【ICDE2013勉強会】

R28: Skyline and Snapshot Query

担当：大島裕明（京都大学）

Skyline and Snapshot Query

[1] On Answering Why-not Questions in Reverse Skyline Queries
- Md. Saiful Islam (Swinburne Univ. of Technogy)
- Rui Zhou (Swinburne Univ. Technology)
- Chengfei Liu (Swinburne Univ. Technology)

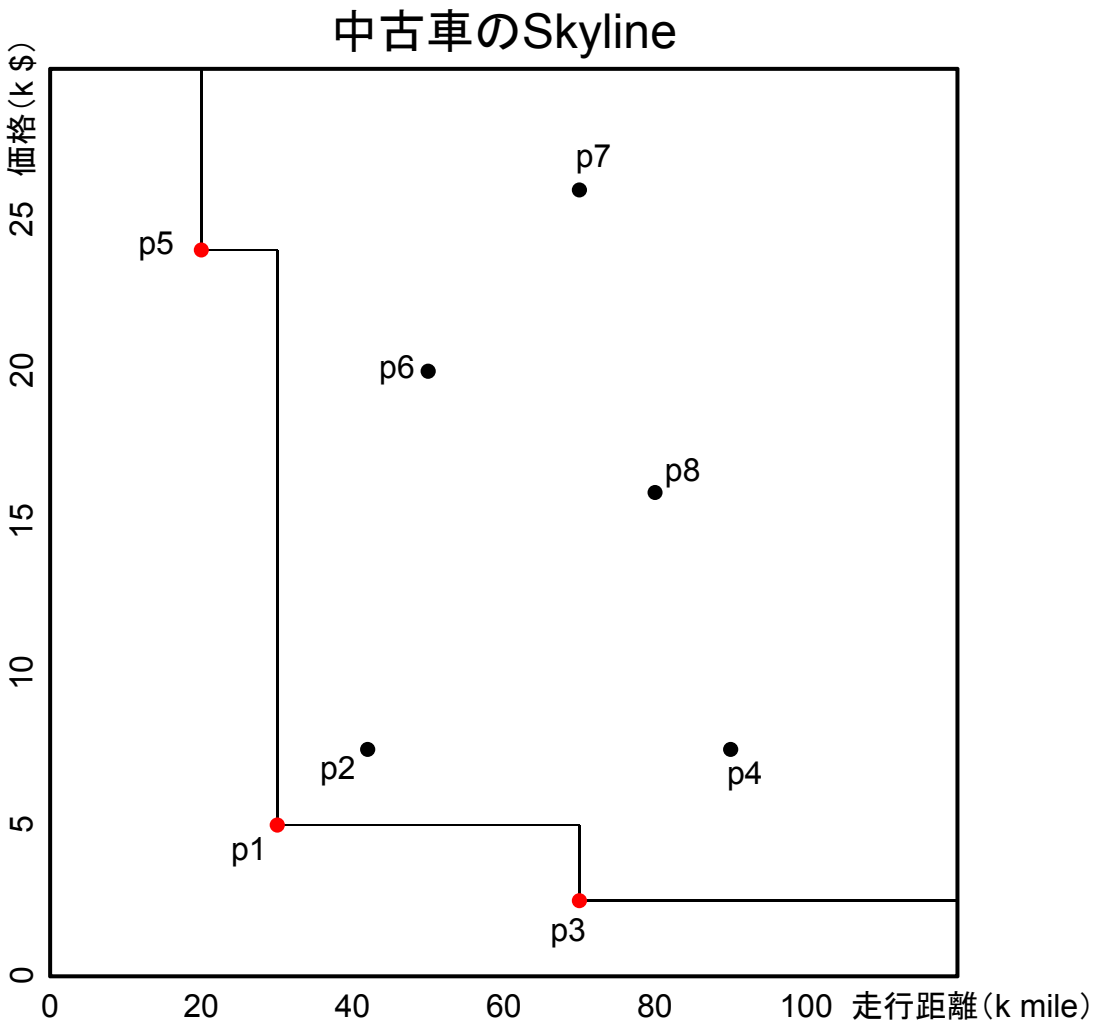
[2] Layered Processing of Skyline-Window-Join (SWJ) Queries using Iteration-Fabric
- Mithila Nagendra (Arizona State University)
- K. Selcuk Candan (Arizona State University)

[3] Efficient Snapshot Retrieval over Historical Graph Data
- Udayan Khurana (University of Maryland)
- Amol Deshpande (University of Maryland)

- 本資料中の図表の引用元はこれらの論文です

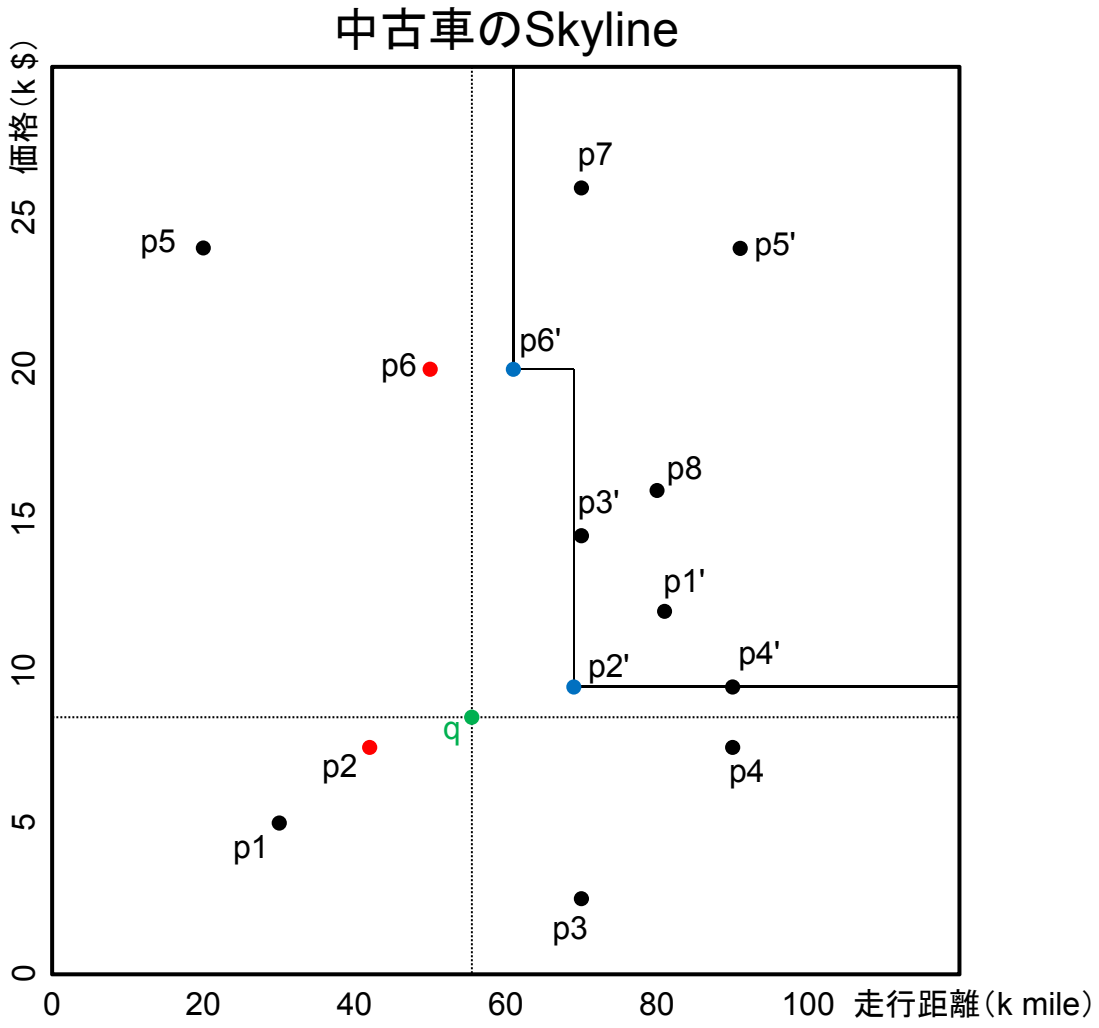
Skylineの概要

- Skyline
  - Dominateされていない点の集合
  - 右図の赤い点集合
- Dominated?
  - p2 is dominated by p1
  - p2はp1より、マジだめ
    - 値段は高い
    - 走行距離は長い
  - p1とp3は互いに not dominate
    - 値段はp1が安い
    - 走行距離はp3が短い



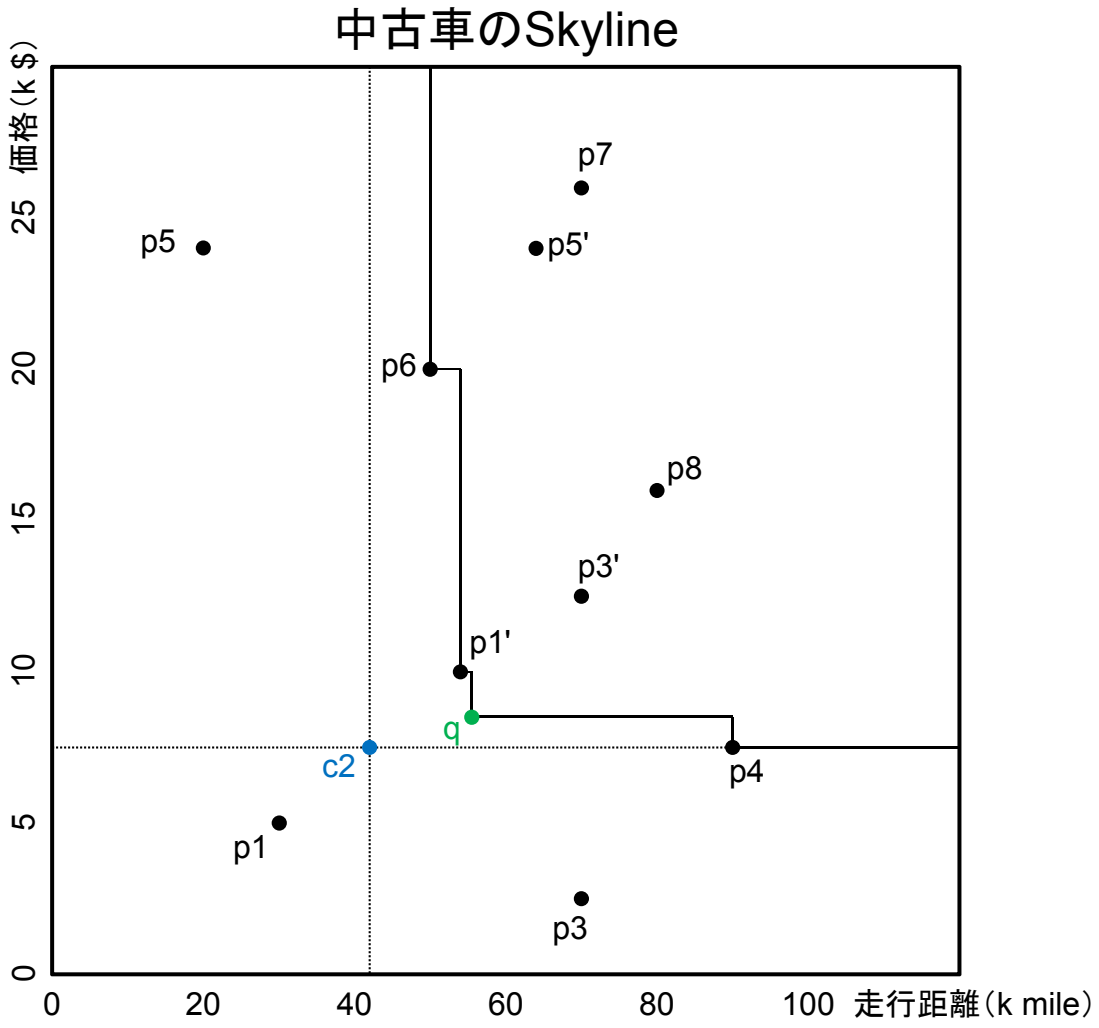
Skylineの概要

- Dynamic Skyline
  - クエリqからの距離に基づくSkyline
  - 右図の赤い点集合
  - ＝右図の青い点集合
  - 求め方
    - qを原点とする第一象限に全ての点をマッピング
    - その後、マッピングされた点のSkylineを取得



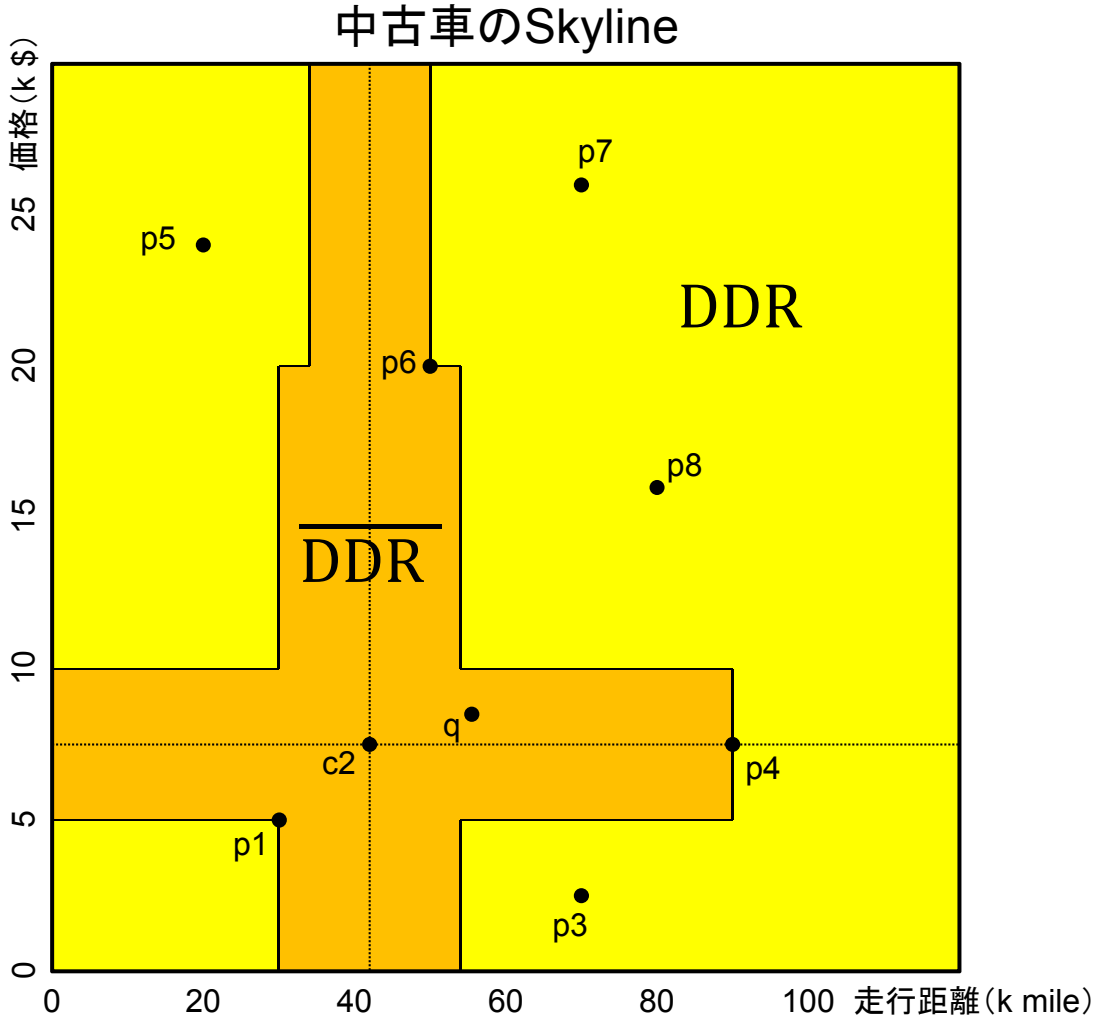
Skylineの概要

- Reverse Skyline
  - 入力
    - 点集合P
    - 点集合C
    - クエリq
  - 出力
    - Cの部分集合
    - c∈CがReverse Skylineに含まれる
      ⇓
      cをクエリとして、P∪{q}を対象としたDynamic Skylineを求めたらそこにqが含まれる
  - 例
    - P={p1,p3,…,p8}
    - C={c2}
    - Reverse Skyline = {c2}



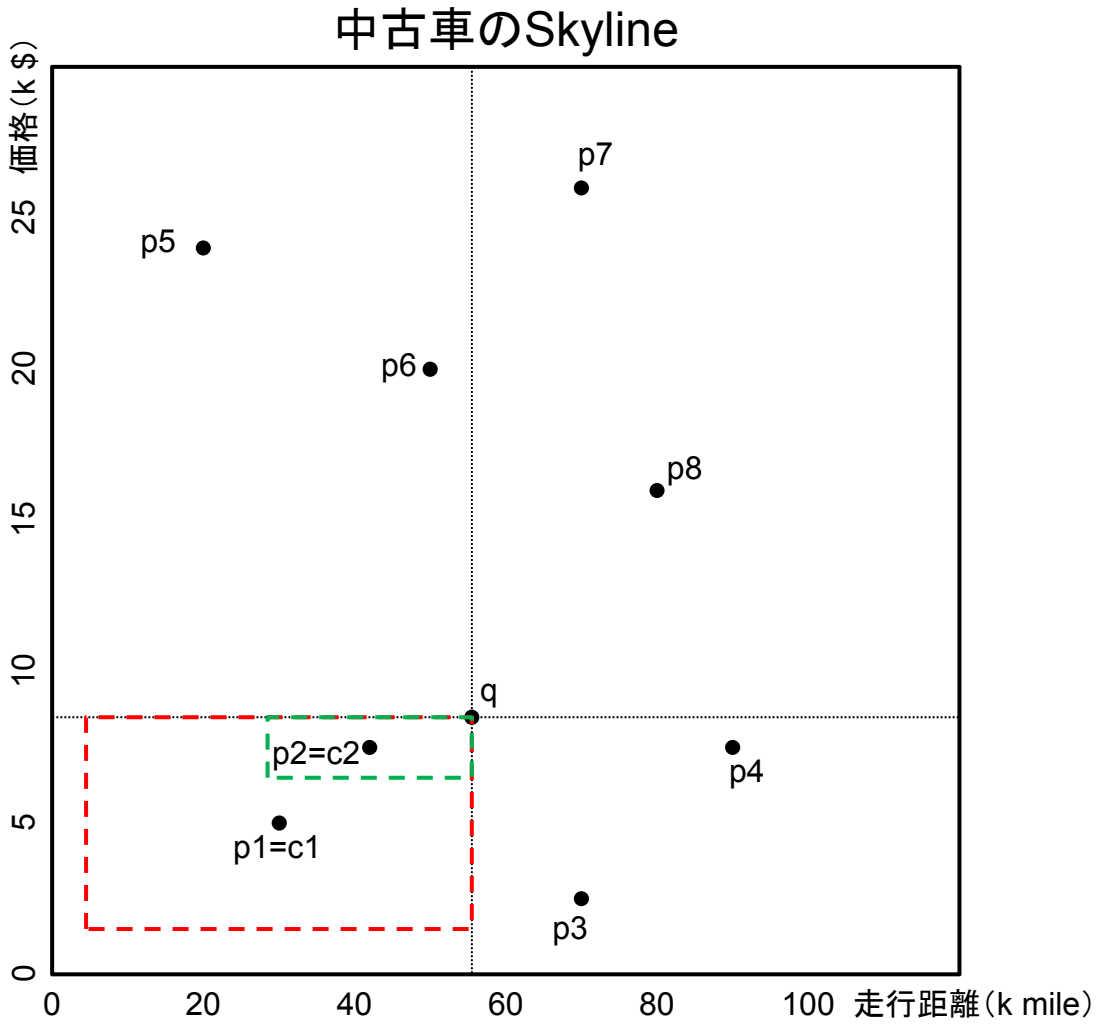
Skylineの概要

- Dynamic Dominance Region (DDR)
  - 入力
    - 点集合P
    - 点c
  - DDR
    - cをクエリとしたとき、Dynamic Skylineのいずれかによってdominateされている領域
  - $\overline{\text{DDR}}$
    - DDRではない領域
  - DDRと点q
    - 適当な点qを考えたとき、それが$\overline{\text{DDR}}$にあれば、点qはDynamic Skylineになる
    - 点qにdominateされる領域は、DDRになる



Skylineの概要

- 手抜きReverse Skyline
  - 入力
    - 点集合P
    - 点集合C
    - クエリq
  - window_query
    - c∈Cを中心とし、qを頂点の1つとする長方形
    - cのwindow_queryに、p∈Pが含まれていない
      ⇓
      cはこの入力のReverse Skylineに含まれる
  - 例
    - c2はReverse Skyline
    - c1はちがう



On Answering Why-not Questions in Reverse Skyline Queries

研究の目的

- 自分の中古車をどうにか客に売りたい

背景

- ある客は、その中古車になぜか興味を持ってくれない

対策

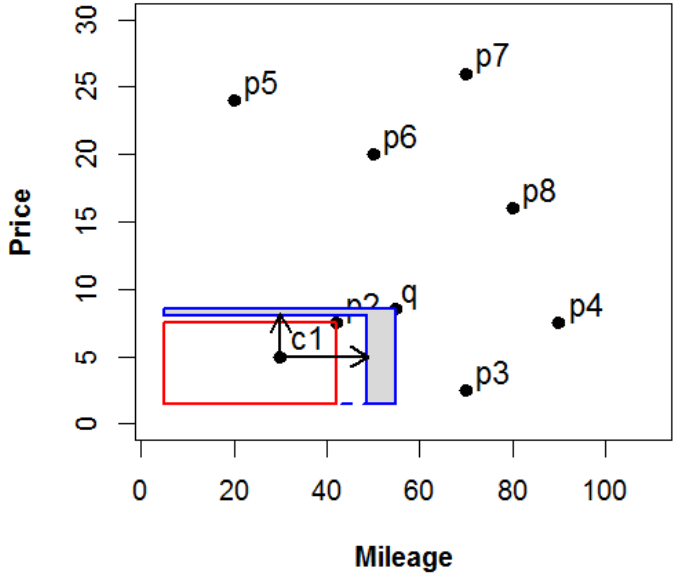
- 客が興味を持ってくれない理由を知る
- 客の好みを変えてしまうと、興味を持ってくれるに違いない
- 中古車を変えると、客が興味を持ってくれるに違いない

例

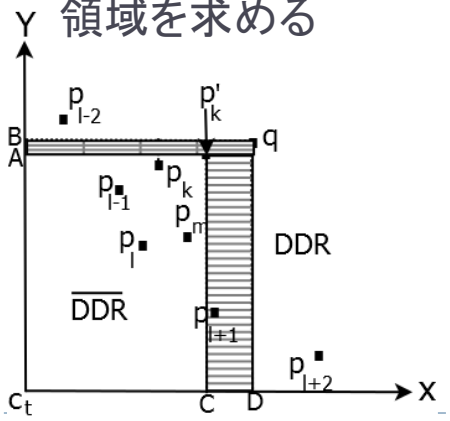
- 客の好み：価格5,000$、走行距離30キロマイル
- 自分の中古車：価格8,500$、走行距離55キロマイル
- ある他の中古車：価格7,500$、走行距離42キロマイル

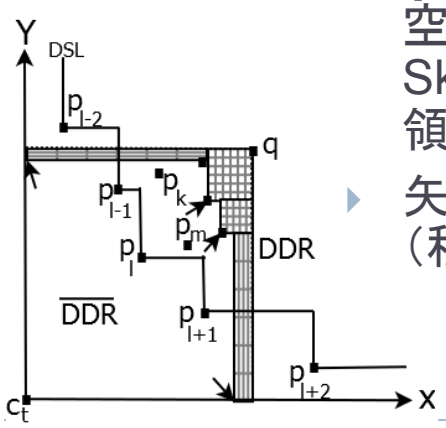
On Answering Why-not Questions in Reverse Skyline Queries

- （なるべく小さく）客を変える
  - 元の好み
    - 価格5,000$、距離30キロマイル
  - 変更後の好み
    - 価格5,000$、距離48.5キロマイル
    - 価格8,000$、距離30キロマイル
- p_kに勝てる領域を求める
  - qに対するc_tのwindow_queryにp_kが含まれない領域にc_tがあること
  - p_kとqの中間点p'_kを基にそのような領域を求める

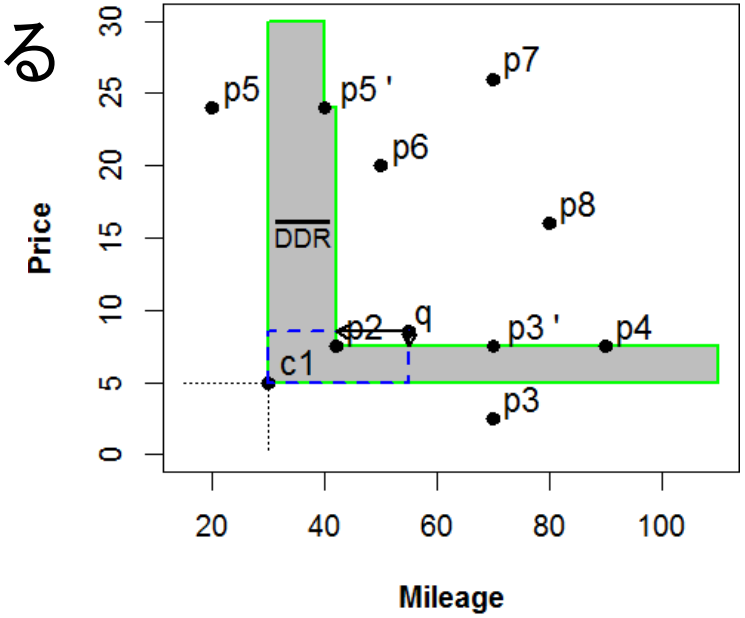


- c_tの移動候補を求める
  - qとc_tを頂点とする長方形の空間の点で、qに対するSkylineの点の全てに勝てる領域を求める
  - 矢印で表されるような点が（移動量が小さい）移動候補

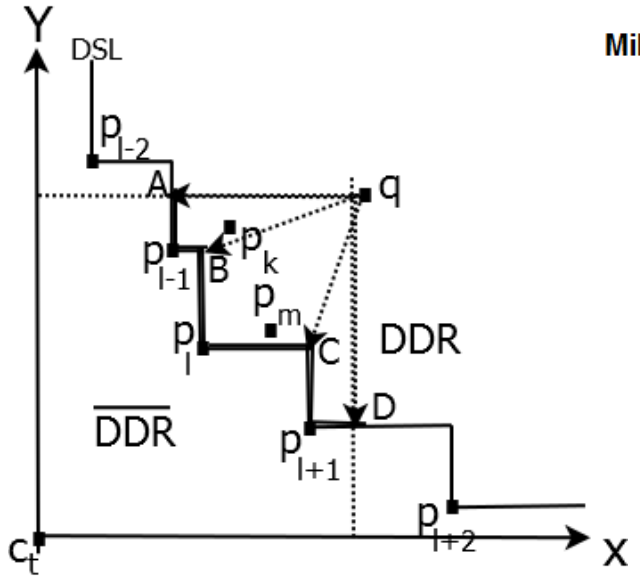




- （なるべく小さく）中古車を変える
  - 元の中古車
    - 価格8,500$、距離55キロマイル
  - 変更後の中古車
    - 価格8,500$、距離42キロマイル
    - 価格7,500$、距離55キロマイル



- $\overline{\text{DDR}}$に移動する
  - 直感的に、右図



Layered Processing of Skyline-Window-Join (SWJ) Queries using Iteration-Fabric

研究の目的

- 静的なデータではなく、データストリームのwindowにおけるSkyline
- 2本のデータストリームのJOIN

skyline-window-join

提案手法

- JOINしながらSkylineを求める
  - 従来手法では、JOINしてからSkylineする

既存手法の組み合わせ

- StabSky：データストリームにおけるSkylineを効率よく求める手法
- Iterative：Skyline-Join（JoinのSkyline）を効率よく求める手法

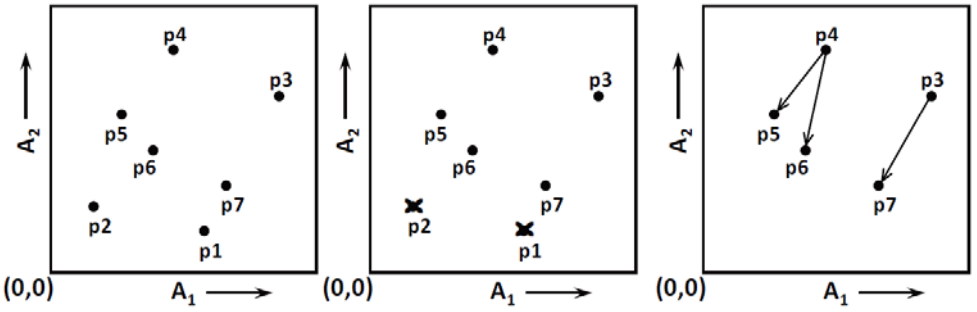
Layered Processing of Skyline-Window-Join (SWJ) Queries using Iteration-Fabric

StabSky [ICDE2005]

- データストリームのSkyline
- 右図：A1、A2でMAXのSkylineはp3とp4
- (a) データストリームのデータ
- (b) Skylineになり得ない点を無視する
  - p1は、より若いp7にdominateされている
  - p2も同様
- (c) ある点をdominateしている点のうち、より若い点との関係をcriticalとする

Iterative [ICDE workshop 2008]

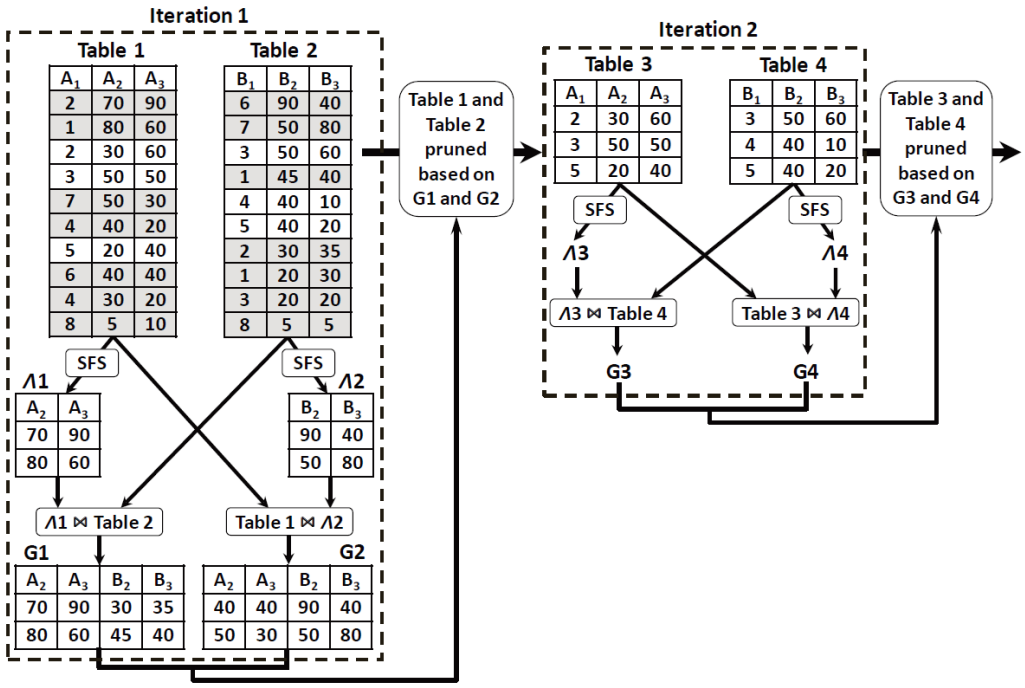
- JoinのSkylineを求める反復的手法
- 繰り返し処理
  - Λ：各テーブルのSkyline
  - G：Λともう一方のテーブルのJoinを求め、その中でdominateされていないもの
- 枝刈り処理
  - Table3：Table1のタプルのうち、
    - G2でのA1の最低値よりも大きいA1を持つ
    - G2でのA2の最低値よりも大きいA2を持つ
  - Table4：Table2のタプルのうち、
    - ……



(a) データストリーム
番号が若い方が古い

(b) Skylineにならない点は無視する

(c) Criticalなdominance関係



Efficient Snapshot Retrieval over Historical Graph Data

研究の目的

- グラフのための分散データベースシステム

提案手法

- 差分グラフの階層的分散インデックス：DeltaGraph

DeltaGraph

- 各ノードは、グラフを表す
- 下位ノードを組み合わせれば上位ノードが構成される
- 実際に記録されるのは、エッジ（edge deltas）の情報
  - エッジは、ノードの差分を表現

Efficient Snapshot Retrieval over Historical Graph Data

ポイント

- 保存されるのはΔ
- 列ベースで保存＝分割可能
  - （グラフの）構造のΔ、
    ノード属性のΔ、
    エッジ属性のΔ、…

差分関数
Intersection
Skewed
Balanced
…