Scalable preference disaggregation: A multiple criteria sorting approach based on the MapReduce framework

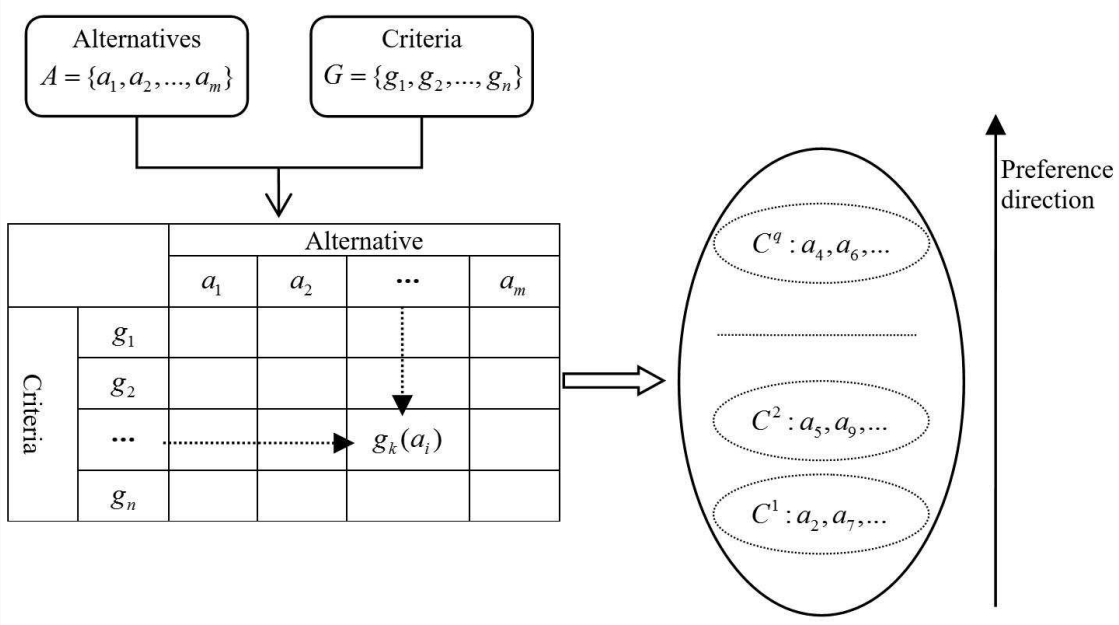
Jiapeng Liu

November 22, 2018

School of Management
Xi'an Jiaotong University
P.R. China

Introduction

- Multiple criteria sorting (MCS) is the practice of assigning a set of alternatives evaluated on multiple criteria to predefined and preference-ordered categories.



Introduction

- In this study, we assume that the preference information is composed of a set of assignment examples on reference alternatives.

Introduction

- In this study, we assume that the preference information is composed of a set of assignment examples on reference alternatives.
- Many approaches have been proposed to deal with MCS problems based on this indirect preference information.
  - methods motivated by value functions, such as the UTADIS method and its variants, and the MHDIS method
  - methods based on outranking relations, such as the ELECTRI Tri-B methods, ELECTRI Tri-C methods, ELECTRI-based methods, and PROMETHEE-based methods
  - rule induction-oriented procedures, such as the DRSA method and its extensions
  - techniques incorporating the weighted Euclidean distance

Introduction

- In the era of big data, developments in information technology have resulted in an explosive growth in data gathered from various fields.

Introduction

- In the era of big data, developments in information technology have resulted in an explosive growth in data gathered from various fields.
- This needs decision analysis methods to scale up well with the requirements of big data.

Introduction

- In the era of big data, developments in information technology have resulted in an explosive growth in data gathered from various fields.
- This needs decision analysis methods to scale up well with the requirements of big data.
- However, it is challenging for existing MCS methods to deal with problems that contain a large set of alternatives and massive preference information.

Introduction

- In the era of big data, developments in information technology have resulted in an explosive growth in data gathered from various fields.
- This needs decision analysis methods to scale up well with the requirements of big data.
- However, it is challenging for existing MCS methods to deal with problems that contain a large set of alternatives and massive preference information.
    - Traditional decision problems usually involve several dozens of alternatives.

Introduction

- In the era of big data, developments in information technology have resulted in an explosive growth in data gathered from various fields.
- This needs decision analysis methods to scale up well with the requirements of big data.
- However, it is challenging for existing MCS methods to deal with problems that contain a large set of alternatives and massive preference information.
  - Traditional decision problems usually involve several dozens of alternatives.
  - These methods require the data to fit into the main memory, in which LP/IP solvers search for the optimal solution.

Introduction

- In the era of big data, developments in information technology have resulted in an explosive growth in data gathered from various fields.
- This needs decision analysis methods to scale up well with the requirements of big data.
- However, it is challenging for existing MCS methods to deal with problems that contain a large set of alternatives and massive preference information.
  - Traditional decision problems usually involve several dozens of alternatives.
  - These methods require the data to fit into the main memory, in which LP/IP solvers search for the optimal solution.
  - This exceeds the processing capabilities of existing MCS methods in terms of the memory consumption and/or the computational time when dealing with huge amounts of data.

Introduction

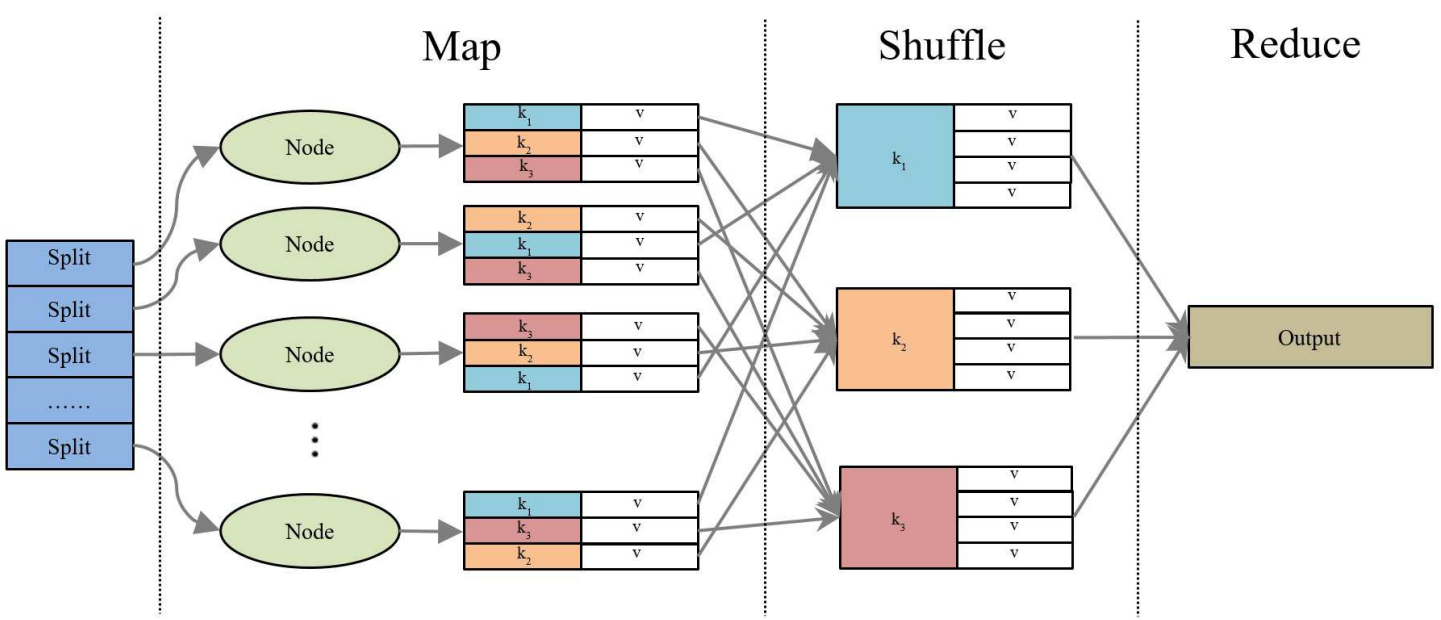
- Thus, new techniques must be introduced to redesign existing MCS methods so that they can scale up well with new storage and time requirements.

Introduction

- Thus, new techniques must be introduced to redesign existing MCS methods so that they can scale up well with new storage and time requirements.
- In this study, we propose a new approach based on the MapReduce framework, in order to address the MCS problem with a large set of alternatives and massive preference information.

Introduction

- Thus, new techniques must be introduced to redesign existing MCS methods so that they can scale up well with new storage and time requirements.
- In this study, we propose a new approach based on the MapReduce framework, in order to address the MCS problem with a large set of alternatives and massive preference information.
- MapReduce is a popular parallel computing paradigm developed by Google Inc., which is designed to process large-scale data sets.



The proposed approach

- The aim of this study is to classify a finite set of m alternatives $A = \{a_1, a_2, ..., a_m\}$ into p categories $\overline{C} = \{C_1, C_2, ..., C_p\}$, such that C_{h+1} is preferred to C_h (denoted by $C_{h+1} \succ C_h$), $h = 1, ..., p-1$.

The proposed approach

- The aim of this study is to classify a finite set of m alternatives $A = \{a_1, a_2, ..., a_m\}$ into p categories $\overline{C} = \{C_1, C_2, ..., C_p\}$, such that C_{h+1} is preferred to C_h (denoted by $C_{h+1} \succ C_h$), $h = 1, ..., p-1$.
- Such a classification decision is built on some preference information provided by the DM, which involves a set of assignment examples concerning a finite set of reference alternatives $A^R = \{a^*, b^*, ...\}$.

The proposed approach

- The aim of this study is to classify a finite set of m alternatives $A = \{a_1, a_2, ..., a_m\}$ into p categories $\overline{C} = \{C_1, C_2, ..., C_p\}$, such that C_{h+1} is preferred to C_h (denoted by $C_{h+1} \succ C_h$), $h = 1, ..., p-1$.
- Such a classification decision is built on some preference information provided by the DM, which involves a set of assignment examples concerning a finite set of reference alternatives $A^R = \{a^*, b^*, ...\}$.
- An assignment example specifies the assignment of a reference alternative $a^* \in A^R$ to a category $C(a^*) \in \overline{C}$.

The proposed approach

- The aim of this study is to classify a finite set of m alternatives $A = \{a_1, a_2, ..., a_m\}$ into p categories $\overline{C} = \{C_1, C_2, ..., C_p\}$, such that C_{h+1} is preferred to C_h (denoted by $C_{h+1} \succ C_h$), $h = 1, ..., p-1$.
- Such a classification decision is built on some preference information provided by the DM, which involves a set of assignment examples concerning a finite set of reference alternatives $A^R = \{a^*, b^*, ...\}$.
- An assignment example specifies the assignment of a reference alternative $a^* \in A^R$ to a category $C(a^*) \in \overline{C}$.
- All the alternatives $a \in A \cup A^R$ are evaluated in terms of n criteria $g_1, g_2, ..., g_n$.

The proposed approach

- The aim of this study is to classify a finite set of m alternatives $A = \{a_1, a_2, ..., a_m\}$ into p categories $\overline{C} = \{C_1, C_2, ..., C_p\}$, such that C_{h+1} is preferred to C_h (denoted by $C_{h+1} \succ C_h$), $h = 1, ..., p-1$.
- Such a classification decision is built on some preference information provided by the DM, which involves a set of assignment examples concerning a finite set of reference alternatives $A^R = \{a^*, b^*, ...\}$.
- An assignment example specifies the assignment of a reference alternative $a^* \in A^R$ to a category $C(a^*) \in \overline{C}$.
- All the alternatives $a \in A \cup A^R$ are evaluated in terms of n criteria $g_1, g_2, ..., g_n$.
- The performance of $a \in A \cup A^R$ on g_j, $j \in G = \{1, ..., n\}$, is denoted by $g_j(a)$.

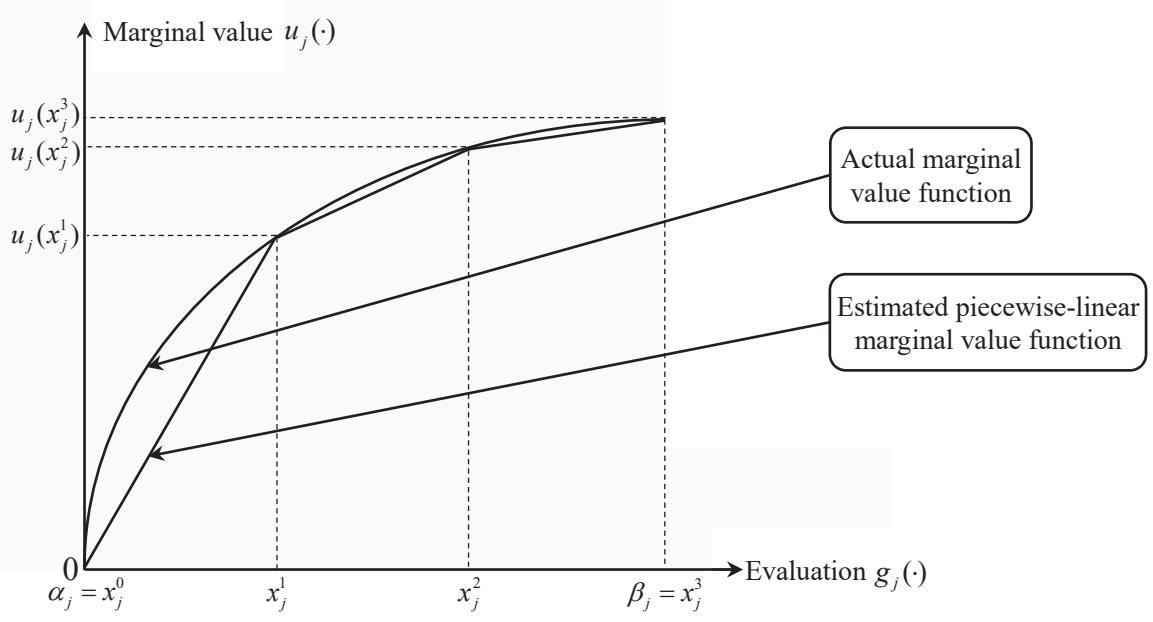
The proposed approach

To perform the assignment of alternative $a \in A \cup A^R$, we shall use as the preference model an additive value function U of the following form:

$$U(a) = \sum_{j=1}^{n} u_j(g_j(a)), \quad a \in A \cup A^R,$$

where $U(a)$ is the comprehensive value of a, and $u_j(g_j(a))$, $j = 1, ..., n$, are marginal value functions for each criterion.

The proposed approach



- In this study, a piecewise-linear function $u_j(\cdot)$ is used to estimate the actual value function of criterion g_j, $j = 1, ..., n$.

The proposed approach

- Defining the characteristic vector $\mathbf{V}(a) \in \mathbb{R}^{\gamma}$ of alternative a by

$$\mathbf{V}(a) = \left(\underbrace{v_1^1(a), ..., v_1^{\gamma_1}(a)}_{\text{criterion } g_1}, ..., \underbrace{v_j^1(a), ..., v_j^{\gamma_j}(a)}_{\text{criterion } g_j}, ..., \underbrace{v_n^1(a), ..., v_n^{\gamma_n}(a)}_{\text{criterion } g_n} \right)^{\mathsf{T}}$$

and denote

$$\mathbf{u} = \left(\underbrace{\Delta u_1^1, ..., \Delta u_1^{\gamma_1}}_{\text{criterion } g_1}, ..., \underbrace{\Delta u_j^1, ..., \Delta u_j^{\gamma_j}}_{\text{criterion } g_j}, ..., \underbrace{\Delta u_n^1, ..., \Delta u_n^{\gamma_n}}_{\text{criterion } g_n} \right)^{\mathsf{T}},$$

we can compute comprehensive value $U(a)$ as follows:

$$U(a) = \mathbf{u}^{\mathsf{T}} \mathbf{V}(a).$$

The proposed approach

The consistency principle for assignment: For any pair of alternatives a and b, a given value function $U(\cdot)$ is said to be consistent with the assignment of a and b (denoted by $C(a)$ and $C(b)$, respectively, and $C(a), C(b) \in \overline{C}$), if and only if

$$U(a) \geqslant U(b) \Rightarrow C(a) \succsim C(b), \tag{1}$$

$$U(a) \leqslant U(b) \Rightarrow C(a) \precsim C(b), \tag{2}$$

where $\succsim$ and $\precsim$ mean "as least as good as" and "as most as good as", respectively. Observe that (1) and (2) are equivalent to

$$C(a) \prec C(b) \Rightarrow \mathbf{u}^{\mathsf{T}} (\mathbf{V}(a) - \mathbf{V}(b)) < 0, \tag{3}$$

$$C(a) \succ C(b) \Rightarrow \mathbf{u}^{\mathsf{T}} (\mathbf{V}(a) - \mathbf{V}(b)) > 0, \tag{4}$$

since $U(a) = \mathbf{u}^{\mathsf{T}}\mathbf{V}(a)$ and $U(b) = \mathbf{u}^{\mathsf{T}}\mathbf{V}(b)$.

The proposed approach

First, let us introduce a set of indicators $t\left(a^{*}, b^{*}\right)$ for any pair of reference alternatives $a^{*}, b^{*} \in A^{R}$ such that $C\left(a^{*}\right) \neq C\left(b^{*}\right)$, which are defined as

$$t\left(a^{*}, b^{*}\right)=\begin{cases} 1, & \text{if } C\left(a^{*}\right) \succ C\left(b^{*}\right), \\ 0, & \text{if } C\left(a^{*}\right) \prec C\left(b^{*}\right). \end{cases}$$

The proposed approach

First, let us introduce a set of indicators $t\left(a^*, b^*\right)$ for any pair of reference alternatives $a^*, b^* \in A^R$ such that $C\left(a^*\right) \neq C\left(b^*\right)$, which are defined as

$$t\left(a^*, b^*\right) = \begin{cases} 1, & \text{if } C\left(a^*\right) \succ C\left(b^*\right), \\ 0, & \text{if } C\left(a^*\right) \prec C\left(b^*\right). \end{cases}$$

According to the above consistency principle for assignment, we aim to find a vector $\mathbf{u}$ such that, for any pair of reference alternatives $a^*, b^* \in A^R$ with $C\left(a^*\right) \neq C\left(b^*\right)$, we have

$$\mathbf{u}^{\mathsf{T}}\left(\mathbf{V}\left(a^*\right) - \mathbf{V}\left(b^*\right)\right) \begin{cases} > 0, & \text{if } t\left(a^*, b^*\right) = 1, \\ < 0, & \text{if } t\left(a^*, b^*\right) = 0. \end{cases} \tag{5}$$

The proposed approach

Then, instead of using slack variables, we can transform $\mathbf{u}^{\mathsf{T}}\left(\mathbf{V}\left(a^{*}\right)-\mathbf{V}\left(b^{*}\right)\right)$ into a value $y\left(a^{*}, b^{*}\right)$ for any pair of reference alternatives $a^{*}, b^{*} \in A^{R}$ so that we can use the difference between $y\left(a^{*}, b^{*}\right)$ and $t\left(a^{*}, b^{*}\right)$ to measure the inconsistency. $y\left(a^{*}, b^{*}\right)$ should satisfy the following conditions:

The proposed approach

Then, instead of using slack variables, we can transform $\mathbf{u}^{\mathsf{T}}\left(\mathbf{V}\left(a^{*}\right)-\mathbf{V}\left(b^{*}\right)\right)$ into a value $y\left(a^{*}, b^{*}\right)$ for any pair of reference alternatives $a^{*}, b^{*} \in A^{R}$ so that we can use the difference between $y\left(a^{*}, b^{*}\right)$ and $t\left(a^{*}, b^{*}\right)$ to measure the inconsistency. $y\left(a^{*}, b^{*}\right)$ should satisfy the following conditions:

- (a) $y\left(a^{*}, b^{*}\right)$ is monotone and increasing with respect to $\mathbf{u}^{\mathsf{T}}\left(\mathbf{V}\left(a^{*}\right)-\mathbf{V}\left(b^{*}\right)\right)$,

The proposed approach

Then, instead of using slack variables, we can transform $\mathbf{u}^{\mathsf{T}}\left(\mathbf{V}\left(a^{*}\right)-\mathbf{V}\left(b^{*}\right)\right)$ into a value $y\left(a^{*}, b^{*}\right)$ for any pair of reference alternatives $a^{*}, b^{*} \in A^{R}$ so that we can use the difference between $y\left(a^{*}, b^{*}\right)$ and $t\left(a^{*}, b^{*}\right)$ to measure the inconsistency. $y\left(a^{*}, b^{*}\right)$ should satisfy the following conditions:

- (a) $y\left(a^{*}, b^{*}\right)$ is monotone and increasing with respect to $\mathbf{u}^{\mathsf{T}}\left(\mathbf{V}\left(a^{*}\right)-\mathbf{V}\left(b^{*}\right)\right)$,
- (b) $y\left(a^{*}, b^{*}\right)$ is bounded within the interval $(0,1)$,

The proposed approach

Then, instead of using slack variables, we can transform $\mathbf{u}^{\mathsf{T}}\left(\mathbf{V}\left(a^{*}\right)-\mathbf{V}\left(b^{*}\right)\right)$ into a value $y\left(a^{*}, b^{*}\right)$ for any pair of reference alternatives $a^{*}, b^{*} \in A^{R}$ so that we can use the difference between $y\left(a^{*}, b^{*}\right)$ and $t\left(a^{*}, b^{*}\right)$ to measure the inconsistency. $y\left(a^{*}, b^{*}\right)$ should satisfy the following conditions:

- (a) $y\left(a^{*}, b^{*}\right)$ is monotone and increasing with respect to $\mathbf{u}^{\mathsf{T}}\left(\mathbf{V}\left(a^{*}\right)-\mathbf{V}\left(b^{*}\right)\right)$,
- (b) $y\left(a^{*}, b^{*}\right)$ is bounded within the interval $(0,1)$,
- (c) $\begin{cases} 0 < y(a^{*}, b^{*}) < 0.5, & \text{if } \mathbf{u}^{T}(\mathbf{V}(a^{*})-\mathbf{V}(b^{*})) < 0, \\ y(a^{*}, b^{*}) = 0.5, & \text{if } \mathbf{u}^{T}(\mathbf{V}(a^{*})-\mathbf{V}(b^{*})) = 0, \\ 0.5 < y(a^{*}, b^{*}) < 1, & \text{if } \mathbf{u}^{T}(\mathbf{V}(a^{*})-\mathbf{V}(b^{*})) > 0. \end{cases}$

The proposed approach

In this study, we use the following Sigmoid function to instantiate the function $y\left(a^*, b^*\right)$:

$$y\left(a^*, b^*\right) = \frac{1}{1 + e^{-\mathbf{u}^\top(\mathbf{V}(a^*) - \mathbf{V}(b^*))}}. \tag{6}$$

The Sigmoid function satisfies all the above requirements of the function $y\left(a^*, b^*\right)$.

The proposed approach

Then, we can consider the following non-linear optimization model to derive a value function that is as both consistent and robust as possible.

$$\max \prod_{a^*, b^* \in A^R \text{ with } C(a^*) \neq C(b^*)} \left[\frac{1}{1 + e^{-\mathbf{u}^\top (\mathbf{V}(a^*) - \mathbf{V}(b^*))}} \right]^{t\left(a^*, b^*\right)} \cdot \left[1 - \frac{1}{1 + e^{-\mathbf{u}^\top (\mathbf{V}(a^*) - \mathbf{V}(b^*))}} \right]^{1 - t\left(a^*, b^*\right)}$$

$$\text{s.t.} \quad \mathbf{u}^\top \mathbf{e} = 1, \ \mathbf{u} \geqslant \mathbf{0}.$$

The proposed approach

- In order to solve the non-linear optimization model for a large-scale problem efficiently, we propose a new parallel implementation for the Zoutendijk's feasible direction method based on the MapReduce framework.

The proposed approach

- In order to solve the non-linear optimization model for a large-scale problem efficiently, we propose a new parallel implementation for the Zoutendijk's feasible direction method based on the MapReduce framework.
- The model can be reformulated as follows

$$\min f(\mathbf{u}) = \sum_{a^*, b^* \in A^R \text{ with } C(a^*) \neq C(b^*)} \left[-t\left(a^*, b^*\right) \mathbf{u}^{\mathsf{T}} \left(\mathbf{V}\left(a^*\right) - \mathbf{V}\left(b^*\right)\right) + \ln\left(1 + e^{\mathbf{u}^{\mathsf{T}}\left(\mathbf{V}(a^*) - \mathbf{V}(b^*)\right)}\right)\right]$$

$$\text{s.t. } \mathbf{u}^{\mathsf{T}}\mathbf{e} = 1, \ \mathbf{u} \geqslant \mathbf{0}.$$

The proposed approach

The Zoutendijk's feasible direction method

Input:

Initial feasible solution $\hat{\mathbf{u}} = (1/\gamma,\ ...,\ 1/\gamma)^{\mathsf{T}}$.

1: Determine Ω and $\bar{\Omega}$ according to the current feasible solution $\hat{\mathbf{u}}$.
2: Calculate $\nabla f\,(\hat{\mathbf{u}})$.
3: Solve the LP model and obtain the optimal solution $\mathbf{d}^*$.
4: **if** $\nabla f(\hat{\mathbf{u}})^{\mathsf{T}}\mathbf{d}^* = 0$ **then**
5: Stop and $\hat{\mathbf{u}}$ is the global optimal solution.
6: **else**
7: Solve the model and obtain the optimal solution λ^*.
8: Update $\hat{\mathbf{u}} \leftarrow \hat{\mathbf{u}} + \lambda^*\mathbf{d}^*$.
9: Go to step 1.
10: **end if**

Output:

The optimal solution $\hat{\mathbf{u}}$.

The proposed approach

$$\min f(\mathbf{u}) = \sum_{a^*, b^* \in A^R \text{ with } C(a^*) \neq C(b^*)} \left[-t\left(a^*, b^*\right) \mathbf{u}^{\mathsf{T}} \left(\mathbf{V}\left(a^*\right) - \mathbf{V}\left(b^*\right)\right) + \ln\left(1 + e^{\mathbf{u}^{\mathsf{T}}\left(\mathbf{V}(a^*) - \mathbf{V}(b^*)\right)}\right)\right]$$

$$\text{s.t. } \mathbf{u}^{\mathsf{T}}\mathbf{e} = 1, \ \mathbf{u} \geqslant \mathbf{0}.$$

- Considering that the number of reference alternatives in a large-scale problem is reasonably large, the objective is composed of a huge number of terms (i.e., $-t\left(a^*, b^*\right) \mathbf{u}^{\mathsf{T}} \left(\mathbf{V}\left(a^*\right) - \mathbf{V}\left(b^*\right)\right) + \ln\left(1 + e^{\mathbf{u}^{\mathsf{T}}\left(\mathbf{V}(a^*) - \mathbf{V}(b^*)\right)}\right)$, $a^*, b^* \in A^R$).
- This inspires us to utilize the MapReduce framework to accelerate the computation of $f(\mathbf{u})$ and $\nabla f(\mathbf{u})$ for the Zoutendijk's feasible direction method.

The proposed approach

Calculate $\nabla f(\hat{\mathbf{u}})$: Map phase.

Input:

$\langle key, value \rangle$ where *key* is the index of subset and *value* is the subset of pairs of reference alternatives $(a^*, b^*) \in A^R \times A^R$ such that $C(a^*) \neq C(b^*)$; the current feasible solution $\hat{\mathbf{u}}$.

1: $\boldsymbol{\rho} \leftarrow \mathbf{0}$.
2: **for** any pair of reference alternatives (a^*, b^*) in this subset **do**
3: $\quad \boldsymbol{\rho} \leftarrow \boldsymbol{\rho} + \left[-t(a^*, b^*) + \frac{e^{\hat{\mathbf{u}}^\top (\mathbf{V}(a^*) - \mathbf{V}(b^*))}}{1 + e^{\hat{\mathbf{u}}^\top (\mathbf{V}(a^*) - \mathbf{V}(b^*))}} \right] (\mathbf{V}(a^*) - \mathbf{V}(b^*))$.
4: **end for**

Output:

$\langle key = \hat{\mathbf{u}}, value = \boldsymbol{\rho} \rangle$.

The proposed approach

Calculate $\nabla f(\hat{\mathbf{u}})$: Reduce phase.

Input:

$\langle key = \hat{\mathbf{u}}, value = list(\boldsymbol{\rho}) \rangle$.

1: $\boldsymbol{\delta} \leftarrow \mathbf{0}$.
2: **for** any $\boldsymbol{\rho}$ in $list(\boldsymbol{\rho})$ **do**
3: $\quad \boldsymbol{\delta} \leftarrow \boldsymbol{\delta} + \boldsymbol{\rho}$.
4: **end for**

Output:

$\langle key = \hat{\mathbf{u}}, value = \boldsymbol{\delta} \rangle$ where $\boldsymbol{\delta}$ is equal to $\nabla f(\hat{\mathbf{u}})$.

Experimental analysis

The experimental analysis is based on Best Chinese Universities Rankings (BCUR) in 2018, which provides the overall ranking of 600 universities in China.

Table 1: Considered criteria and corresponding indicators in BCUR.

Dimension	Criteria	Indicator
Teaching and learning	(g_1) Quality of incoming students	Average score of incoming freshmen in national college entrance exam
	(g_2) Education outcome	Employment rate of bachelor degree recipients
	(g_3) Reputation	Income from donations
Research	(g_4) Scale of research	Number of papers in Scopus
	(g_5) Quality of research	Field weighted citation impact
	(g_6) Top research achievements	World top 1% most cited paper
	(g_7) Top scholars	Chinese most cited researchers
Social service	(g_8) Technology service	Research income from industry
	(g_9) Technology transfer	Income from technology transfer
Internationalization	(g_{10}) International student ratio	International students as a percentage of total students

Experimental analysis

- We divide all the 600 universities into five categories according to their total scores. Each category is composed of 120 universities and Cl_5 and Cl_1 are the best and worst categories, respectively.
- The performance of the proposed approach is compared with that of the classical UTADIS method through *cross-validation*.

Table 2: Accuracy of different methods for the problem of university classification.

Method	$\gamma_j = 1$	$\gamma_j = 2$	$\gamma_j = 3$	$\gamma_j = 4$	$\gamma_j = 5$
UTADIS	0.9050	0.8917	0.9033	0.9100	0.9083
Proposed approach	0.9246	0.9285	0.9462	0.9598	0.9514
t-test	0.0000^*	0.0000^*	0.0000^*	0.0000^*	0.0000^*

Thank you!