

移动群智服务关键技术研究

Key Techniques in Mobile Crowd Service



开题人：吴 垚
导 师：陈 红
专 业：计算机应用技术
时 间：2016年9月9日



中国人民大学信息学学院

提纲

- 选题意义
- 研究现状
- 研究内容
- 研究创新
- 论文大纲

提纲

■ 选题意义

□ 研究现状

□ 研究内容

□ 研究创新

□ 论文大纲

■ 选题意义

◆ 背景

◆ 意义

◆ 挑战

Mobile location based service

Crowd power of crowdsourcing

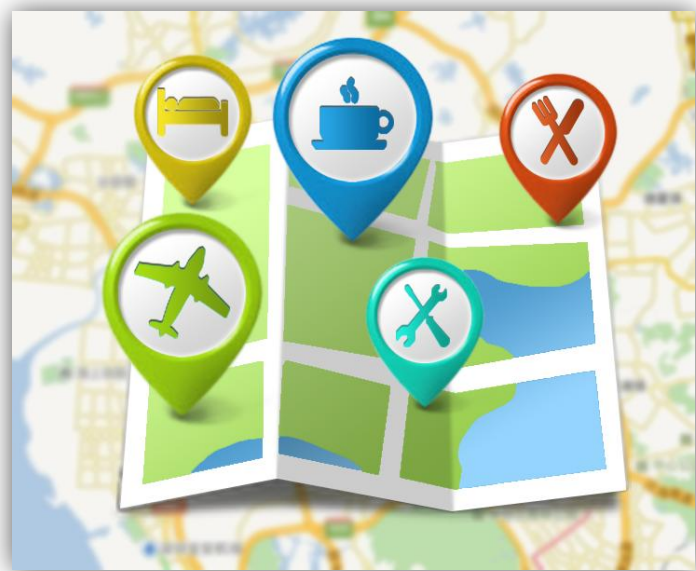
Service instant intelligent search

Mobile location based service

移动设备的普及(4.61B users from Statista)

基于位置的服务兴起

广泛应用于餐馆酒店电影院等的查找



Crowd

power of crowdsourcing

机器难以处理的问题

参与者的普适性

灵活经济覆盖广泛



Service

instant intelligent search

传统搜索引擎检索

新型共享服务搜索

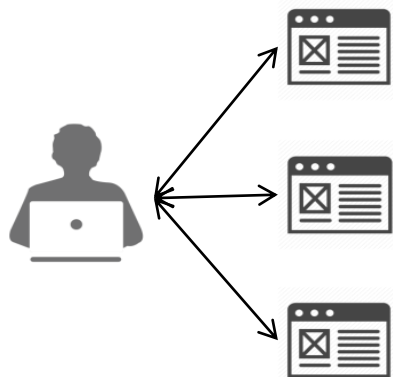


◆ 意义

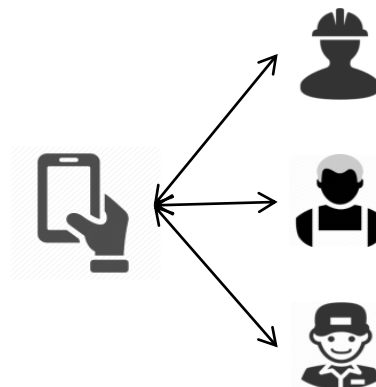


移动群智服务是众包技术与基于位置服务相结合，通过智能搜索技术实现移动个体的专业化服务。

- Best restaurants near me
- Users can label these picture
- Best selling books this month
- Get me a cab
- Fix my computer
- I need a house cleaner right now



Search Engine



Service Engine

◆ 意义

Magic

Magic is a service provided by Magic, Inc.



It can perform tasks that none of the others can.



空下来，格外精彩。



谷歌也要做移动服务：提供疏通马桶和开锁等。

◆ 意义

Mobile

静态性、单一性

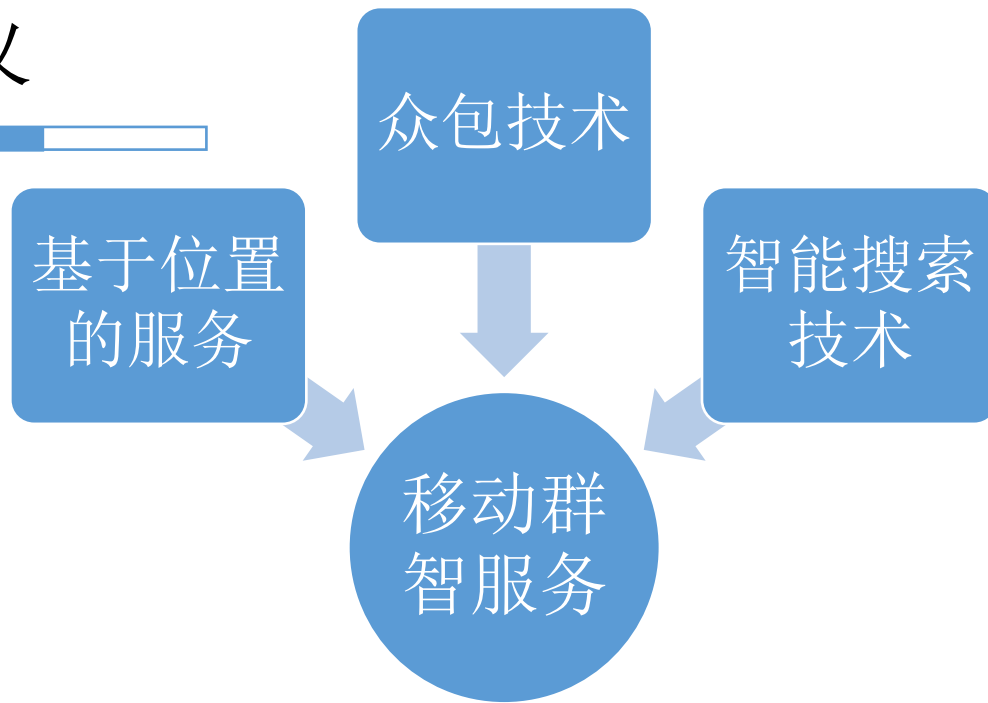
Crowd

专业性、一对一

Service

文本性、智能性

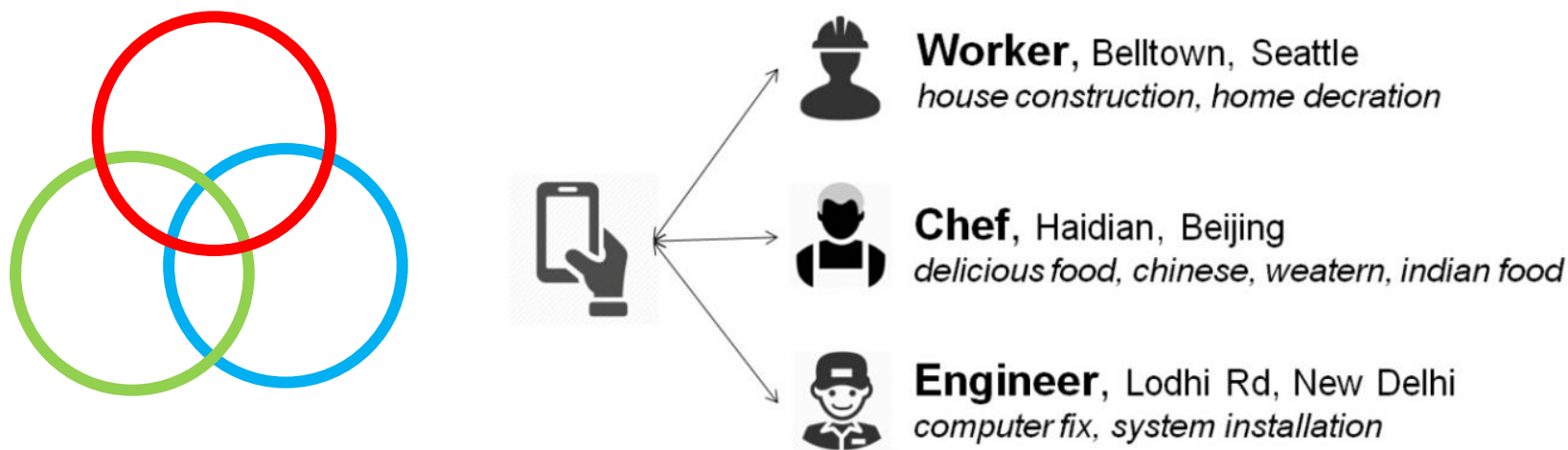
◆ 意义



- **Mobile**: It takes the advantages of mobile devices and it is offline and location based;
- **Crowd**: Any user can join in this platform to serve or be served, especially for the long-tail users;
- **Service**: It provides available and flexible service labor for users rather than cyber information and It guarantees the time-efficiency of service query key techniques;

◆ 意义

将众包技术与基于位置的服务相结合，通过智能搜索技术实现移动服务，从而达到设备计算能力与人类专业知识的相结合。



充分发挥人类智慧劳动资源；

解决大规模问题的有效方式；

实时的个体专业化服务搜索。

◆ 挑战

随着移动设备的不断普及移动群智服务有望超越传统的众包服务和网页搜索服务而变得更加普及。然而实现移动群智服务不是一件简单的事情。

- 与传统的基于位置服务不同，移动群智服务具有很强的移动性和不确定性；
- 相比于现有的众包服务，移动群智服务不简简单单是的一对多的任务发布完成，还包括一对一的个体服务任务，而且对专业性要求较高；
- 不同于传统的搜索引擎技术，移动群智服务中查询的对象是具有移动性的用户个体，同时更强到服务的智能性，即高效、准确、探索性，实现移动服务的智能化。



提纲



■ 选题意义

■ 研究现状

□ 研究内容

□ 研究创新

□ 论文大纲

■ 研究现状

- ◆ 众包技术(crowdsourcing)
- ◆ 基于位置的服务(LBS)
- ◆ 搜索技术(search engine)

◆ 众包

➤ 激励机制

➤ 任务分配

➤ 应用技术

◆ 众包➤激励机制

激励方式	评价
报酬支付激励	通过金钱货币的方式回报参与者的感知数据，是目前主要的激励方式。
娱乐游戏激励	指通过将游戏策略引入到群智感知系统中，利用游戏的娱乐性和吸引性来激励用户完成感知任务。
社交关系激励	指在参与者所处的现有的或者服务器平台构建的某种社会网络关系中，参与者为保持在社交关系中的归属感而受到激励。
虚拟积分激励	指参与者会从感知任务中得到虚拟积分做为回报，由虚拟积分转换成的真实货币，某种实物或者由其带来的虚拟回报感会激励用户参与感知任务。

Gao H, Liu C, Wang W, Zhao J, Song Z, Su X, Crowcroft J, Leung KK. A survey of incentive mechanisms for participatory sensing. Communications Surveys Tutorials, 2015,17(2):918–943.

Zhang X, Yang Z, Sun W, Liu Y, Tang S, Xing K, Mao X. Incentives for mobile crowd sensing: A survey. Communications Surveys Tutorials, 2015,18(1):54–67.

◆ 众包 ➤ 任务分配

分类	模型
一个任务分配给多个人	<p>1. Each task is done by k participants evenly.</p> $\max \left\{ \min_j \left\{ \sum_{i=1}^n x_{i,j} \right\} \right\}$ $\forall i: \sum_{j=1}^m c_{i,j} x_{i,j} \leq b_i$
一个人分配给多个任务	<p>2. Each participant performs w tasks evenly.</p> $\min \left\{ \max_i \left\{ \sum_{j=1}^m c_{i,j} x_{i,j} \right\} \right\}$ $\forall j: \sum_{i=1}^n x_{i,j} \geq k_j$

He S, Shin D H, Zhang J, et al. Toward optimal allocation of location dependent tasks in crowdsensing. IEEE INFOCOM, 2014: 745-753.

Ho C J, Vaughan J W. Online Task Assignment in Crowdsourcing Markets. AAAI. 2012, 12: 45-51.

◆ 众包 ➤ 应用技术

采用众包技术辅助计算机难以处理的问题。

Data Mining, Yael, SIDMOD2013	mining association rules from the crowd
Find Algorithm, UIUC, ICDE2014	Given a set of items, property P, and a number k, use humans to find k items that satisfy P
Table Match, NUS, ICDE2014	sometimes values of two web table columns may be disjoint and incompleteness in the column values
Entity Resolution, Stanford, VLDB2014	identify all records in a database that refer to the same underlying entity, and are therefore duplicates of each other
Database, UCB, SIGMOD2011	CrowdDB uses human input via crowdsourcing to process queries that neither database systems nor search engines can adequately answer
...	...

◆ 位置

- 隐私保护
- 空间文本对象查询

◆ 位置 ➤ 隐私保护

位置隐私保护主要是让用户的唯一识别信息和位置信息的关联性减弱或者消除，难点在于服务质量和隐私安全的权衡。

分类	评价
法规政策Regulation&Policies	依靠法律强制和道德约束力
模糊泛化Obfuscation	不准确，不精确，模糊化 无需中间件且轻量化，但难以解决泛化域内单个用户的推测攻击
基于匿名Anonymity-Based	k-Anonymity (l-diversity, t-closeness, p-sensitive)、Mix Zone、Using Dummies、Path Confusion
加密方法	安全性强，但是计算量大

Liu Y, Yang Z, Wang X, et al. Location, localization, and localizability[J]. Journal of Computer Science and Technology, 2010, 25(2): 274-297.

曾菊儒, 陈红, 彭辉, 等. 参与式感知隐私保护技术[J]. 计算机学报, 2016, 39(3): 595-614.

◆ 位置 ➤ 空间文本对象查询

空间文本对象查询是对具有位置信息和文本信息的对象进行索引和查询。

主要分类		经典技术
空间关键词查询	基于R-tree的索引查询	IR-tree是将R-tree的每个节点赋予一个文本内容的摘要，变种如DIR-tree, CIR-tree, CDIR-tree和WIRtree等。其他基于R-tree的索引如KR*-tree, IR2 - tree 和SKI。
	基于Quad-tree的索引查询	基于空间填充曲线的SFC-QUAD，基于倒排索引的IQ-tree，线性的ILQ-tree等。
移动对象/查询	移动查询	将移动查询模型成 Safe Region 进行查找。
	移动对象	根据对象移动模型，进行划分，如基于对象划分的 TPR-tree 和基于空间划分的 B^x-tree 。

L. Chen, G. Cong, C. S. Jensen, and D. Wu. Spatial keyword query processing: An experimental evaluation. PVLDB, 6(3):217–228, 2013.

S. Chen, C. S. Jensen, and D. Lin. A benchmark for evaluating moving object indexes. PVLDB, 1(2):1574–1585, 2008.

◆ 搜索

➤ 表示模型

~~➤ 索引技术~~

➤ 网页排序

◆ 搜索 ➤ 表示模型

表示（排序）模型主要是定义文档、查询和相关关系的表示方法。

Model	Perspective
Boolean model	Set theory and Boolean algebra
Vector space model	Vector and linear algebra
Probabilistic model	Probabilistic
Bayesian network model	
Language model	
Page rank model	Link structure
...	...

◆ 搜索 ➤ 网页排序

基于表示模型的排序算法。

排序算法	描述
Distance Rank	main goal is computed on the basis of the shortest distance between two pages and ranked. Calculating the Minimum Average Distance Between two pages and more pages.
HITS	two different forms of Web pages called hubs and authorities. 'N' highly relevant pages are computed and find values on the fly.
PageRank	if a page contains important links towards it then the links of this page towards the other page are also to be considered as important pages. Computes values at index time and results are sorted on the priority of pages.
EigenRumor	Use the adjacency matrix which is constructed from agent to object link not page to page.
...	...

提纲



■ 选题意义

■ 研究现状

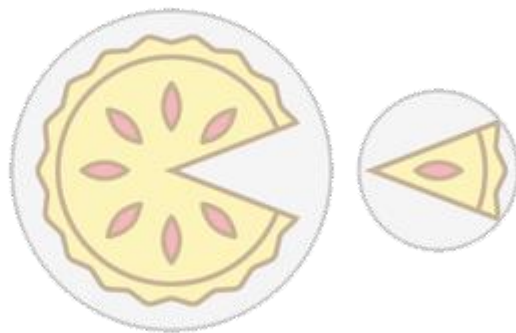
■ 研究内容

□ 研究创新

□ 论文大纲

■ 研究内容

- ◆ 体系结构(Platform)
- ◆ 激励机制(Incentive)
- ◆ 隐私保护(Privacy)
- ◆ 智能搜索(3E)



$P^2I E^3$

◆ 体系结构

现有工作不足：

- 在基于位置服务中，查询对象是静态的，不能满足移动群智服务参与者的不确定性和可移动性；
- 现有的众包服务，主要是一对多的任务发布匹配完成，在移动群智服务中还包括一对一位置关联的个体服务，而且对专业性有要求；
- 不同于传统的搜索引擎技术，移动群智服务中查询的对象是具有移动性的用户个体而不是网页，同时更强到服务的智能性，即高效、准确、探索性，实现移动服务的智能化。

◆ 体系结构

Components:

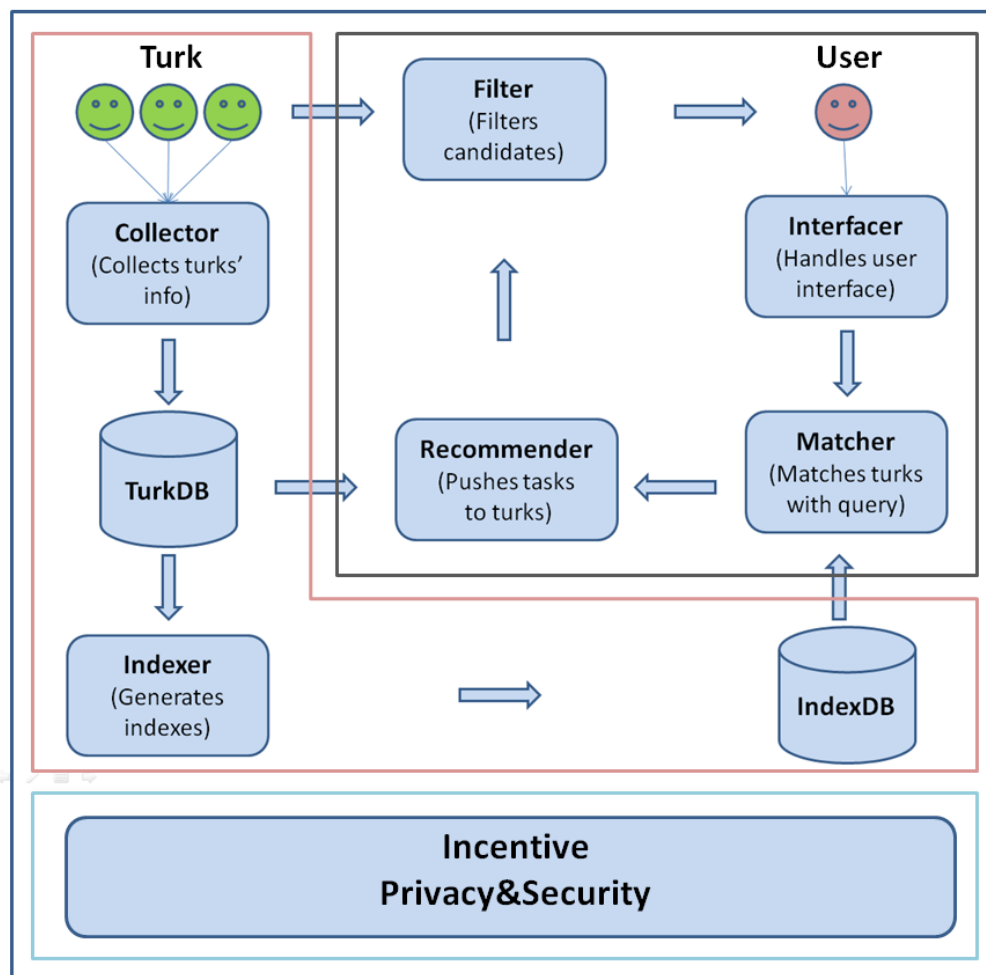
Turk Collector
Service Indexer
Query Matcher
Task Recommender
Candidate Filter
User Interfacer

Key Techs:

Indexing
Matching
Recommendation

Goals:

Efficiency
Effectiveness
Exploration



◆ 体系结构

Service query (a)

Service notification (b)

Service message (c)

Profile panel (d)

将众包技术与基于位置的服务相结合，通过搜索技术实现移动群智服务，从而达到计算设备与人类智慧的相结合。



(a) Service query



(b) Service notification



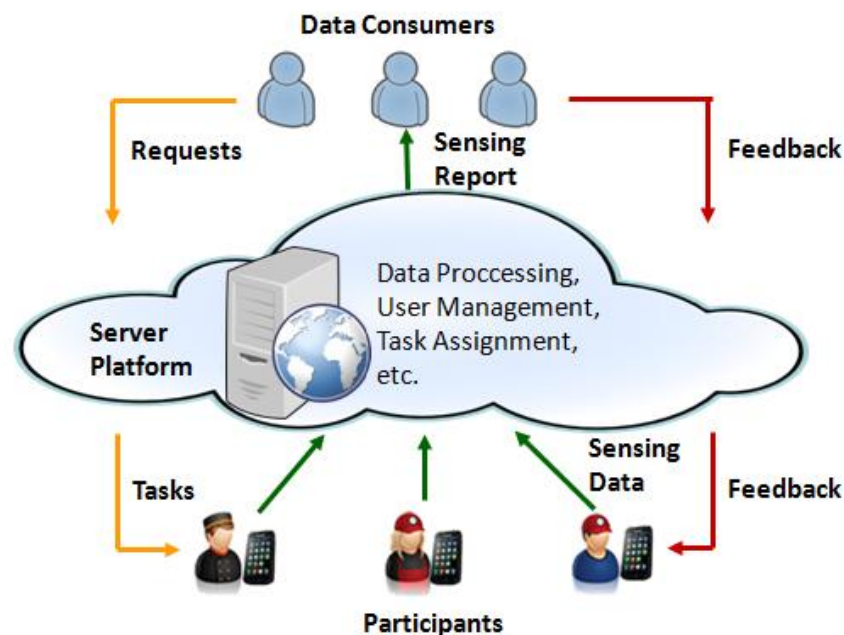
(c) Service message



(d) Profile panel

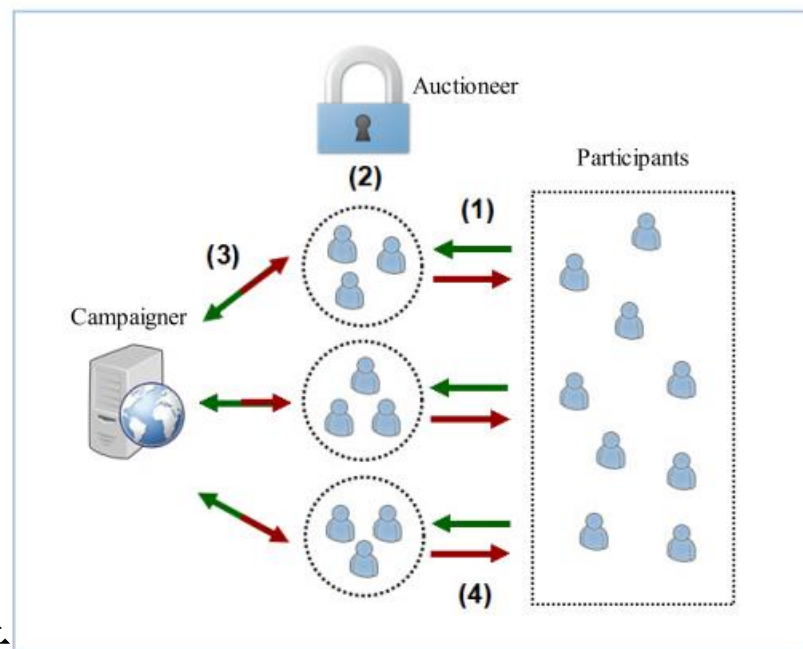
◆ 激励机制

- 受限于感知任务参与者数量不足和提供数据质量不高，群智服务的发展受到了严重的影响；
- 现有的激励机制主要是集中在以拍卖模型为基础，研究参与者个体与服务器之间的博弈；
- 如何更好的平衡服务器与参与者之间的效益、提高参与率、保证数据质量和参与者的可信性值得研究。



◆ 激励机制

- 提出基于纳什讨价还价模型(Nash Bargain Theory)的将参与者划分为群(Crowd)的激励机制；
- 提出K-least算法，具有K-anonymity的聚类算法，将参与者划分为群；
- 提出均匀质量评估模型和惩罚性报酬分配机制；
- 该方法能够提高参与率和参与者位置隐私，同时保证数据质量和参与者的可信性。



◆ 隐私保护

- 现有的位置隐私保护主要集中在客户端服务器的不可信模型；
- 用户之间的隐私泄露问题没有得到相应的关注；
- 如何保证用户隐私同时不降低服务质量是隐私保护面临的难题。



◆ 隐私保护

- 提出了基于好友关系强度的概率隐私保护模型；
- 文章在TF-IDF的基础上，结合信息论熵的理论，采用熵权法来衡量好友之间的双向关系强度；
- 根据好友关系强度，建立了个性化的灵活的概率位置隐私保护模型；
- 引入基于柯西分布和拉普拉斯分布的泛化(Cloaking)模型，很好的平衡了位置信息的可用性(Utility)和隐私性(Privacy)。



$$Pr[|\Delta| \leq \ell] \geq \xi$$

◆ 群智搜索

随着移动设备的不断普及和共享经济的不断发展，移动群智服务逐渐成为一种新的信息服务方式。如何从大量的参与者中，找到合适的对象完成相应的服务任务是移动群智服务需要解决的问题。因此，考虑到：

1. 参与者的不确定移动性
2. 对象文本信息的稀疏性
3. 服务任务的上下文信息



如何提高移动群智感知搜索效率(Efficiency)、效果(Effective)和效益(Exploration)，是本研究的核心问题。

◆ 群智搜索

- **Efficiency**: The results of service query should be returned immediately, which needs efficient indexing technique for localized frequent moving spatial-textual objects;
- **Effectiveness**: The results should return turks around just as the users query, which needs accurate spatial- textual matching techniques;
- **Exploration**: The results should also return potential turks that are qualified for the query, which needs context-aware recommendation.

◆ 群智搜索

不确定移动对象的搜索

(Mobile Spatial Textual Objects Search)

基于远距监督的对象匹配

(Objects Matching with Distance Supervision)

基于众包排序的任务推荐

(Crowd Rank for Task Recommendation)

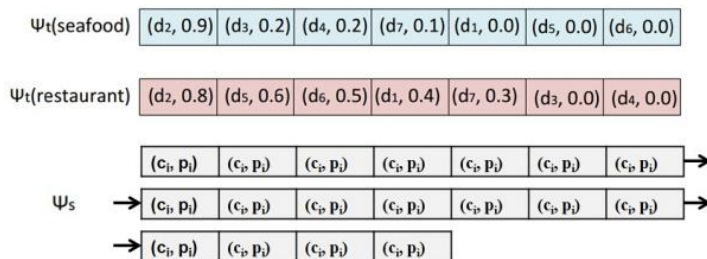
◆ 群智搜索

Efficiency

- 现有的基于位置的服务主要是通过空间关键词技术来搜索查找对象，即给定一组兴趣点(POIs)集合或者兴趣域(ROIs)集合和一个包括位置和关键词的查询，返回相关的兴趣点或者兴趣域；
- 现有的研究主要集中在查询静态对象上，如餐馆、酒店、景点等。很少研究对可移动性物体的查询，诸如查询附近可以修电脑的人；
- 现有的方法没有能很好的适应这个情况的，这需要新的索引技术和查询处理算法。

◆ 群智搜索

- 提出了不确定可移动对象的查询问题，基于概率实例的移动模型和带有时间渐进的空间文本相似度度量模型；
- 提出了基于有序列表的线性四分树和倒排文件的混合索引技术BIGtree；
- 并在此基础上提出改进的阈值算法(TA)，通过延迟调整和优先终止来提高查询处理速度；
- 实验表明，我们提出的索引技术和优化查询技术在物理存储效率和查询效率上都相对于已有技术有所提高。



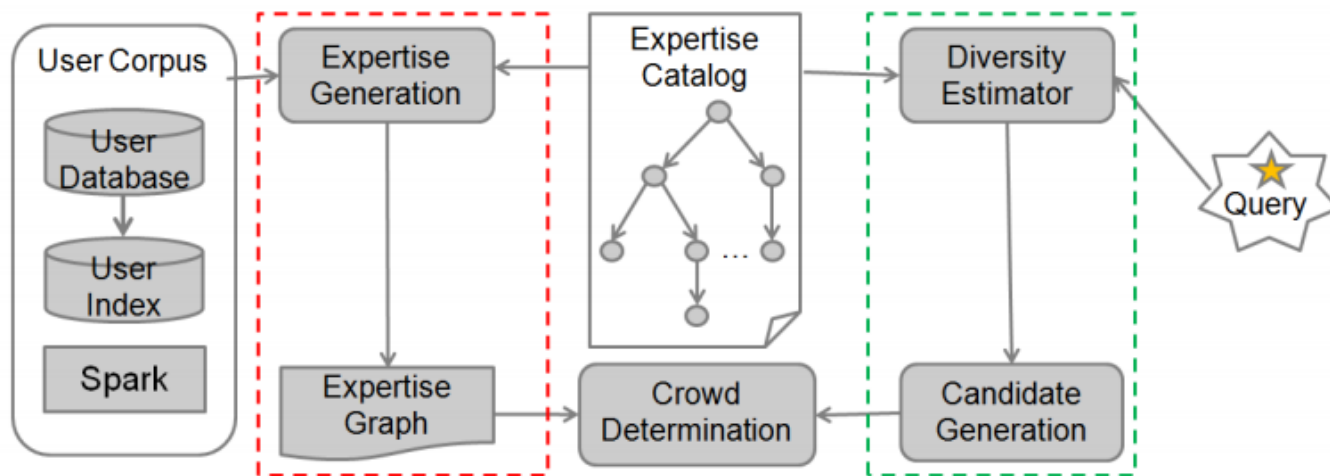
◆ 群智搜索

Effectiveness

- 现有的移动众包服务受限的原因主要有以下三个方面：参与者人数不够、付费服务提供、缺少内在的激励；
- 现有众包服务中参与者专业差别没有体现，在搜索服务质量上，相似性是一个重要的衡量标准，同时多样性也不容忽视；
- 现有的话题专家的发现主要是基于文本内容、网络特征或者社会活动。这些研究的不足在于：往往集中在一个特定领域，如学术领域，从算法而使得适应性变差；基于关键词的匹配往往存在失误的情况；基于链接的方法往往存在话题漂移问题。

◆ 群智搜索

- 设计并实现了从现有社交网络中发现潜在用户的框架，并具有高效和可扩展性；
- 通过远距监督和树形匹配的模式，用层次性知识表达模式进行相似性匹配和多样性度量；
- 提出了针对多样性专家群组发现NP问题的贪婪算法。



◆ 群智搜索

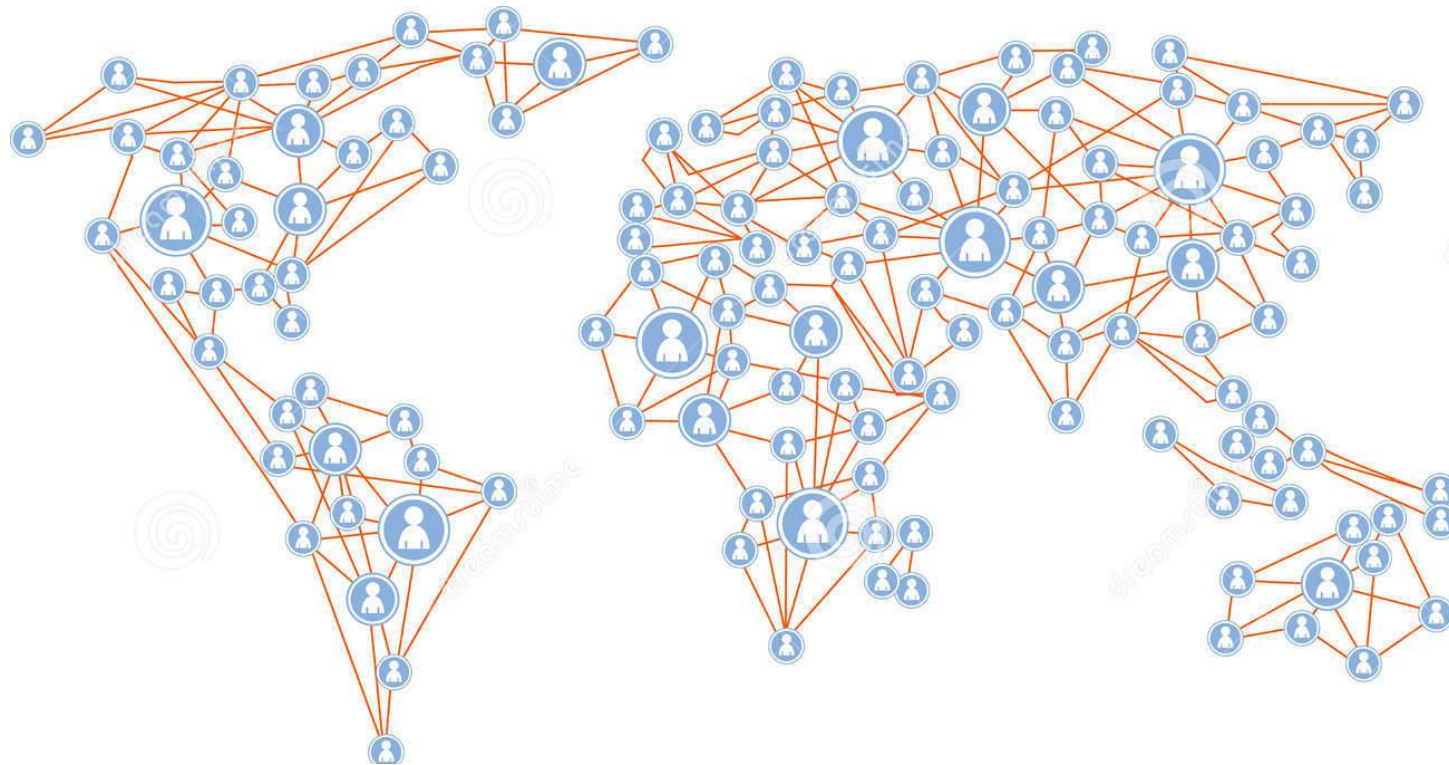
Exploration

移动群智服务往往具有文本、位置、时间等多属性信息，为基于上下文的用户任务匹配提供了基础。我们将结合算法和众包的模式，实现用户任务的推荐。有些复杂任务难以单单的用关键词表示，因此，对于计算推荐度低的用户和任务，我们提出众包推荐，来提高准确度。



Turk	User	Location	Skill	Time	Rating
T1	U2	Home	Delivery	Morning	4
T2	U2	Home	Fixing	Morning	5
T3	U2	Office	Delivery	Afternoon	4
T4	U2	Home	Delivery	Morning	?

◆ 群智搜索



提纲



■ 选题意义

■ 研究现状

■ 研究内容

■ 研究创新

□ 论文大纲

■ 研究创新

移动群智服务通用服务平台

- 移动微服务可扩展智能平台
- 搜索与推荐技术相结合

移动群智服务群组博弈激励机制

- 基于纳什讨价还价的群组博弈模型
- 具有隐私保护的K-least聚类方法
- 均匀质量评估模型和惩罚性报酬分配机制

移动群智服务社交好友位置隐私保护

- 基于好友关系强度的概率隐私保护模型
- 基于TF-IDF的信息论熵权法来衡量好友之间的双向关系强度
- 基于柯西分布和拉普拉斯分布的泛化模型

■ 研究创新

可移动空间文本对象搜索

- 基于概率实例和时间渐进移动模型空间文本相似度度量模型
- 基于有序列表的线型四分树和倒排文件的混合索引技术**BIGtree**
- 提出改进的阈值算法，通过延迟调整和优先终止来提高查询速度

潜在智慧群体发现

- 通过远距监督和树形匹配的模式，进行相似性匹配和多样性度量
- 提出了针对多样性专家群组发现**NP**问题的贪婪算法

群智排序推荐技术

- 多维上下文信息推荐模型
- 最难推荐任务规约
- 基于同行众包进行排序推荐

提纲



■ 选题意义

■ 研究现状

■ 研究内容

■ 研究创新

■ 论文大纲

■ 论文大纲

Mobile	Crowd	Service
静态性、单一性	专业性、一对一	文本性、智能性
PFP	MagiCrowd&Cents	HelPal
	AXE	
PIN		
	ROCK	

	已发表	工作中
	已投稿	仅计划

■ 论文大纲

Publish me and Protect me: Personalized and Flexible Location Privacy Protection in Social Mobile Network[C], Wu Yao, Peng Hui, Zhang Xiaoying, Chen Hong, Li Cuilping, *IWQoS 2015*, Portland, USA. **Accept.**

MagiCrowd: A Crowd based Incentive for Location-aware Crowd Sensing[C], Wu Yao, Wu Yuncheng, Peng Hui, Chen Hong, Li Cuilping, *WCNC 2016*, Doha, Qatar. **Accept.**

A Survey on Incentive Mechanisms for Crowd Sensing[J], Yao Wu, Juru Zeng, Hui Peng, Hong Chen, Cuiping Li, *Journal of Software*, 2016. **Accept.**

HelPal: A General System for Diversified Mobile Crowds[C], Wu Yao, Ziyi Xiong, Tianzhen Wu, Yuncheng Wu, Hong Chen, Cuiping Li, Xiaoying Zhang. *NIPS2017*, **Submit.**

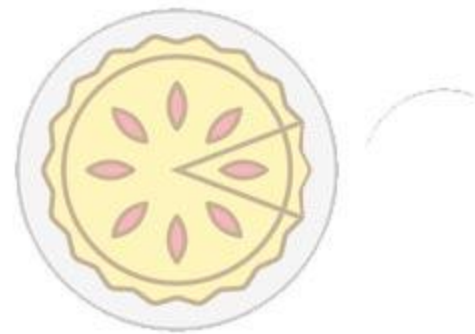
AXE: Mobile Spatial Textual Objects Search[C], Yao Wu, Yuncheng Wu, Juru Zeng, Wenjuan Liang, Hong Chen, Cuiping Li, Xiaoying Zhang. *EDBT2017*, **Submit.**

PIN: Potential Local Diverse Expert Crowd Finding From Billion Grassroots by Distant Supervision[C], Yao Wu, Wenjuan Liang, Hong Chen, Cuiping Li. **In work.**

ROCK: Crowd Rank for Task Recommendation, **Envisioned.**

■ 论文大纲

- 第一章 研究背景介绍
- 第二章 移动群智服务体系结构
- 第三章 移动群智服务激励机制
- 第四章 移动群智服务隐私保护
- 第五章 移动群智服务智能搜索
- 第六章 总结与展望



■ 论文大纲



■ 论文大纲

2016.12 完成PIN

2017.06 完成ROCK

2017.12

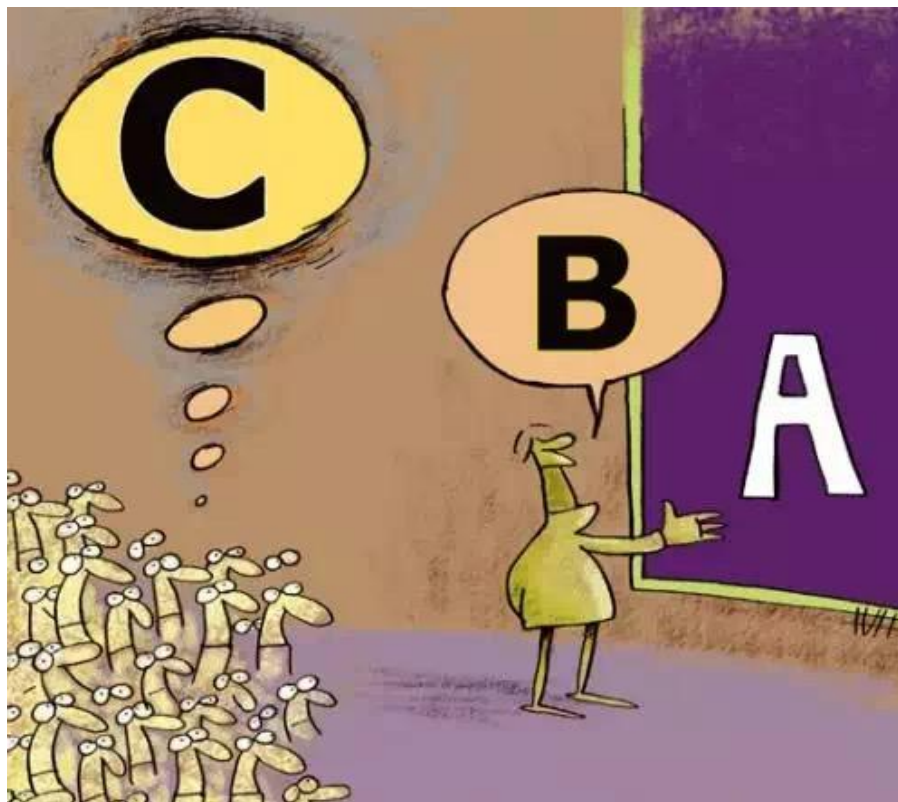
毕业论文 2018.06

完成所有



提纲

- 选题意义
- 研究现状
- 研究内容
- 研究创新
- 论文大纲





Thanks! Question?

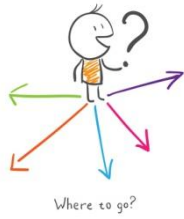
A life hit the biggest lucky,
in his way of life midway,
namely in the prime of life in he found his mission.

Sternstunden der Menschheit, Stefan Zweig





Q1



WSN →

MCS →

IOT



DBIR



Q2



Zaarly

Magic

Magic is a service provided by Magic, Inc.

人工智能

共享经济

搜索引擎

应用前景



It can perform tasks that none of the others can.



空下来，格外精彩。



谷歌也要做上门服务：疏通马桶和开锁。

Gigwalk



CrowdFlower

Q3

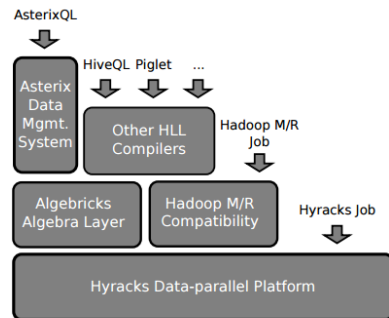


Figure 1: ASTERIX layers and entry points

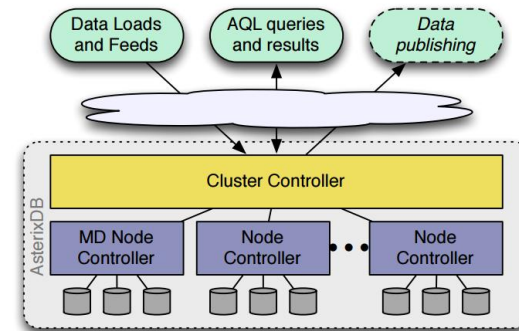


Figure 1: AsterixDB system overview

MobiBAD OptiX

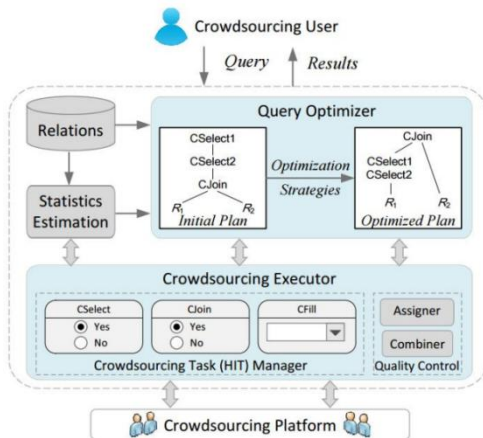


Fig. 1. Workflow of crowdsourced query.

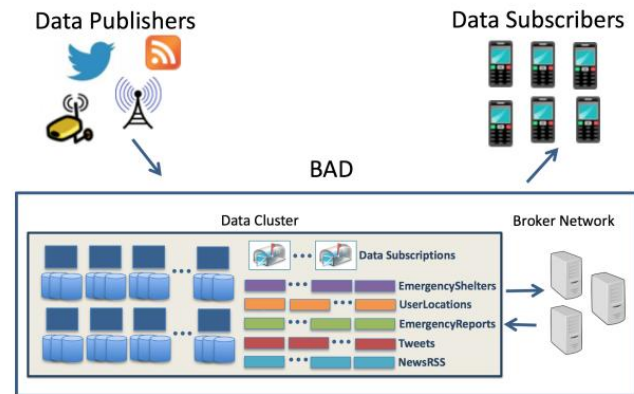


Figure 1: Big Active Data – System Overview.

Qx