Creation Cost Minimization

11612126 李可明

Problem 1

- Minimize Creation Cost, discrete user preference set W:

Given a dataset D, a user preference set W, a cover percentage τ, a positive integer k and a creation cost function C, CCM introduces a new product p such that the cover ratio

$$cp(p,k) = \frac{|\{w|\forall w \in W, p\ is\ in\ TopK(w)\}|}{|W|} > \tau$$

and the creation cost $C(p)$ is minimized.

Problem 2

- Minimize Creation Cost, user distribution θ:

Given a dataset D, a user preference distribution θ, a cover percentage τ, a positive integer k and a creation cost function C, CCM introduces a new product p such that the cover ratio

$$cp(p,k) = \frac{|\{w|\forall w \in \theta, p\ is\ in\ TopK(w)\}|}{|\theta|} > \tau$$

and the creation cost $C(p)$ is minimized.

For this week: reading 2 papers

- $\boldsymbol{kSPR}$: [Tang et al., 2017] Tang, B., Mouratidis, K., and Yiu, M. L. (2017). Determining the impact regions of competing options in preference space. In Proceedings of the 2017 ACM International Conference on Management of Data, pages 805–820. ACM.

- $\boldsymbol{TopRR}$: [Tang et al., 2019] Tang, B., Mouratidis, K., Yiu, M. L., and Chen, Z. (2019). Creating top ranking options in the continuous option and preference space. Proceedings of the VLDB Endowment, 12(9):xxx–yyy.

kSPR

$kSPR$-Problem Definition

- The k-Shortlist Preference Region problem ($kSPR$) takes as input

  a dataset D,

  a focal record $p = (p_1, p_2, \cdots, p_d)$,

  and an integer k.

It reports all the regions in preference space where if the weight vector lies, p ranks among the top-k records.

A straight forward solution is to find

$$\cup\, C_i$$

For all C_i that $C_i = \mathrm{h}_{i,1}^{-} \cap h_{i,2}^{-} \cap \cdots \cap h_{i,|D|-k}^{-}$,

$h_{i,j}^{-}$ is the half-space of $S(r_{i,j}) < S(p)$ and $S(r) = w \cdot r$

$kSPR$-Problem Definition

A straight forward solution is to find

$$\cup\, C_i$$

For all C_i that $C_i = \mathrm{h}_{i,1}^{-} \cap h_{i,2}^{-} \cap \cdots \cap h_{i,|D|-k}^{-}$,

$h_{i,j}^{-}$ is the half-space of $S(r_{i,j}) < S(p)$ and $S(r) = w \cdot r$

在一个半空间内表明p比一个r分数高， 只要p在至少|D|-k个半空间内它必能排前k(D包含p)

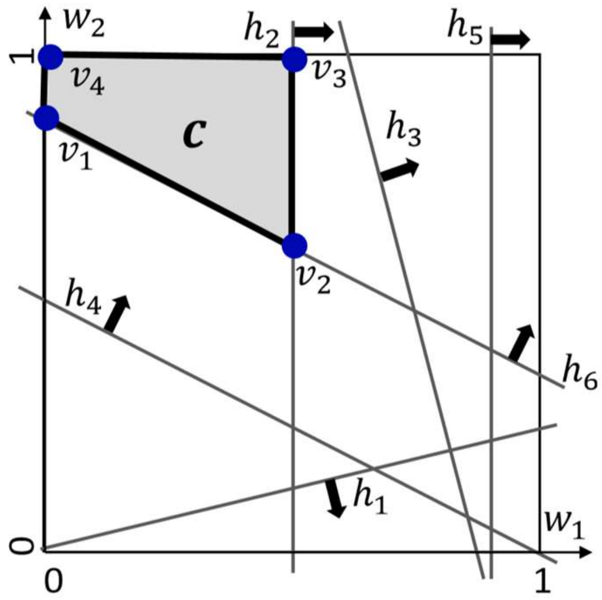
而这|D|-k个半空间是从|D|-1个半空间组合， 只要将所有结果取并集就是最后结果， 时间复杂度为组合数

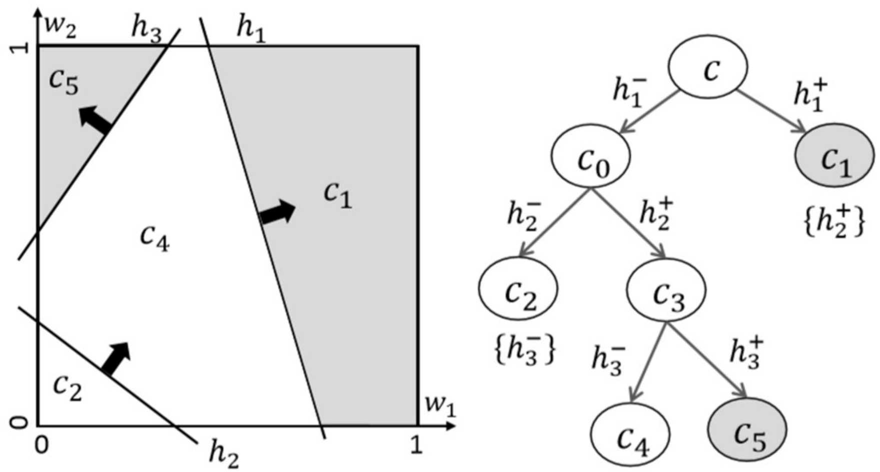
$kSPR\text{-}CTA$

- Cell c is defined as

  $h_1^- \cap h_2^- \cap h_3^- \cap h_4^+ \cap h_5^- \cap h_6^+$

- 剪枝:
  - 数halfspaces数
  - 利用option之间的domination关系
  - …





(a) Arrangement (b) $CellTree$

TopRR

TopRR: problem definition

- Given a dataset D,

  a positive integer k,

  and a preference region $wR \subset W$,

  compute the maximal region $o\mathrm{R}$ in option space where a new option o should lie, so that it is a top-ranking option.

TopRR: Notation

Notation	Description
$\mathcal{D}$	Option dataset
$\boldsymbol{p}_i$	An (existing) option in $\mathcal{D}$
$\mathcal{W}$	Continuous preference space
$\boldsymbol{w}$	A weight vector in $\mathcal{W}$
wR	Region in preference space
$\mathcal{O}$	Continuous option space
$\boldsymbol{o}$	New option to be placed in $\mathcal{O}$
oR	Region in option space (output of TopRR)
$S_{\boldsymbol{w}}(\boldsymbol{p}_i)$	The score of $\boldsymbol{p}_i$ according to $\boldsymbol{w}$ (equal to $\boldsymbol{p}_i \cdot \boldsymbol{w}$)
$TopK(\boldsymbol{w})$	k-th highest score in $\mathcal{D}$ according to $\boldsymbol{w}$
$\mathsf{oH}(\boldsymbol{w})$	Impact halfspace for $\boldsymbol{w}$ (Definition 2)
$\mathsf{wHP}(\boldsymbol{p}_i, \boldsymbol{p}_j)$	Hyperplane for $S_{\boldsymbol{w}}(\boldsymbol{p}_i) = S_{\boldsymbol{w}}(\boldsymbol{p}_j)$ in preference space
$\mathsf{wH}(\boldsymbol{p}_i, \boldsymbol{p}_j)$	Halfspace for $S_{\boldsymbol{w}}(\boldsymbol{p}_i) \geq S_{\boldsymbol{w}}(\boldsymbol{p}_j)$ in preference space

TopRR:

- **Impact halfspace:**

$$\boldsymbol{oH}(\boldsymbol{w}) = \{\boldsymbol{o} \in \boldsymbol{\mathcal{O}}: \boldsymbol{S_w}(\boldsymbol{o}) \geq \boldsymbol{TopK}(\boldsymbol{w})\}$$

where $\boldsymbol{\mathcal{O}}$ is the option region and $S_w(o)$ is the score of o respect to w

For example,

$w = (0.4, 0.6), TopK(w) = 0.5,$

then $oH(w) = \{(x, y)|0.4x + 0.6y \geq 0.5, x \in [0, 1], y \in [0,1]\}$

w的impact halfspace里的所有options都是能在w这个权重下排前k的，也就是说所有w的impact halfspace的交集就是我们TopRR问题下要的能满足在任意w下都能排前k的region

TopRR

LEMMA 1. *Let* $\mathbf{p}$ *and* $\mathbf{p}'$ *be two options, and* wR_i *be a convex polytope in preference space. Let also* $\mathcal{V}$ *be the vertex set that defines* wR_i*, i.e., the set of the polytope's extreme vertices. If it holds that* $S_{\mathbf{v}}(\mathbf{p}) \geq S_{\mathbf{v}}(\mathbf{p}')$ *for every vertex* $\mathbf{v} \in \mathcal{V}$*, then* $S_{\mathbf{w}}(\mathbf{p}) \geq S_{\mathbf{w}}(\mathbf{p}')$ *also holds for every weight vector* $\mathbf{w} \in \mathsf{wR}_i$*.*

PROOF. Consider a weight vector $\boldsymbol{w} \in \mathsf{wR}_i$. Since wR_i is a convex polytope, $\boldsymbol{w}$ can be expressed as:

$$\boldsymbol{w} = \sum_{\boldsymbol{v} \in \mathcal{V}} \lambda_{\boldsymbol{v}} \boldsymbol{v} \tag{1}$$

for some positive $\lambda_{\boldsymbol{v}}$ coefficients, where $\sum_{\boldsymbol{v} \in \mathcal{V}} \lambda_{\boldsymbol{v}} = 1$.

Since $S_{\boldsymbol{v}}(\boldsymbol{p}) \geq S_{\boldsymbol{v}}(\boldsymbol{p}')$ for each vertex $\boldsymbol{v} \in \mathcal{V}$, it follows that $\lambda_{\boldsymbol{v}} S_{\boldsymbol{v}}(\boldsymbol{p}) \geq \lambda_{\boldsymbol{v}} S_{\boldsymbol{v}}(\boldsymbol{p}')$ too. By summing across all $\boldsymbol{v} \in \mathcal{V}$, we get:

$$\begin{aligned}
&\sum_{\boldsymbol{v} \in \mathcal{V}} \lambda_{\boldsymbol{v}} S_{\boldsymbol{v}}(\boldsymbol{p}) \geq \sum_{\boldsymbol{v} \in \mathcal{V}} \lambda_{\boldsymbol{v}} S_{\boldsymbol{v}}(\boldsymbol{p}') \\
\Rightarrow\ &\sum_{\boldsymbol{v} \in \mathcal{V}} \lambda_{\boldsymbol{v}} (\boldsymbol{v} \cdot \boldsymbol{p}) \geq \sum_{\boldsymbol{v} \in \mathcal{V}} \lambda_{\boldsymbol{v}} (\boldsymbol{v} \cdot \boldsymbol{p}') \\
\Rightarrow\ &\boldsymbol{p} \cdot \sum_{\boldsymbol{v} \in \mathcal{V}} \lambda_{\boldsymbol{v}} \boldsymbol{v} \geq \boldsymbol{p}' \cdot \sum_{\boldsymbol{v} \in \mathcal{V}} \lambda_{\boldsymbol{v}} \boldsymbol{v} \qquad \triangleright \text{ applying Equation (1)} \\
\Rightarrow\ &\boldsymbol{p} \cdot \boldsymbol{w} \geq \boldsymbol{p}' \cdot \boldsymbol{w}
\end{aligned}$$

Thus, $\forall\, \boldsymbol{w} \in \mathsf{wR}_i$, it holds that $S_{\boldsymbol{w}}(\boldsymbol{p}) \geq S_{\boldsymbol{w}}(\boldsymbol{p}')$. $\square$

TopRR:

- **kIPR**, RANK-k INVARIAN PREFERENCE REGION:

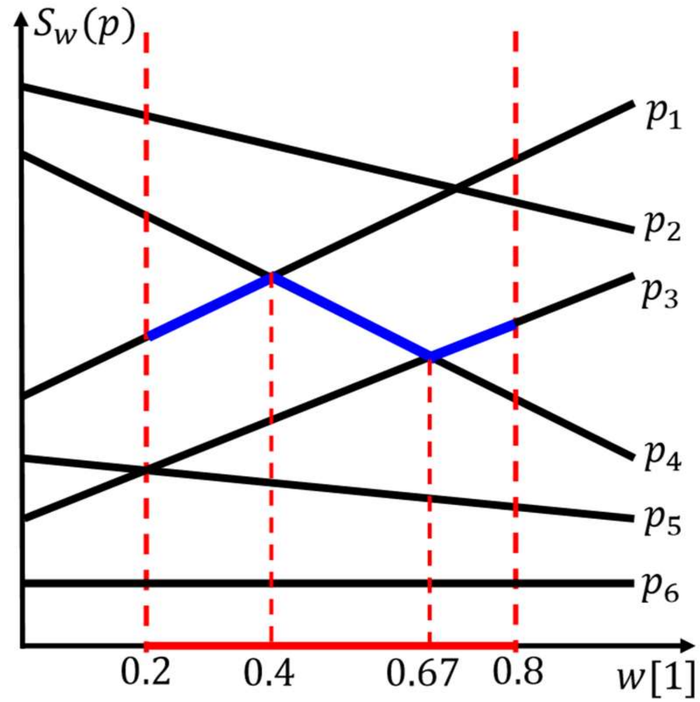
A region wRi in preference space is a rank-k invariant preference region (**kIPR**) if the top-k result for every weight vector

$$w \in wR_i$$

(i) comprises the same k options, and

(ii) the top-k-th option is always the same.

As shown in the figure, region [0.2, 0.8] and [0.2, 0.67] are not $kIPR$ but region [0.67, 0.8] is.



TopRR:

LEMMA 2: The $TopRR$ solution for a $kIPR$ convex polytope wR_i is option region

$$oR_i = \bigcap_{v \in V} oH(v)$$

Theorem: Given the partitioning of wR into $kIPRs$, let V_{all} be the set of all vertices that define those $kIPRs$. The $TopRR$ solution for wR is option region

$$oR_i = \bigcap_{v \in V_{all}} oH(v)$$

TopRR:

TAS, test-and-split:

(i) Testing whether a region is $kIPR$

(ii) Splitting a $non-kIPR$ region

(iii) Recursively repeat (i) and (ii) until all regions are split as $kIPR$

(iv) Find the result oR by

$$oR_i = \bigcap_{v \in V_{all}} oH(v)$$

where V_{all} is the set of all vertices that define those $kIPRs$.