

【ICDE2015勉強会】

Session 9, 10, 21

担当: 董 テイテイ(名大)

紹介内容

- ▶ Session 9: **Keyword and Top-K Processing**
 - ▶ 4: **Interactive Top-k Spatial Keyword Queries**
- ▶ Session 10: **Query Processing 2**
 - ▶ 3: Linear Path Skylines in Multicriteria Networks
- ▶ Session 21: **Trajectories**
 - ▶ 1: Making Sense of Trajectory Data: A Partition-and-Summarization Approach

Interactive Top-k Spatial Keyword Queries

- ▶ Kai Zheng (The University of Queensland) et al.
- ▶ Top-k spatial keyword queries



Fig. 1: An example of spatial keyword query # 図は論文から引用

- ▶ 問題点: $Score = \beta S_{spatial} + (1 - \beta) S_{keyword}$
 - ▶ 好みパラメータ β の指定は難しい
 - ▶ TF-IDFなどの方法で $S_{keyword}$ を計算することは直感ではない

[VLDB'09]の方法を使うと,

1. $\beta = 0.5$ (default)
2. TF-IDF を使う
 $\omega_{fish\&chips} = 0.25 < \omega_{music} = 0.5$
3. Result: o_6 ($k = 1$)

o_1 seems better than o_6 !



User interactionsで重み(好みとkeywordの重要性)を学習

Interactive Top-k Spatial Keyword Queries

▶ Solution overview

▶ Candidate generation

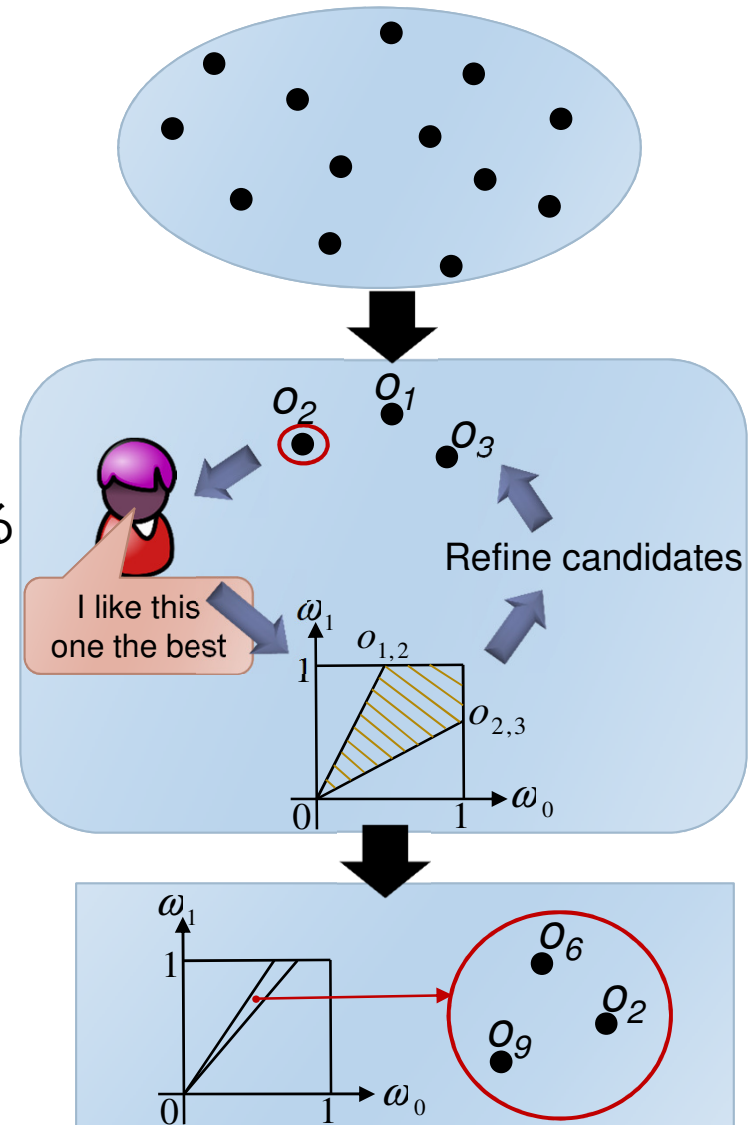
- ▶ 全ての候補オブジェクトを見つける

▶ Interaction

- ▶ 反復して重みを学習
 - 候補の部分集合を決めユーザに見せる
 - ユーザは好きな候補を選ぶ
 - 重みを絞り込む(線形制約を計算)
 - 次の候補部分集合を決める

▶ Finalisation

- ▶ 重みを推定し, トップk結果を計算



Interactive Top-k Spatial Keyword Queries

▶ Candidate generation

- ▶ 候補 = **(k-1)-skyband**: データベースの中で, **最大k-1** 個のオブジェクトに**支配**するオブジェクト集合

- ▶ a は b を**支配**(dominate): a は b に比べて, q により近い, より多いキーワードを含む

- (0.1, music) は (0.2, music) を**支配**
(0.1, fish&chips, music) は (0.1, music) を**支配**

- ▶ **Branch-and-bound** search using **IR²-tree** [ICDE'08]

- ▶ **IR²-tree**: R-treeとsignature fileの組合せ

▶ Interaction

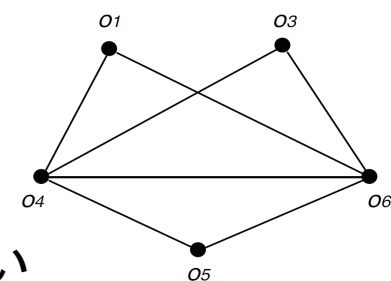
- ▶ 候補の部分集合を決める

- ▶ Strategies: random selection (**RS**), densest subgraph (**DS**), uncertainty reduction (**UR**)

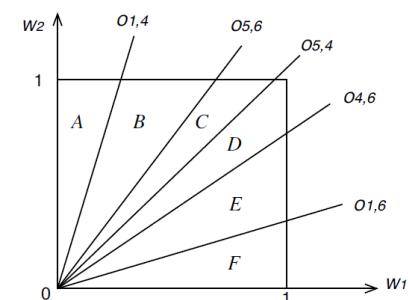
Interactive Top-k Spatial Keyword Queries

▶ Interaction

- ▶ 候補の部分集合を決める (**RS**, **DS**, **UR**)
 - ▶ Random selection (**RS**): ランダムに決める
 - ▶ Densest subgraph (**DS**): 有効な制約の**数**を最大化
 - 支配関係のあるオブジェクトから, 有効な制約が得られない
 - 支配関係のないオブジェクトを**グラフ**で表現し, **発見的な手法**で近似 **densest subgraph** を求める
 - ▶ Uncertainty reduction (**UR**): 各制約の**有効さ**を**区別**
 - 各重み空間の **uncertainty** を **entropy** (volume) で表現し, volume を最小化
 - **貪欲な手法**で近似解を求める



(a) Dominance graph



(b) Geometric representation of constraints

図は論文から引用

▶ Finalisation

- ▶ **Interaction**の終了条件: 閾値を設定
- ▶ 重みの推定: **確信度**が最も高い重み

紹介内容

- ▶ **Session 9: Keyword and Top-K Processing**
 - ▶ 4: Interactive Top-k Spatial Keyword Queries
- ▶ **Session 10: Query Processing 2**
 - ▶ 3: [Linear Path Skylines in Multicriteria Networks](#)
- ▶ **Session 21: Trajectories**
 - ▶ 1: Making Sense of Trajectory Data: A Partition-and-Summarization Approach

Linear Path Skylines in Multicriteria Networks

- ▶ Michael Shekelyan (Ludwig-Maximilians-University Munich) et al.
- ▶ **多基準経路計算 (Multicriteria routing)**
 - ▶ グラフで始点ノードから終点ノードまでの経路の計算
 - ▶ グラフの各エッジには**多数のコスト基準**がある
 - ▶ 例: 移動時間, 移動距離, エネルギー消費量, 交通信号の数など
 - ▶ 部分コスト基準を最適化する多数の経路結果 (**path skyline**) を考える
 - ▶ 例: 経路1 (15分, 10km), 経路2 (12分, 12km)

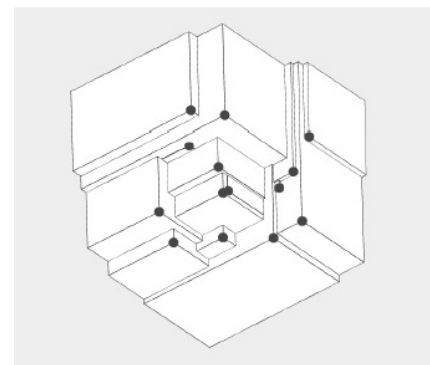
- ▶ **Linear path skyline**

- ▶ コスト基準の**線形的組合せ**のもとで最適である経路の集合

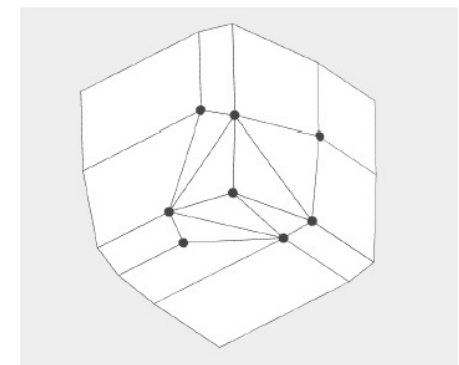
$$\{ \arg \min_{x \in \mathcal{X}} \omega^T x \mid \omega \in W^d \}$$

- ▶ 結果数を減らす

図は論文から引用



(c) Conventional Skyline 3D

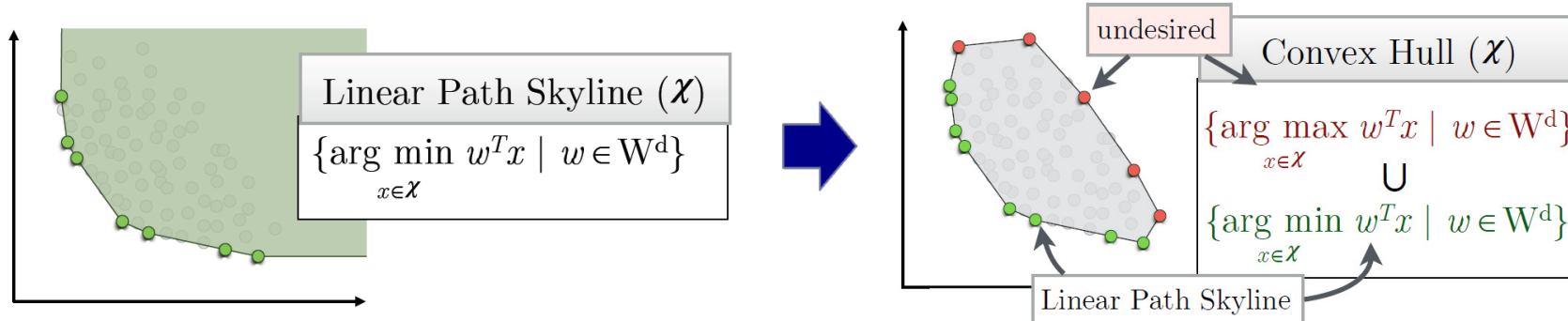


(d) Linear Skyline 3D

Linear Path Skylines in Multicriteria Networks

▶ Linear path skylineの計算

▶ 凸包 (Convex hull) の計算に転換



▶ 拡張ベクトル **INF** で計算を簡単化

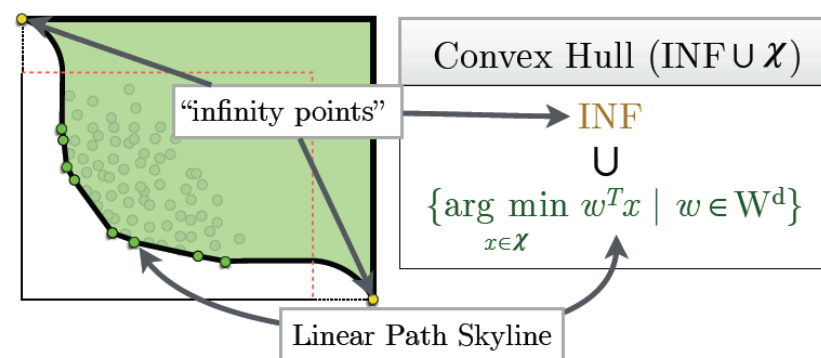
▶ **INF**: infinity points

▶ 最後の計算結果から **INF** を除く

▶ Incremental 計算アルゴリズム

▶ Beneath-and-Beyond 手法と A*-search 手法を用いる

計算しやすいため



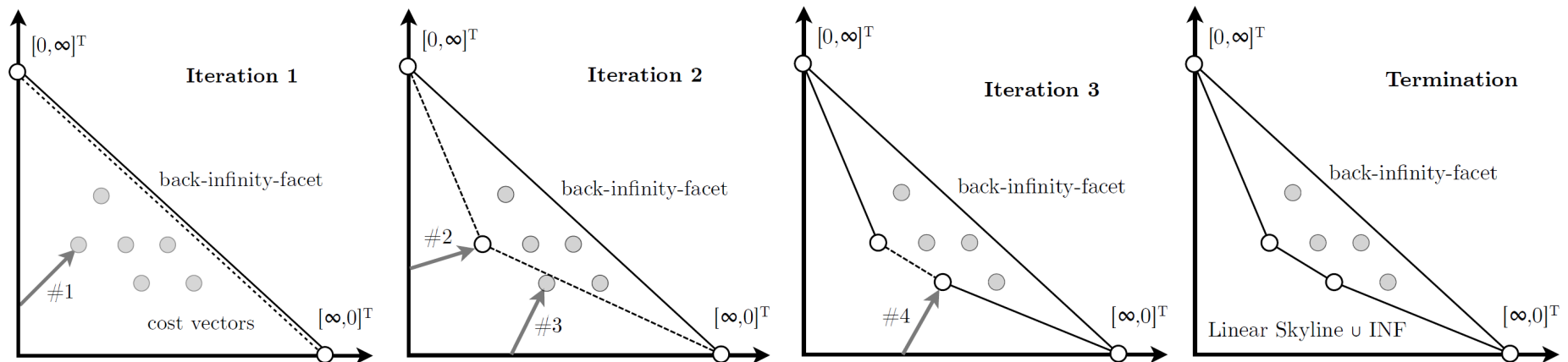
図はポスターから引用

Linear Path Skylines in Multicriteria Networks

▶ Linear path skylineの計算

▶ Incremental計算アルゴリズム

- ▶ Beneath-and-Beyond手法とA*-search手法を用いる凸包計算



▶ ϵ -linear path skyline

- ▶ ϵ -条件でさらに結果数を減らす

▶ Linear path skylineの部分集合

- ▶ 計算方法: 拡張したIncremental計算アルゴリズム

紹介内容

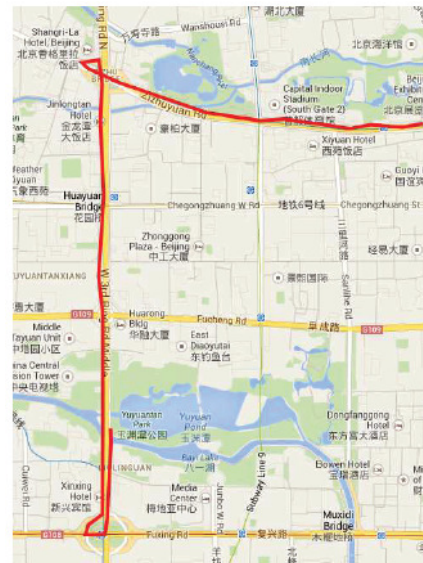
- ▶ **Session 9: Keyword and Top-K Processing**
 - ▶ 4: Interactive Top-k Spatial Keyword Queries
- ▶ **Session 10: Query Processing 2**
 - ▶ 3: Linear Path Skylines in Multicriteria Networks
- ▶ **Session 21: Trajectories**
 - ▶ 1: Making Sense of Trajectory Data: A Partition-and-Summarization Approach

Making Sense of Trajectory Data: A Partition-and-Summarization Approach

- ▶ Han Su (The University of Queensland) et al.
- ▶ 目的: 生の軌跡データをテキストで表現

- ▶ 利点

- ▶ 人間にとって納得がいく
- ▶ 出力テキストは管理しやすい
- ▶ フォーマットやスキーマが異なるデータはスタイルが同じなテキストに転換できる
- ▶ テキストマイニングアルゴリズムで重要なsemantic情報を発見



The car started from the Beijing Exhibition Center and moved along Zizhuyuan Street passing by the Beijing Shangri-la Hotel. Then it moved from the Beijing Shangri-la Hotel to the Yuyuantan Park along W 3rd Ring Road Middle highway. Along this road the speed of the car is 15 km/h which is 14km/h slower than usual.

(a) Trajectory aligned on a map (b) Summary of the trajectory

- ▶ Challenge

- ▶ 行動パターンの変化性により, 適切に軌跡データを分割(partition)
- ▶ 各分割に対する適切なテキスト要約(summarization)

Making Sense of Trajectory Data: A Partition-and-Summarization Approach

▶ System overview: **Partition-and-Summarization**

- ▶ 生軌跡データからsymbolic軌跡データに**転換** [SIGMOD'13]
 - ▶ Symbolic軌跡: **ランドマーク**と時間の系列
 - ▶ 軌跡の特徴量の抽出(6, 拡張できる)
 - **Routing** features: grade of road, road width, direction
 - **Moving** features: speed, number of stay points, number of U-turn
- ▶ **分割** (partition)
 - ▶ 各分割の始点と終点の**ランドマークの重要性**を最大化
 - ▶ 各分割内の**類似度**を最大化
 - ▶ 各分割間の**類似度**を最小化
- ▶ **要約** (summarization)
 - ▶ 各分割に対して, **最も重要な特徴量**を選択し, 前定義したテンプレートに入れて, テキストを生成

Making Sense of Trajectory Data: A Partition-and-Summarization Approach

▶ 分割 (partition)

- ▶ Conditional Random Field (CRF)を用いる
- ▶ 各分割の始点と終点の**ランドマークの重要性**を最大化
 - ▶ **ランドマークの重要性**: HITS-likeアルゴリズム [WWW'09] で推測
- ▶ 各分割内の**類似度**を最大化, 各分割間の**類似度**を最小化
 - ▶ 軌跡segmentの**類似度**: 特徴量ベクトルのcosine類似度
 - ▶ Dynamic programmingで最適な分割を求める
 - 分割の粒度が指定できる**最適なk-分割**を求める

▶ 要約 (summarization)

- ▶ 各分割に対して, **最も重要な特徴量**を選択し, 前定義したテンプレートに入れて, テキストを生成
 - ▶ **特徴量の重要さ**: **不規則率 (irregular rate)**
 - Routing features: 履歴的に最も人気なルートの特徴量との距離
 - Moving features: 履歴的に規則なパターンの特徴量との距離