THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

On Generalizing Collective Spatial Keyword Queries (Extended Abstract)

Harry Kai-Ho Chan[1], Cheng Long[2], Raymond Chi-Wing Wong[1]

[1]The Hong Kong University of Science and Technology, [2]Nanyang Technological University

Introduction

- **Collective spatial keyword query (CoSKQ):**
  Given: A query consisting a location and a set of keywords
  Find: A set *S* of objects which
  (1) covers the query keywords
  (2) ***cost(S)*** is minimized
- **Our contributions**:
  (1) A unified cost function - generalizes existing cost functions
  (2) A unified approach - solves CoSKQ problem in a unified way

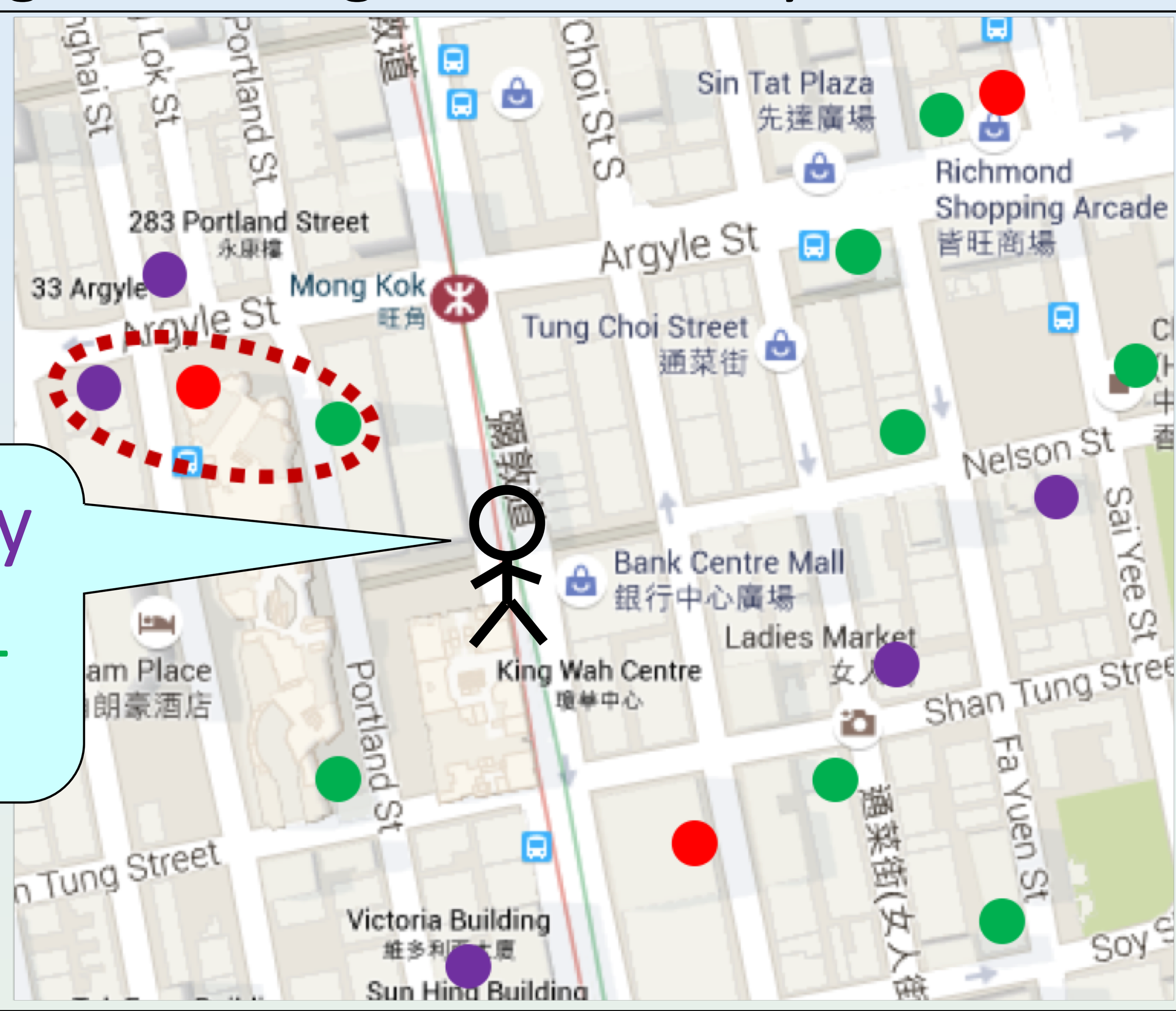


TABLE 1: $cost_{unified}$ under different parameter settings

	Parameter $\alpha \in (0,1]$	$\phi_1 \in \{1, \infty, -\infty\}$	$\phi_2 \in \{1, \infty\}$	$cost_{unified}(S\|\alpha, \phi_1, \phi_2)$	Existing/New
a	0.5*	1	1	$\sum_{o \in S} d(o,q) + \max_{o_1,o_2 \in S} d(o_1,o_2)$	$cost_{SumMax}$ [2]
b	0.5*	1	∞	$\max\{\sum_{o \in S} d(o,q), \max_{o_1,o_2 \in S} d(o_1,o_2)\}$	$cost_{SumMax2}$ (New)
c	0.5*	∞	1	$\max_{o \in S} d(o,q) + \max_{o_1,o_2 \in S} d(o_1,o_2)$	$cost_{MaxMax}$ [3], [17], [2]
d	0.5*	∞	∞	$\max\{\max_{o \in S} d(o,q), \max_{o_1,o_2 \in S} d(o_1,o_2)\}$	$cost_{MaxMax2}$ [17]
e	0.5*	$-\infty$	1	$\min_{o \in S} d(o,q) + \max_{o_1,o_2 \in S} d(o_1,o_2)$	$cost_{MinMax}$ [2]
f	0.5*	$-\infty$	∞	$\max\{\min_{o \in S} d(o,q), \max_{o_1,o_2 \in S} d(o_1,o_2)\}$	$cost_{MinMax2}$ (New)
g	1	1	-	$\sum_{o \in S} d(o,q)$	$cost_{Sum}$ [3], [2]
h	1	∞	-	$\max_{o \in S} d(o,q)$	$cost_{Max}$ (New)
i	1	$-\infty$	-	$\min_{o \in S} d(o,q)$	$cost_{Min}$ (New)

* Following the existing studies, $\alpha = 0.5$ is used to illustrate the case of $\alpha \in (0,1)$ for simplicity

A Unified Cost Function

$$cost_{unified}(S|\alpha, \phi_1, \phi_2)$$

$$= \left\{ \left[\alpha \cdot D_{q,o}(S|\phi_1)\right]^{\phi_2} + \left[(1-\alpha) \max_{o_1,o_2 \in S} d(o_1,o_2)\right]^{\phi_2} \right\}^{\frac{1}{\phi_2}}$$

where $\alpha \in (0,1]$, $\phi_1 \in \{1, \infty, -\infty\}$, $\phi_2 \in \{1, \infty\}$

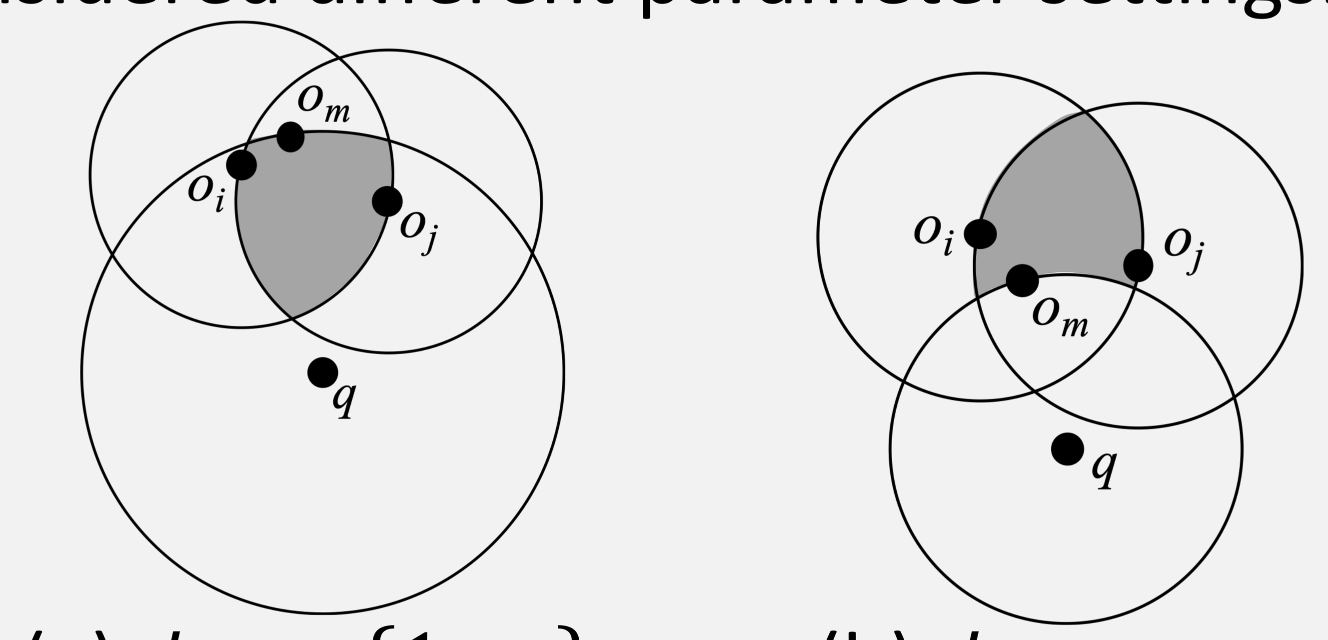
The *query-object distance component*

$$D_{q,o}(S|\phi_1) = \left[\sum_{o \in S} (d(o,q))^{\phi_1}\right]^{\frac{1}{\phi_1}} = \begin{cases} \sum_{o \in S} d(o,q), & \text{if } \phi_1 = 1 \\ \max_{o \in S} d(o,q), & \text{if } \phi_1 = \infty \\ \min_{o \in S} d(o,q), & \text{if } \phi_1 = -\infty \end{cases}$$

A Unified Approach

- An object set S is a **feasible set** if S covers all query keywords
- Two objects $o_i, o_j \in S$ are **object-object distance contributors** wrt S if $d(o_i, o_j)$ contribute to $\max_{o,o' \in S} d(o,o')$

Unified-E. An exact algorithm that iterates through the object-object distance contributors. It has pruning techniques that considered different parameter settings.



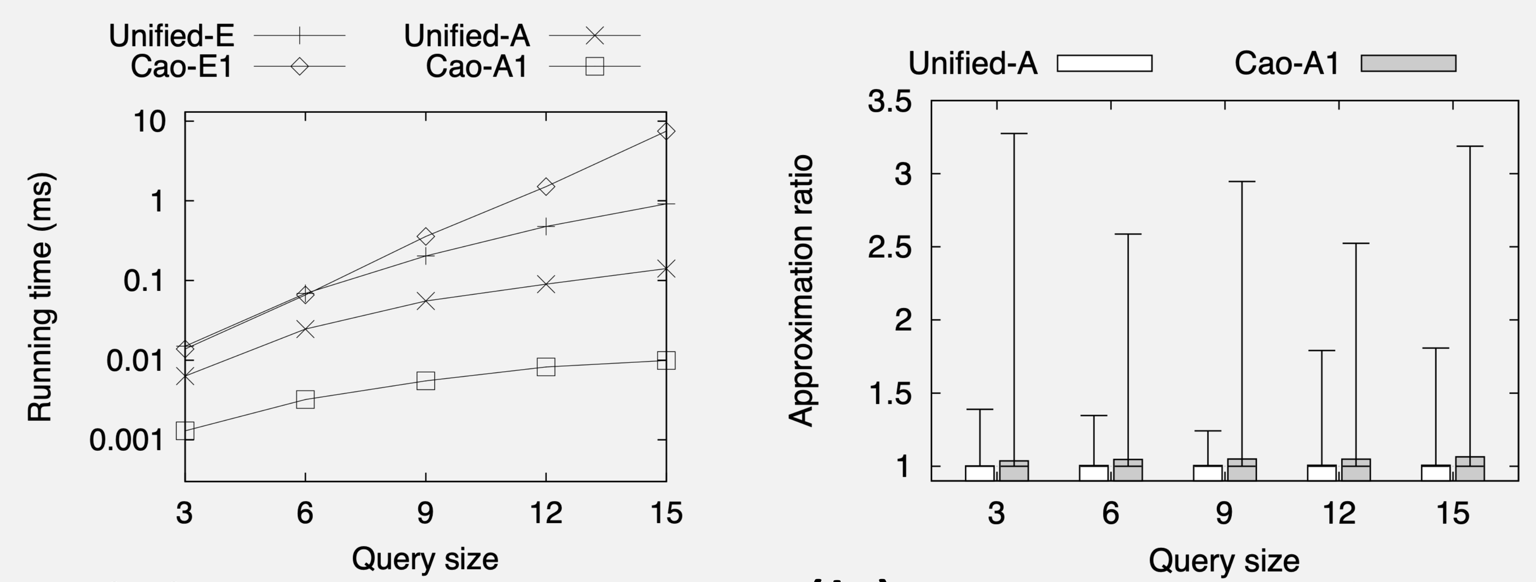
(a) $\phi_1 \in \{1, \infty\}$ (b) $\phi_1 = -\infty$

Fig. Search space in Unified-E

Unified-A. An approximate algorithm that replaces the step of constructing the best feasible set with an efficient step of constructing the (arbitrary) feasible set, and thus it enjoys better efficiency.

Experiment

Dataset	Hotel	GN	Web
# of objects	20,790	1,868,821	579,727
# of unique words	602	222,409	2,899,175
# of words	80,645	18,374,228	249,132,883



(a) Running time (b) Approximation ratio

Fig. Effect of query size on $cost_{MinMax2}$