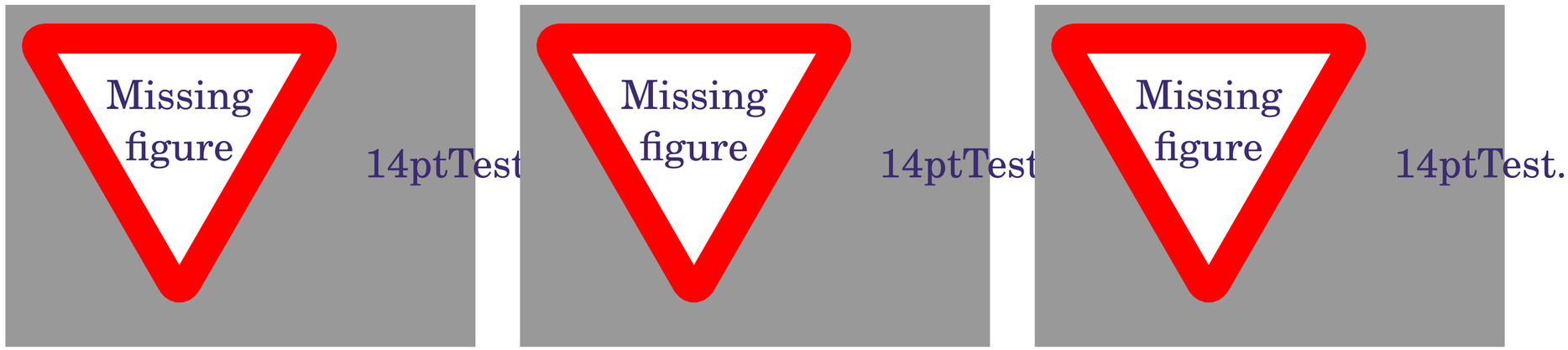
Step One - Group Feature Extraction

- Suppose f_1, f_2, f_3 are three features of G_q.

  f_1: $\{x_1, x_2, x_3, x_4, x_5, x_2, x_3, x_4, x_1, x_2\}$

  f_2: $\{y_2, y_2, y_1, y_2, y_3, y_3, y_5, y_4, y_4, y_2\}$

  f_3: $\{z_1, z_4, z_2, z_4, z_5, z_3, z_1, z_2, z_4, z_2\}$



(a) f_1 (b) f_2 (c) f_3

Figure 1: Histogram of G_q on three features

Step Two - Outlying Degree Scoring

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

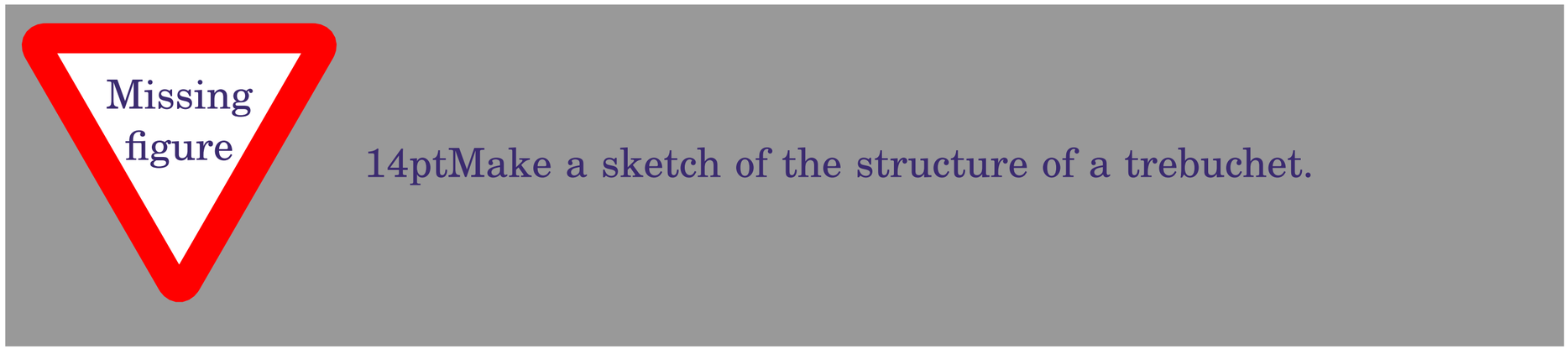


Figure 2: EMD of one feature

Step Two - Outlying Degree Scoring

- Calculate the outlying degree

$$OD(G_q) = \sum_1^n EDM(h_{q_s}, h_{k_s})$$

  - n ⇔ the number of contrast groups.
  - h_{k_s} ⇔ the histogram representation of G_k in the subspace s.

Step Three - Outlying Aspects Identification

- 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.

Illustration

Table 1: Original Dataset

G_1	F_1	F_2	F_3	F_4	G_2	F_1	F_2	F_3	F_4
	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

G_3	F_1	F_2	F_3	F_4	G_4	F_1	F_2	F_3	F_4
	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

Evaluation Results

Synthetic Dataset

- Synthetic Dataset and Ground Truth

Table 2: Synthetic Dataset and Ground Truth

Query group	$\mathbf{F_1}$	$\mathbf{F_2}$	F_3	$\mathbf{F_4}$	F_5	F_6	F_7	F_8
i_1	**10**	**8**	9	**7**	7	6	6	8
i_2	**9**	**9**	7	**8**	9	9	8	9
i_3	**8**	**10**	8	**9**	6	8	7	8
i_4	**8**	**8**	6	**7**	8	8	6	7
i_5	**9**	**9**	9	**7**	7	7	8	8
i_6	**8**	**10**	8	**8**	6	6	8	7
i_7	**9**	**9**	7	**9**	8	8	8	7
i_8	**10**	**9**	10	**7**	7	7	7	7
i_9	**9**	**10**	8	**8**	7	6	7	7
i_{10}	**9**	**9**	7	**7**	7	8	8	8

Synthetic Dataset Results

Table 3: The experiment result on synthetic dataset

Method	Truth Outlying Aspects	Identified Aspects	Accuracy
GOAM	$\{F_1\}$, $\{F_2F_4\}$	$\{F_1\}$, $\{F_2F_4\}$	100%
Arithmetic Mean based OAM	$\{F_1\}$, $\{F_2F_4\}$	$\{F_4\}$, $\{F_2\}$	0%
Median based OAM	$\{F_1\}$, $\{F_2F_4\}$	$\{F_2\}$, $\{F_4\}$	0%

Conclusion

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.

Thanks For Your Listening

Made by MC

QingDao technological university