ARTS1422 Data Visualization

Lecture 2

A Life Cycle of User-Centric Visualization Design

Quan Li

Spring 2024
2024. 02.29

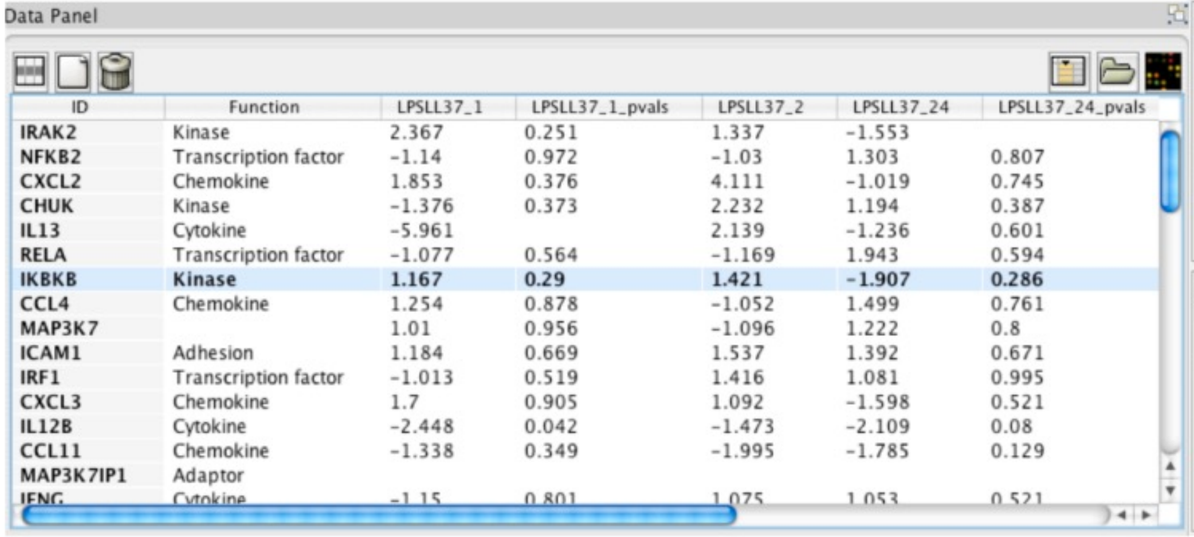
上海科技大学
ShanghaiTech University

Definition of Visualization

- **Computer-based** visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively

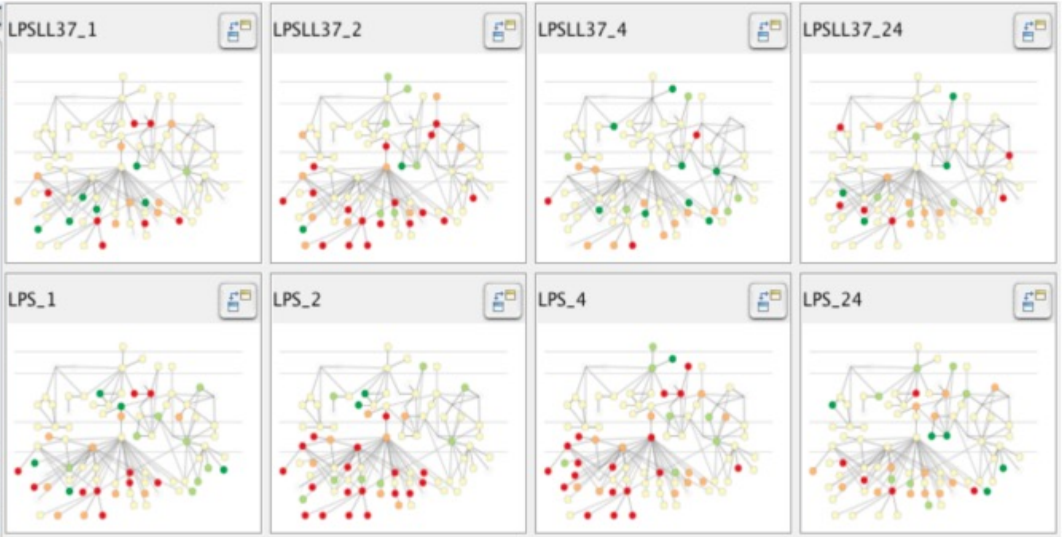
Why Having Human-in-the-loop?

- Visualization is suitable when there is a need to **augment human capabilities** rather than replace people with computational decision-making methods
- Beyond human patience
- Scale to large datasets
- Support interactivity



Data Panel

ID	Function	LPSLL37_1	LPSLL37_1_pvals	LPSLL37_2	LPSLL37_24	LPSLL37_24_pvals
IRAK2	Kinase	2.367	0.251	1.337	-1.553	
NFKB2	Transcription factor	-1.14	0.972	-1.03	1.303	0.807
CXCL2	Chemokine	1.853	0.376	4.111	-1.019	0.745
CHUK	Kinase	-1.376	0.373	2.232	1.194	0.387
IL13	Cytokine	-5.961		2.139	-1.236	0.601
RELA	Transcription factor	-1.077	0.564	-1.169	1.943	0.594
IKBKB	Kinase	1.167	0.29	1.421	-1.907	0.286
CCL4	Chemokine	1.254	0.878	-1.052	1.499	0.761
MAP3K7		1.01	0.956	-1.096	1.222	0.8
ICAM1	Adhesion	1.184	0.669	1.537	1.392	0.671
IRF1	Transcription factor	-1.013	0.519	1.416	1.081	0.995
CXCL3	Chemokine	1.7	0.905	1.092	-1.598	0.521
IL12B	Cytokine	-2.448	0.042	-1.473	-2.109	0.08
CCL11	Chemokine	-1.338	0.349	-1.995	-1.785	0.129
MAP3K7IP1	Adaptor					
IFNG	Cytokine	-1.15	0.801	1.075	1.053	0.521



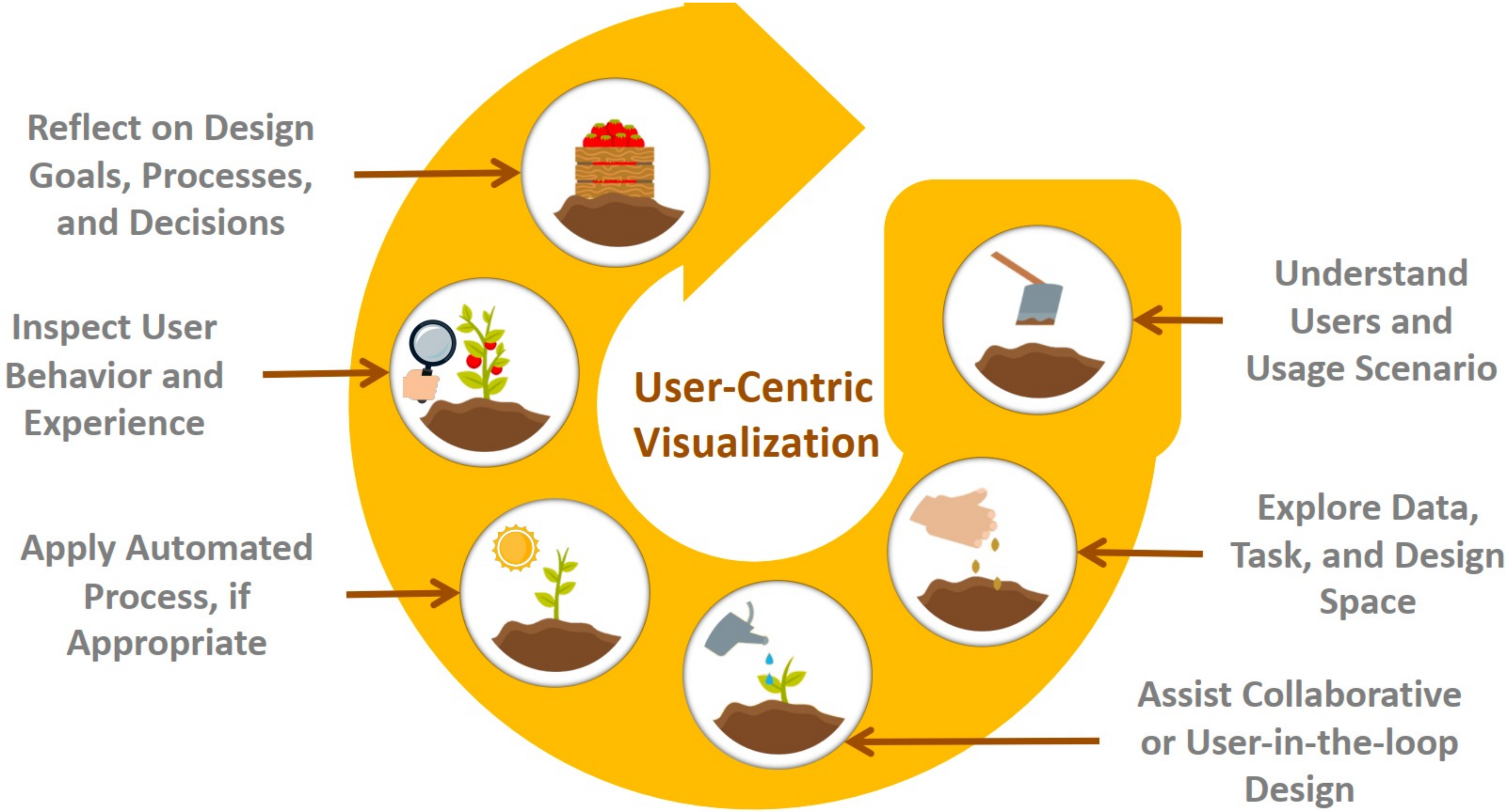
When to Use Visualization?

A good visualization enables **users** to complete **tasks** effectively on the **data**

Cognition → Perception (→ Cognition)

When NOT to Use Visualization

Do not need vis when fully automatic solution exists and is trusted


Reflect on Design Goals, Processes, and Decisions
Inspect User Behavior and Experience
Apply Automated Process, if Appropriate
User-Centric Visualization
Understand Users and Usage Scenario
Explore Data, Task, and Design Space
Assist Collaborative or User-in-the-loop Design

Area 1: Theoretical & Empirical

This area focuses on theoretical and empirical research topics that aim to establish the foundation of VIS as a scientific subject.

Theoretical & Empirical →

Area 2: Applications

This area encompasses all forms of application-focused research.

Applications →

Area 3: Systems & Rendering

This area focuses on the themes of building systems, algorithms for rendering, and alternate input and output modalities.

Systems & Rendering →

Area 4: Representations & Interaction

This area focuses on the design of visual representations and interaction techniques for different types of data, users, and visualization tasks.

Representations & Interaction →

Area 5: Data Transformations

This area focuses on the algorithms and techniques that transform data from one form to another to enable effective and efficient visual mapping as required by the intended visual representations.

Data Transformations →

Area 6: Analytics & Decisions

This area focuses on the design and optimization of integrated workflows for visual data analysis, knowledge discovery, decision support, machine learning, and other data intelligence tasks.

Analytics & Decisions →

Understanding Users & Usage Scenarios

- Insights from literature
- Insights from interactions with target users
  - Observational study
  - Surveys
  - Interviews
- Possible Outcome
  - User characteristics
  - User practices
  - User needs

Image Source: https://www.stratascratch.com/blog/how-to-become-a-data-scientist-from-scratch/

Understanding Users

Process	Role	Role Description	In prior Studies	Level of Expertise: Statistics	Computer Science	Domain Knowledge	Human Centered Design
Preparation	Data Steward	Domain expert responsible for governing access and use of data	Data Broker [87]; Data Owners [84]	little to none	little to none	proficient	little to none
	Data Shaper	Developer responsible for supporting the curation and preparing data for analysis	Data Shaper [38]; Data Preper [38];	working	knowledgeable	knowledgeable	little to none
Deployment & Engineering	Data Engineer	Engineer proficient in developing Data Science technologies, including data preparation and analysis pipelines	Platform Builder [38]; Data Developer [27]; Hacker [35]; Scripter [34]; Engineer [57]; BI engineer [84]	working	proficient	knowledgeable	little to none
	ML / AI Engineer	Engineer proficient in developing and deploying machine learning / artificial intelligence methods to support data science processes	Hacker [33]; Modeling Specialists [38]	proficient	knowledgeable	working	little to none
Analysis	Generalist	Multidisciplinary individual focused solely on data science	Polymath [39]; Data Creative [27]	knowledgeable	knowledgeable	knowledgeable	working
	Research Scientist	A domain expert involved in research typically with technical expertise in 'Data Science' technologies	Data Researcher [27, 57]; X-informatician [12]	knowledgeable	knowledgeable	proficient	working
	Technical Analyst	A technical individual from whom data science is not core to their job but occurs only at the margins of other work	Data Analyst [4, 38, 39, 57, 84]; Application User [35]; Business Analyst [36, 57]; Data Business Person [27]; Analysis Team Members [84]	working	working	proficient	little to none
	Moonlighter	Non-technical individual tasked to perform data science duties, either voluntarily or through necessity	Moonlighter [39]	little to none	little to none	proficient	little to none
Communication	Evangelist	Manager, team leader, or analyst tasked with disseminating findings from data science work	Data Evangelist [39]; Communicator [87]; Insight Actor [39]	working	working	knowledgeable	little to none

Table 2. Summary of data science roles and an illustration of their skill sets. Data science skills were classified along four axes: statistics, computer science, domain knowledge, and human centred design. We use a color gradient to illustrate a level of expertise: proficient ■; knowledgeable ■; working ■; and little to none.

Crisan, A., Fiore-Gartland, B. and Tory, M., 2020. Passing the Data Baton: A Retrospective Analysis on Data Science Work and Workers. IEEE Transactions on Visualization and Computer Graphics, 27(2), pp.1860-1870.

Explore Data, Task, and Design Space (1)

- How to draw it: **Visual encoding** idiom
  - Many possibilities for how to create
- How to manipulate it: **interaction** idiom
  - Even more possibilities
  - Make single idiom dynamic
  - Link multiple idioms together through interaction
- Tasks serve as constraint on design (as do data)
  - Idioms do not serve all tasks equally
  - Challenge: recast tasks from domain-specific vocabulary to abstract forms
- Most possibilities ineffective
  - Validation is necessary, but risky
  - Increases chance of finding good solutions if you understand full space of possibilities

Explore Data, Task, and Design Space (1)

- Data Sources
  - Tabular, corpora, notebook…
- Data Characteristics
  - Number, text, image, audio, video…
  - Temporal, spatial, categorical, numerical…
  - Structured, unstructured…
  - Archived, real-time…
- Data Facts
  - Value, distribution, difference, extreme…

- Low-level Analytic Tasks
  - Retrieve Value
  - Filter
  - Compute Derived Value
  - Find Extreme
  - Sort
  - Determine Range
  - Characterize Distribution
  - Find Anomalies
  - Cluster
  - Correlate

Chen, Y., Yang, J. and Ribarsky, W., 2009, April. Toward effective insight management in visual analytics systems. In 2009 IEEE Pacific Visualization Symposium (pp. 49-56). IEEE.

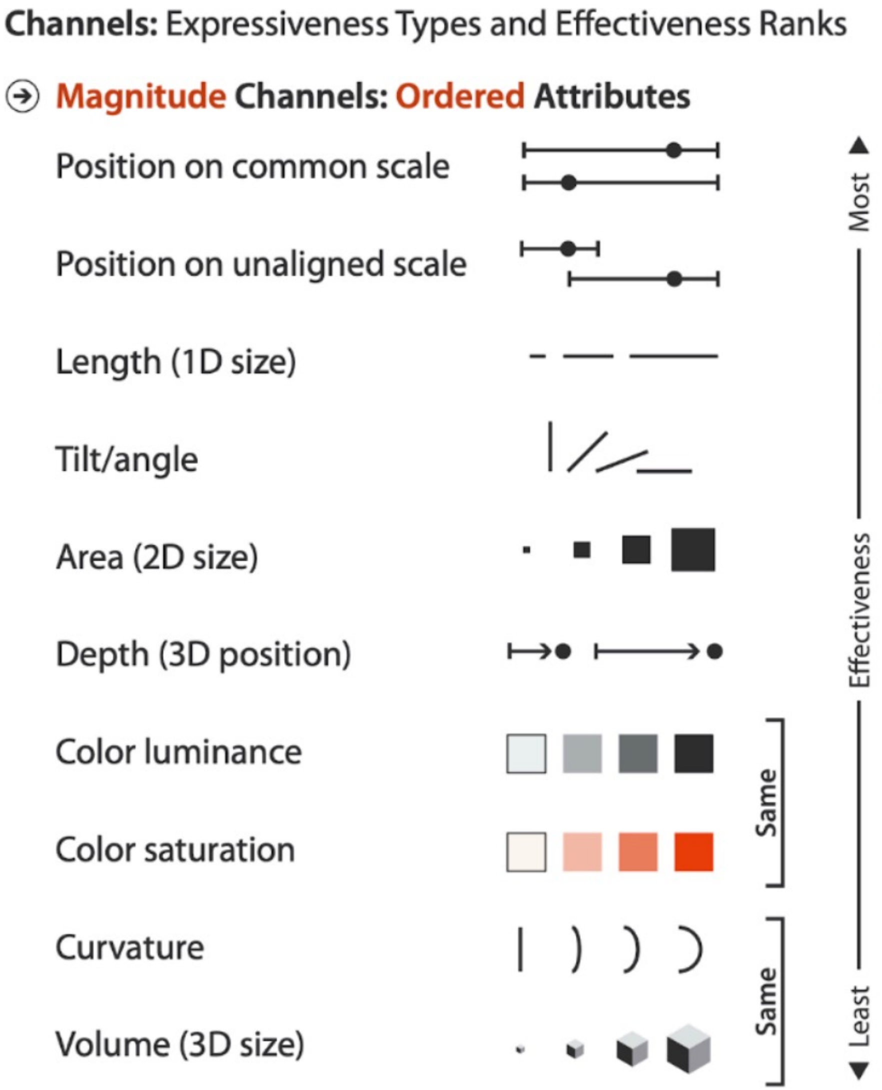
Amar, R., Eagan, J. and Stasko, J., 2005, October. Low-level components of analytic activity in information visualization. In IEEE Symposium on Information Visualization, 2005. INFOVIS 2005. (pp. 111-117). IEEE.

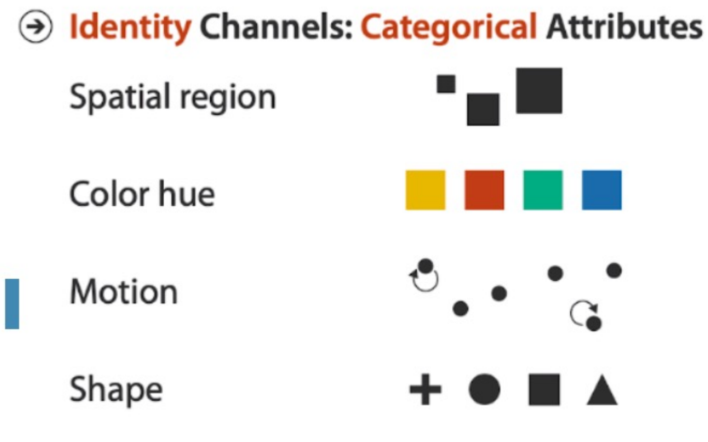
Explore Data, Task, and Design Space (2)

- Chart Type
  - Bar chart, line chart, pie chart…
  - Infographics, scientific rendering…
- Style
  - Dashboard, slideshow, comic…
- Medium
  - Image, video, sound, object…
- Encoding
  - Properties: color, shape, size…
  - Dimension: 2D, 3D, …
- Layout
  - Arrangement and navigation
- Interaction
  - Operation: zoom, pan, rotate…
  - Control: mouse + keyboard, gesture, touch, voice, …
- Space
  - Digital vs. physical: paper, screen, XR …
  - Location: hand, table, wall, room…
  - Scale: inch, foot, yard …

Vis designers must take into account three very different kinds of resource limitations:

- Computers: processing time, system memory, …
- Humans: attention, memory, retention, …
- Displays: pixels, information density, …





"A visualization represents data using a collection of graphical marks such as bars, lines, and point symbols. The attributes of a mark — such as its position, shape, size, or color — serve as channels in which we can encode underlying data values. With a basic framework of data types, marks, and encoding channels, we can concisely create a wide variety of visualizations."

-- Jeffrey Heer

- Narrow down Design Space
  - Requirements
  - Constraints
  - User preference

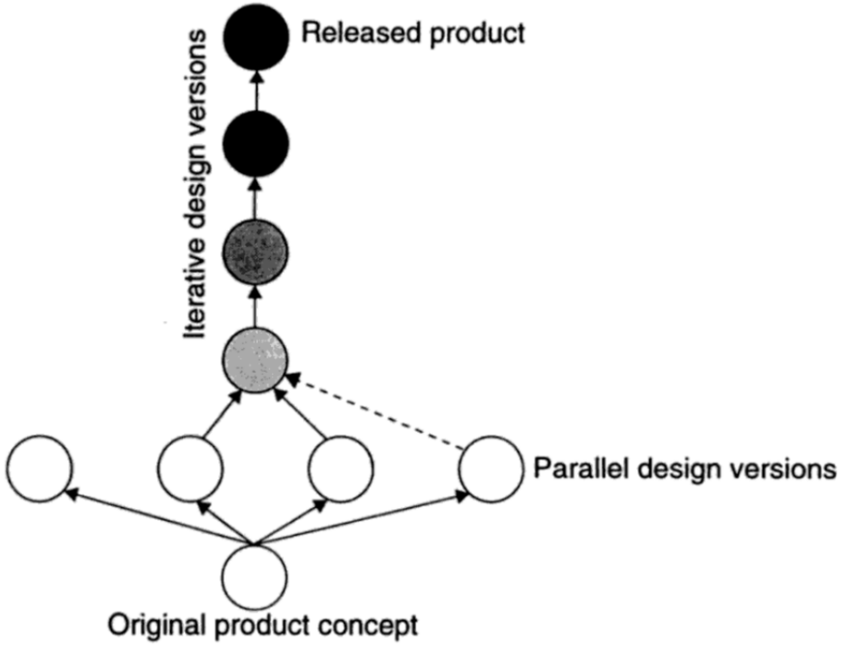


Figure 8 *Conceptual illustration of the relation between parallel and iterative design. Normally, the first prototype would be based on ideas from several of the parallel design sketches.*

Nielsen, J., 1994. *Usability engineering*. Morgan Kaufmann.

- Suggest and/or Compare Design Alternatives

- Feedback on Rapid Prototypes

Figure 1. V1 shows how good students and struggling students spend their time on solving the problem and reviewing before the exam in order to convey gaming the system by submitting frequently may need to spend much more time reviewing the question before an exam.

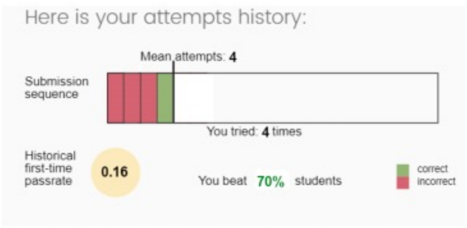


Figure 2. V2 shows the attempt times of peers by mean attempts and first-time pass rate from last semester in order to convey that difficult problems take considerable effort for other students to solve and that it is unnecessary to give up quickly by resorting to gaming the system.

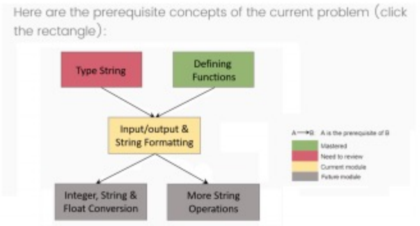


Figure 3. V3 shows one example of the prerequisite relationship among learning concepts in order to convey that learning concepts are connected and that the negligence of one concept may hinder the mastery of later concepts due to the cumulative nature of the course.

Xia, M., Asano, Y., Williams, J.J., Qu, H. and Ma, X., 2020, August. Using Information Visualization to Promote Students' Reflection on" Gaming the System" in Online Learning. In *Proceedings of the Seventh ACM Conference on Learning@ Scale* (pp. 37-49).

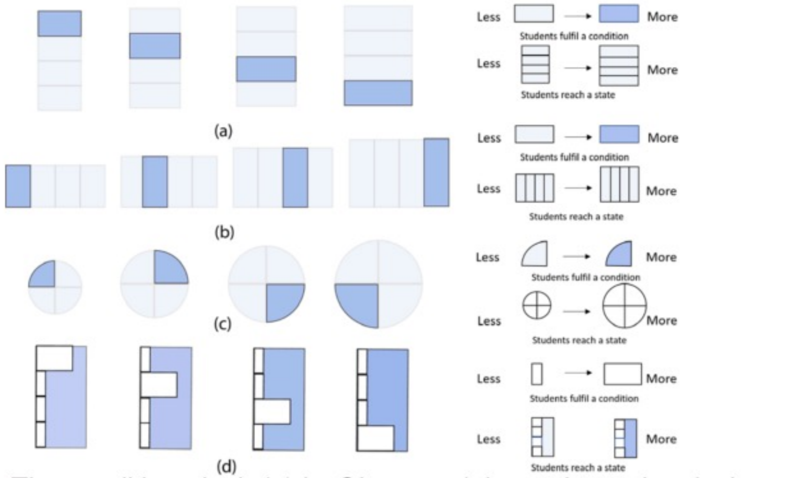
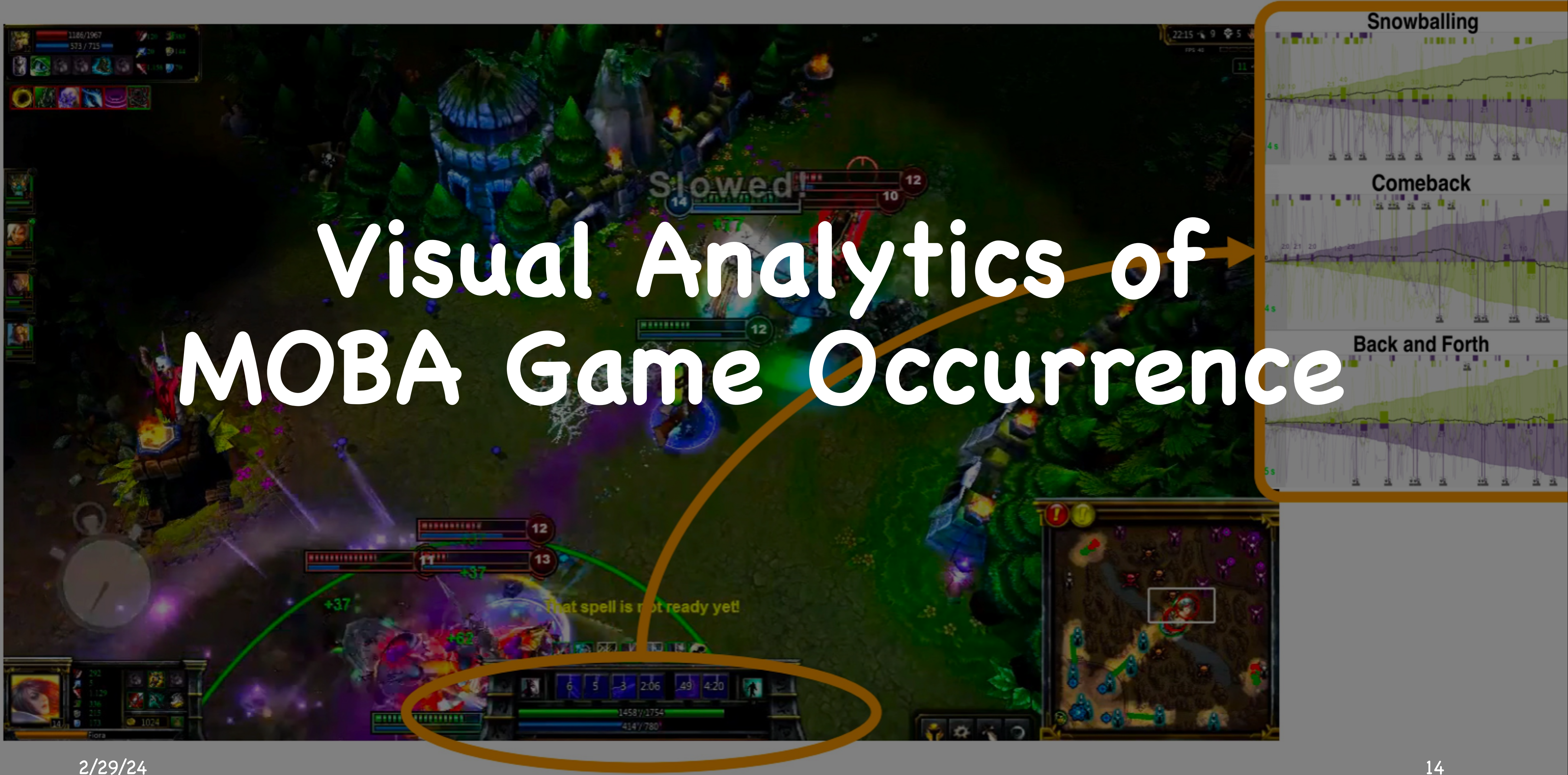


Fig. 7. The condition glyph (a) in *QLens* and three alternative designs (b) (c), and (d).

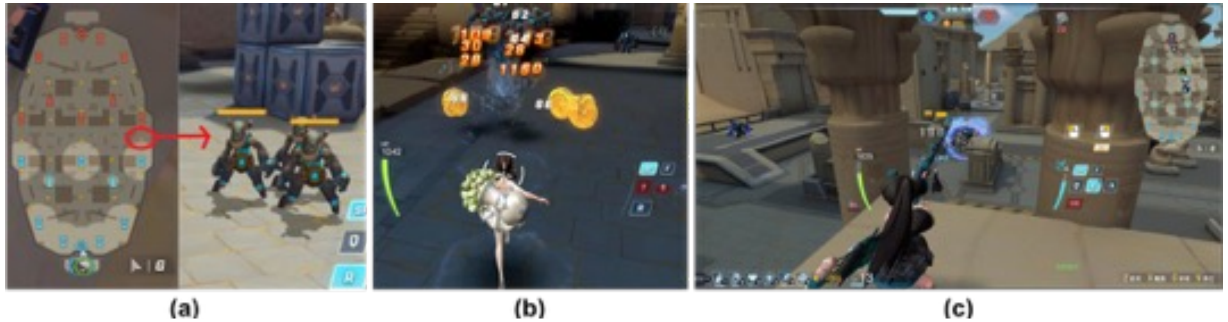
Xia, M., Velumani, R.P., Wang, Y., Qu, H. and Ma, X., 2020. Qlens: Visual analytics of multi-step problem-solving behaviors for improving question design. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), pp.870-880.

Assist Collaborative or User-in-the-loop Design

Visual Analytics of MOBA Game Occurrence

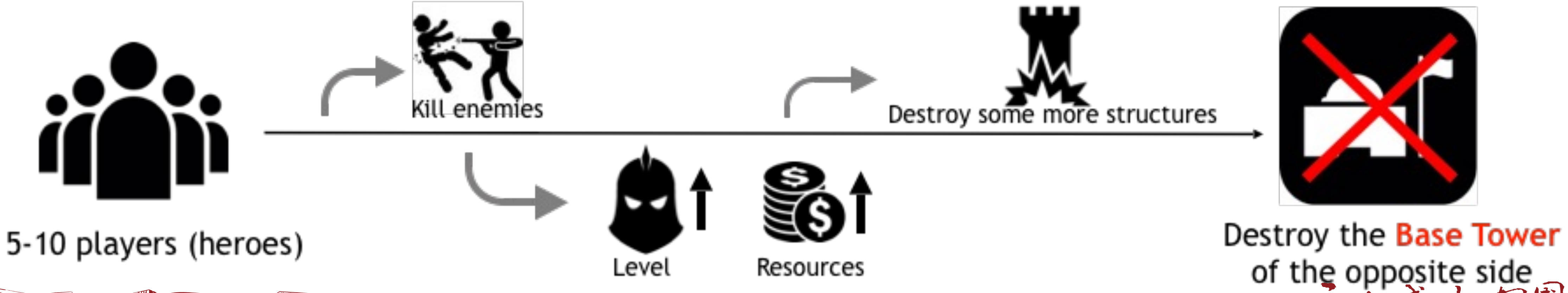


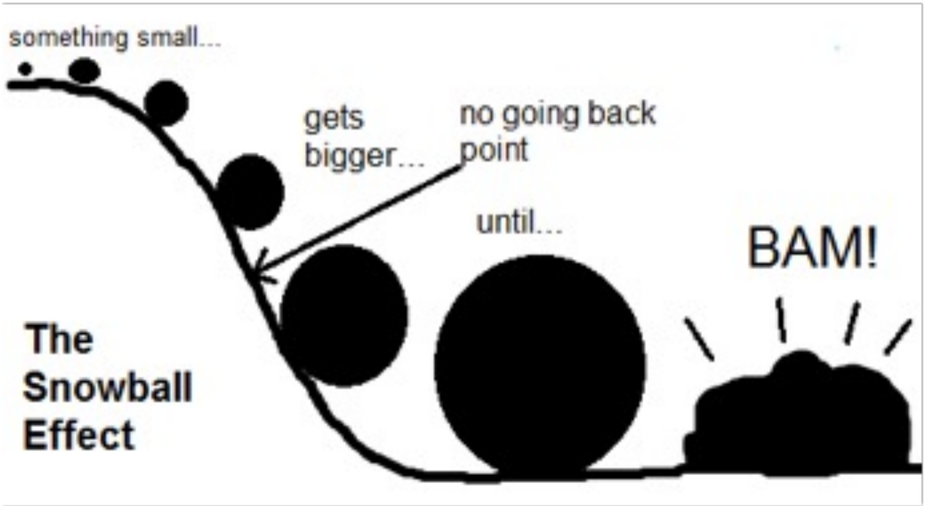
MOBA (Multiplayer Online Battle Arena) Games



MOBA game is a type of video games
in which two teams launch coordinated attacks on each other's base

Golds are rewarded for killing hostile computer-controlled units (b) (i.e. yellow points in the map) (a). Heroes in two teams are in a combat (c).





Snowballing 碾压局

A team achieves and maintains advantages over their opponents **without much effort** throughout the remaining game.

Back and Forth 有来有回局

There is **no clear winner** until the very end.

Comeback 翻盘局

A team **overcomes a substantial disadvantage**, particularly when this results in the disadvantaged team winning.

Game Occurrences

1

Too much snowballing 碾压局 ✗

- Winners get bored
- Undergo a limited set of actions
- Encourage a player to droput mid-game

2

Too much comeback 翻盘局 ✗

- Losers are not happy
- Game is not stable
- Misalignment between expectation and reality

3

Too much back and forth 有来有回局 ✗

- No adequate differentiation
- No adequate sense of mastery

A Good Balance

Approach

A Multi-Phased Co-design Process

1 Observational Study

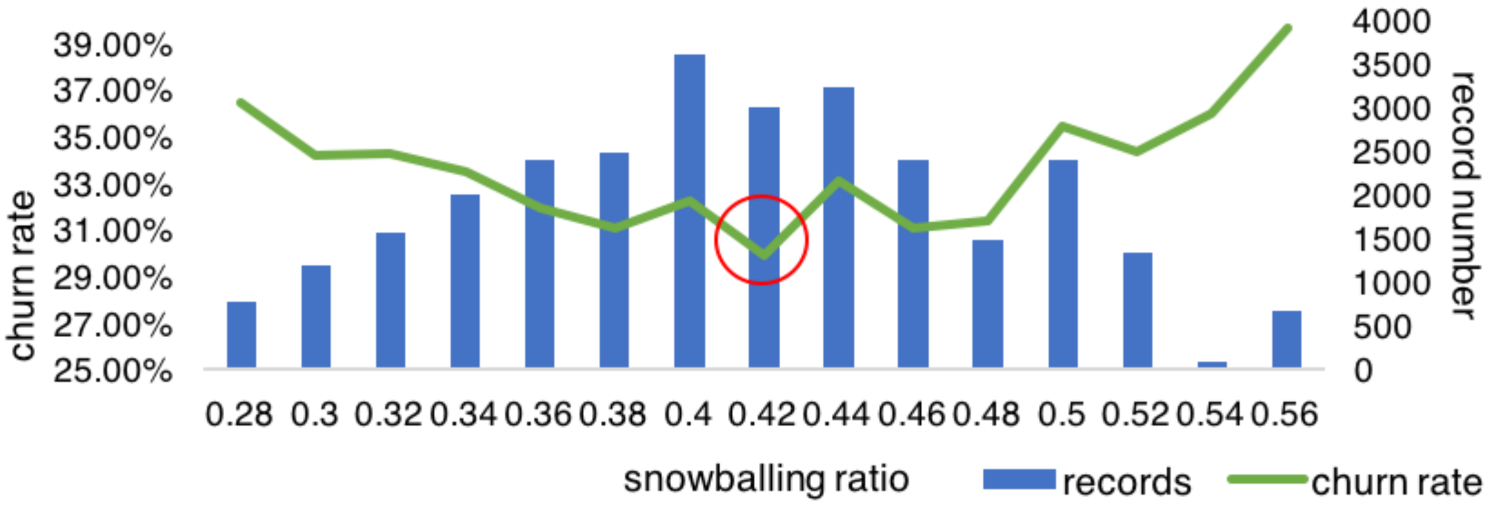
2 Design of the Single-Match Module

3 Design of Full System

1

Phase 1: Observational Study

About Game Team



15-day churn rate first decreases and then increases as the percentage of snowballing occurrences goes up

- Work with a team of experts: One User Experience Analyst (E.1), One Data Analyst (E.2) and two Game Designers (E.3-4)
- The target is to **reduce the churn rate** (prevent players to dropout from the game)
- The **lowest churn rate** can be achieved when the **ratio of snowballing** occurrence is about 0.42

Study Procedure

Shadow the team's daily working process

- **Observe players playing game**
- Conduct **on-site interviews** with players
- Calculate gameplay **statistics**
- **Infer** factors that result in game occurrences
- **Improve** game desgin accordingly

Drawbacks

- Design rules to classify occurrences is **expensive**

  Uncertainity exists whether more efficient rules could be extracted to cover more occurrences
- Depend on **expertise** to understand reasons for occurrences (snowballing, comeback, etc.)

The game team envisions Machine Learning (ML) to reason game occurrences !!!

Study Procedure

Difficulties by involving ML

Communication Barrier and Vague Requirements

1

- Articulate **factors of interests** to ML experts
- **Ill-defined and abstract** concepts

Skepticism about Performance

2

- **Unable** to understand ML model **outputs**
- Don't **trust** ML models

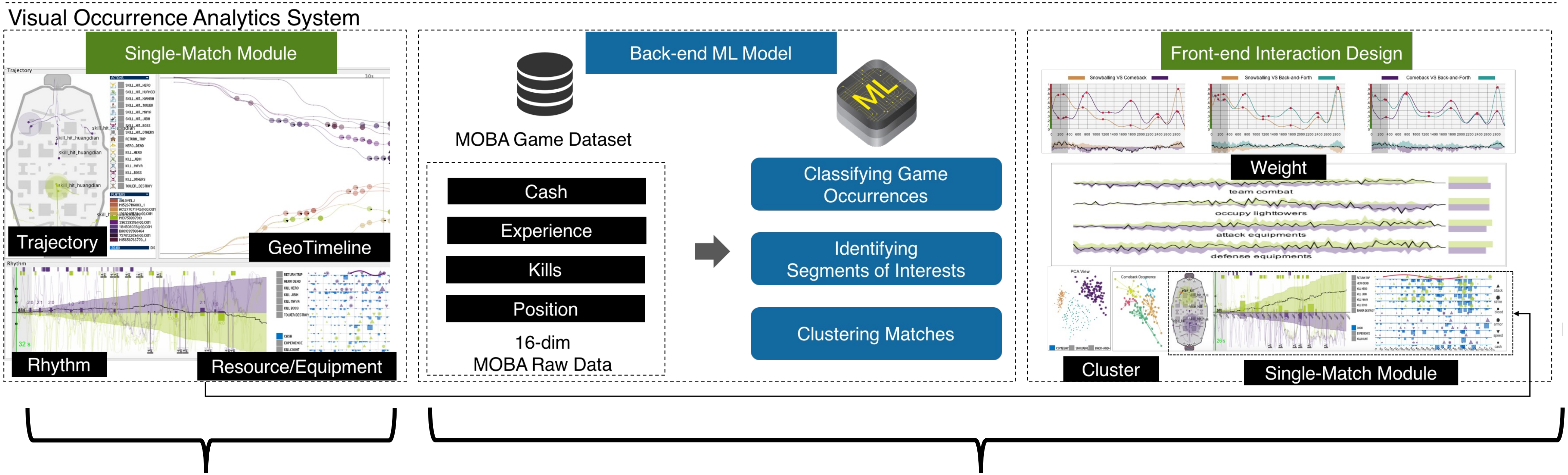
Unclear Division of Responsibilities

3

- **Unable to extract** ill-defined behavior patterns
- **Little** experience working with ML

Needs and Expectations of Collaborating with ML

Multi-phased co-design process



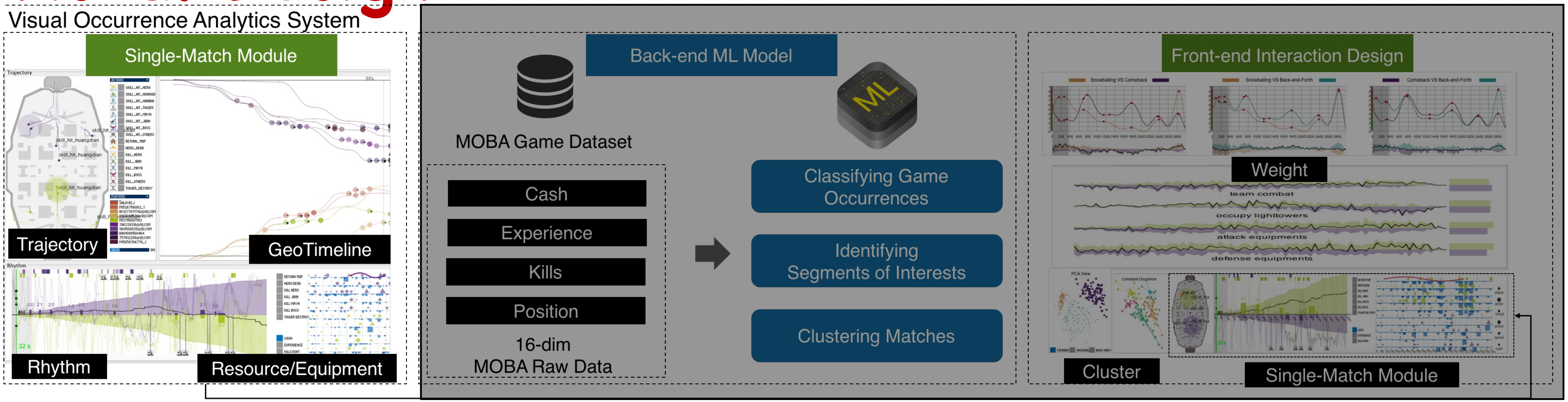
Design of the single-match module (Phase 2)

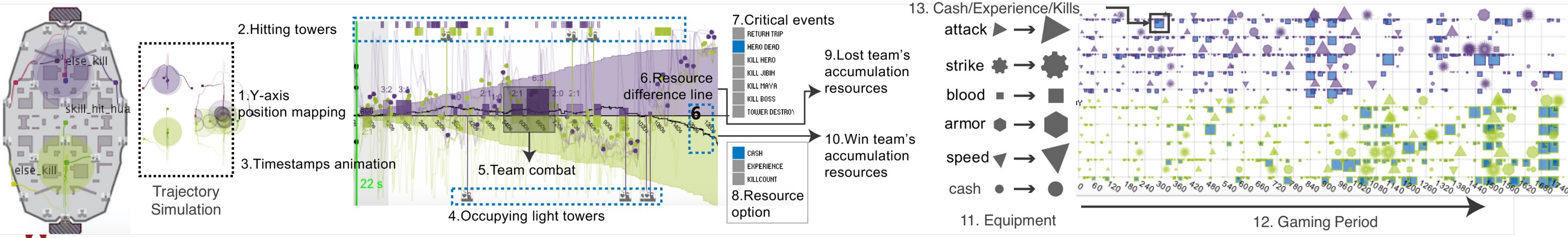
Design of the full system with ML embedded (Phase 3)

Phase 2
Design of the Single-Match Module

"Visualization prototype"

Visual Design



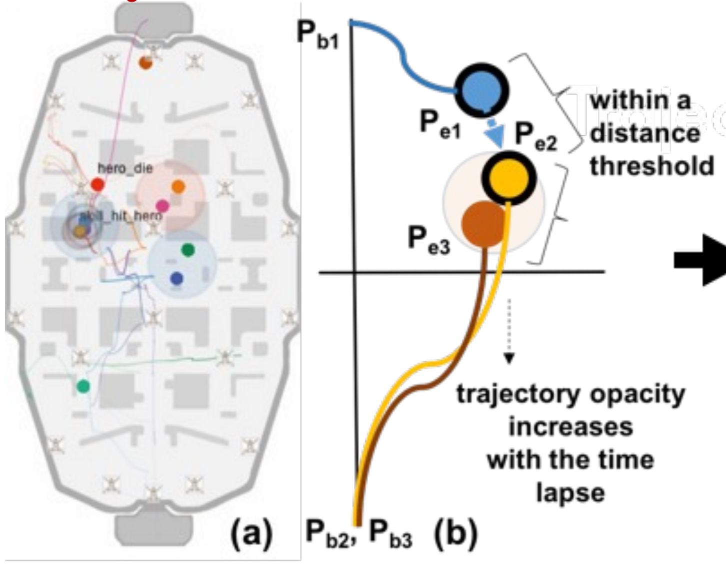


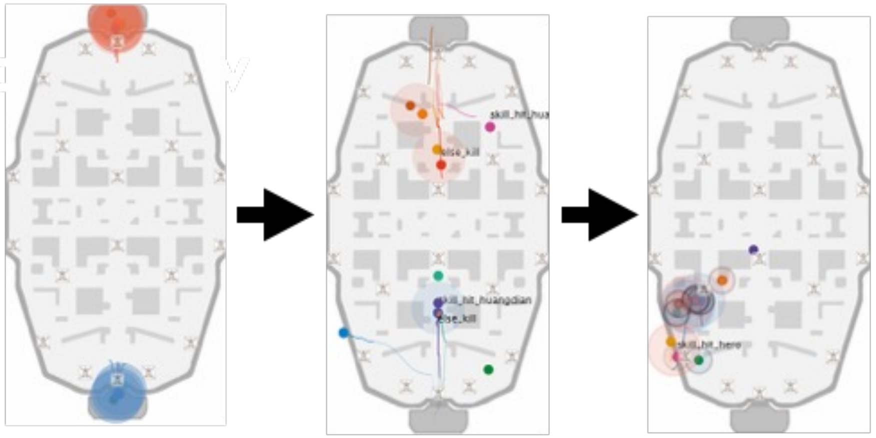
1. Trajectory

2. Rhythm

3. Resources/Equipment

Trajectory



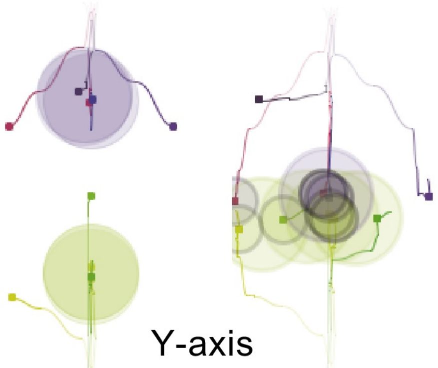


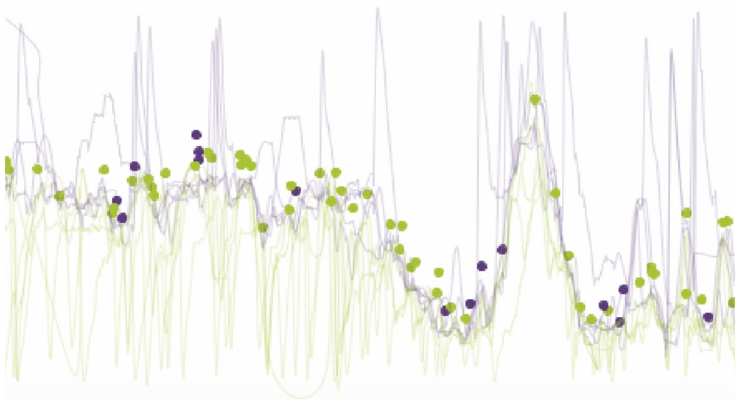
Moving by a smooth animated transition

Players from the same camp form a bubble if they are close

Border intensifies when opponents meet

Use animation to simulate how players move

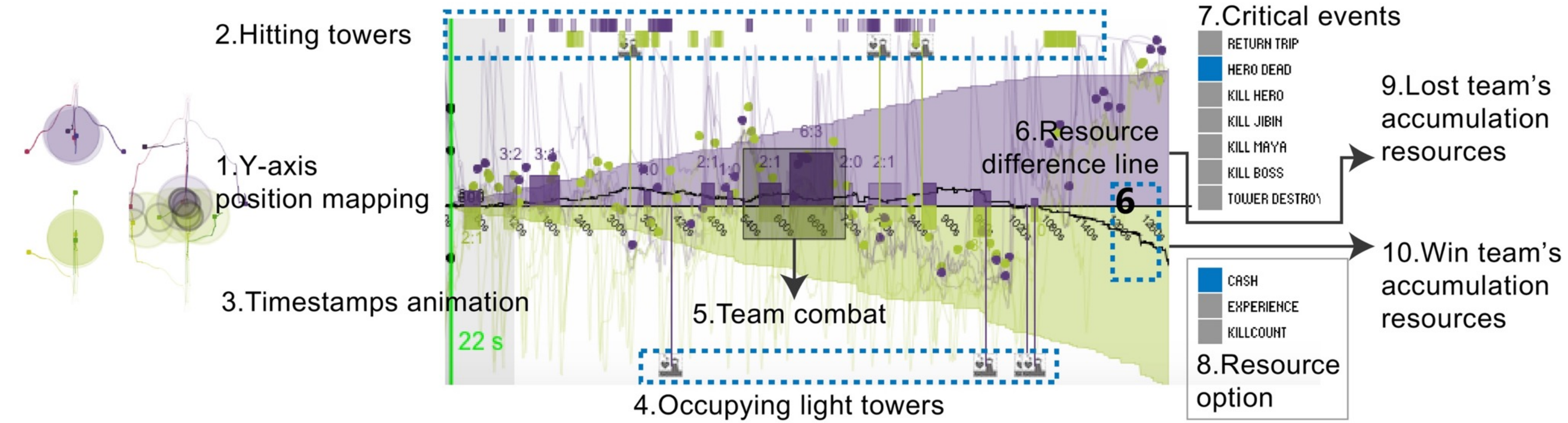




Y-axis position mapping over time
Y轴坐标映射沿时间轴展开

Serve as cues on which team is marching

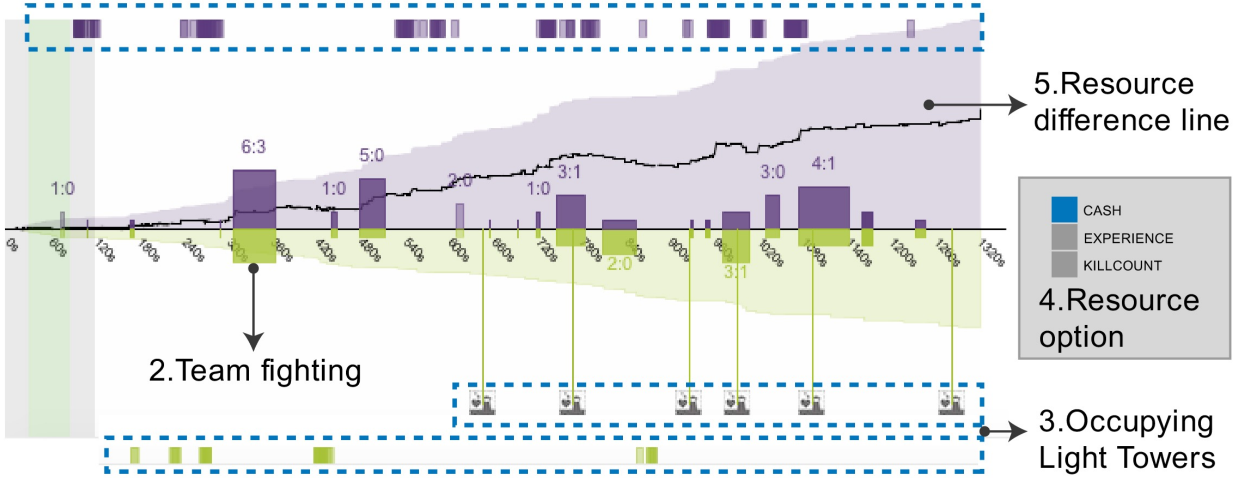
Rhythm



- **Hitting towers events** encoded by rectangles along timeline
- **Brush and drag a time slider to select interesting time**
- Glyphs indicate when occupying **light towers** happen
- **Team combats** encoded by vertical bars with height indicating the results
- **Critical events can be clicked and** displayed on the trajectories

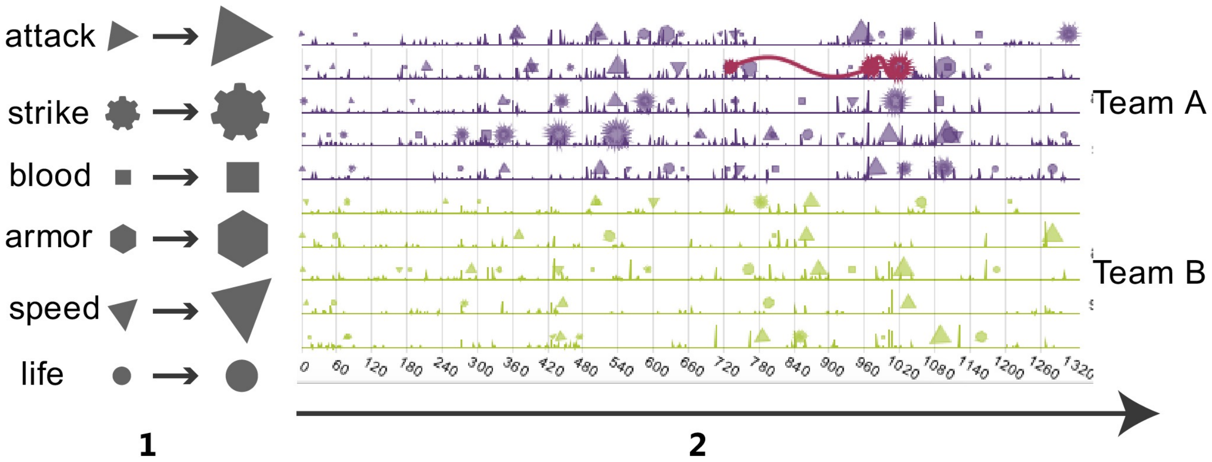
立志成才 报国裕民

Resources/Equipment



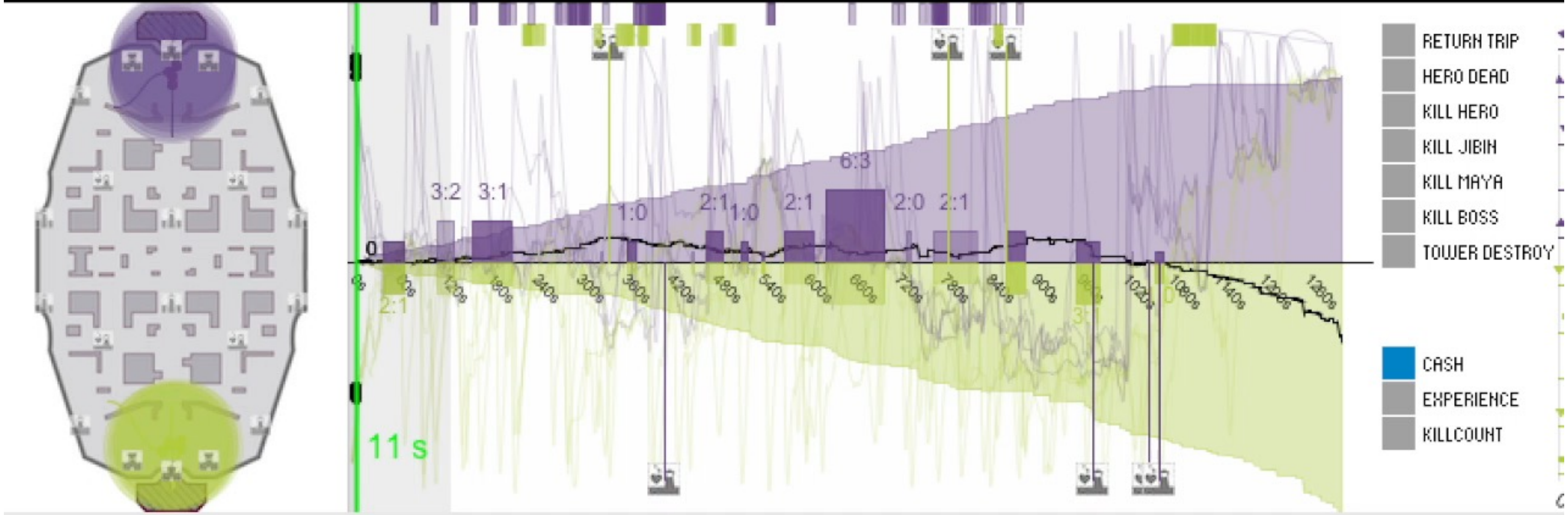
- **Cumulative resources (双方累计资源)** are visualized by two areas with a **black curve** indicating the **difference (双方累计资源差)** between them

- **Different equipment (装备)** encoded by different **glyphs** and the **level of equipment (装备等级)** visualized by the **size** of the glyphs



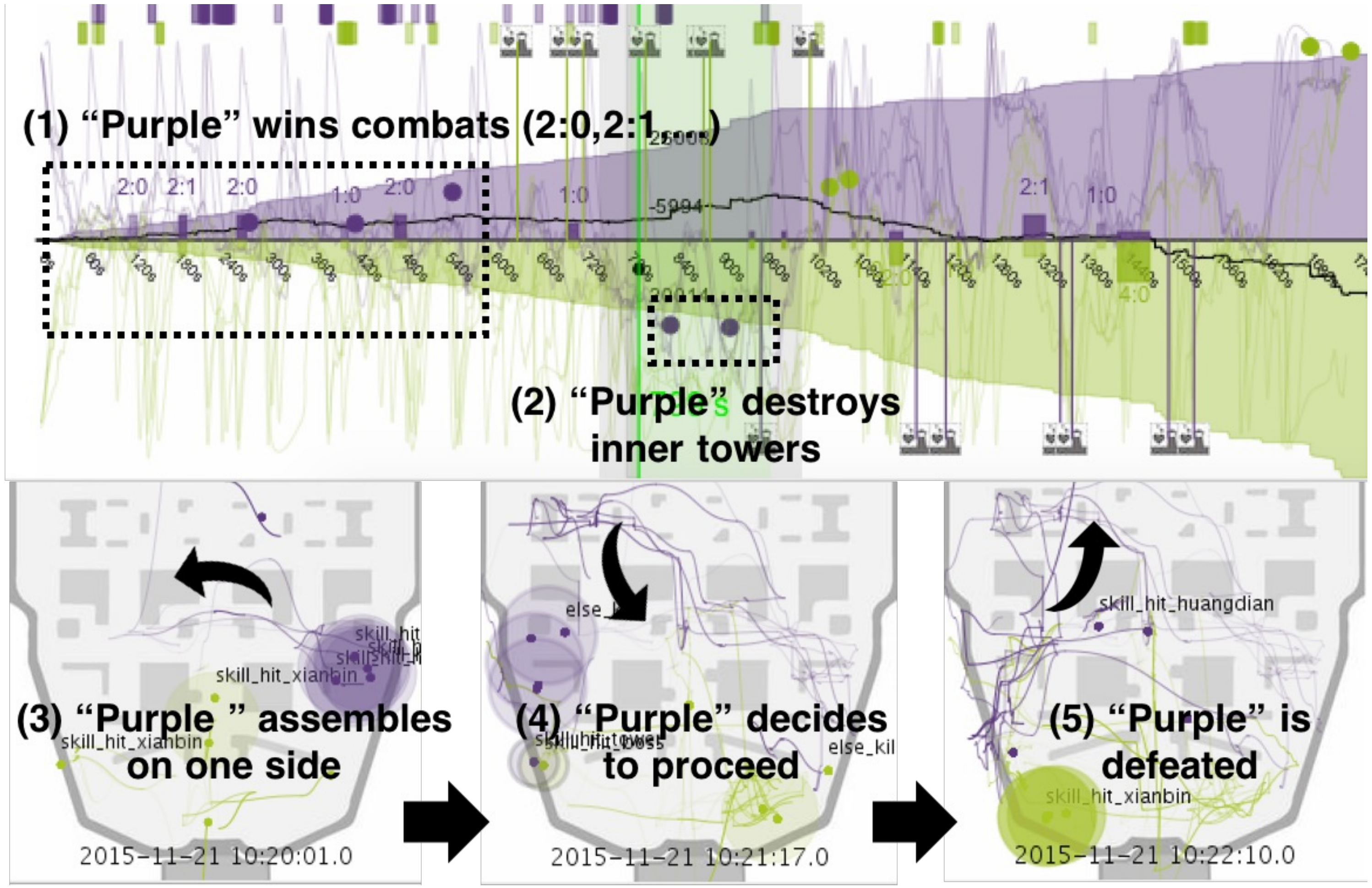
Interaction

Interaction



Phase 2 Feedback and Discussion

A case study of comeback occurrence by the single-match module
Before comeback
翻盘前的情况

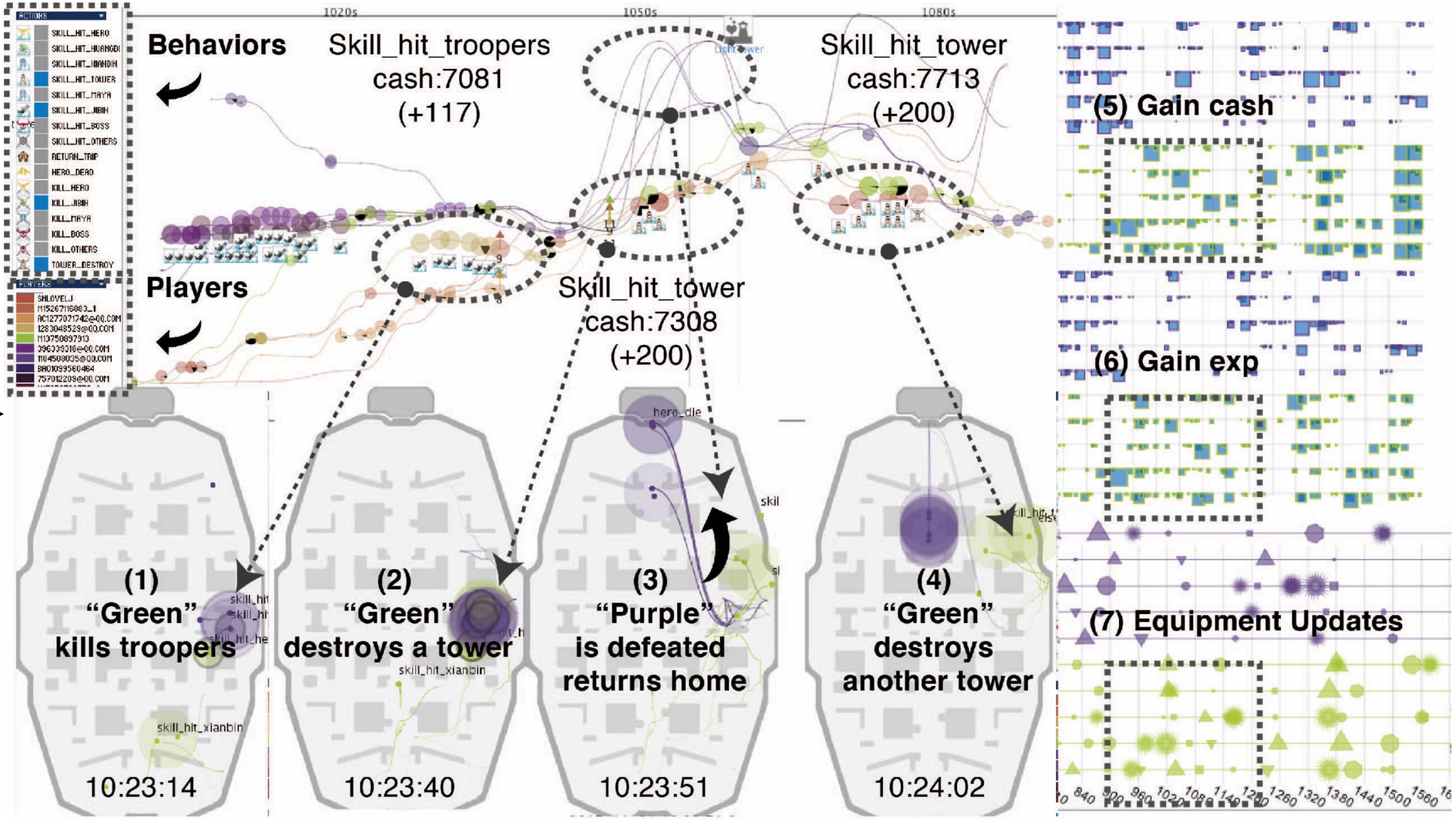


Before comeback: (1) The early advantaged team won in team combat. (2) Two inner towers were destroyed. (3-5) The "purple" players assembled and proceeded, but were defeated.

Phase 2 Feedback and Discussion

A case study of comeback occurrence by the single-match module

Comeback
翻盘时



Comeback: The comeback team (1) killed the troopers, (2) destroyed one tower and (3) the early advantaged team was defeated and returned. The comeback team grasped the chance, (4) destroyed another outer tower, (5, 6) obtaining resources and (7) upgrading equipment.

Experts' feedback

➢ Takeaway message

- Imbalanced rewards between different towers 不同塔的推塔收益不一样
- The gain from previous stage of the match could not compensate the loss, and this let the comeback able to outweigh in the following battles
- "Can quickly digest a match and easily detect when a team retreats or proceeds"
- "Make the findings more compelling compared with statistical data report"
- Case by case analysis 基于一场一场的分析
- Insufficient for large-scaled gameplay analysis 不适用于大量比赛数据的分析
- Still have to manually go through a match

Phase 2 Expectations for the Full System

上海科技大学 ShanghaiTech University

*"It would be great to know **which matches are more likely to contain typical patterns…**"*

*"…obtain occurrence patterns **based on many matches…**" – From the game team*

⬇

1. Suggest
match segments of interest

2. Cover
sufficient patterns of game occurrences

How to do?

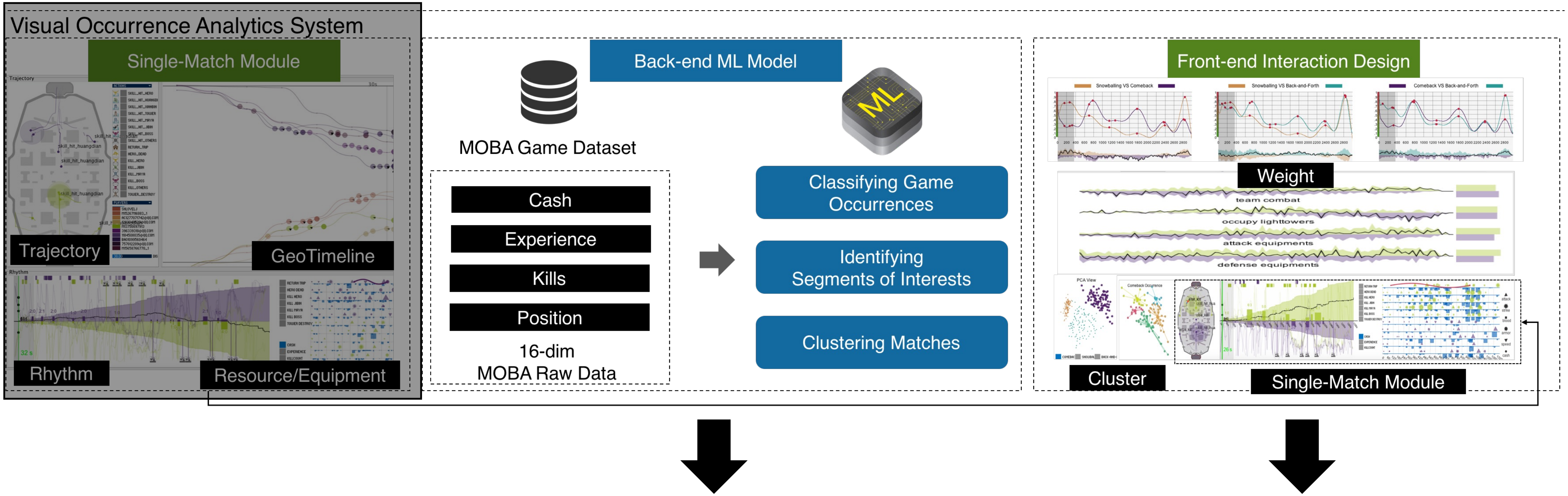
立志成才 报国裕民

3

Phase 3
Design of Full System

"Incorporates ML models with interactive analysis"

Phase 3 Design of Full System



Back-end ML Model

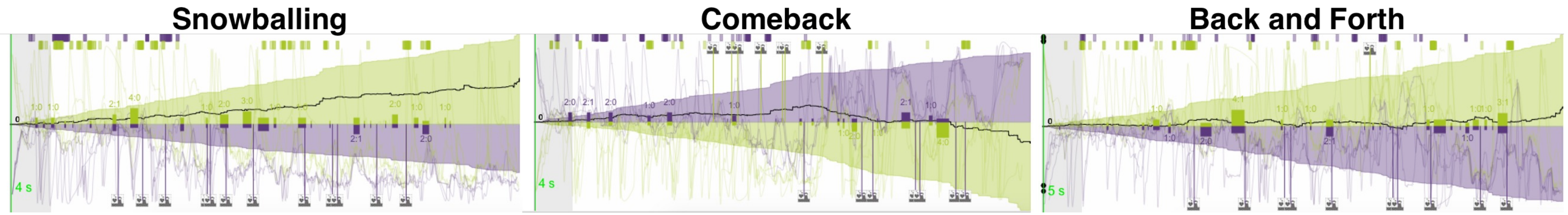
- **Extract** critical match segments
- **Cluster** similar matches

Front-end Interface

- Assist users to **understand ML output**

Back-end ML Model

1. Infer the ML model input   2. Identify critical segments   3. Cluster matches



Patterns of resources (cash, experiences, kills) and trajectories (positions) of **snowballing**, **comeback** and **back and forth** are visually **different**

Infer the ML model input

1. Preparing MOBA Raw Data for Models

- **16 dimensions** (cash, experiences, kills of 2 teams and positions of 10 players) = 3*2+10 = 16
- **Normalized** the match length to 3000 seconds
- Raw data is a **16 * 3000 matrix**

2. Training Models to Classify Occurrences

- 2500 matches for training while the rest for testing

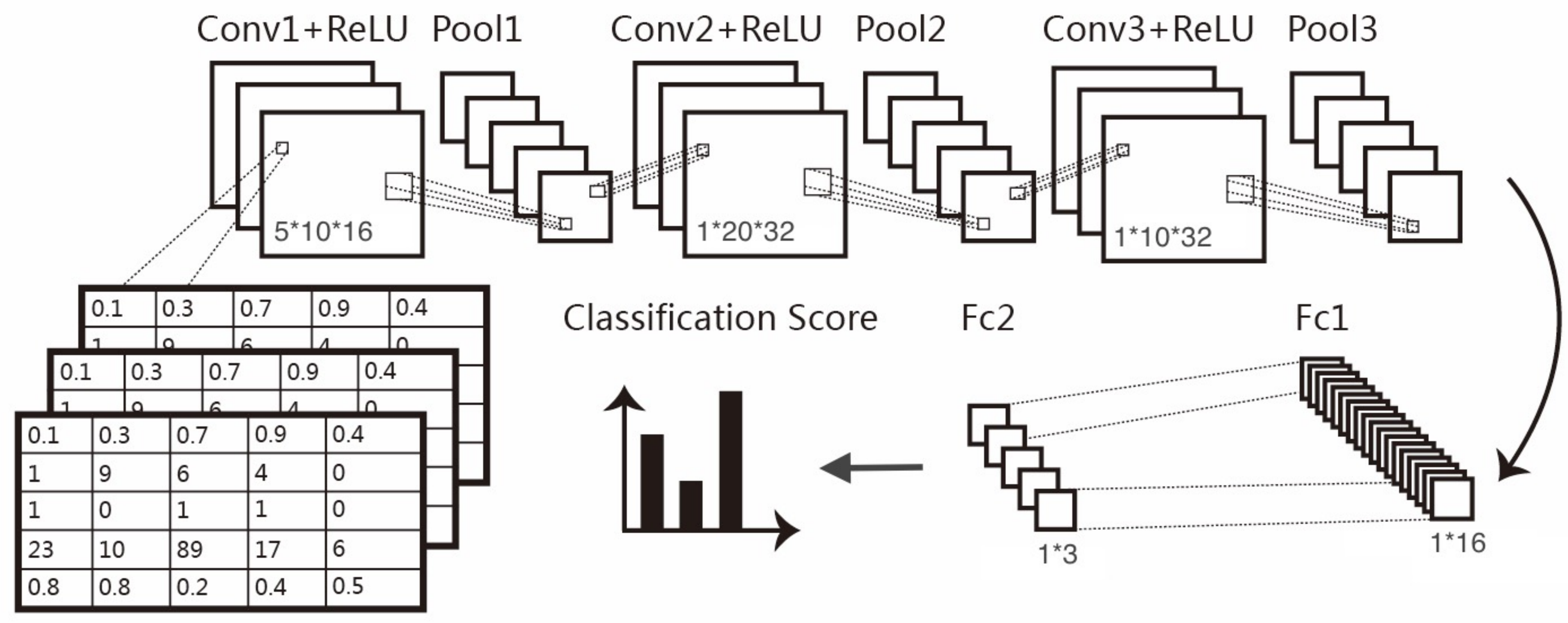
Feature \ Classifier	*KNN*	*Decision Forest*	*RBF-SVM*
Full (16-dim)	**85.5%**	85.5%	**87.2%**
Cash (2-dim)	84.2%	83.4%	84.8%
Exp (2-dim)	78.5%	80.2%	82.5%
Kills (2-dim)	81.0%	82.3%	**88.0%**
Pos (10-dim)	84.5%	**85.8%**	86.2%

Back-end ML Model

1. Infer the ML model input　　2. Identify critical segments　　3. Cluster

- 局势分类
- 深度神经网络框架
- 深度模型训练
- 时空片段区域检测

- 使用类似于Alexnet的架构，代替人工提取特征进行分类的机器学习方法
- 五层网络、四个阶段：convolution卷积，rectification修正，汇集pooling
- 80%数据进行训练，20%进行测试（数据分为三类：碾压、翻盘及有来有回）
- 基于噪声卷积策略，检测哪部分输入数据区域能够引起分类结果发生较大变化



CNN框架：

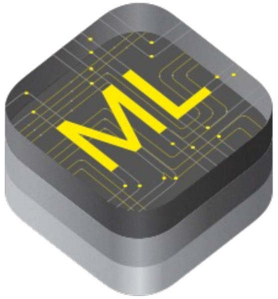
包含三层卷积层及两层充分连接层，层与层之间是修正及汇集操作

Back-end ML Model

1. Infer the ML model input matches   2. Identify critical segments   3. Cluster

Obtain 3-dim classification confidence score

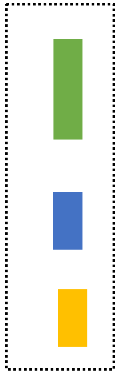
A Match (input) ➡



CNN

➡

Snowballing
Comeback
Back and Forth



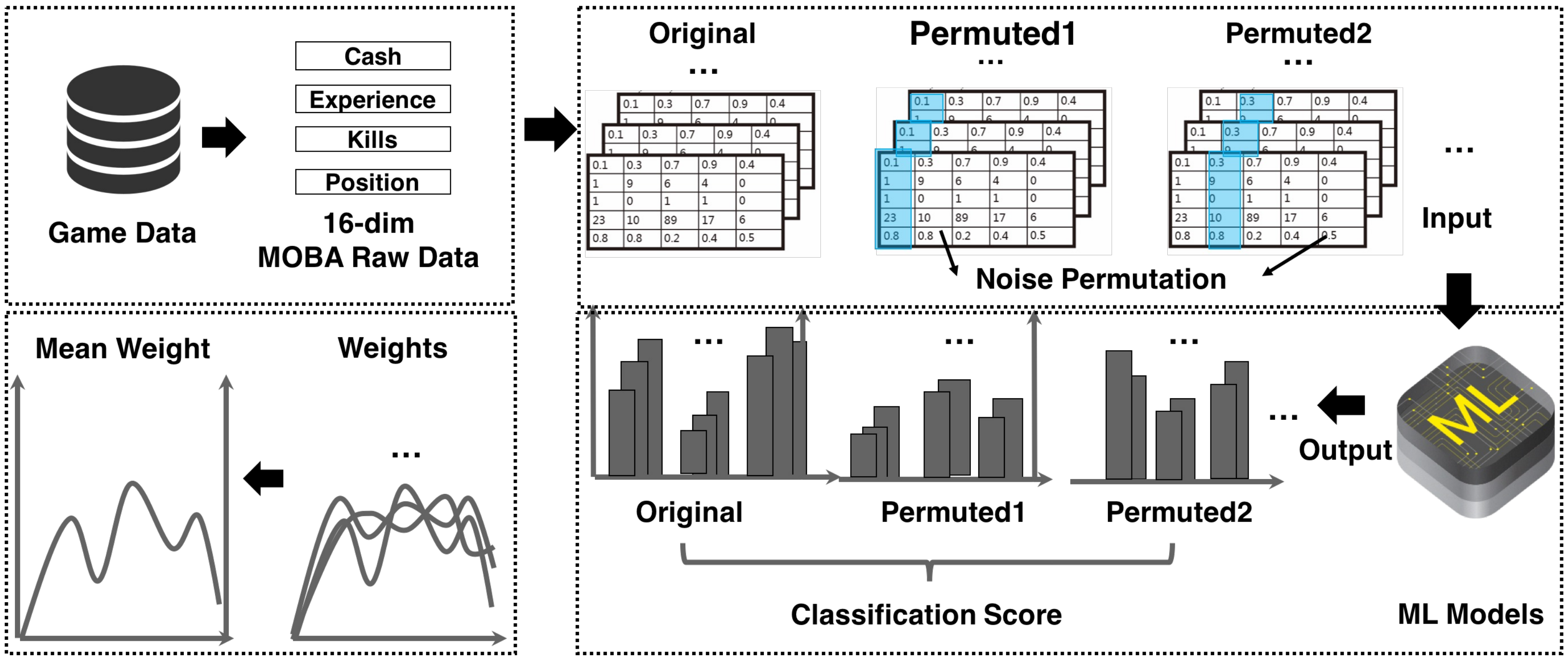
0.7
0.2
0.1

➡ Classification score (3-dim)

Confidence that the input **belongs to a category** (snowballing, comeback, back and forth)

Back-end ML Model

1. Infer the ML model input 2. Identify critical segments 3. Cluster matches



1. Apply a patch-based noise perturbation

2. Identify patches that lead to severe performance decay in the classification task

3. Average the segments of all matches of the same occurrence

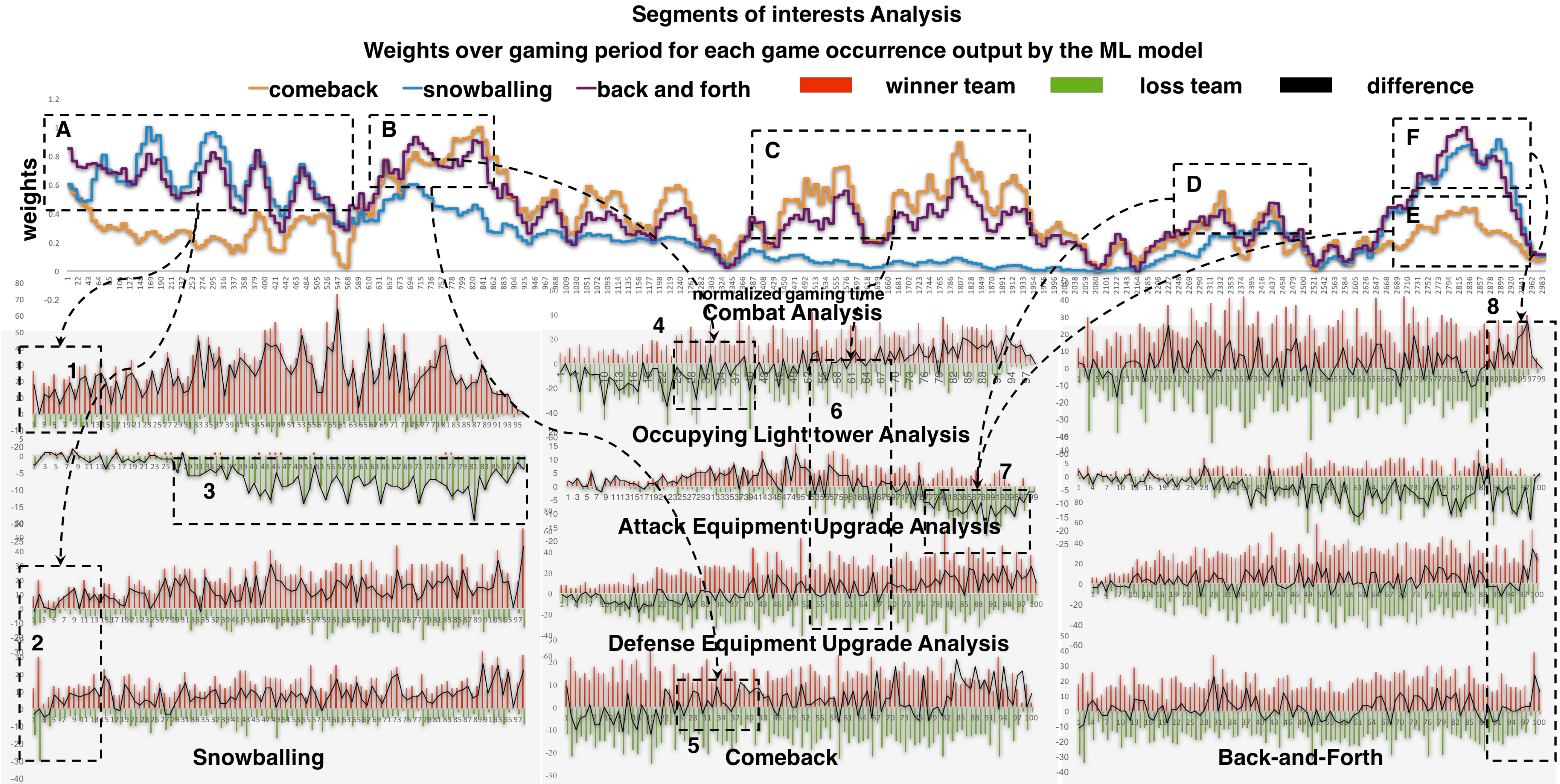
4. Output the weight of segments of interest

Weight indicating how much the performance decay resulted from a permuted patch

Weight View

Breadth: Segment Analysis Across Occurrences

上海科技大学 ShanghaiTech University

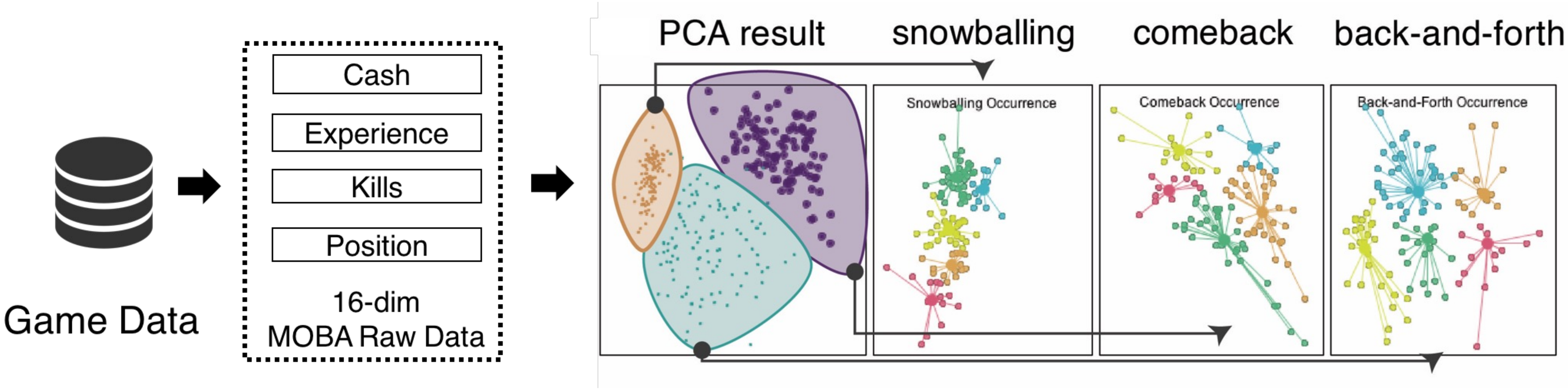


对于碾压局势，前1/6阶段和后1/10阶段对于局势结果影响重大；对于翻盘局势，检测到了四个翻盘点；
有来有回局势与翻盘较相似，除了最后时刻区别较大

立志成才 报国裕民

Back-end ML Model

1. Infer the ML model input   2. Identify critical segments   **3. Cluster matches**



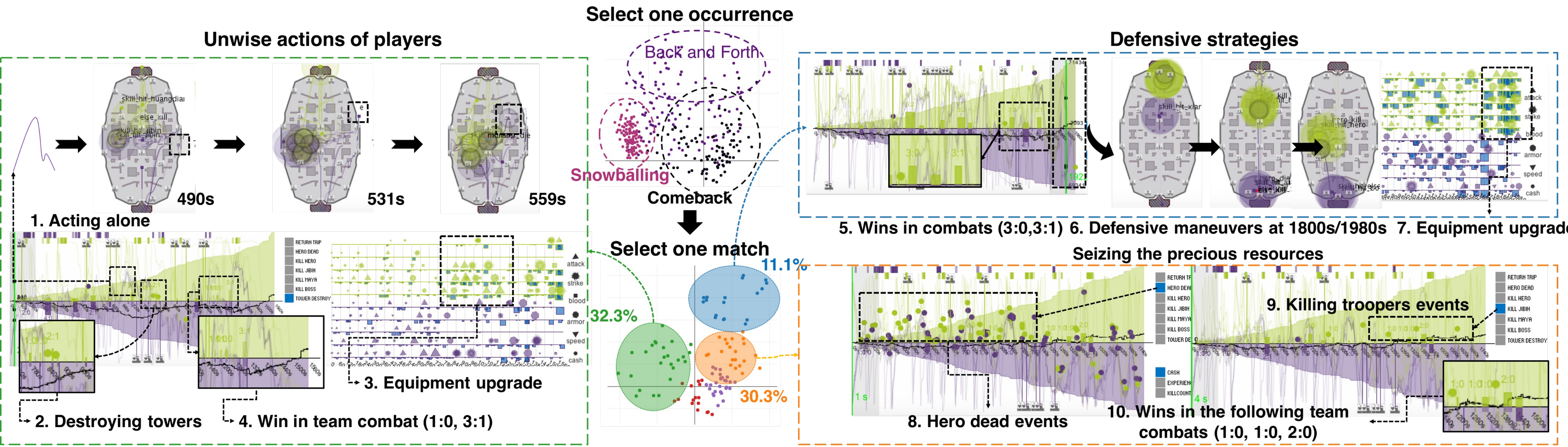
The raw data is taken as the **feature representation** of a match, which is then projected to a two-dimensional space **PCA** for visualization (*snowballing, comeback, back-and-forth* are well separated). Each occurrence is further clustered by **K-Means** (k=*5* in default).

Cluster View

Depth: Cluster Investigation for Each Occurrence

上海科技大学
ShanghaiTech University

Comeback Investigation



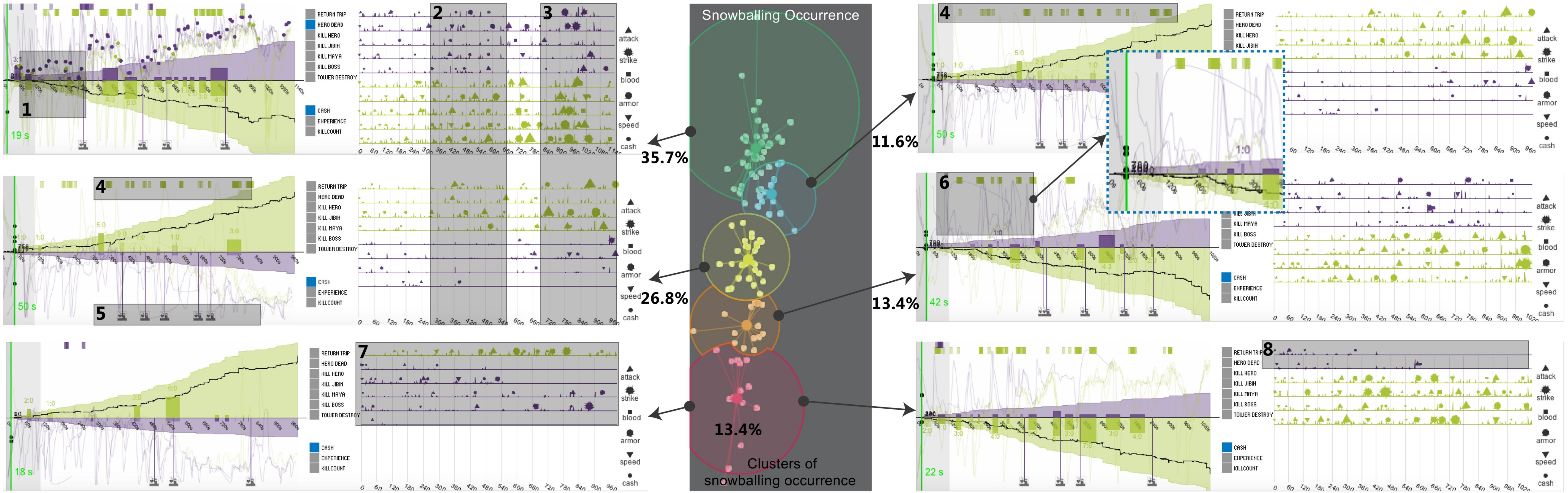
Typical comeback pattern set includes "**unwise actions of players**", "**defensive strategies of comeback teams**" and "**seizing the precious resources**"

Cluster View

Depth: Cluster Investigation for Each Occurrence

上海科技大学 ShanghaiTech University

Snowballing Investigation



- Players are "**Away From Keyboard**"
- **All *hitting towers* events** are conducted by *snowballing* teams
- **Unbalanced number of players** in the beginning of match

Modify game design

- **Modify** the player matching mechanisms
- **Offer weaker team incentives** at early stage
- **Remove** the design of light towers

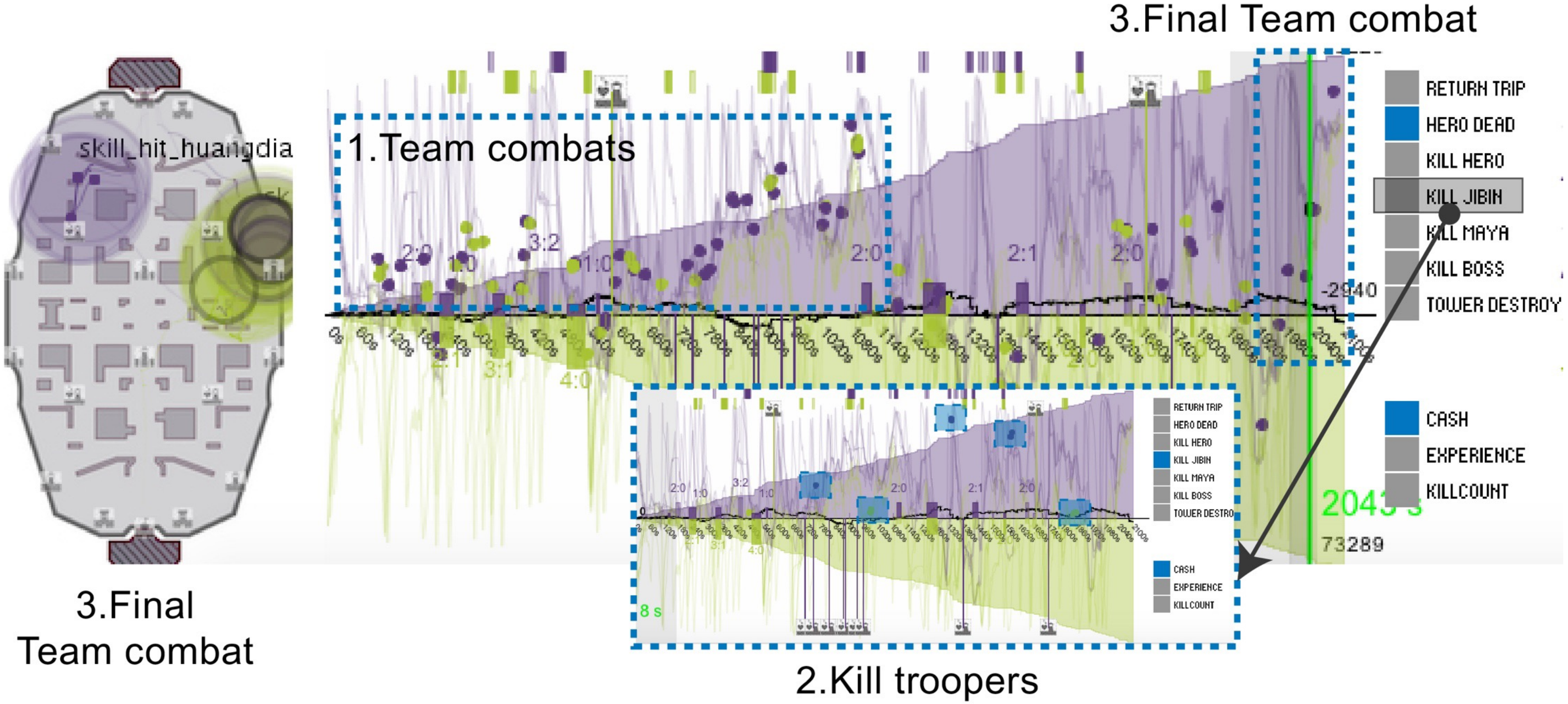
立志成才 报国裕民

Cluster View

Depth: Cluster Investigation for Each Occurrence

上海科技大学 ShanghaiTech University

Back and Forth Investigation



3.Final Team combat

*"Both teams have a good chance for winning the game, and the **final pushes** determine the match"*

立志成才 报国裕民

4

Discussion and Reflection

Discussion and Reflection

System Performance and Experience

- *"Provide designers with **concrete** information to explain segments"*
- *"Display the visualization of gameplay in real-time is **intuitive"***
- ***Streamlines** process as the experts no longer need to randomly select matches*
- *Extend **scale** of analysis and cover **sufficient** patterns for each occurrence*

➡ Satisfaction of online gaming experience improved: "the weekly retention rate rose by 1.1%

Discussion and Reflection

- ✓ With the collected data becoming increasingly large and complex, only automated algorithms are not able to support deep and convenient data exploration
- ✓ By carefully considering the detailed analysis tasks and design requirements, design and implement interactive visualization systems to enable users, even novice users with no background in computer science, to conduct insightful and efficient data exploration in various application domains
- ✓ These interactive systems combine the capabilities of both people and machines and can help users easily gain deep insights into the data of different applications

- Level of Automation
  - L0: No automation
  - L1: Human-directed automation
  - L2: System-suggested, human-directed automation
  - L3: System-directed automation
  - L4: Full automation

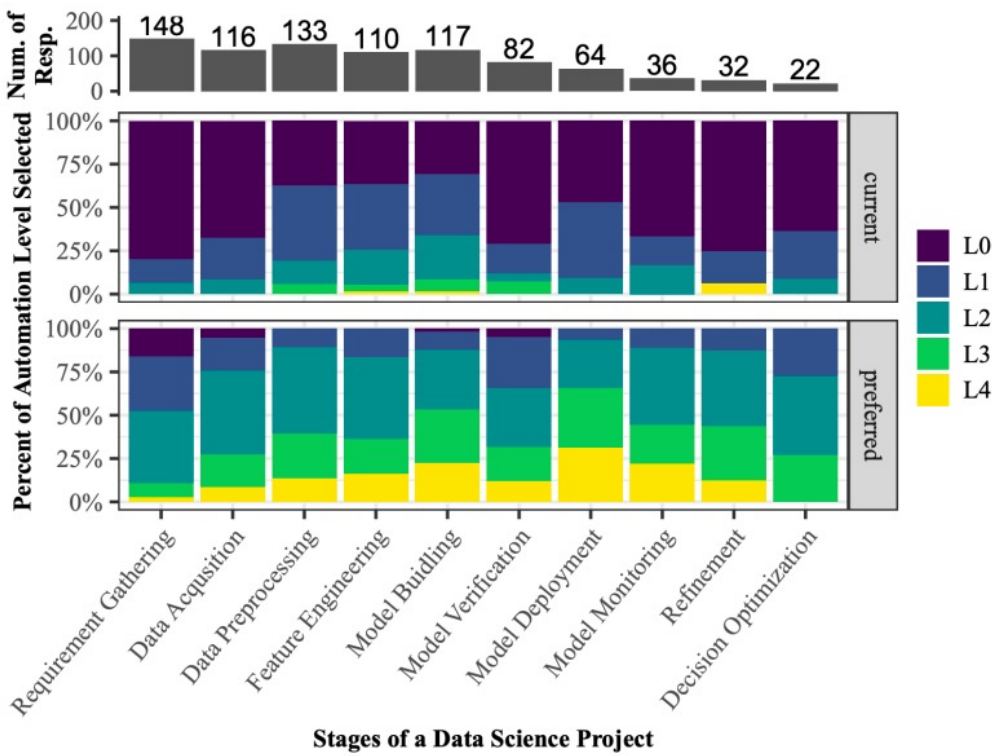


Fig. 4. Comparison of respondents' current practice's level of automation and their desired levels of automation in each stage. The small bar chart at the top shows the number of responses collected for each stage.

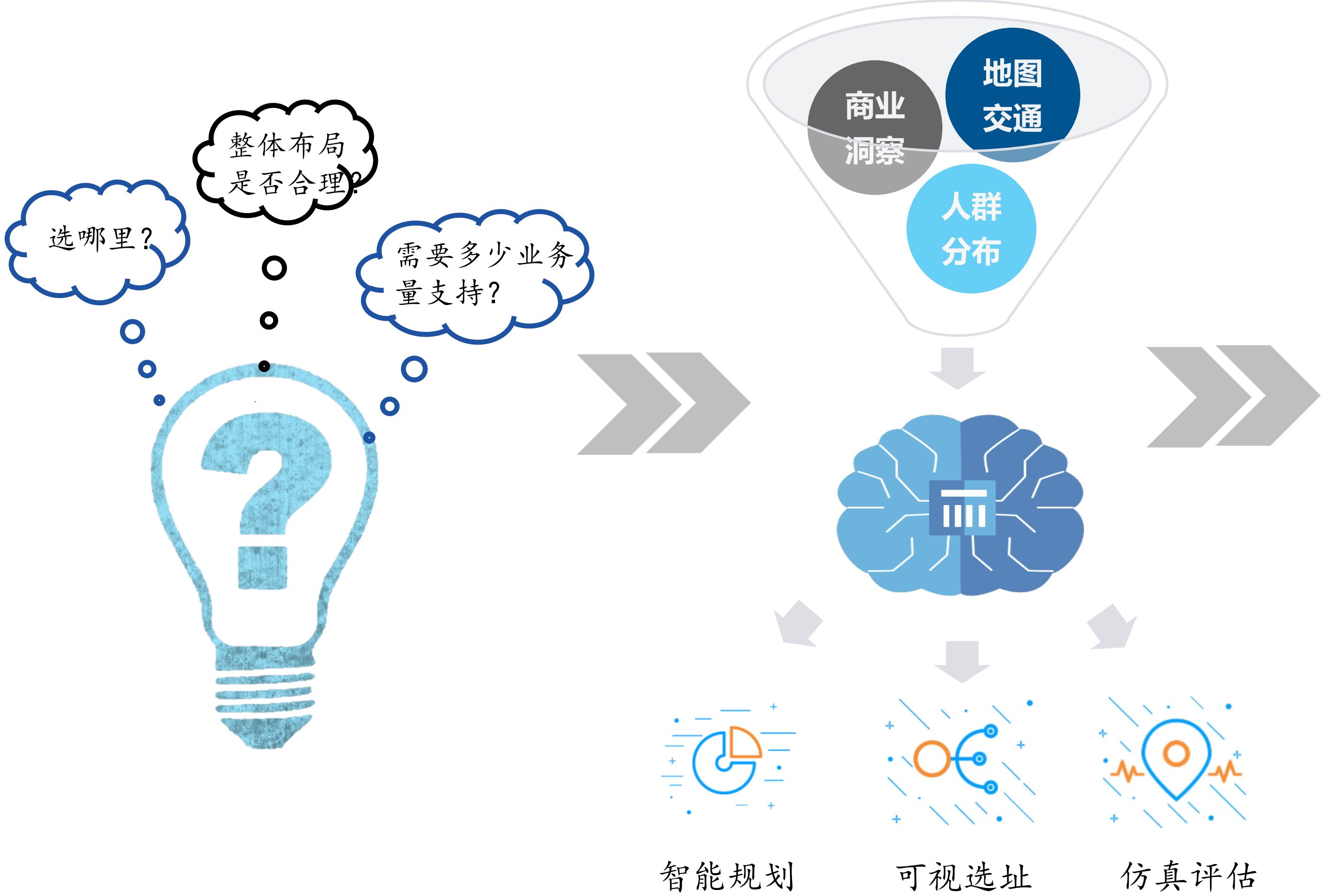
Apply Automated Process, if Appropriate

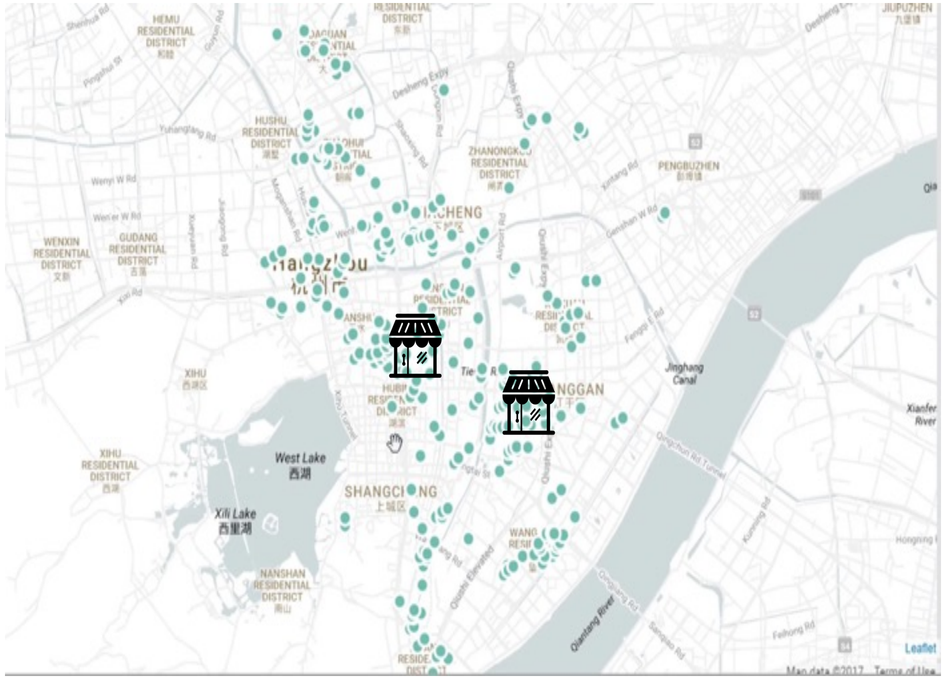

iThink Logistics
Choosing A Warehouse Location
10 Factors to consider


MUST-HAVE SELECTION CRITERIA
WHEN CHOOSING YOUR
WAREHOUSE STORAGE

智能仓储选址

Smart Warehouse Site Selection





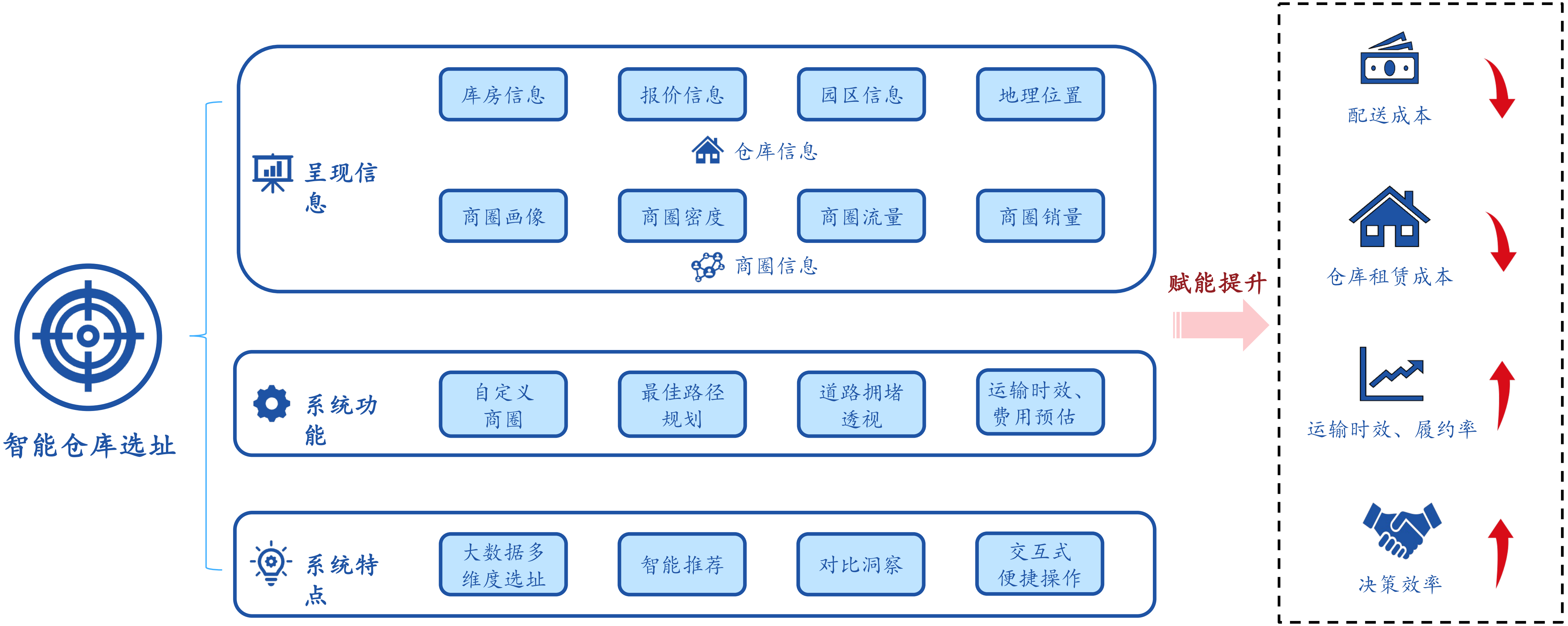
实现价值

- 解决如何最优化店仓选址
- 为市场拓展规划提供管理平台
- 为后续运营改进提供依据

智能仓储选址

Smart Warehouse Site Selection

- 智能选址通过仓库与商圈的信息聚合分析、 数据深度透视，帮助降低物流配送成本、提高选址效率
- 人与机器的和谐分工，充分利用计算机强大的处理能力与人的视觉决策能力，形成可视分析闭环

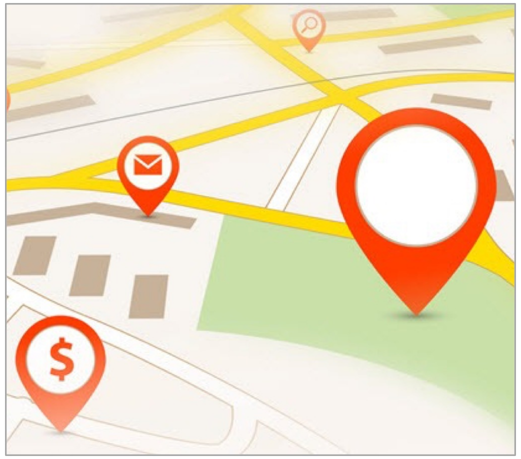


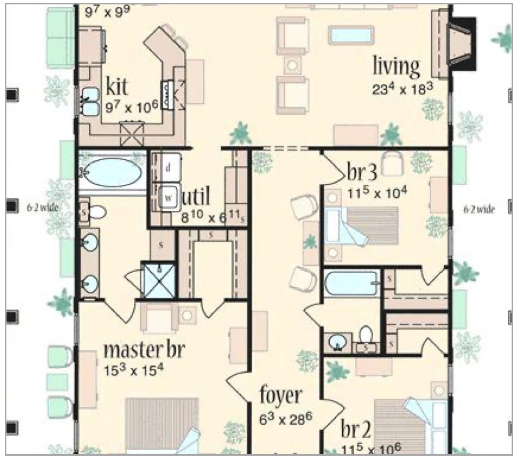
智能仓储选址

Smart Warehouse Site Selection

选择一个合适的仓库位置来满足某个商业区的货物需求，对一个成功的零售企业非常重要

- 我希望找到一个理想的仓库，具备
- 良好的仓库资质
- 区域功能
- 良好的交通条件
- 较低的租赁成本





挑战

Challenges

巨大的解决方案空间

- 异质性数据
- 巨大的候选解决方案空间
- 理解现有的信息

不确定的交通状况

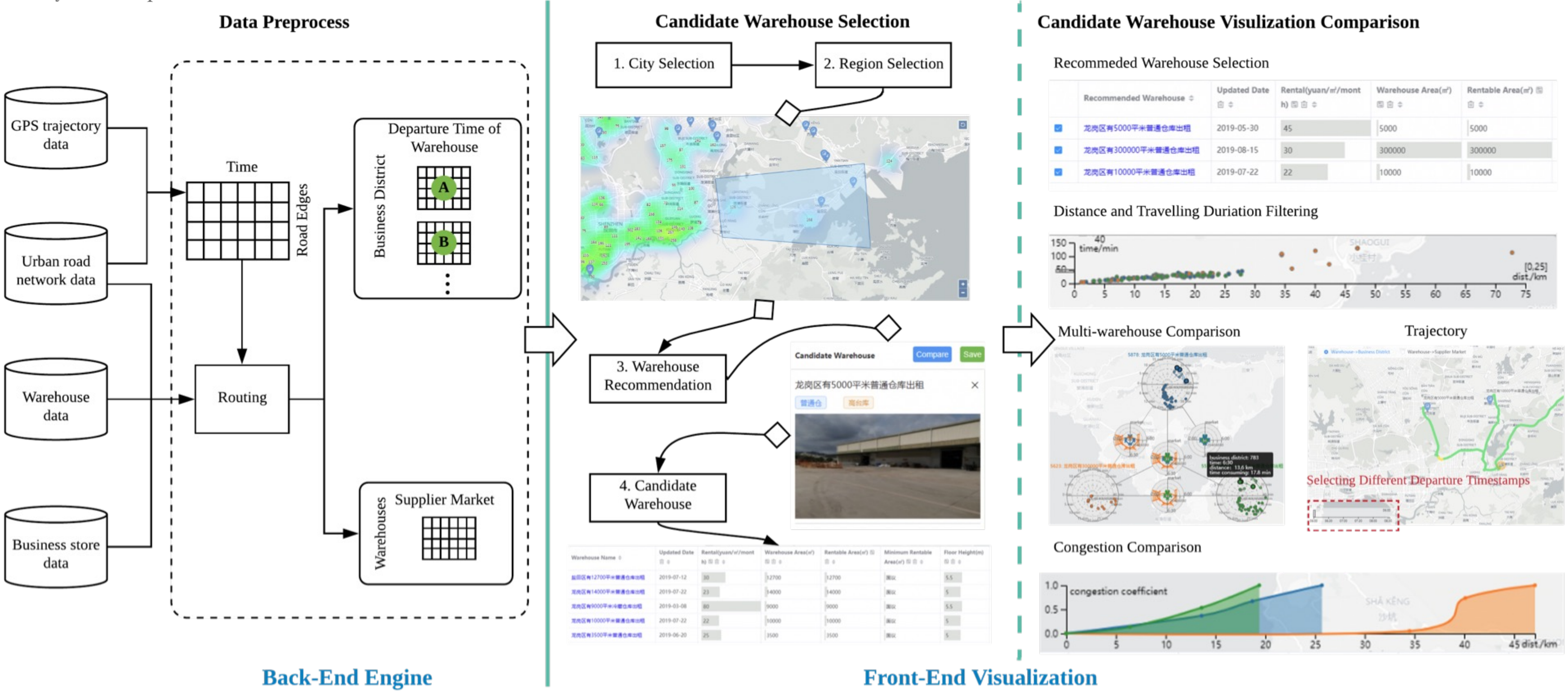
- 地域位置
- 交通状况
- 操纵各种数据来源

不同的商业偏好

- 对偏好的不同要求
- 在表达偏好方面具有良好的灵活性

系统流程

System Pipeline



运输时间估计

Travel-Time Estimation

步骤1

- 路网匹配算法和划分运输车辆行驶轨迹

步骤2

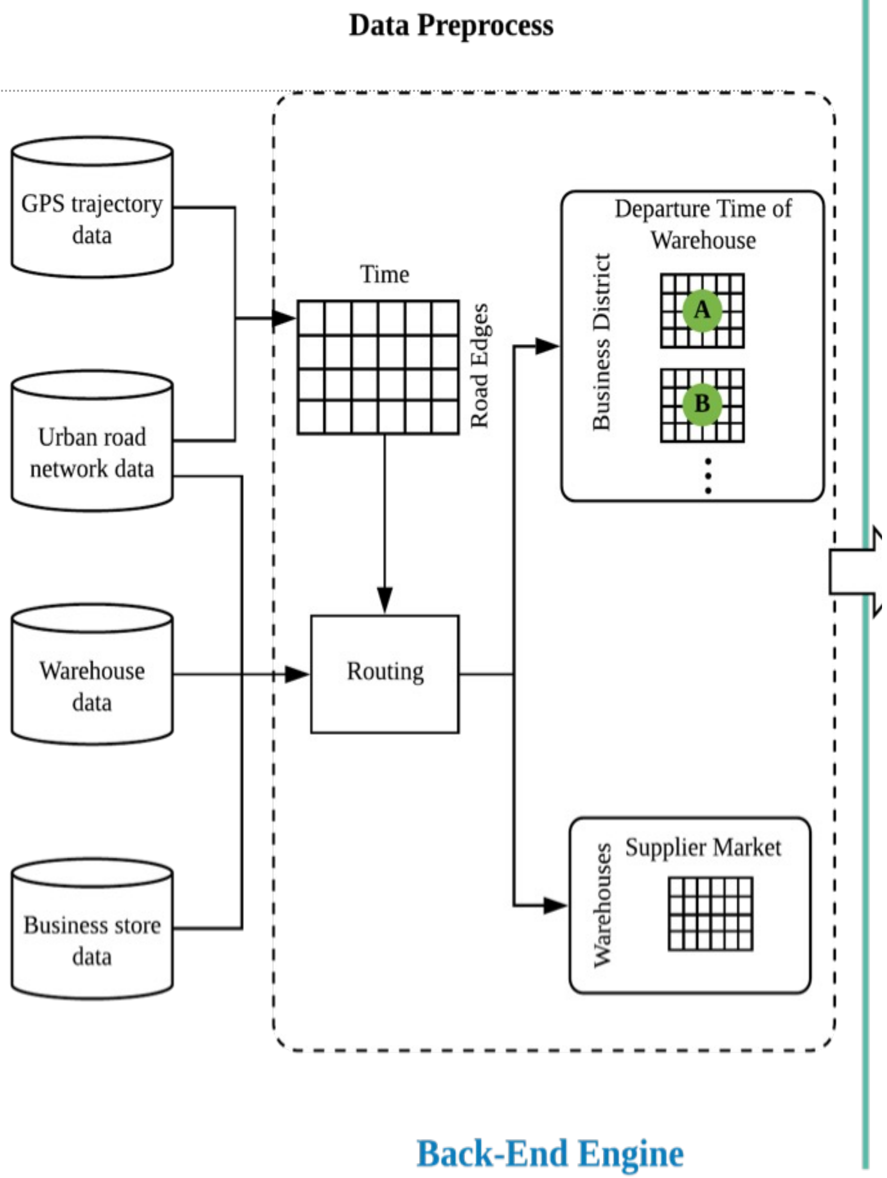
- 将一整天分为48个时间片，并计算每半小时路段的车辆行驶时间

步骤3

- 利用矩阵分解来计算缺失路段的车辆行驶时间

步骤4

- 从一个特定的时间戳开始，计算从一个地方到另一个地方持续时间最短的路径



自动推荐候选仓库列表

Recommending Candidate Warehouse

评估标准

- 单位价格
- 库存容量
- 到商店的平均距离
- 与主要供应商的平均距离
- 附近的主要供应商数量
- 附近高速公路入口的数量

推荐候选仓库

- 通过Simos对标准进行加权
- 通过TOPSIS对候选仓库进行排名

Minimize $\sum_{j\in J} f_j x_j + \sum_{j\in J}\sum_{i\in I} c_{ij} y_{ij} + \sum_{j\in J}\sum_{k\in K} e_{kj} + \sum_{j\in J}\sum_{m\in M} g_{mj}$

subject to

$\sum_{j\in J} y_{ij} = 1, \forall i \in I$

$y_{ij} \le x_j, \forall i \in I, \forall j \in J$

$y_{ij} \ge 0, \forall i \in I, \forall j \in J$

$x_j \in \{0,1\}, \forall j \in J$

目标函数

场景描述

- 商圈地址确定，仓库地址不确定，即有一批候选的仓库可供选址，但不确定要在哪开几个仓库
- 开许多仓库可以降低从仓库到商圈运输的平均距离，从而减少运输成本
- 开许多仓库会增加开仓成本

模型定义

I表示商圈集合，J表示候选的仓库集合目标是选择在哪个位置开仓，因此，对于每个位置我们自定义一个 binary 变量：

$x_j = 1$ (当我们在j位置开仓)，$x_j = 0$(不在j位置开仓)，开仓的成本定义为f_j

- 定义y_{ij}是从仓库j运输到商圈i的货物占比，且$y_{ij} >= 0$，c_{ij}是从候选仓库j到商圈i的运输成本，通常和距离d_{ij}成正比：$c_{ij} = a\, d_{ij}$, a为开车每公里的成本
- e_{kj}是从候选仓库j到高速公路入口k的运输成本，通常和距离f_{kj}成正比: $e_{kj} = a\, f_{kj}$
- g_{mj}是从候选仓库j到批发市场m的运输成本，通常和距离h_{mj}成正比：$g_{mj} = a\, h_{mj}$

自动推荐候选仓库列表

Recommending Candidate Warehouse

灵活定义商圈范围

候选仓深入对比洞察

城市： 上海 / 上海市

选址方案列表 新增方案

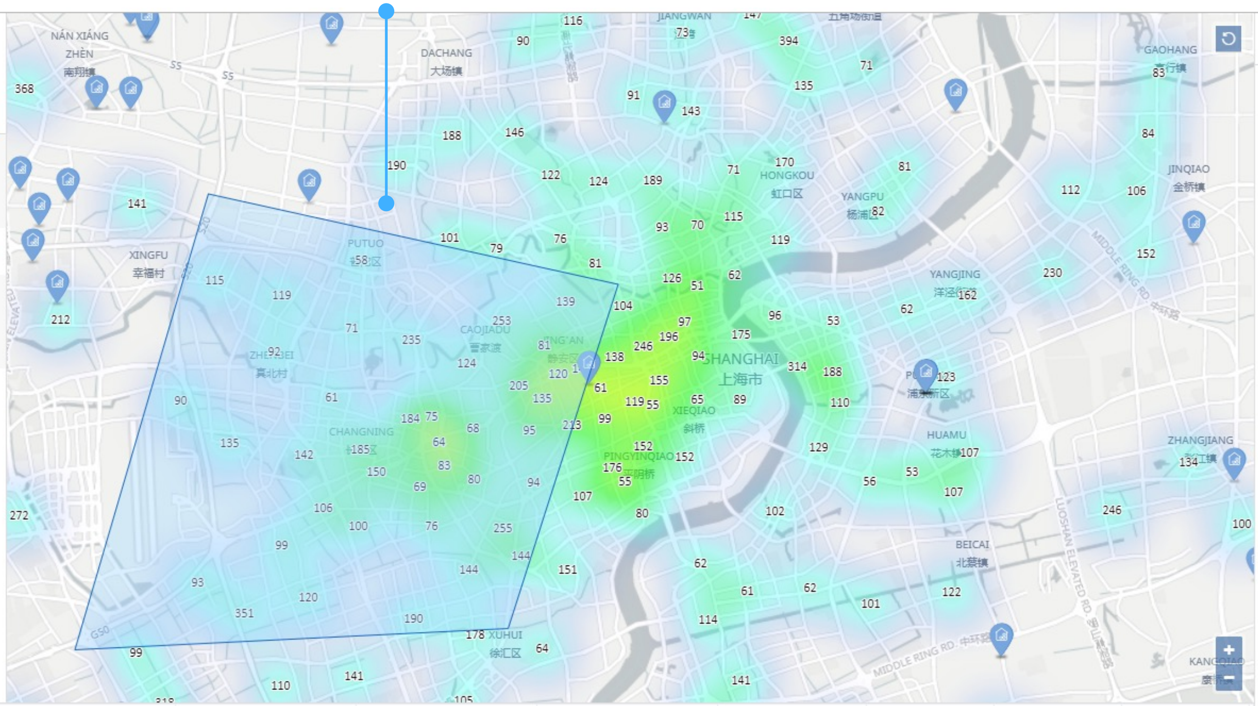
深圳市-ceshi

上海市-113333

上海市-上海西南区域

北京市-测试

广州市-11122222



候选仓 对比洞察 保存方案

南汇区有4000平米恒温仓库出租

恒温仓(5.... 楼库

优质冷冻冷藏仓出租

冷冻仓 高台库

仓库预览 加入候选

松江区有5000平米冷冻仓库出租

冷冻仓(-... 高台库、...

智能推荐合适仓库

多维数据辅助决策

仓库名称	更新日期	租金(元/㎡/月)	仓库面积(㎡)	可租面积(㎡)	起租面积(㎡)	层高(米)	仓库类型	建筑类型	卸货平台
青浦区有13500平米普通仓库出租	2019-06-21	面议	13500	13500	面议	10	普通仓	高台库	装卸平台
嘉定区有33000平米普通仓库出租	2019-07-22	41.4	33000	33000	面议	6	普通仓	楼库	电梯
南汇区有4000平米恒温仓库出租	2019-04-08	90	4000	4000	面议	5	恒温仓(5.00-25...	楼库	装卸平台、电梯
浦东新区有8000平米普通仓库出租	2019-06-20	48	8000	8000	面议	12	普通仓	平库	装卸平台
浦东新区有5500平米普通仓库出租	2019-02-27	48	5500	5500	1900	12	普通仓	单层厂房仓库	吊机、起重葫芦
浦东新区有6000平米普通仓库出租	2019-07-16	45	6000	6000	面议	20	普通仓	其他	龙门吊
浦东新区有6500平米普通仓库出租	2019-07-29	面议	6500	6500	面议	9	普通仓	其他	装卸平台

注意：1. 仓库信息存在更新不及时的情况，准确信息需与仓库主确认。2. 部分页面信息转自网络，如有侵权请联系删除。

仓库推荐模型

↕

交互式可视化

Task #1: 以交互方式协调条件约束

Task #2: 直观地探索商圈与仓库

Task #3: 交互对比候选仓库

Task #4: 有效过滤和排序候选集合

利用仓库周边的道路轨迹数据，推导出可达性，关注不同时间段的可达性

智能仓储选址 – 仓库洞察

Smart Warehouse Site Selection

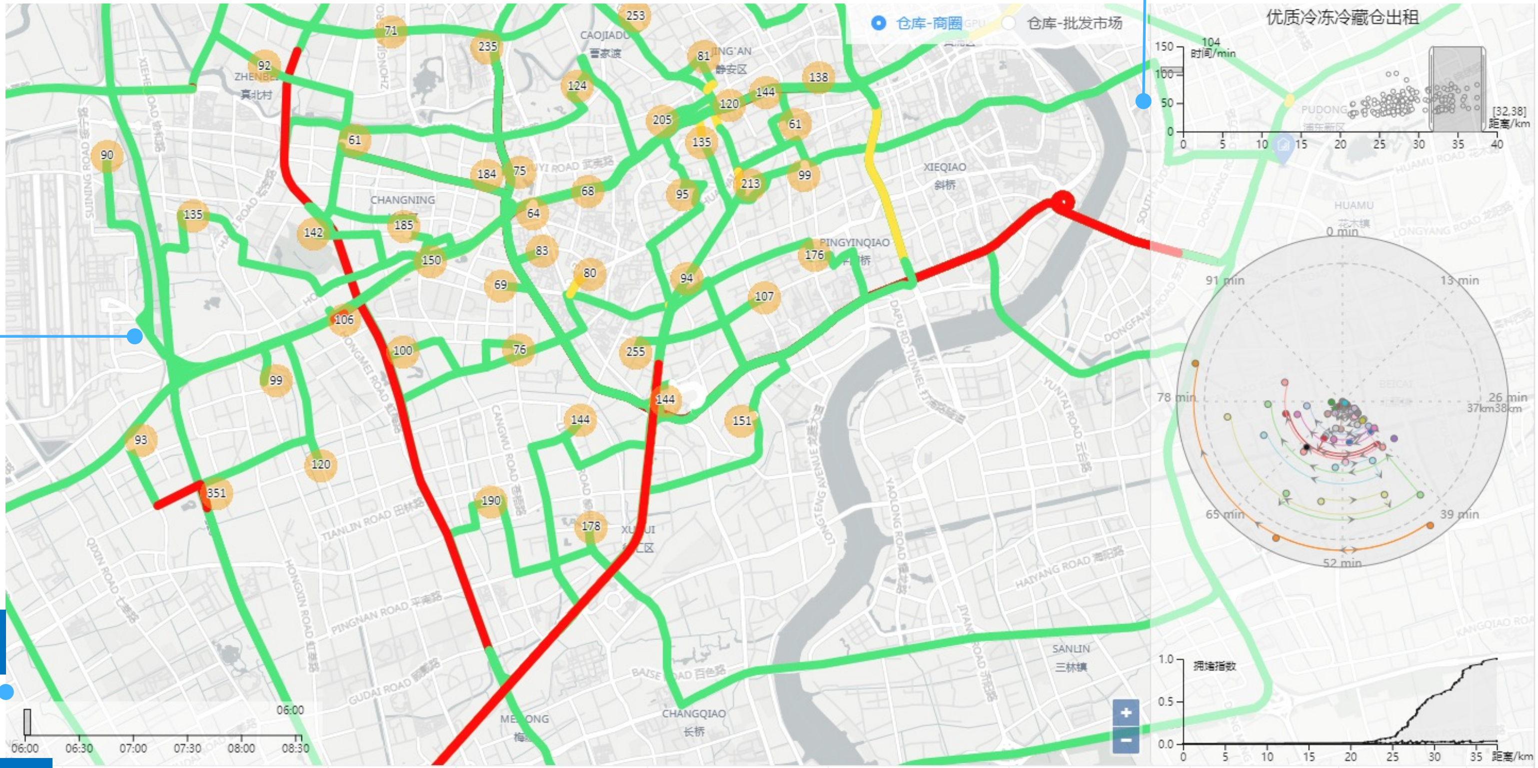
仓库画像，洞察距离与路况对仓库->商圈可达性的影响

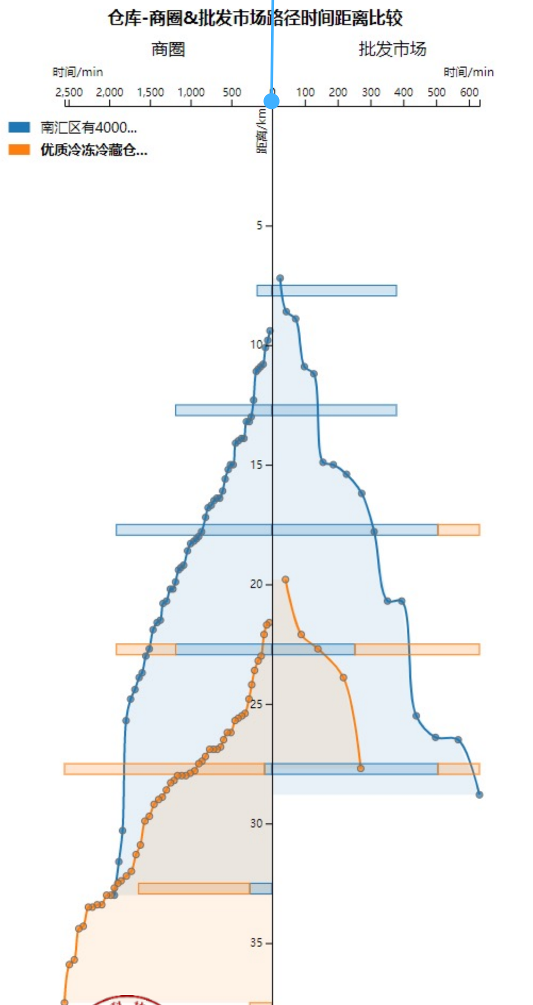
仓库到商圈、高速路口、批发市场的距离、耗时透视

仓库交通优劣势一目了然

根据业务灵活设置时间

更多维仓库信息辅助决策





	推荐仓库	更新日期	租金(元/㎡/月)	仓库面积(㎡)	可租面积(㎡)	起租面积(㎡)	层高(米)	仓库类型	建筑类型
☑	南汇区有4000平米恒温仓库出租	2019-04-08	90	4000	4000	面议	5	恒温仓(5.00-25...	楼库
☑	优质冷冻冷藏仓出租	2019-05-06	114	10000	10000	100	9	冷冻仓	高台库
☐	闵行区有20000平米冷冻仓库出租	2019-03-14	150	20000	5000	面议	6	冷冻仓(0.00--2...	楼库
☐	松江区有5000平米冷冻仓库出租	2019-06-10	90	5000	5000	面议	8	冷冻仓(-18.00-...	高台库、

上海科技大学 ShanghaiTech University

智能仓库选址

[EuroVis2020论文投稿版]

WarehouseVis: A Visual Analytics Approach to Facilitating Warehouse Location Selection for Business Districts

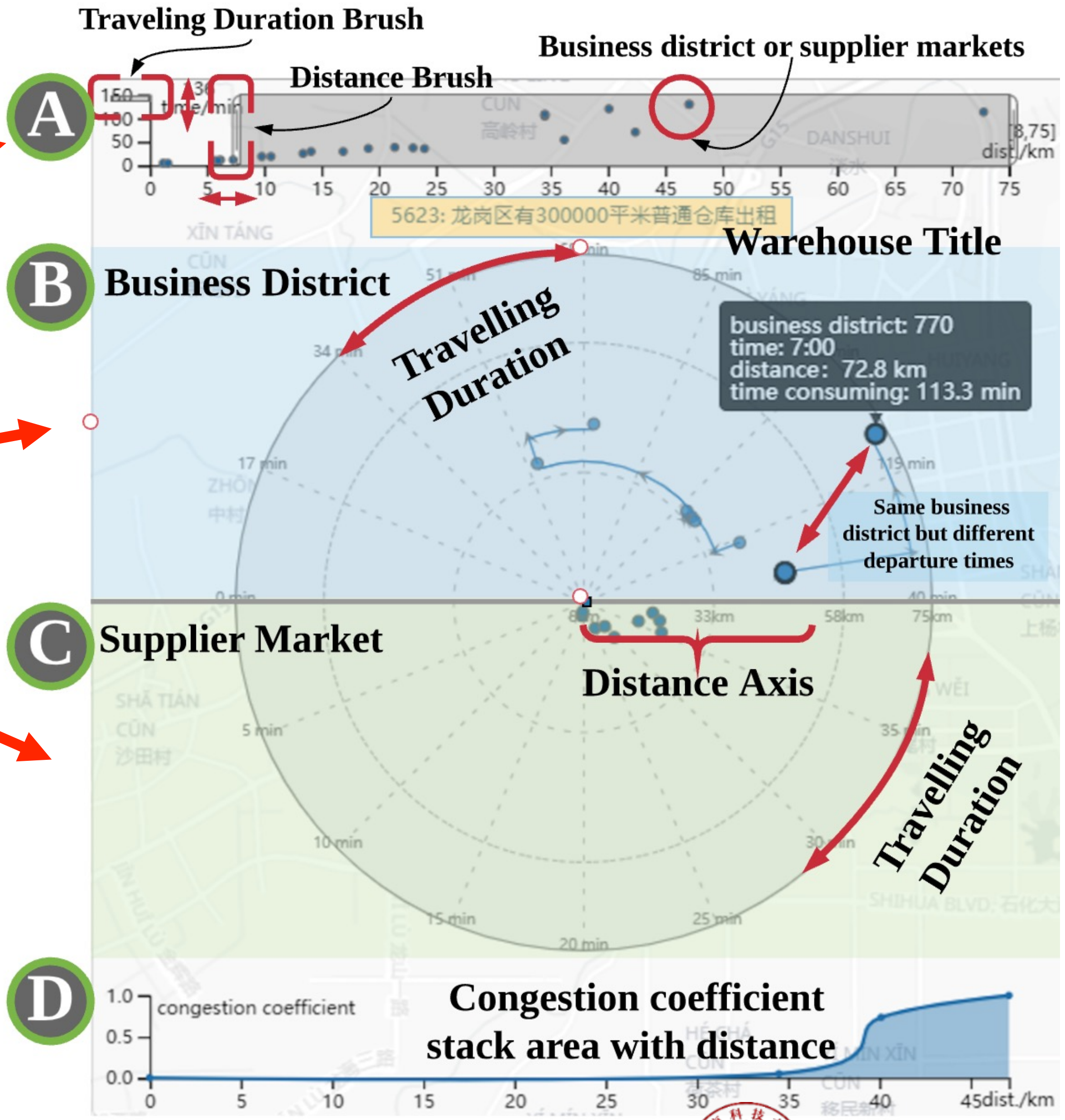
智能仓储选址 - 仓库洞察

Smart Warehouse Site Selection

为用户提供交互，过滤商业区或供应商市场

B和C分别表示商业区和供应商市场的分布情况

仓库的累计拥堵系数与距离的关系



上海科技大学
ShanghaiTech University

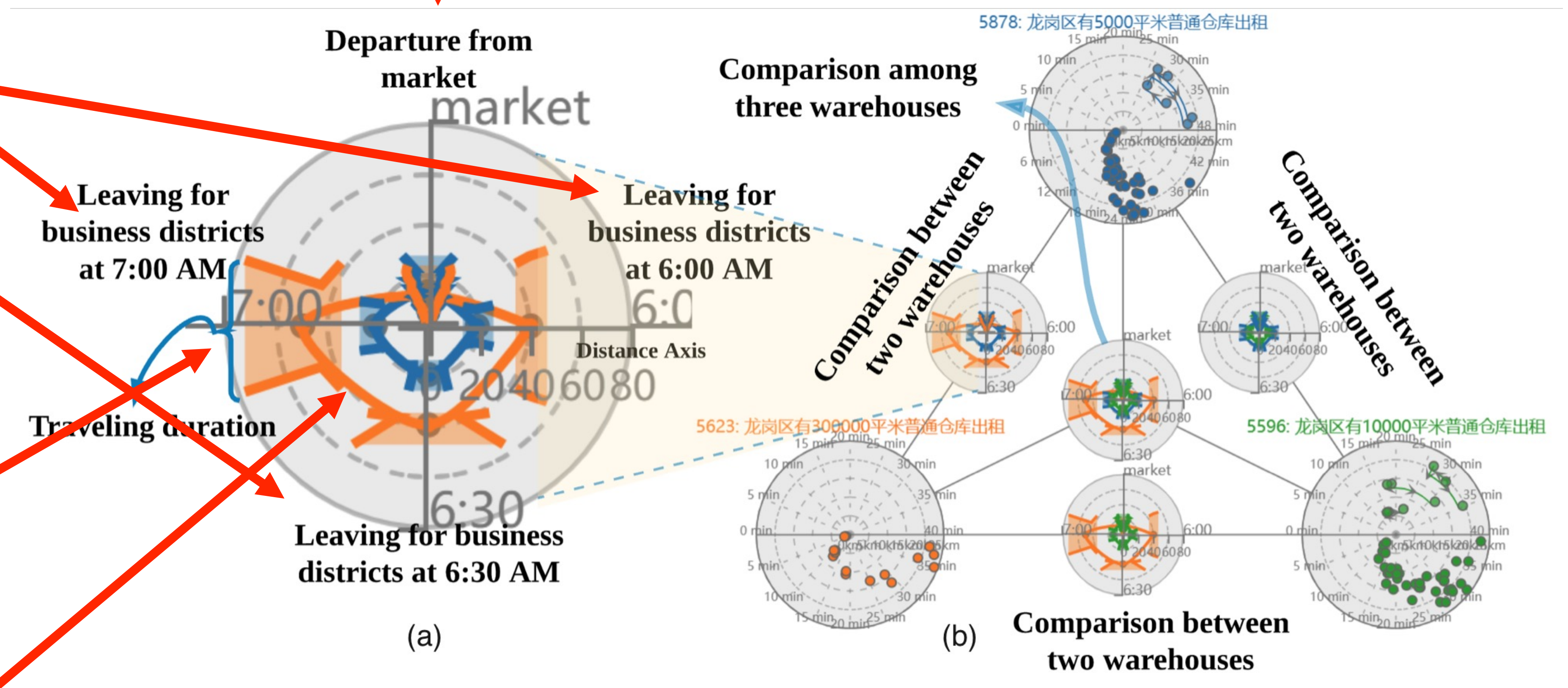
智能仓储选址 - 仓库洞察

Smart Warehouse Site Selection

到供应商市场的距离

从仓库出发的三个时间点到商业区的运输距离

- 从仓库到商业区或市场的归一化后的运输时间
- 相应轴上运输时间的中值



选定的候选仓库的因素

最多可以同时比较三个仓库

上海科技大学 ShanghaiTech University

智能仓库选址

[EuroVis2020论文投稿版]

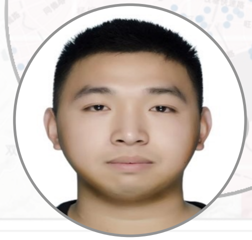
WarehouseVis: A Visual Analytics Approach to Facilitating Warehouse Location Selection for Business Districts

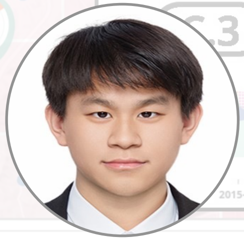
VIS 2023

上海科技大学 ShanghaiTech University | ViSeer LAB 交互智能与可视分析实验室

FSLens: A Visual Analytics Approach to Evaluating and Optimizing the Spatial Layout of Fire Stations


Longfei Chen[1]


He Wang[1]


Yang Ouyang[1]


Yang Zhou[2]


Naiyu Wang[2]


Quan Li[1]

1. School of Information Science and Technology, ShanghaiTech University
2. College of Civil Engineering and Architecture, Zhejiang University

E-mail: chenlf@shanghaitech.edu.cn

Website: https://chenlf126.github.io/

Background

life, property and

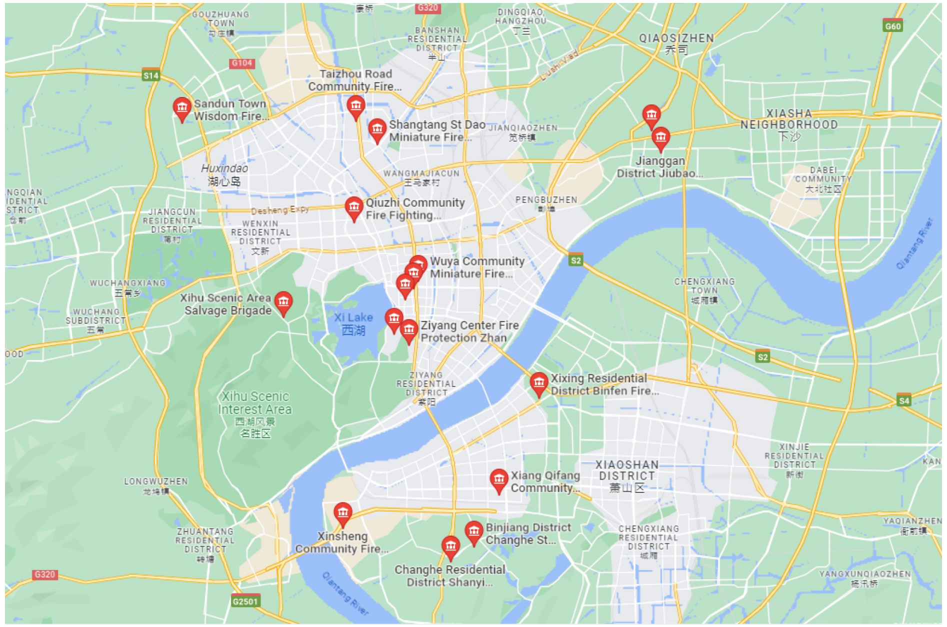
Fire Station

Fire Scene

ViSeer LAB

Ideal

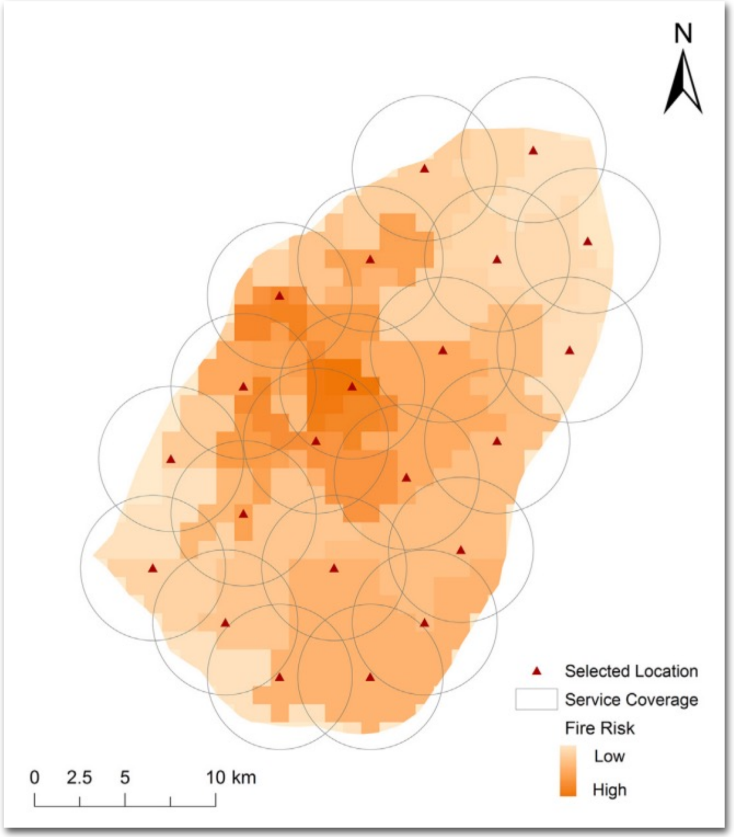
Reality

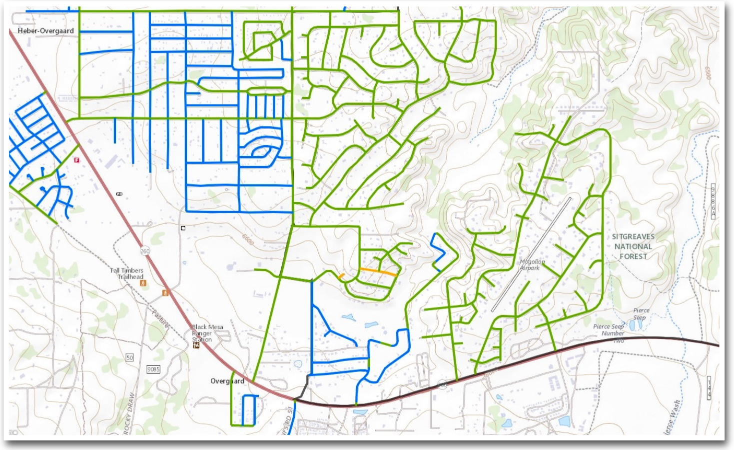


VIS 2023

Previous Studies

- Based on the **Location Set Covering Problem (LSCP)**
- Considering **multiple criteria**, such as
  - Distance
  - Time
  - Overlap
  - ……
- Using **Geographic Information System (GIS)** as an aid





Limitations of Previous Approaches

- **Interpretability** is needed to support decision-making.

- **Inherent conflicts** in multi-criteria planning.

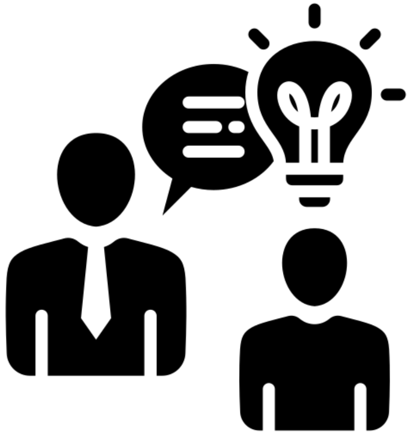
- **Tug-of-war** between the whole and parts.

Our Approach

Observational Study

Expert Team

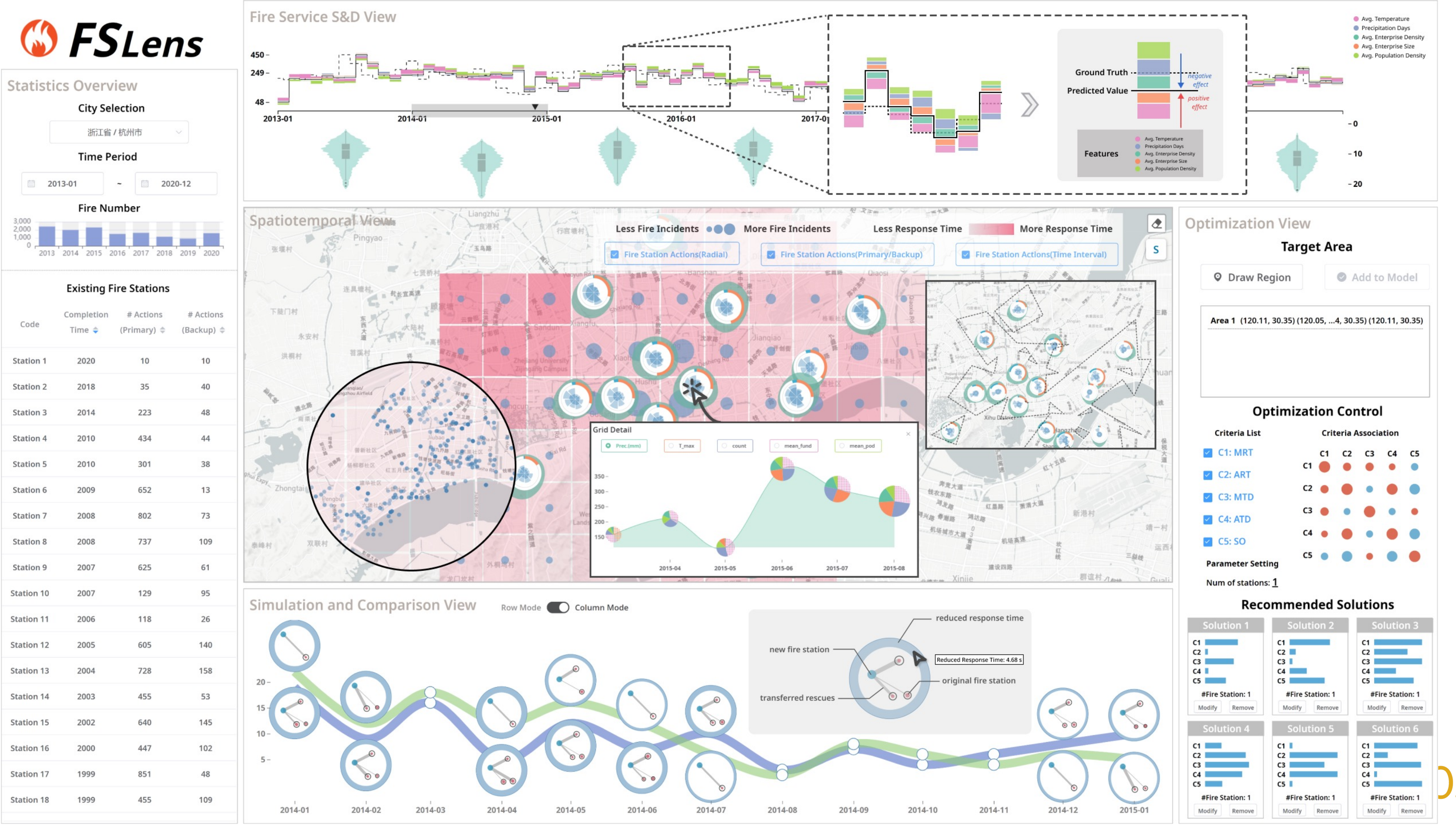
- One fire service manager (E1, Male, Age: 31)
- Two researchers (E2, Female, Age: 37 & E3, Male, Age: 26)



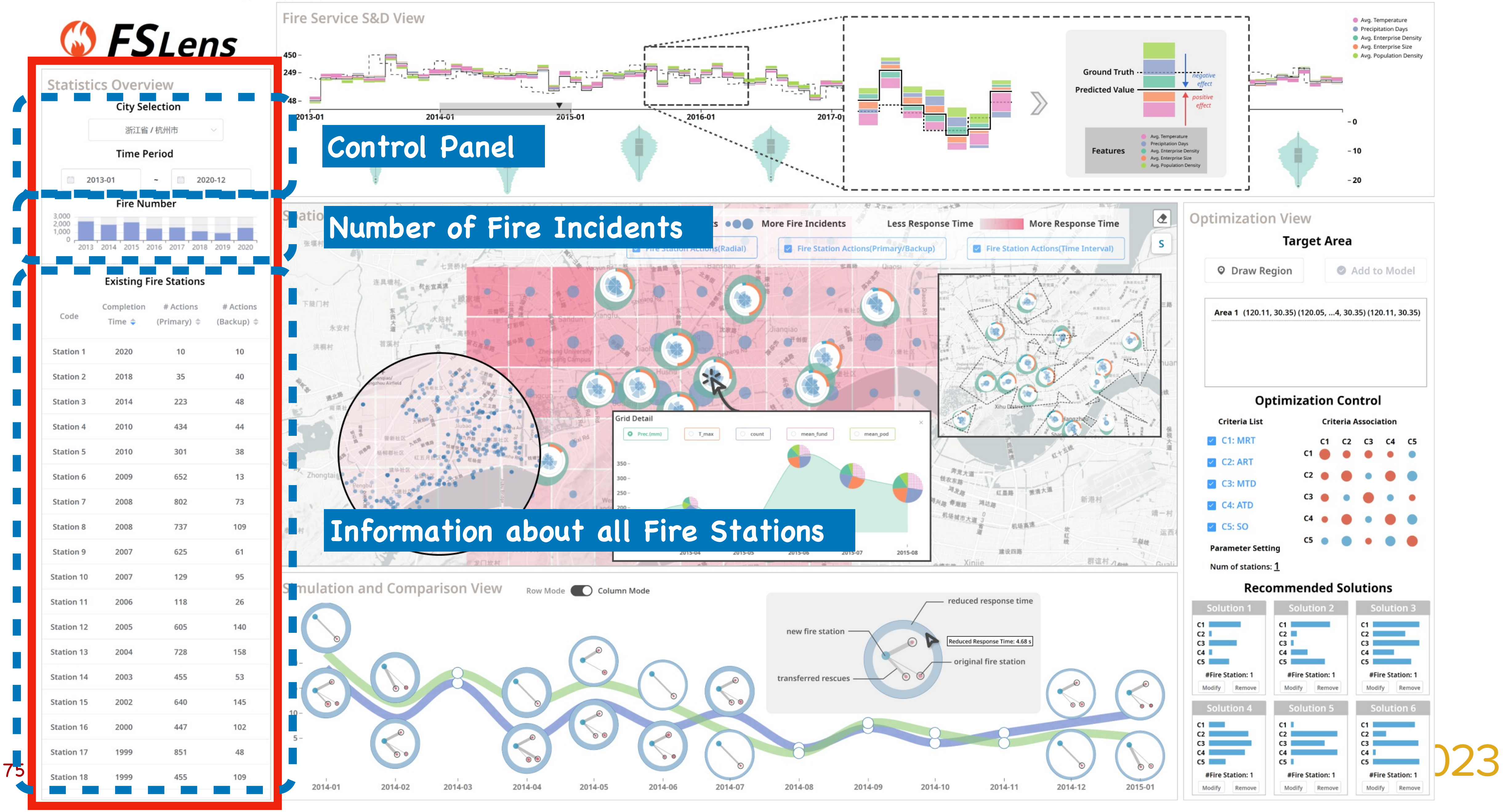
Experts' Needs and Expectations

- **R.1** Reveal patterns and underlying factors contributing to fire incidents.
- **R.2** Assess the rescue capability of the current fire station layout.
- **R.3** Identify and explore regions where fire service resources are scarce.
- **R.4** Provide viable solutions for optimizing the fire station layout.
- **R.5** Model the effect of additional fire stations on the overall layout.

System Overview

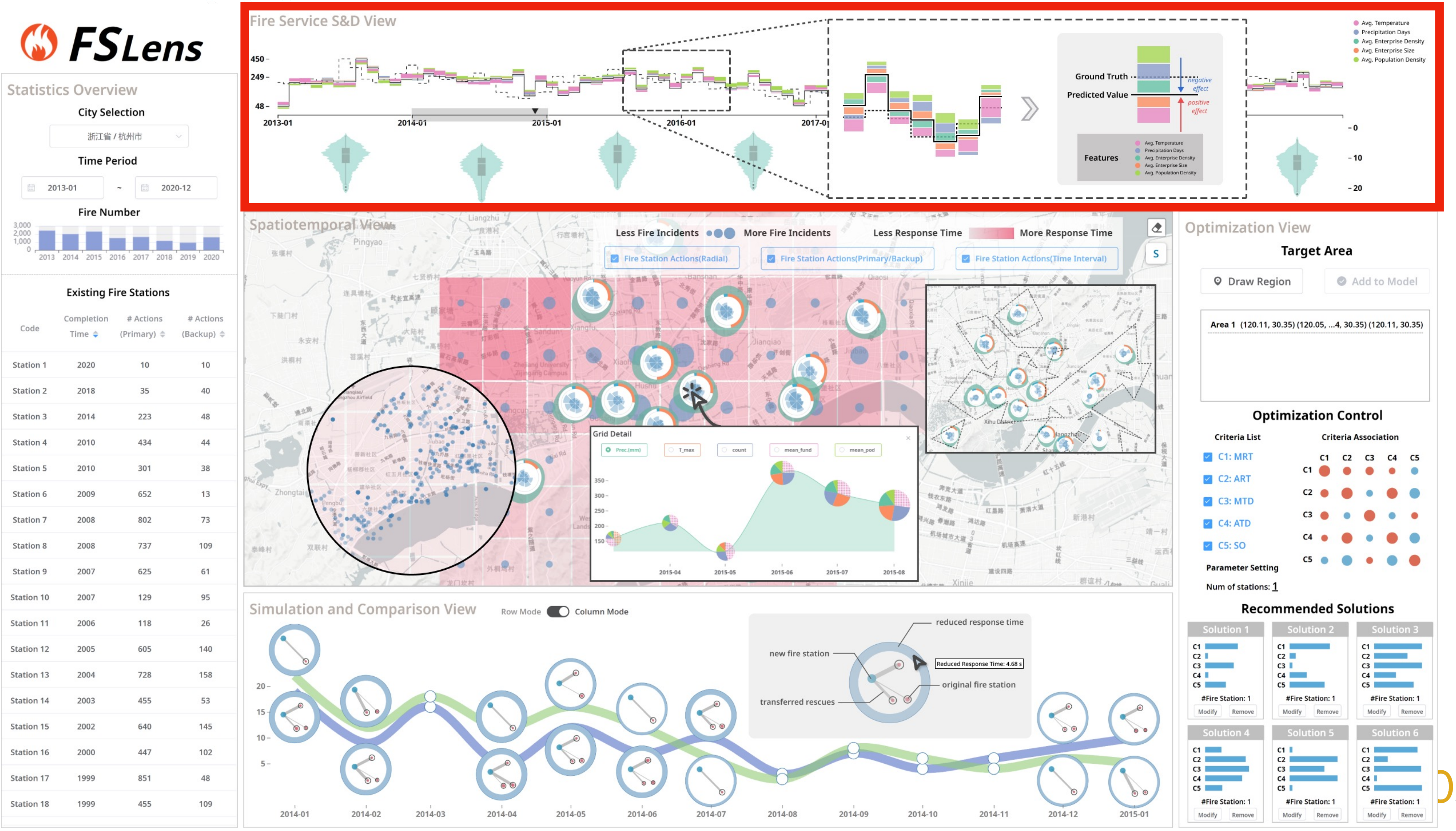


System Overview



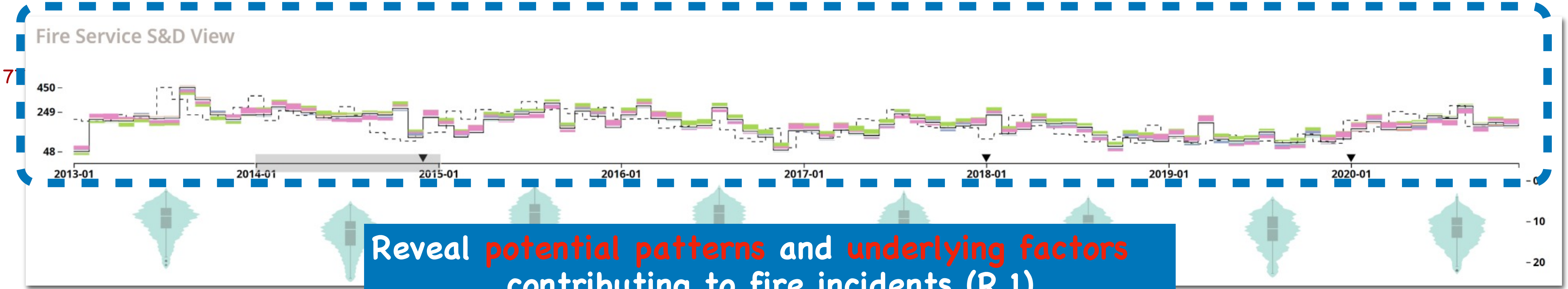
Our Approach

System Overview



76

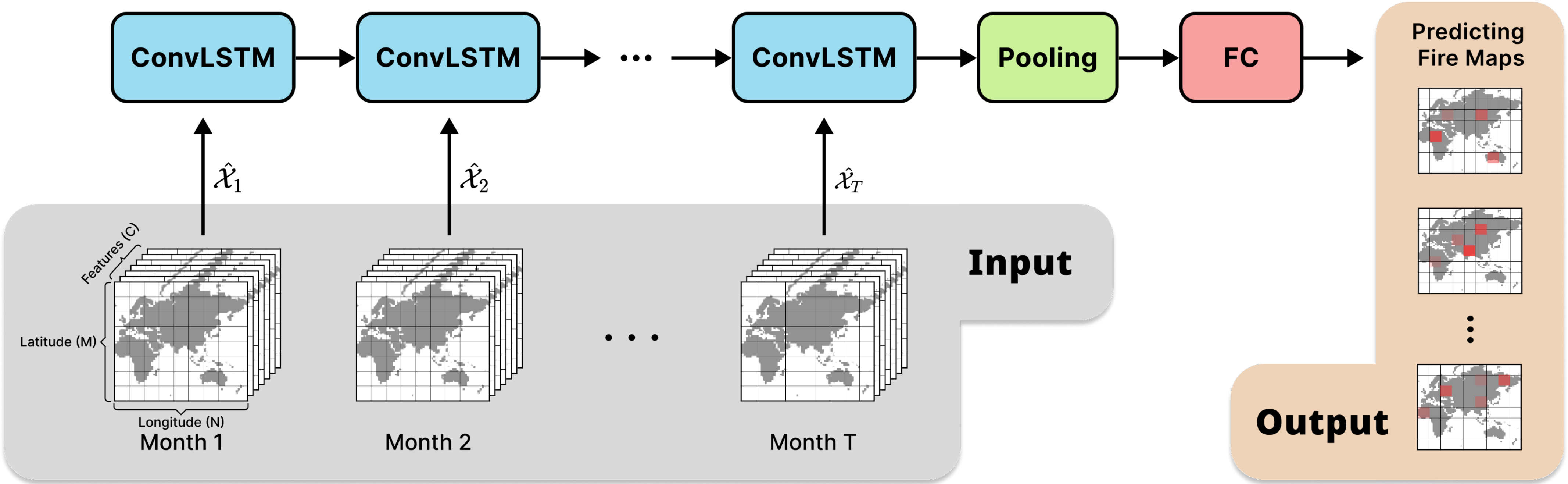
Fire Service Supply & Demand View



Reveal potential patterns and underlying factors contributing to fire incidents (R.1)

Spatiotemporal Sequence Forecasting

78



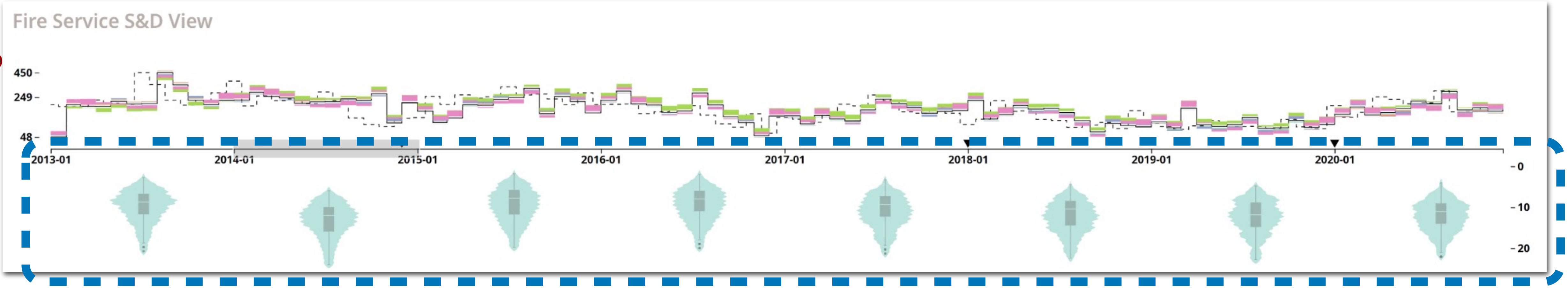
Interpretability Visualization

79

SHAP

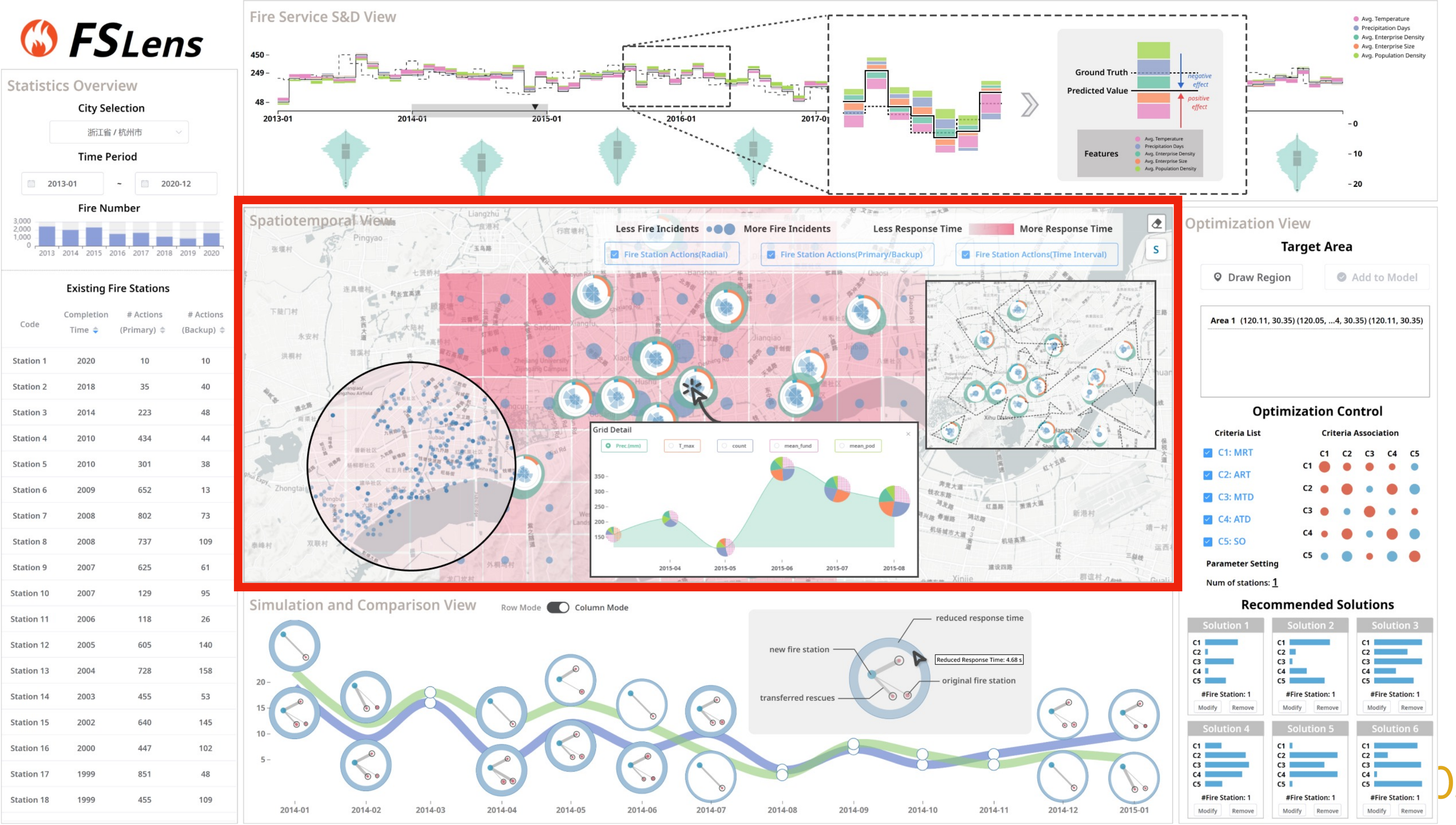
Fire Service Supply & Demand View

80



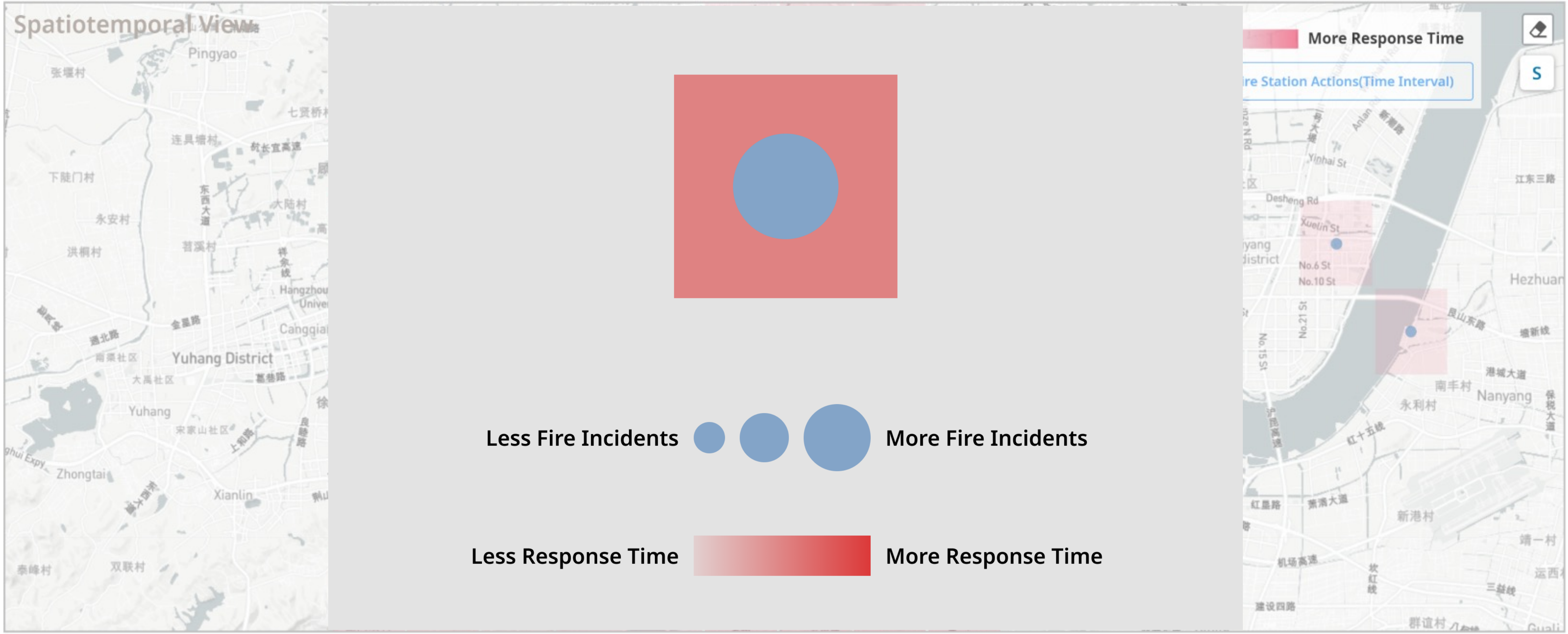
Demonstrate the distribution of response times_ taken by fire stations to reach fire scenes

System Overview



2023

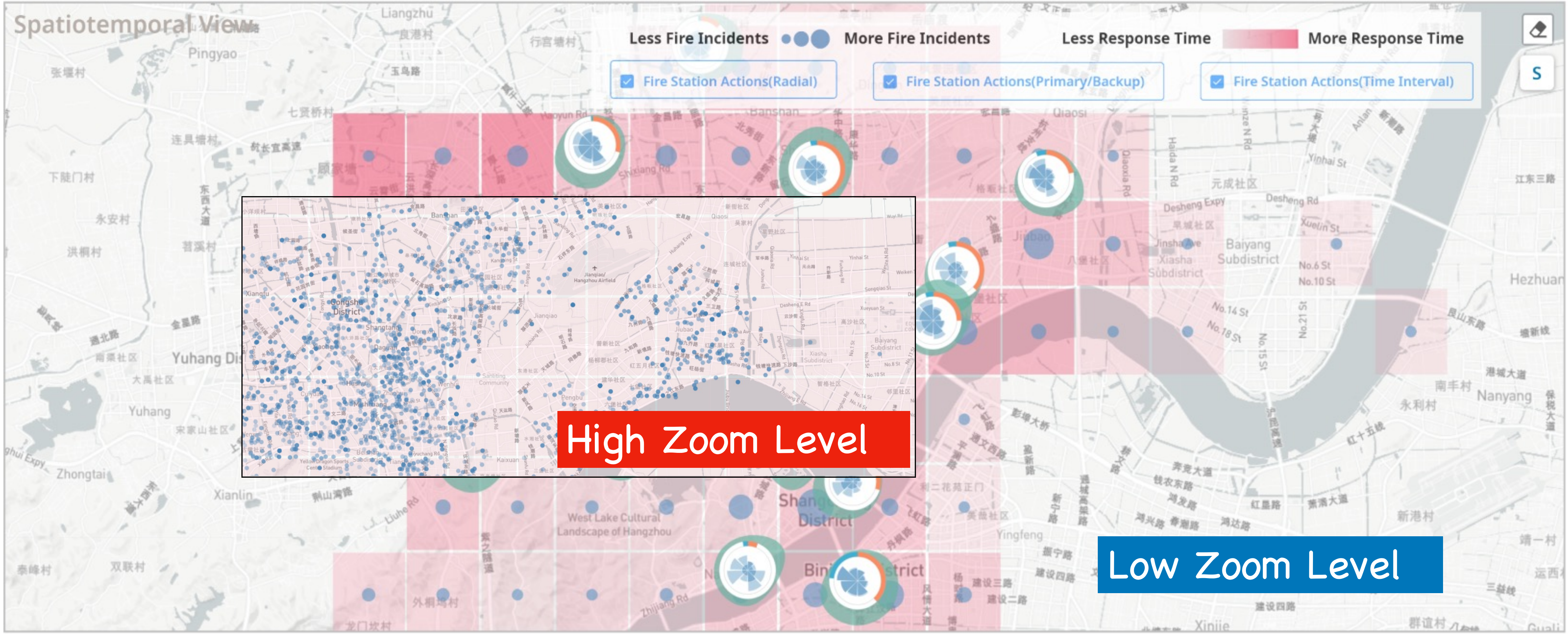
Hierarchical Grid Layer



Spatiotemporal View

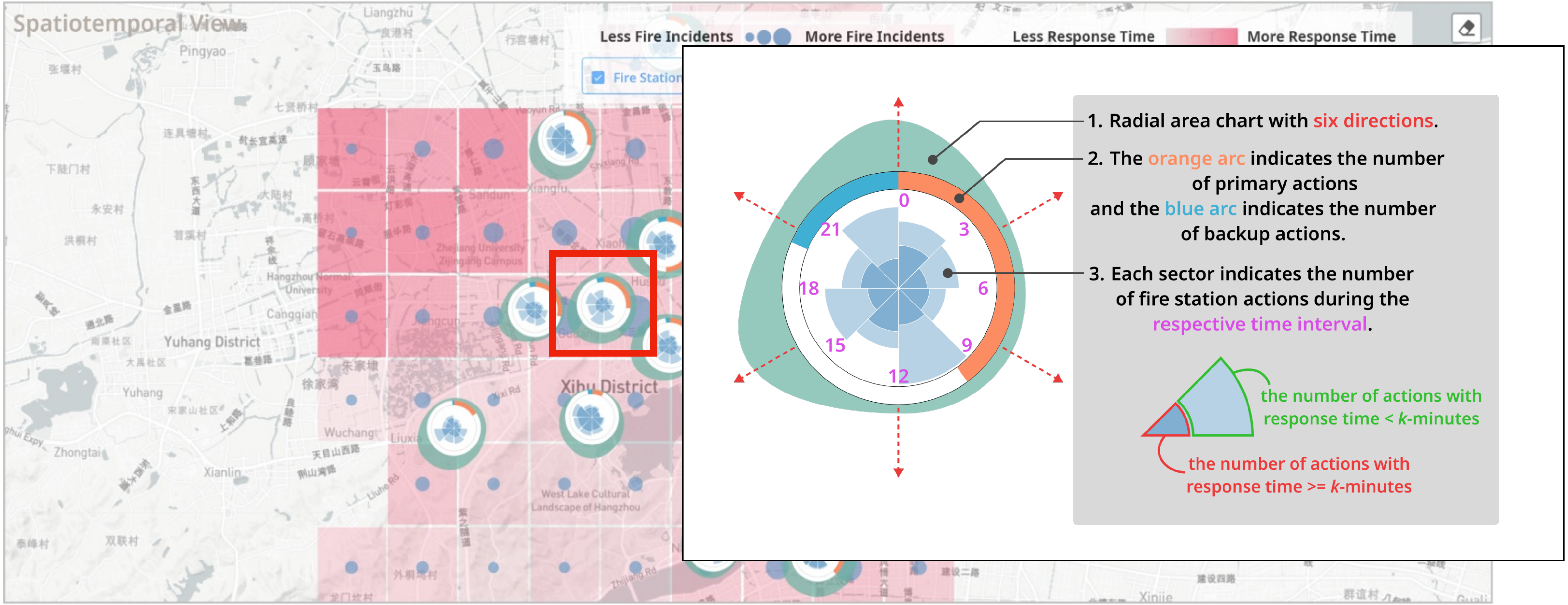
Hierarchical Grid Layer

83



VIS 2023

Enhanced Radial Glyph



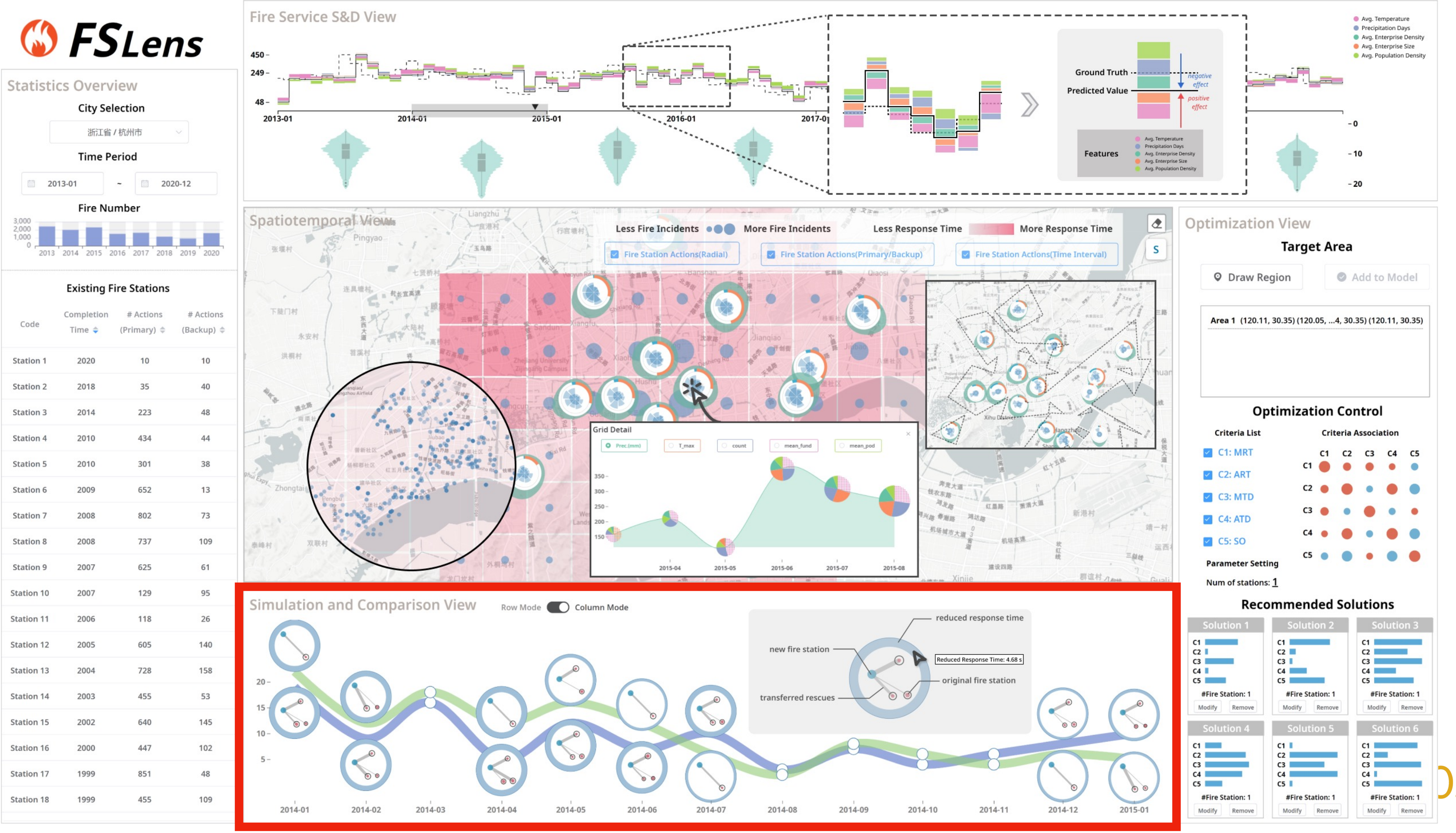
Our Approach

System Overview

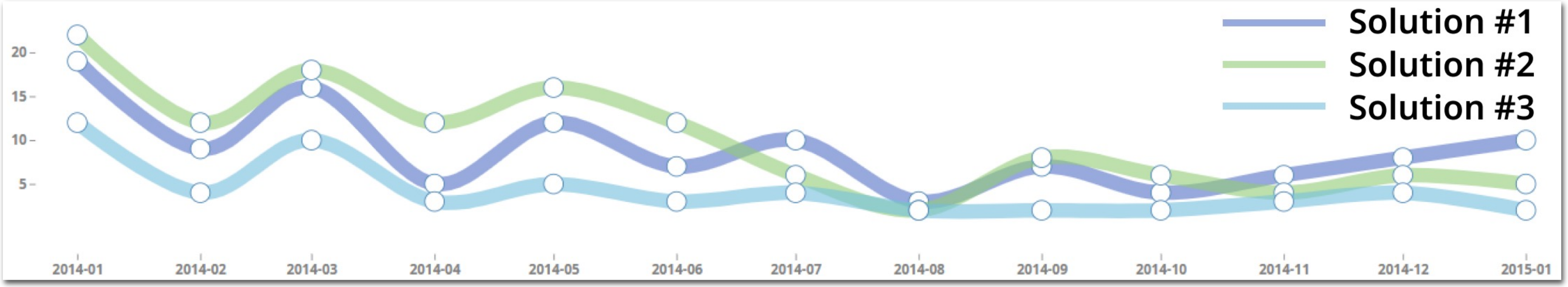
85

Our Approach

System Overview



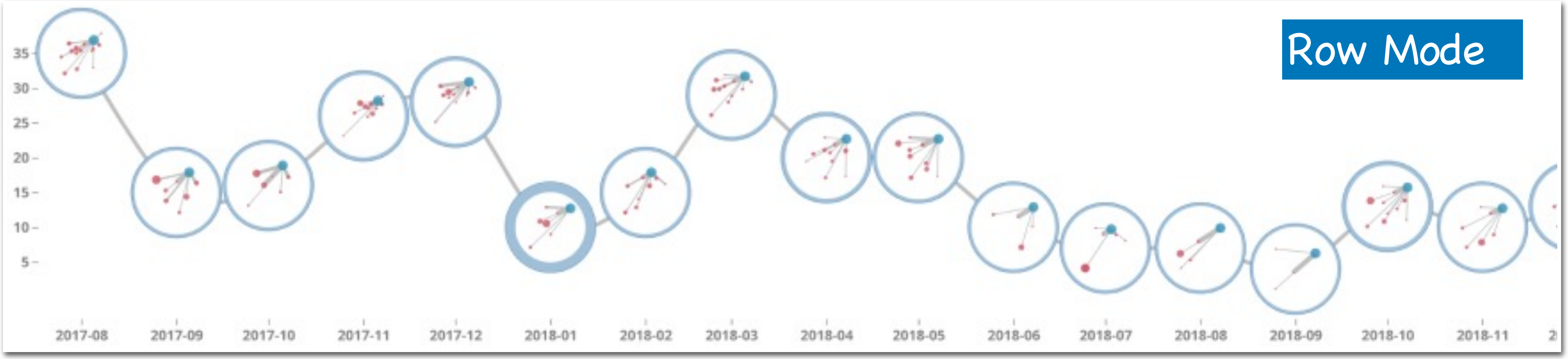
Simulation and Comparison View

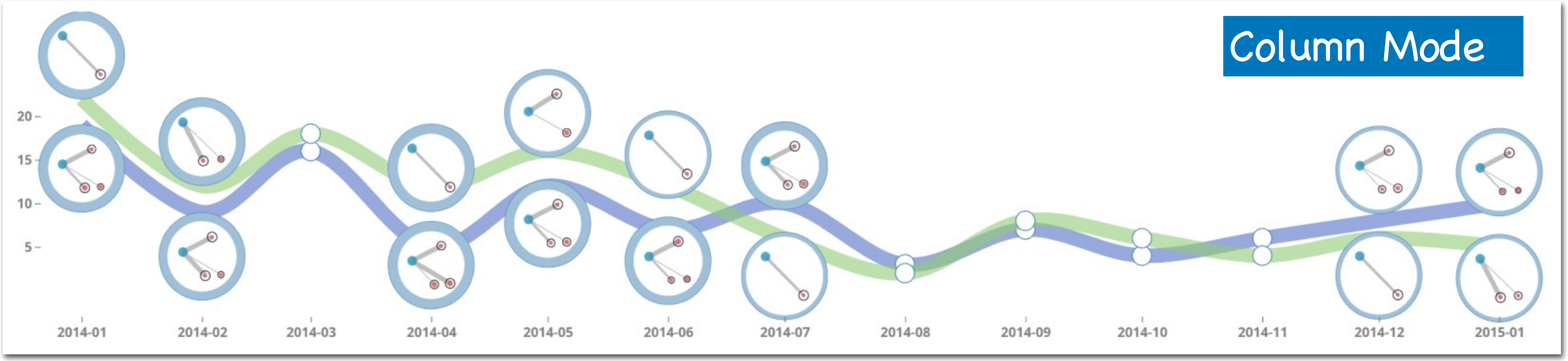


VIS 2023

Simulation and Comparison View

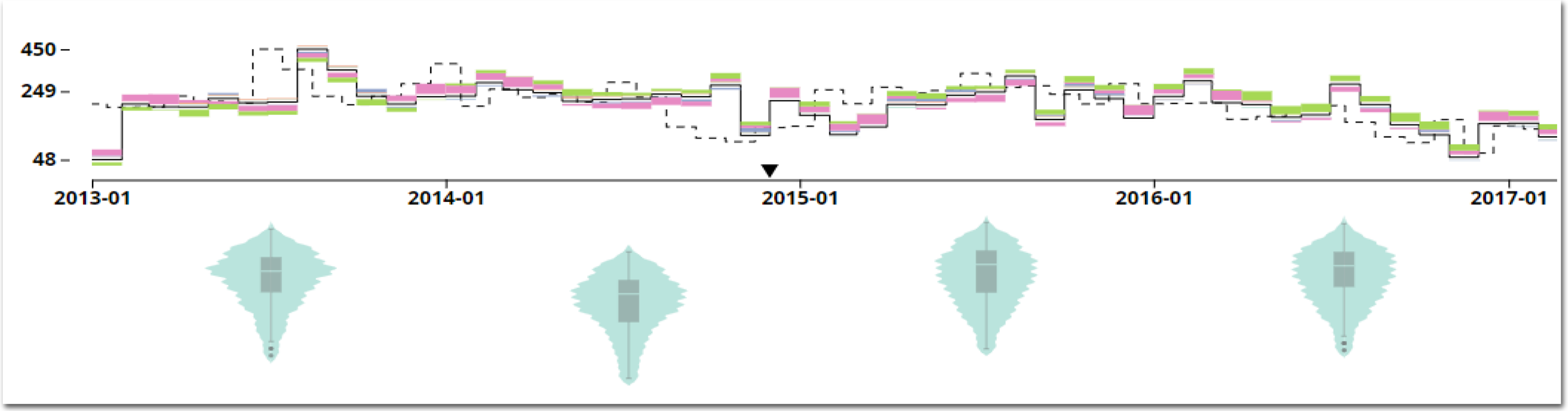
Simulation and Comparison View





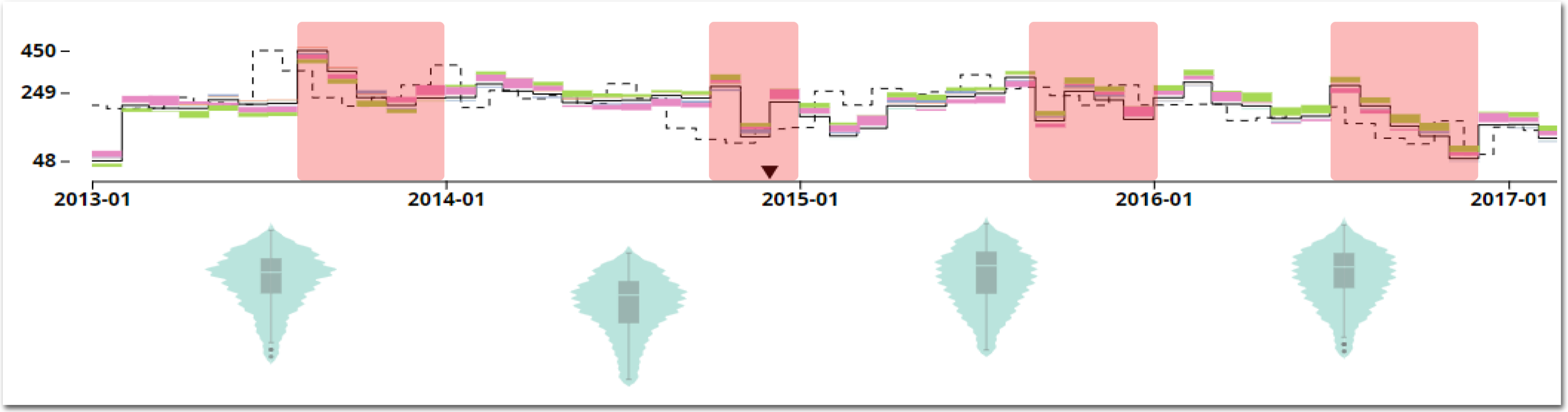
Case Study

Case Study

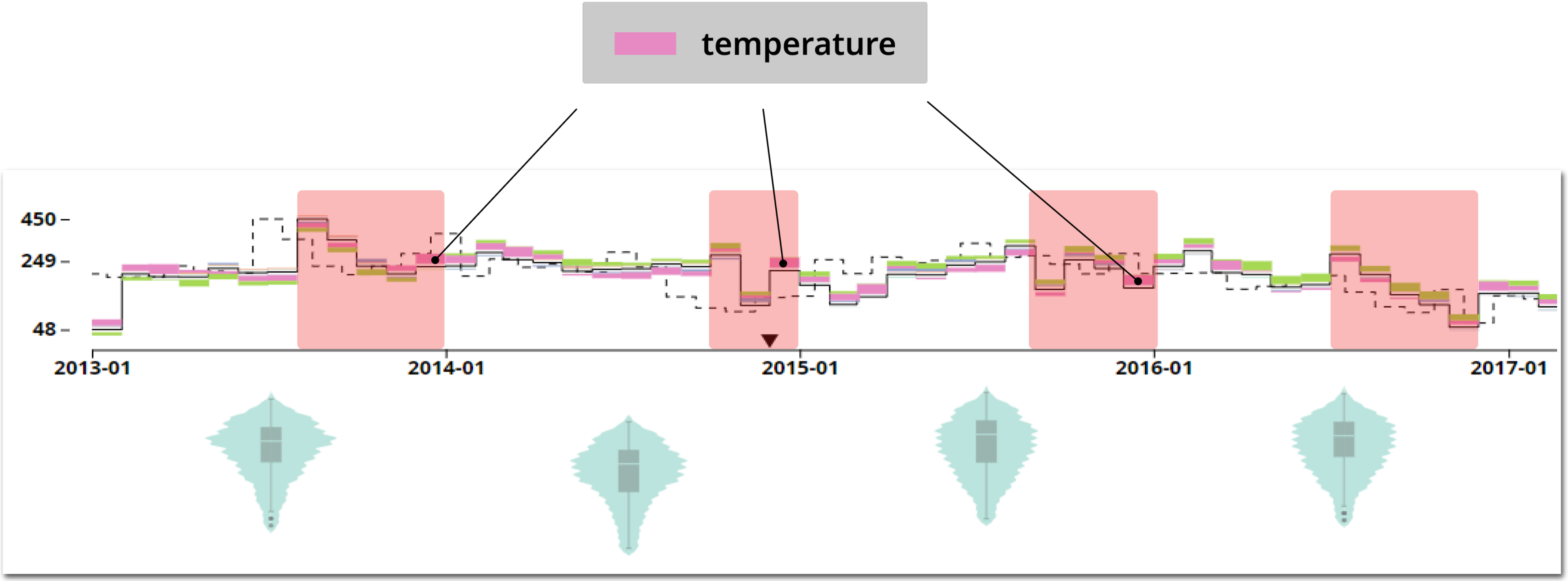


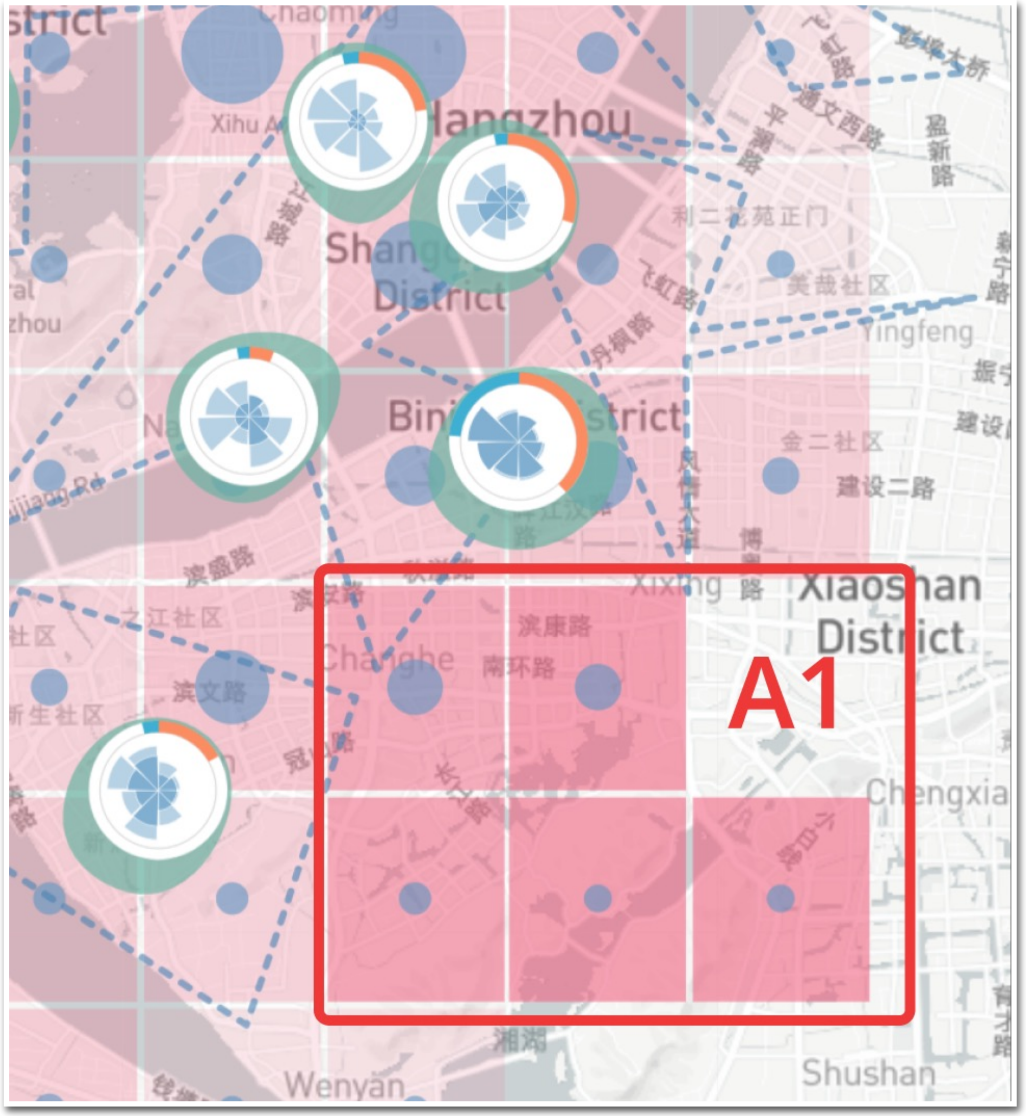
92

Annual Dip

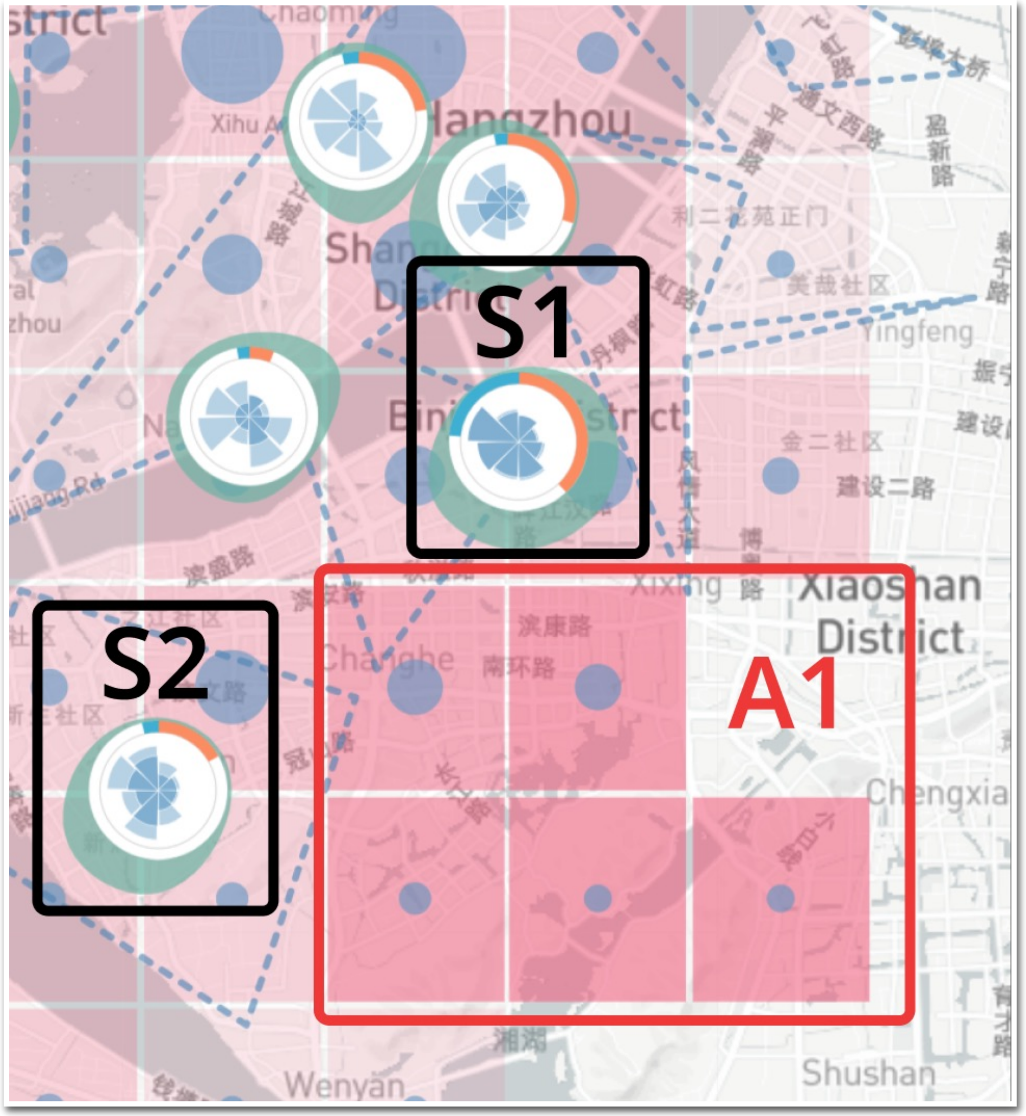


93



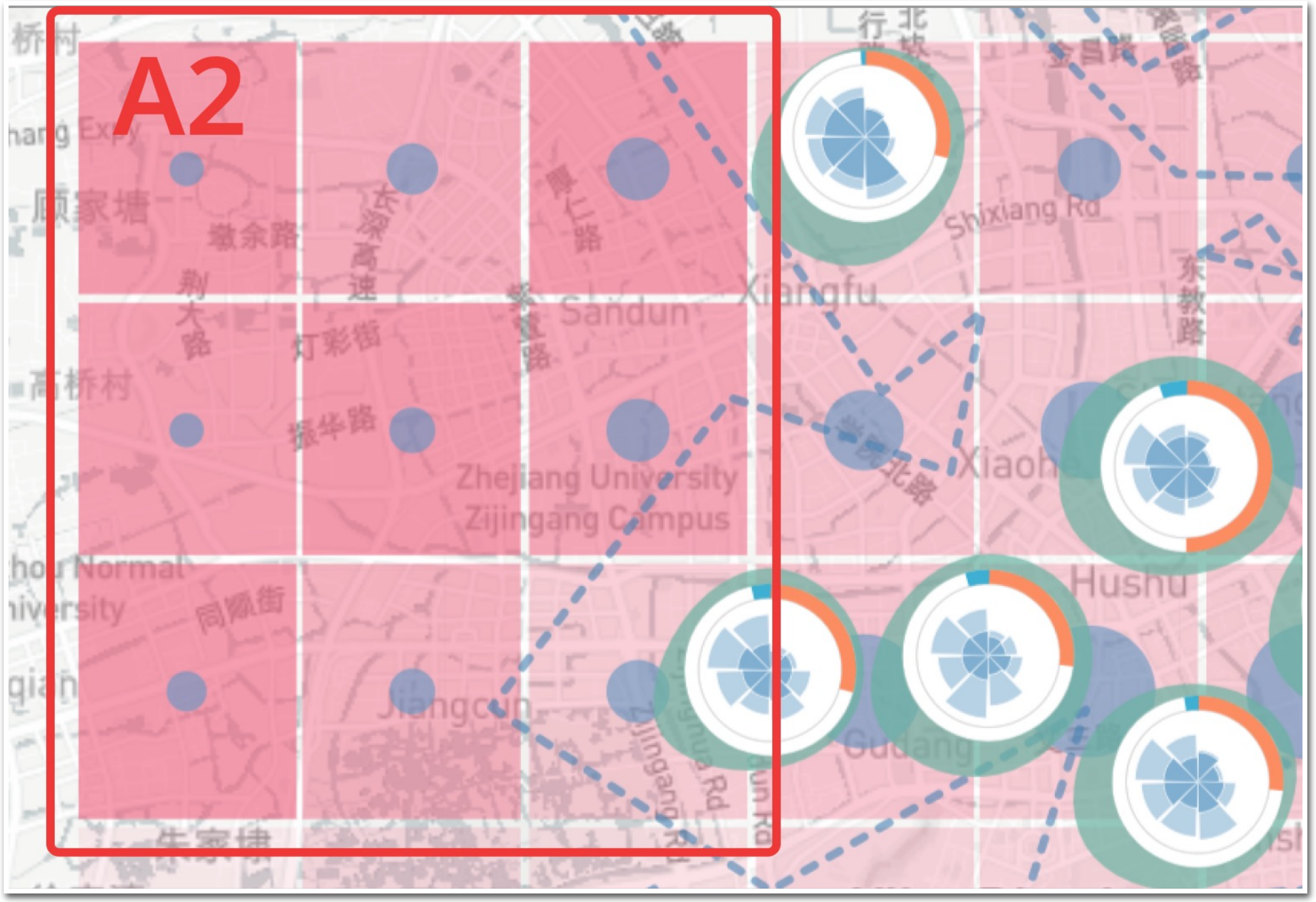


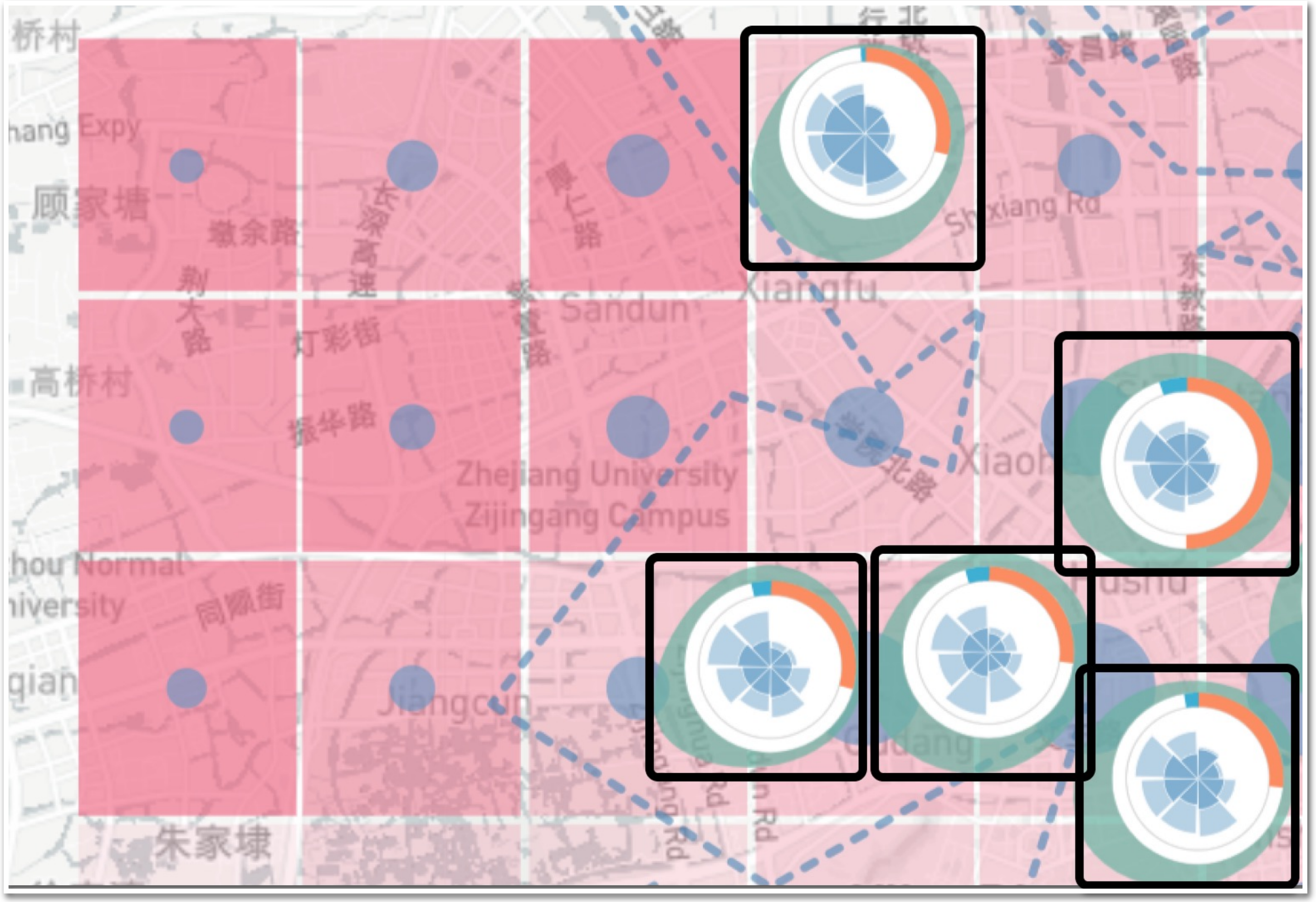
Hangzhou
Shangc
District
Bin
strict
Xiaoshan
District
A1
Changhe
Chengxia
Shushan
Wenyan
Yingfeng
Xixing
滨康路
南环路
湘湖
滨文路
滨盛路
建设二路


S1
S2
A1
Hangzhou
Xiaoshan
District
Yingfeng
Chengxia
Shushan
Wenyan
Changhe
Xixing
Xihu
Chaoming
湘湖
滨康路
南环路
建设二路
金二社区
美哉社区

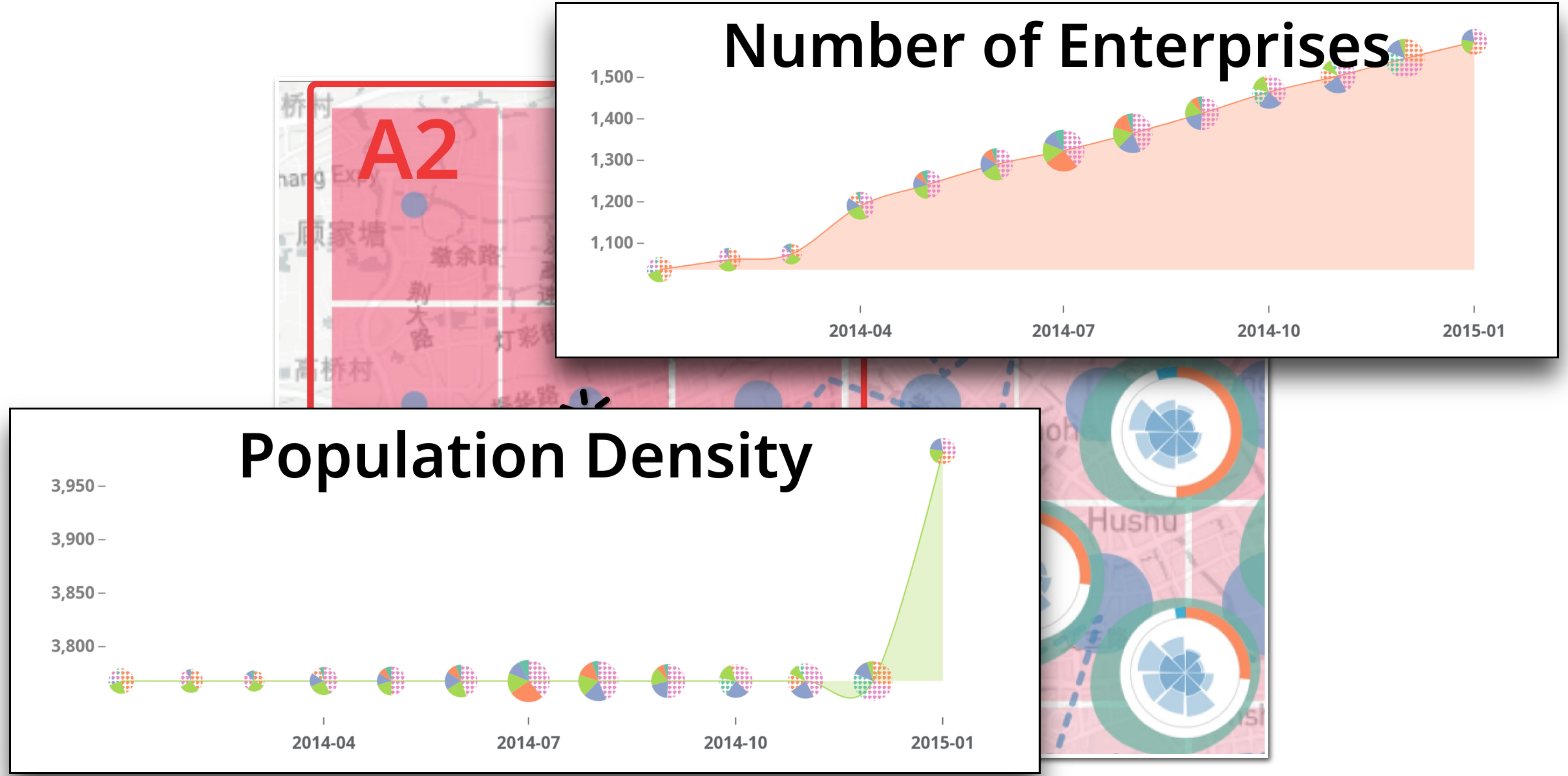

S1
S2
A1
Hangzhou
Xiaoshan District
Chengxia
Shushan
Wenyan
Yingfeng
Binjiang District
Changhe
Xixing

97


A2
Sandun
Xiangfu
Shixiang Rd
Zhejiang University
Zijingang Campus
Xiaoh
Hushu
Jiangcun
Gudang
Zijingang Rd

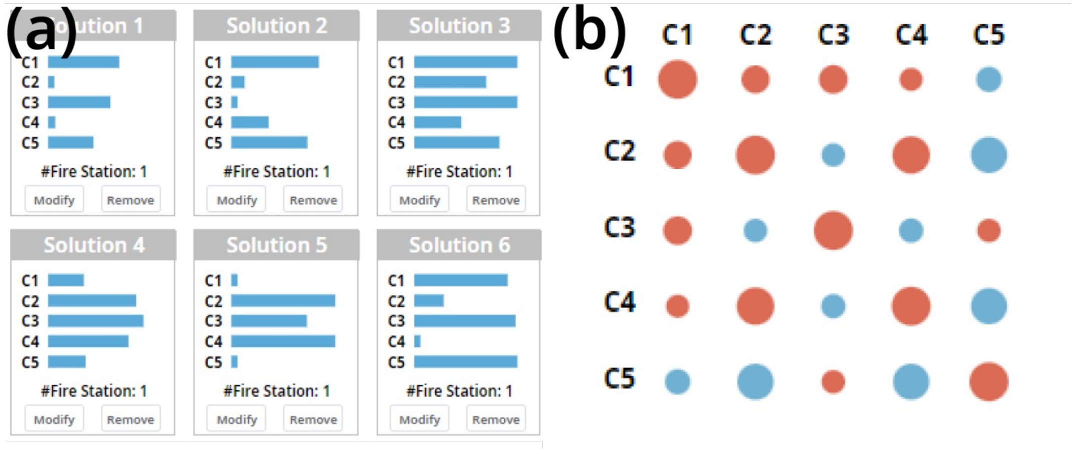

桥村
hang Expy
顾家塘
高桥村
Sandun
Xiangfu
Zhejiang University
Zijingang Campus
hou Normal
iiversity
qian
Jiangcun
朱家埭
灯彩街
振华路
Sh xiang Rd

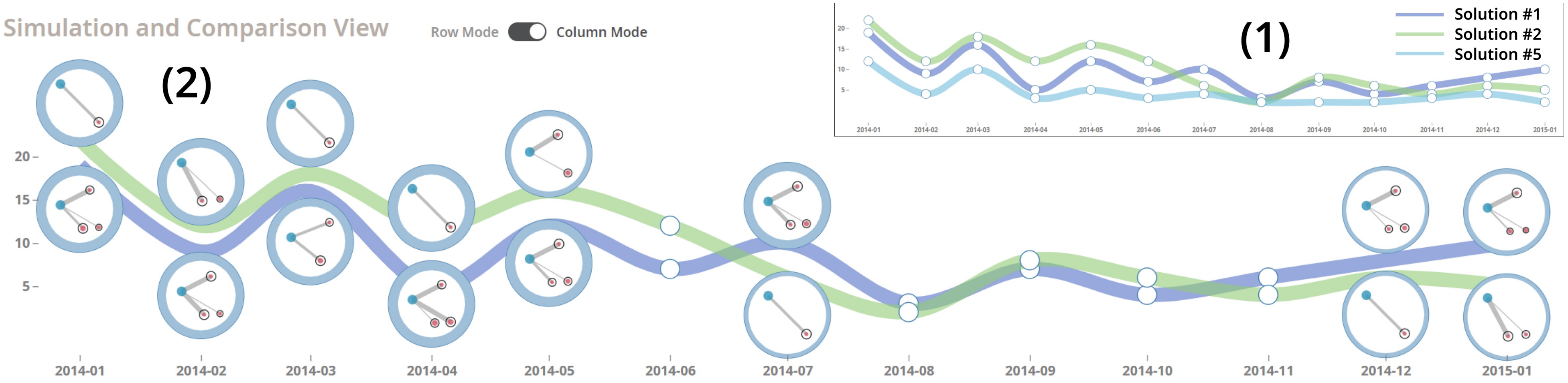
98

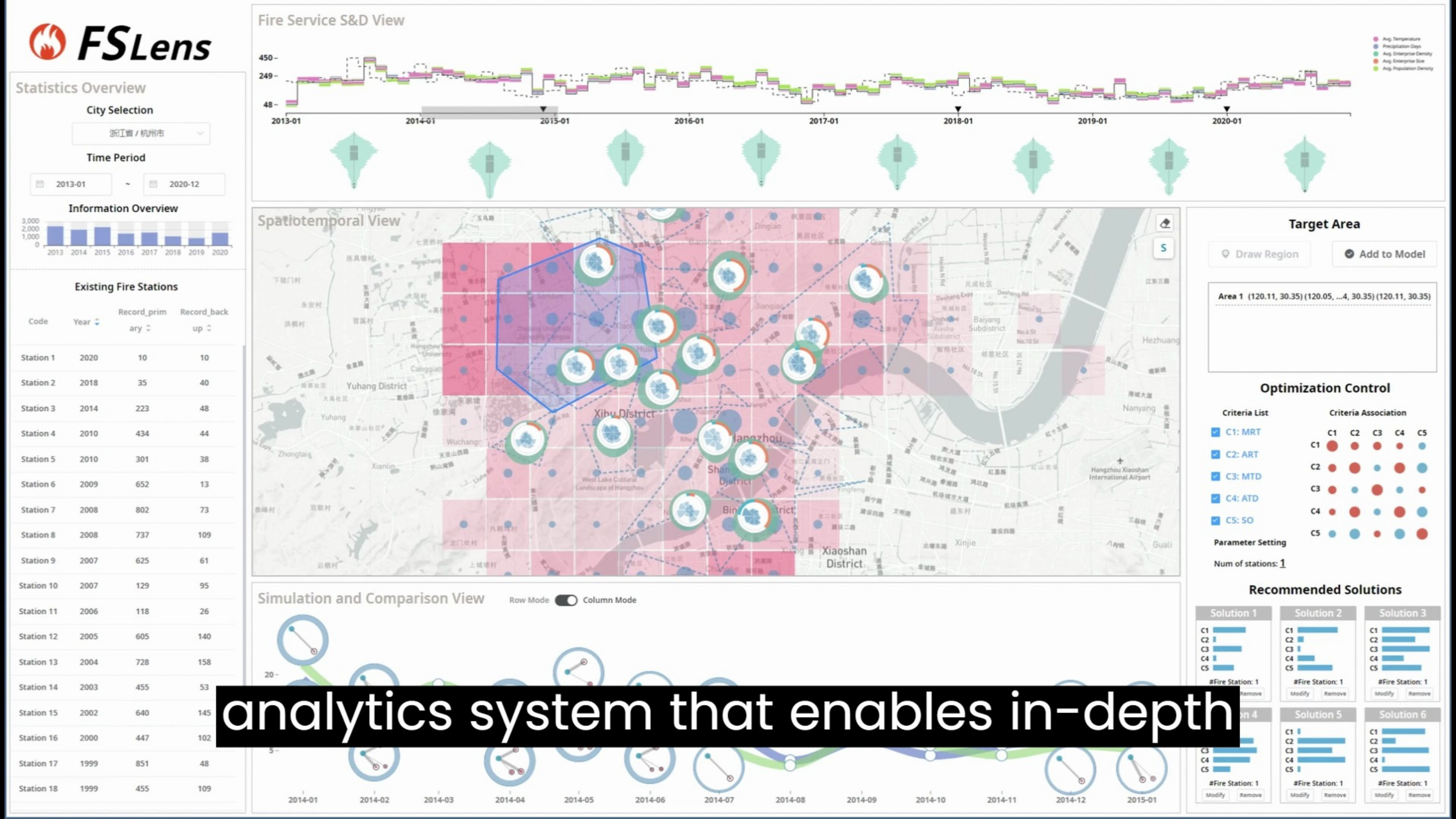


A2
Number of Enterprises
1,500
1,400
1,300
1,200
1,100
2014-04
2014-07
2014-10
2015-01
Population Density
3,950
3,900
3,850
3,800
2014-04
2014-07
2014-10
2015-01
Hushu

Case Study


(a)
Solution 1
Solution 2
Solution 3
Solution 4
Solution 5
Solution 6
C1
C2
C3
C4
C5
#Fire Station: 1
Modify
Remove
(b)
C1 C2 C3 C4 C5


Simulation and Comparison View
Row Mode
Column Mode
(2)
(1)
Solution #1
Solution #2
Solution #5
20
15
10
5
2014-01
2014-02
2014-03
2014-04
2014-05
2014-06
2014-07
2014-08
2014-09
2014-10
2014-11
2014-12
2015-01



analytics system that enables in-depth

Discussion

Contributions

- Begin by evaluating the supply and demand of firefighting resources, taking a "holistic-to-local" perspective to address regional imbalances and focus on specific localities.
- Considering the impact on the existing layout when determining the placement of new fire stations.
- Incorporating a spatiotemporal sequence forecasting model to predict fire occurrence patterns and utilizes model features to elucidate the spatiotemporal distribution of fires.

Limitations

- Historical records of early fire departments may contain inconsistencies and insufficiencies in fire data and standards across different regions.
- Disregarding historical context and real-time traffic conditions.
- Lack of relevant data on fire station characteristics, such as size, staffing, and equipment availability.

- The evaluation of FSLens lacks numerical comparisons due to the unavailability of publicly accessible firefighting datasets for benchmarking and the scarcity of comparable open-source systems.

物流场景中的线路优化

Route Optimization in Logistics Scenarios

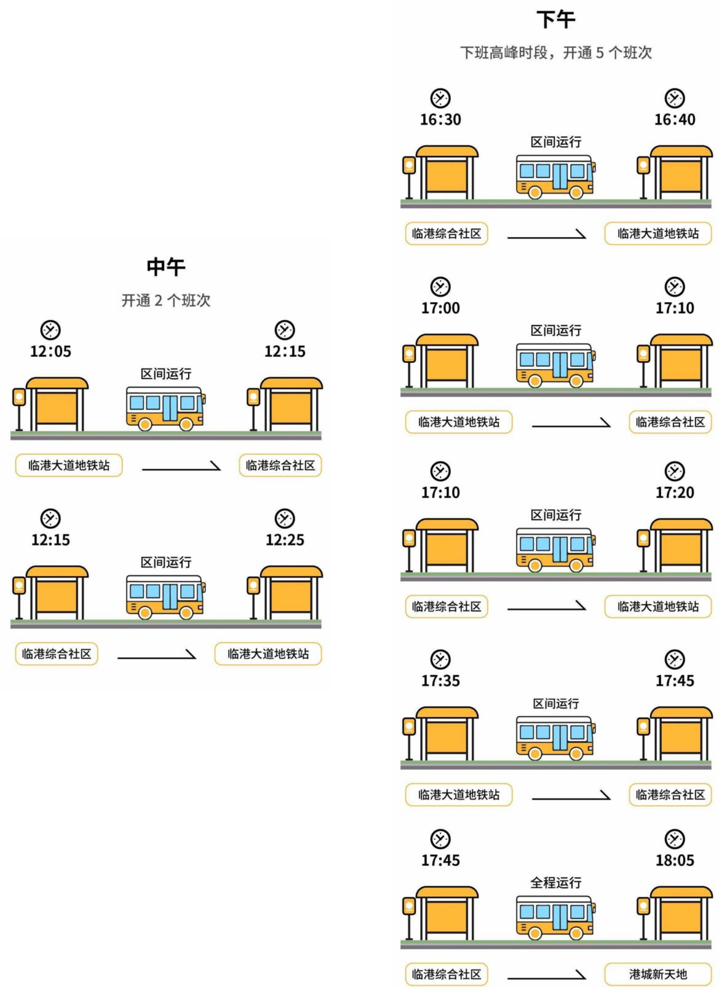
- 电商的快速发展与物流的基础能力和信息化水平不成正比，导致国内物流成本偏高，服务水平低，成为影响电子商务以及信息消费的核心因素之一
- 通过整合物流公司等相关数据，如**即时成本**、**地图POI数据**、**商家订单数据**、**配送网点数据**、**地址信息**、**物流包裹数据**、**快递服务数据**等等，致力于实现物流过程的数字化、可视化，使物流公司和商家的信息对称化程度获得极大提升，实现数据驱动的云供应链协同平台
- 优化物流路线和合理的设置快递站点能够帮组减少物流成本

通过感知乘客的旅行需求优化定制班车路线 – 背景与挑战

Scheduling Customized Shuttle Buses via Perceiving Passengers' Travel Demands – Background and Challenge

- 作为一种"移动性即服务"，穿梭巴士按照预先确定的通勤时间表沿固定路线行驶，由于其缓解拥堵、环境友好和更好的用户体验等优势，在旅行需求旺盛的时期，已经成为移动具有相似出发点和目的地的通勤者的流行手段

- 以可靠和具有成本效益的方式规划班车线路有以下三个挑战：
    - 对旅行需求的收集不可靠且有偏见
    - 费时费力的路径规划和设计
    - 动态和变化的出行需求



Liu Q, Li Q, Tang C, et al. A visual analytics approach to scheduling customized shuttle buses via perceiving passengers' travel demands[C]//2020 IEEE Visualization Conference (VIS). IEEE, 2020: 76-80.

通过感知乘客的旅行需求优化定制班车路线 – 传统做法与痛点

Scheduling Customized Shuttle Buses via Perceiving Passengers' Travel Demands – Conventional Practice and Pain Points

- 定制通勤车的计划分为四个阶段
  - 愿意乘坐巴士的乘客需要提交一份旅行需求调查
  - 收集旅行需求数据，并对旅行目的地相似的乘客进行分类，根据OD旅行方向获得候选的定制巴士路线
  - 对每条备选班车线路进行实地调查，检查道路状况，并对班车线路的初步设计进行相应调整
  - 对定制的班车线路进行了实际的试运行，并对线路的运行时间进行了估算，以便进行调度

- 相对于白天的班车，夜间加班的班车还没有得到很好的实施
  - 大部分加班的员工（即21:30以后）都喜欢自己打车
  - 自愿提交加班出行需求问卷的员工数量非常有限，导致大量的潜在出行需求得不到满足
- 如果直接按照白天的常规班车线路进行夜间运营，由于需求和交通状况不同，线路可能会有偏差

对出行需求的传统认知和处理方式
将不可避免地影响到班车的班次和路线

Liu Q, Li Q, Tang C, et al. A visual analytics approach to scheduling customized shuttle buses via perceiving passengers' travel demands[C]//2020 IEEE Visualization Conference (VIS). IEEE, 2020: 76-80.

通过感知乘客的旅行需求优化定制班车路线 – 需求

Scheduling Customized Shuttle Buses via Perceiving Passengers' Travel Demands - Requirements

R.1 了解员工的加班旅行需求

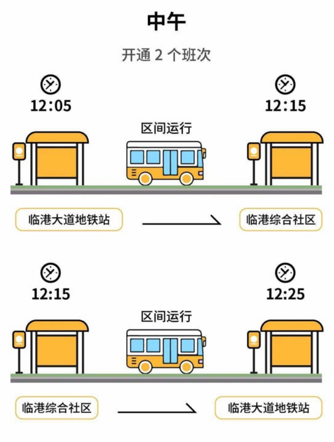
- 提供夜间定制公交线路，需要对员工加班的出行需求有一个清晰的认识

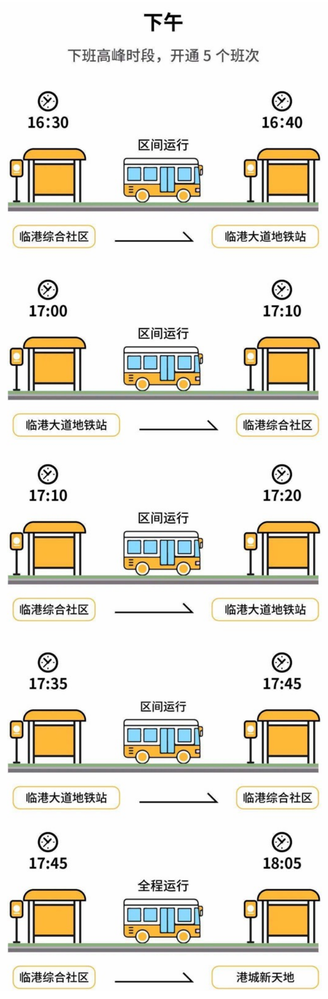
R.2 生成并比较班车路线

- 在可预测的时间内考虑到可能影响交通体验的因素，结合自动推荐连接起点和到目的地的公交站点序列的公交线路

R.3 比较候选的穿梭巴士站

- 如何优化通勤者家庭目的地的可达性，并提供最佳的步行体验，是放置候选定制班车站的一个重要问题。候选的定制班车站点





Liu Q, Li Q, Tang C, et al. A visual analytics approach to scheduling customized shuttle buses via perceiving passengers' travel demands[C]//2020 IEEE Visualization Conference (VIS). IEEE, 2020: 76-80.

上海科技大学 ShanghaiTech University

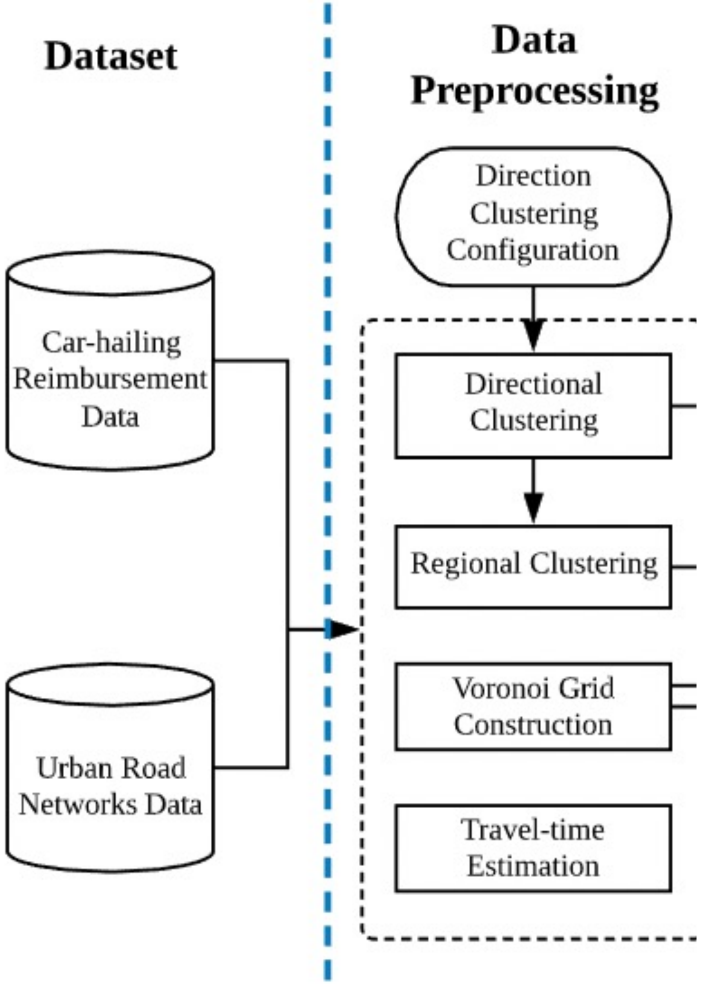
通过感知乘客的旅行需求优化定制班车路线 – 多维数据处理

Scheduling Customized Shuttle Buses via Perceiving Passengers' Travel Demands

1. 城市路网数据。顶点代表道路交叉口，边代表道路（253,890个顶点和314,234条边）

2. 打车报销数据。收集了93,050人次打车报销记录，去掉了身份识别，范围从4月1日到8月1日，2019年的加班。每条打车报销记录包括、出发时间、到达时间、出发地、目的地和支付金额。

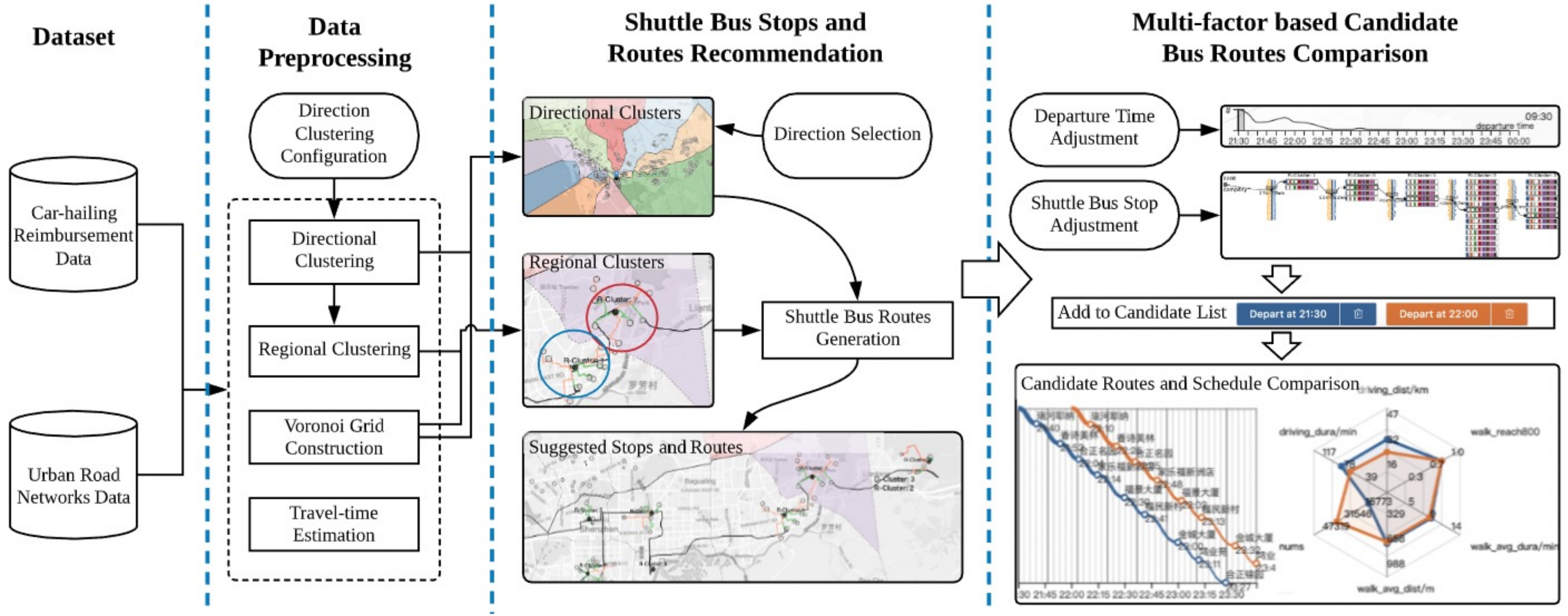
3. 日间班车路线和时间表。通过动态抓取交通状况和两站之间的推荐路线。从21:30开始，以5分钟为间隔，得到了从工作地点到第一个公交站点的推荐公交线路，以及第一个和第二个站点之间以1分钟为间隔的线路，以此类推。



上海科技大学 ShanghaiTech University

通过感知乘客的旅行需求优化定制班车路线

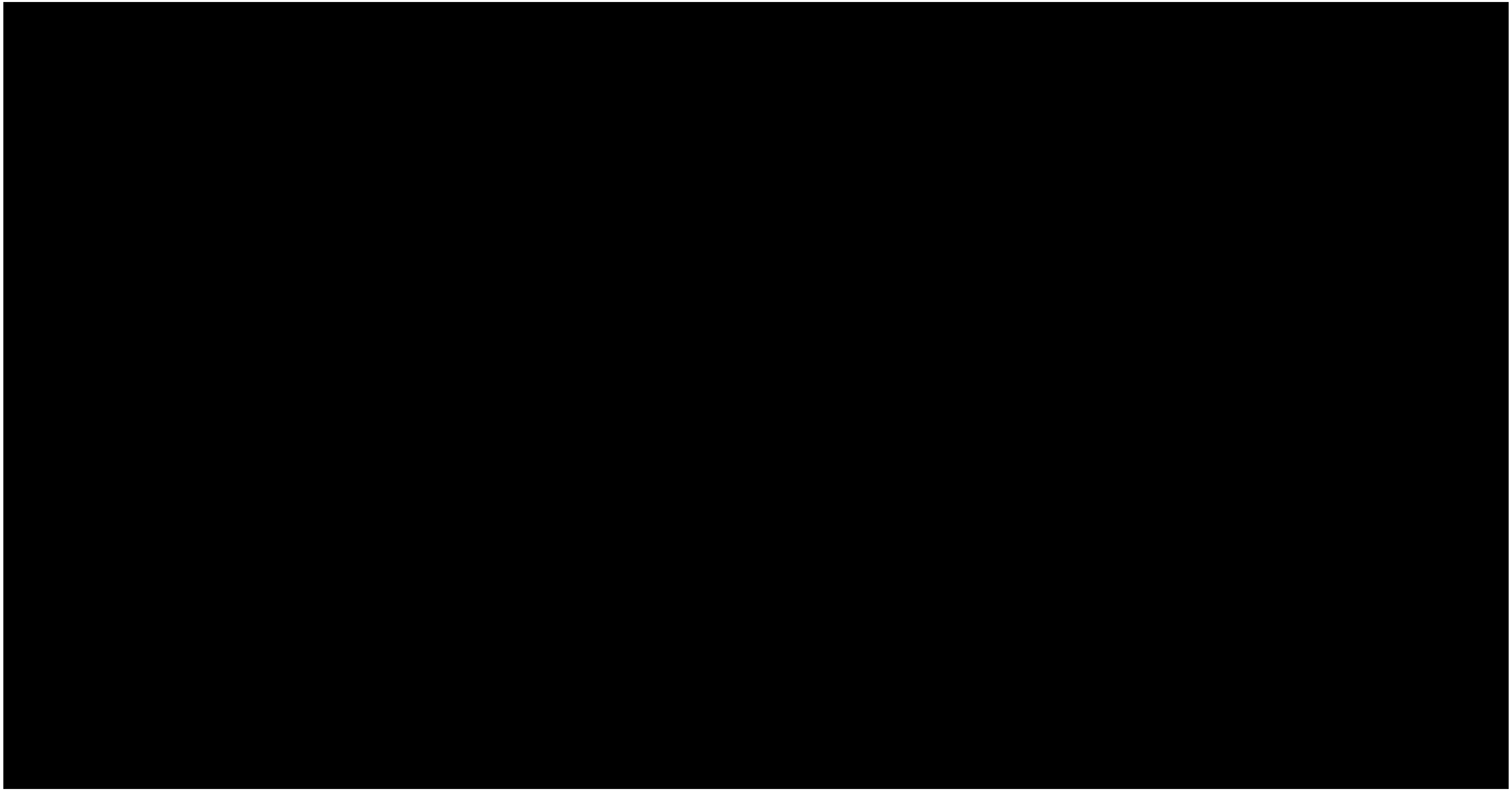
Scheduling Customized Shuttle Buses via Perceiving Passengers' Travel Demands



通勤线路优化及站点配置流程图

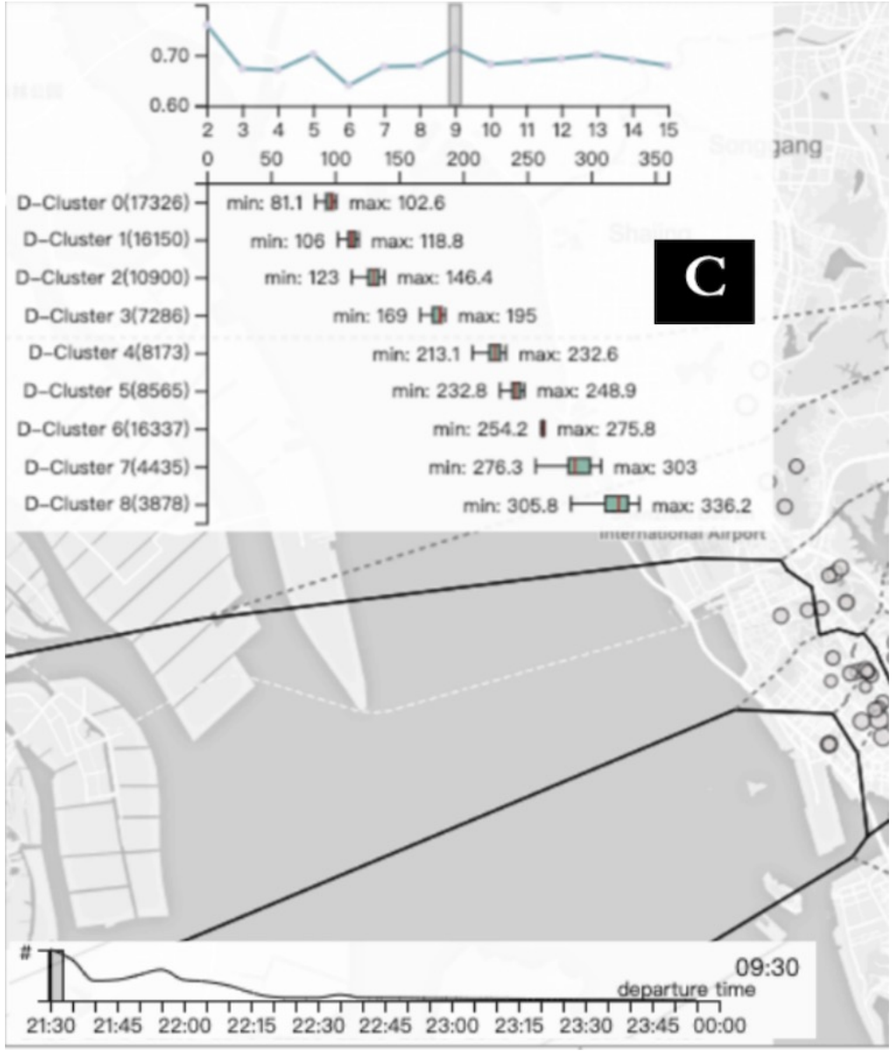
通过感知乘客的旅行需求优化定制班车路线 – 系统交互

Scheduling Customized Shuttle Buses via Perceiving Passengers' Travel Demands – System Overview



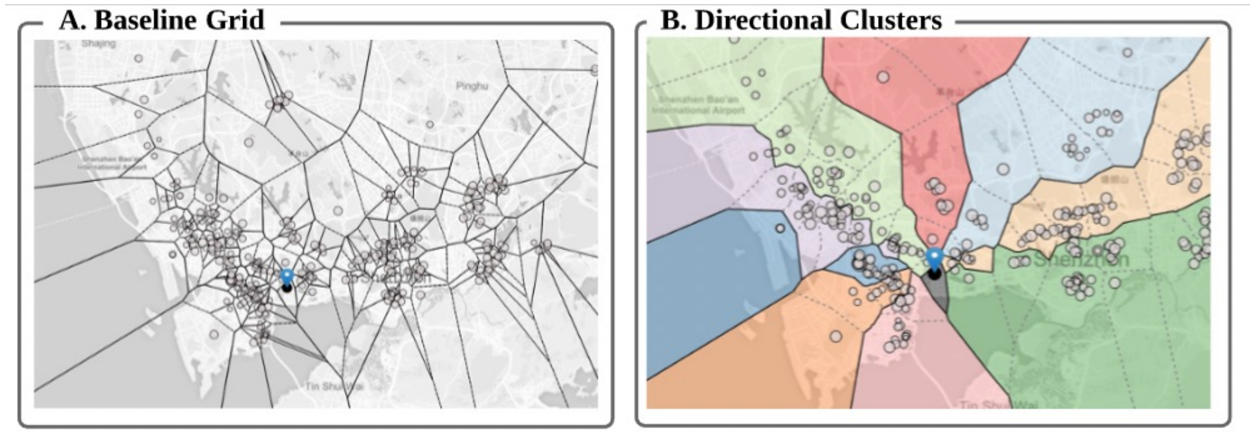
通过感知乘客的旅行需求优化定制班车路线

Scheduling Customized Shuttle Buses via Perceiving Passengers' Travel Demands



第一步：通过方向聚类确定方向

- 使用K-means来获得最初的旅行方向的数量
- 通过在生成的聚类中找到聚类数和Silhouette系数之间关系的峰值来确定K的值
- X轴代表旅行方向的数量，Y轴表示旅行方向数量所对应的Silhouette系数的值
- 综合来看，本案例选择了9作为方向聚类数，其角度分布和各方向聚类中的报销记录分布都比较合理
- 基于上述处理，每条打车记录都有一个唯一的方向聚类ID。方向聚类的结果如图中的彩色部分所示

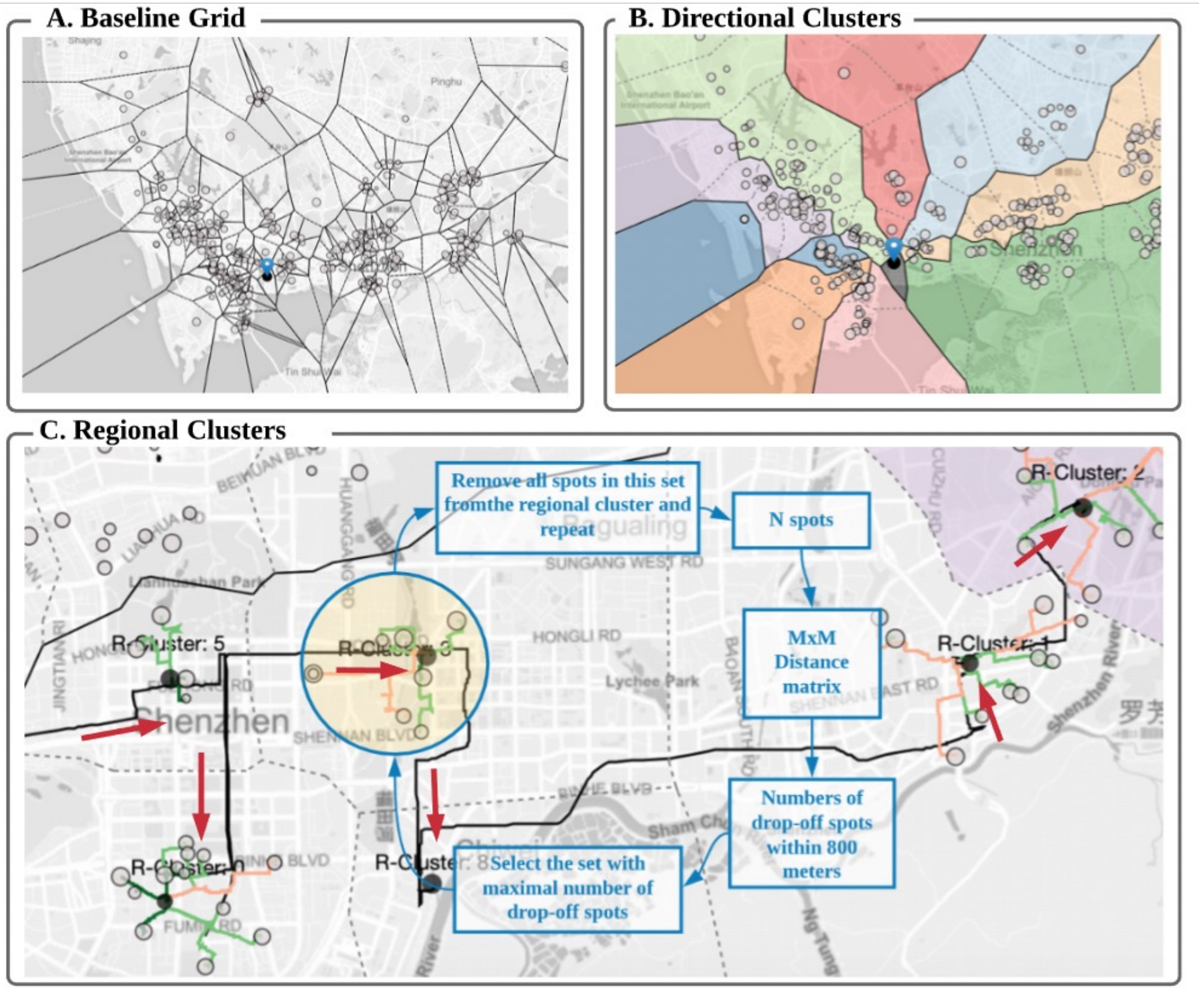


通过感知乘客的旅行需求优化定制班车路线

Scheduling Customized Shuttle Buses via Perceiving Passengers' Travel Demands

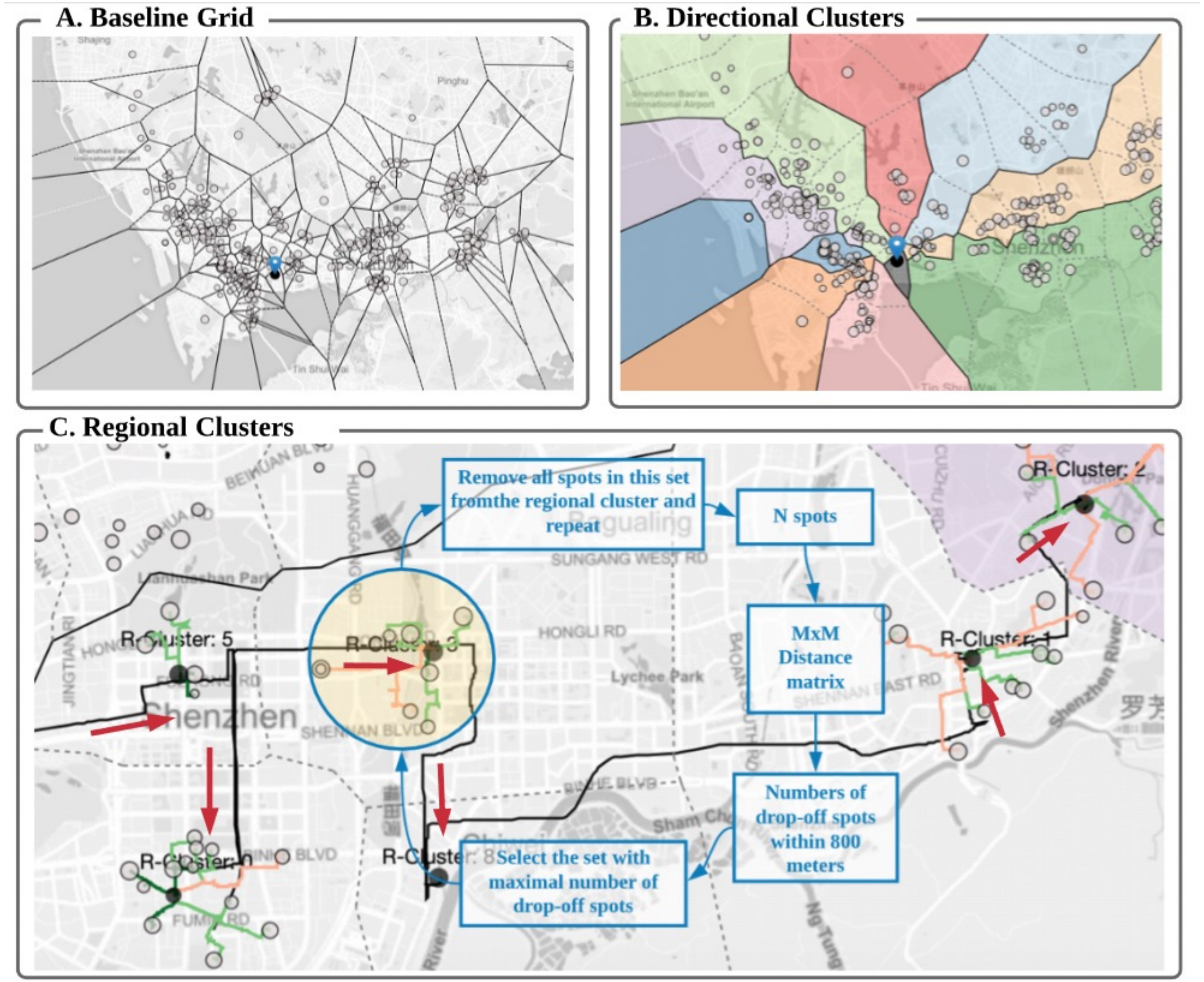
第二步：通过对每个行驶方向的区域聚类来初始化班车站点

- 对于每个出行方向，使用区域聚类来初始化候选班车站点，该站点可以覆盖每个出行方向一个聚类内的附近下车点。因此，每条叫车报销记录都有一个唯一的方向性和区域性集群ID
- 提出了一个基于Voronoi网格的地图视图，以帮助专家理解所识别的方向和区域聚类结果
- 在构建Voronoi网格后，用不同的视觉线索来表示属于相同或不同方向聚类的区域聚类的边界
- 对于一个特定的Voronoi边，如果该边两侧的班车站属于不同的方向聚类，我们将该边描绘成实线；否则，如果该边两侧的班车站属于同一方向聚类但不同的区域聚类，我们将该边描绘为虚线。在其他情况下。我们只是删除该边缘



通过感知乘客的旅行需求优化定制班车路线

Scheduling Customized Shuttle Buses via Perceiving Passengers' Travel Demands



第三步：推荐班车路线和步行路径

- 为了直观地表示每个方向聚类的班车路线以及从初始化的班车站到其覆盖的家庭目的地的路径，我们根据之前抓取到的数据集来估计到每个家庭目的地的班车路线和步行路径
- 如图(C)所示，黑线表示完整的班车路线，班车站（黑点）和家庭目的地（灰点）之间的多条线表示步行路径。路径，颜色表示不同距离的可达性

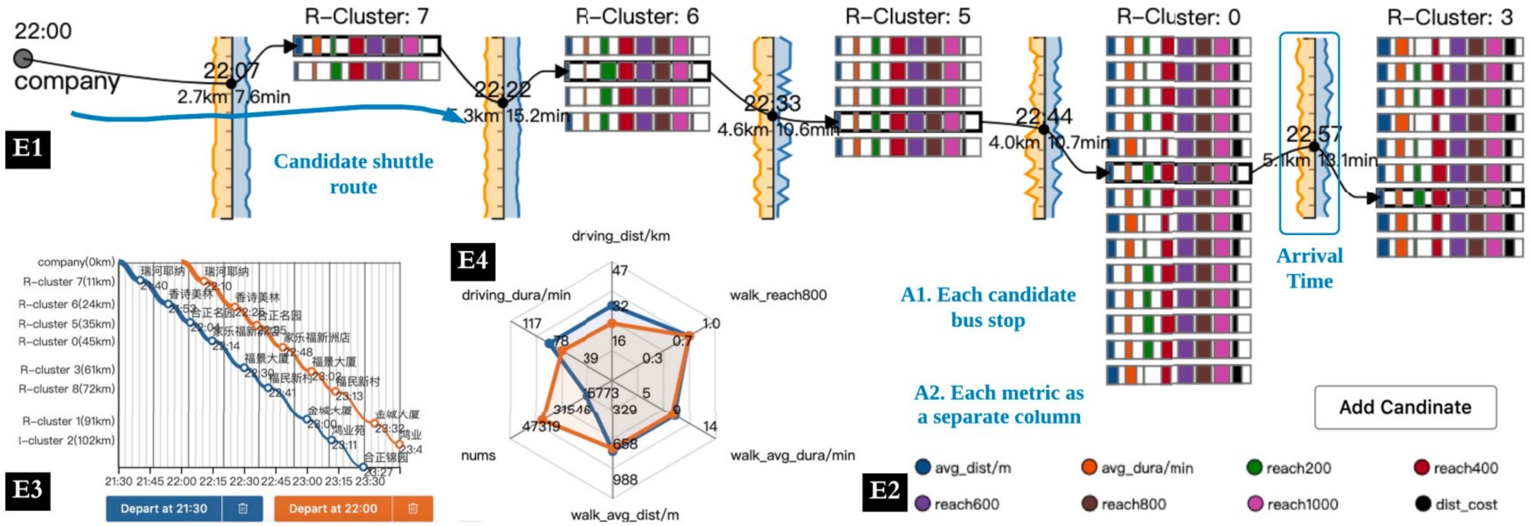
通过感知乘客的旅行需求优化定制班车路线

Scheduling Customized Shuttle Buses via Perceiving Passengers' Travel Demands

第四步：完善班车站点和路线

考虑了以下可能影响交通体验的因素来评估候选班车站：

- **avg dist**：从一个选定的班车站到同一区域集群中其他目的地的加权平均距离
- **avg dura**：从一个选定的班车站到同一区域集群中其他目的地的加权平均步行时间
- **reach200**：从一个班车站到200米内的其他目的地的叫车订单数量与同一区域集群内所有叫车订单的比率。该定义也适用于reach400、reach600等
- **dist_cost**：通过将一个选定的班车站i到同一区域集群中其他目的地的距离乘以订单数量来获得costi，dist_cost是costi的累加值

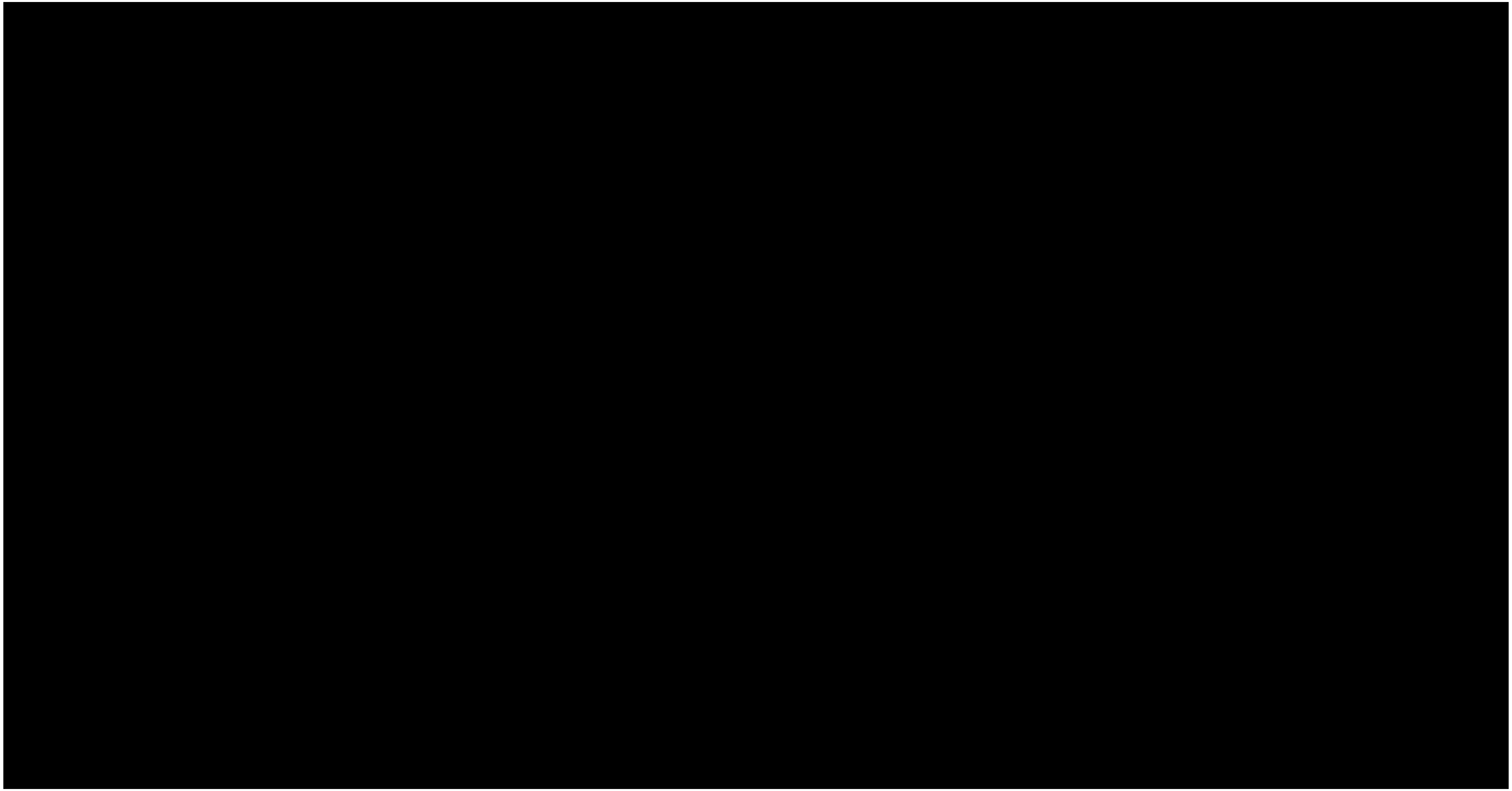


- 每个候选的班车站呈现为一个组合条，其中带有颜色的单个条的长度表示相应班车站的归一化度量值
- 将所有的区域集群在一个旅行方向上横向排列，并将所有区域集群中选定的班车站用一条曲线连接起来，形成一条候选班车线路。
- 班车路线添加到候选名单中，它们将出现在雷达图中
- E3中显示了所选路线的时间表，X轴代表到达时间，Y轴代表区域集群和到工作地点的距离。

上海科技大学
ShanghaiTech University

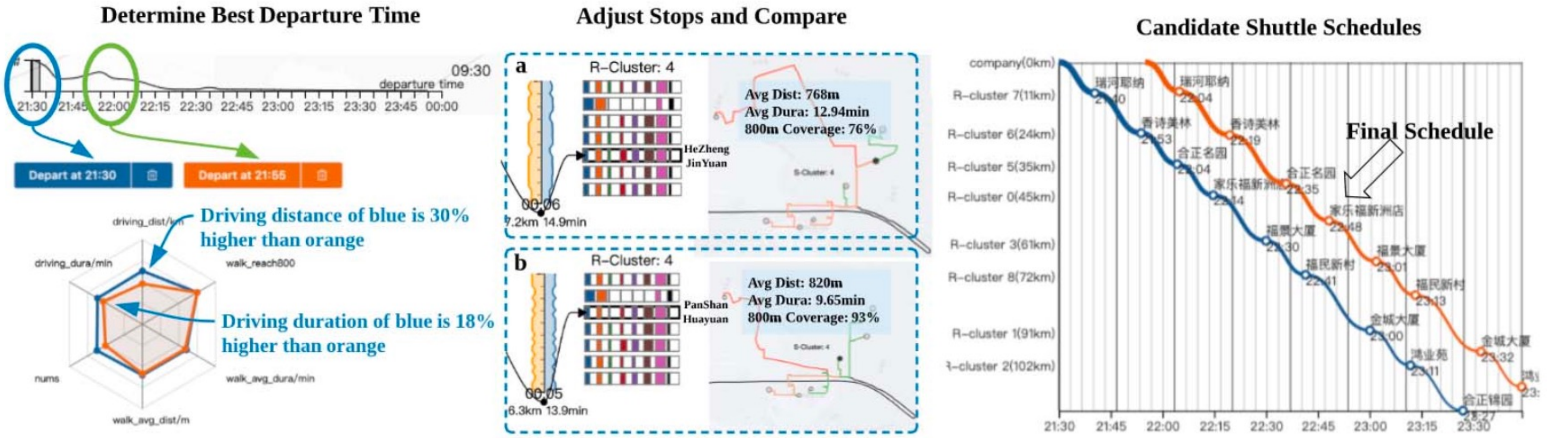
通过感知乘客的旅行需求优化定制班车路线 – 案例分析

Scheduling Customized Shuttle Buses via Perceiving Passengers' Travel Demands – Case Study



通过感知乘客的旅行需求优化定制班车路线 – 案例分析

Scheduling Customized Shuttle Buses via Perceiving Passengers' Travel Demands – Case Study

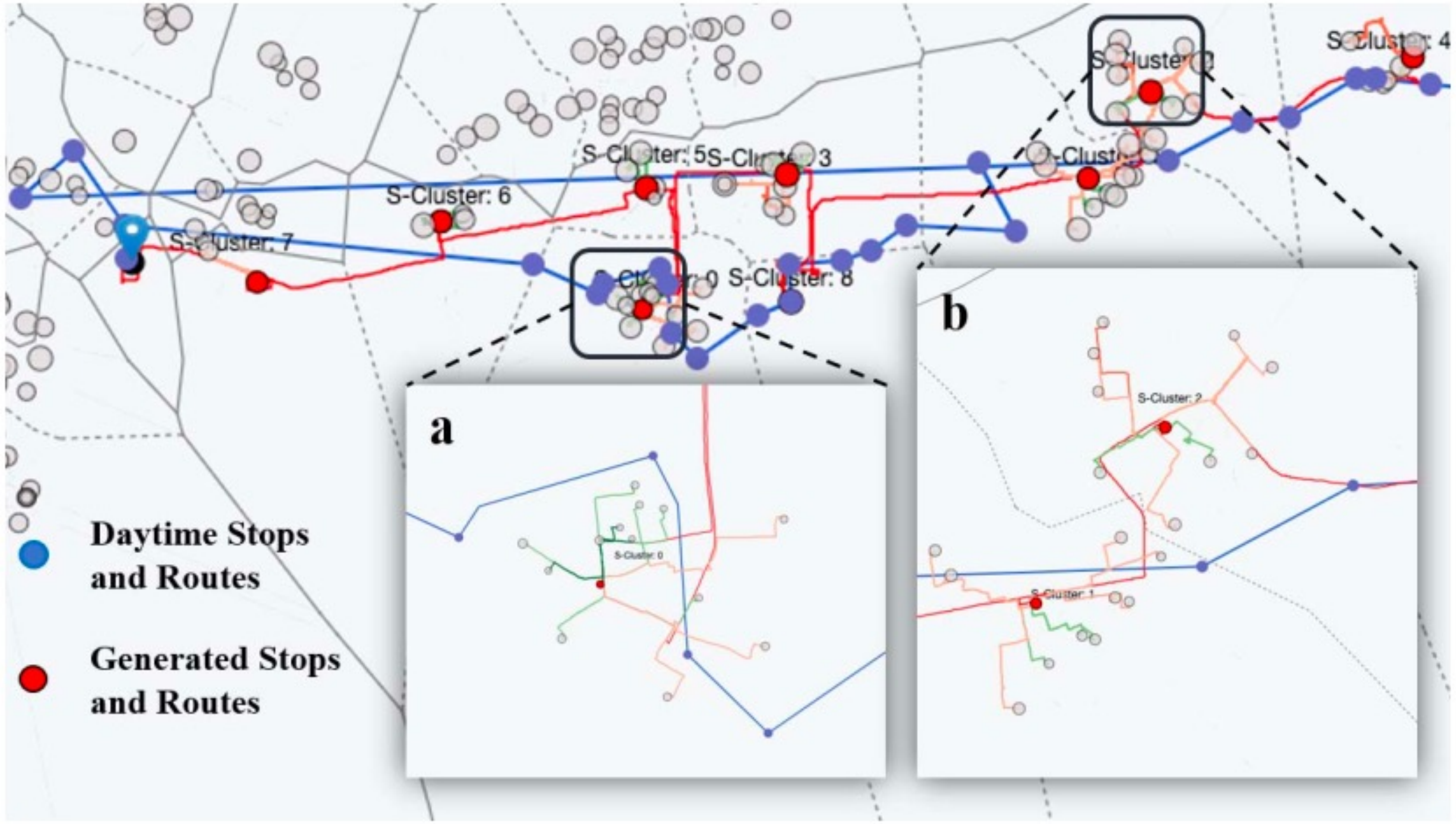


(1) (2) (3)

- 两个峰值分别对应于21:30和21:55的出发时间
- 虽然21:30出发的路线比其他路线覆盖了更多的乘客，但是21:30出发的路线的行驶距离和行驶时间都高于21:55出发的路线
- 系统推荐"和政金苑"为班车站点，但另一个下车点"盘山华苑"位于该区域集群的中间部分
- 点击"盘山华苑"的矩形，发现其平均距离、平均时间和800米内的覆盖率分别为820米、9.65分钟和93%
- "和政金苑"的平均距离为768米，但平均时间更长（12.94分钟），覆盖率更小（76%）

通过感知乘客的旅行需求优化定制班车路线 – 案例分析

Scheduling Customized Shuttle Buses via Perceiving Passengers' Travel Demands – Case Study



- 红线表示ShuttleVis的站点和路线，蓝线表示白天的站点和路线
- 蓝色路线连接了每一个下车点，但在夜间不能很好地覆盖下车点（图b）
- 在图a中，尽管白天的蓝色路线覆盖了大部分夜间的下车点，但它可能会增加驾驶时间

Area 1: Theoretical & Empirical

This area focuses on theoretical and empirical research topics that aim to establish the foundation of VIS as a scientific subject.

Theoretical & Empirical →

Area 2: Applications

This area encompasses all forms of application-focused research.

Applications →

Area 3: Systems & Rendering

This area focuses on the themes of building systems, algorithms for rendering, and alternate input and output modalities.

Systems & Rendering →

Area 4: Representations & Interaction

This area focuses on the design of visual representations and interaction techniques for different types of data, users, and visualization tasks.

Representations & Interaction →

Area 5: Data Transformations

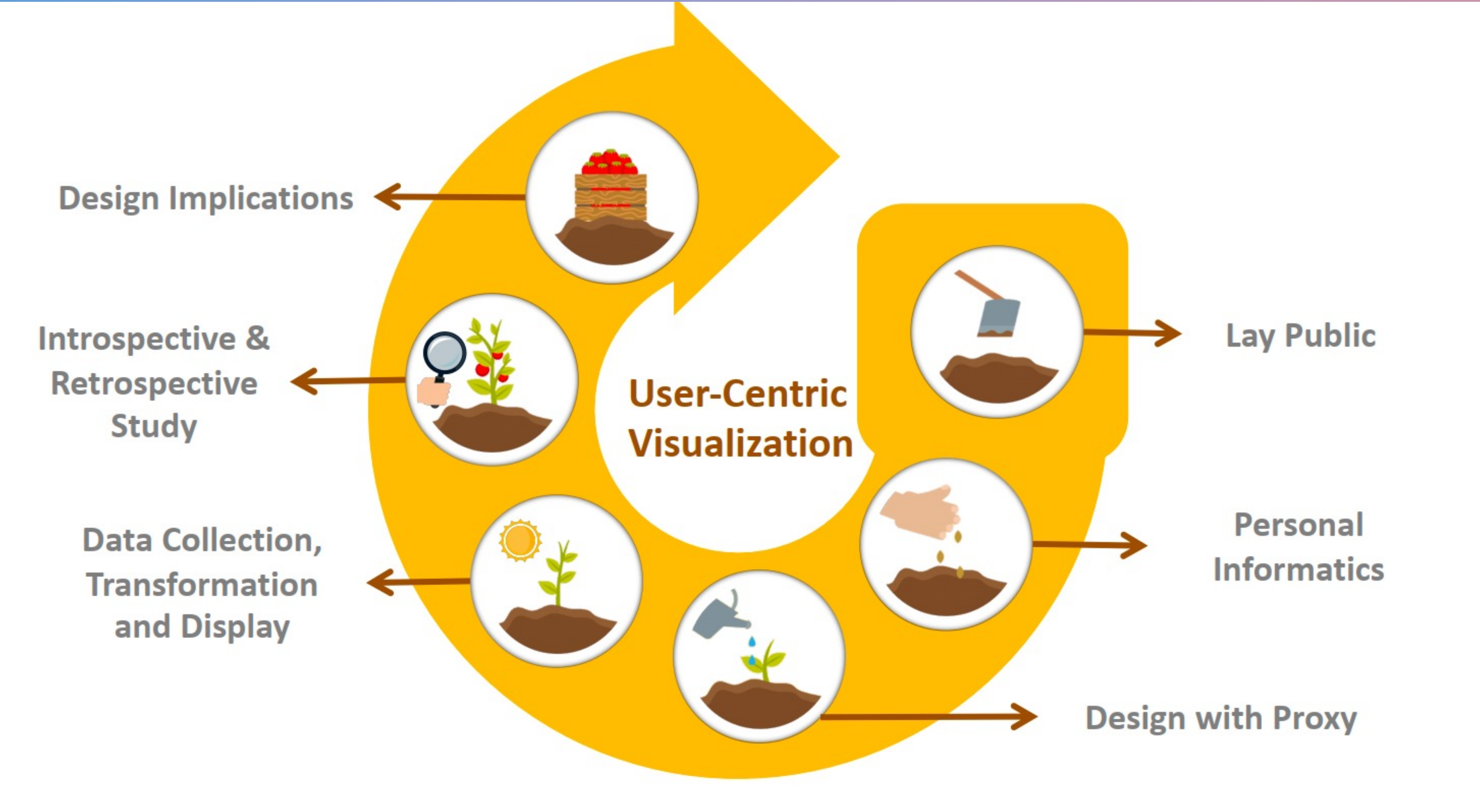
This area focuses on the algorithms and techniques that transform data from one form to another to enable effective and efficient visual mapping as required by the intended visual representations.

Data Transformations →

Area 6: Analytics & Decisions

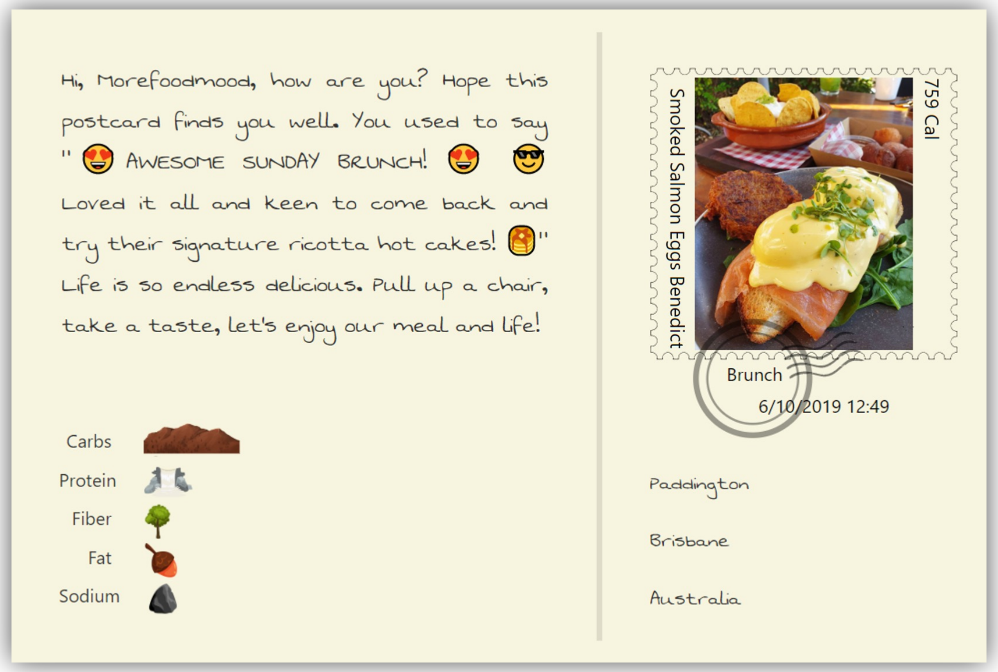
This area focuses on the design and optimization of integrated workflows for visual data analysis, knowledge discovery, decision support, machine learning, and other data intelligence tasks.

Analytics & Decisions →



Design Implications
Introspective & Retrospective Study
Data Collection, Transformation and Display
User-Centric Visualization
Lay Public
Personal Informatics
Design with Proxy

Visual Metaphors: Encourage Retrospective Reflections

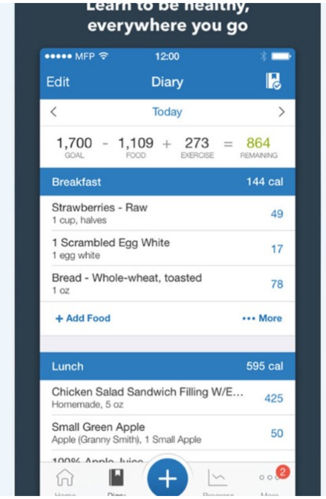


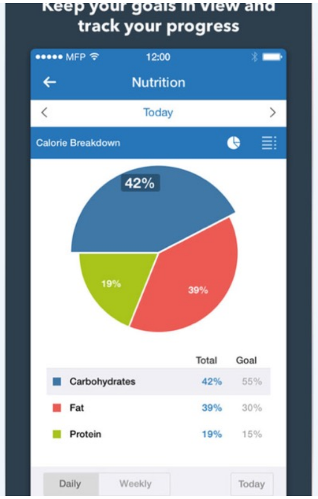
Zhida Sun, Sitong Wang, Wenjie Yang, Onur Yürüten, Chuhan Shi, and Xiaojuan Ma.
In *Proceedings of the 2020 Designing Interactive Systems Conference* (DIS '20).

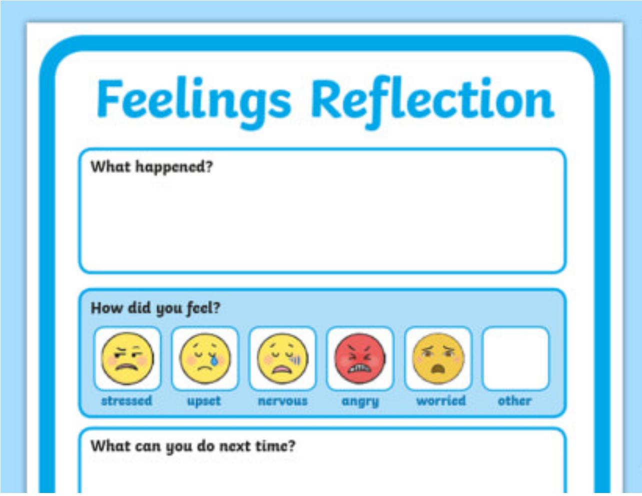
Reflections

- ***Raw numbers**:* concrete, precise, but tedious
- ***Bar charts**:* intuitive but may put users in a judgmental mindset
- ***Waves**:* fluent visual data processing, but hard to balance intuitiveness and interestingness.
- ***Butterfly drawings**:* experience-oriented, but pleasance to the eyes may distract the mind
- ***Table lamps**:* related and immersive, but common encodings may not be available
- Other issues
  - Data and display: sensitivity and latency
  - Retention: novelty effect
  - Personalization: learning curve and privacy

Motivation







Research Question

- How to support users' self-reflection on their physical health and mental well-being with social media food posts?

Research Goal

- Propose an auto-transcription pipeline to easy the tracking process
- Reducing the gulf of communicating the information obtained from historical records to facilitate self-reflection

Challenges

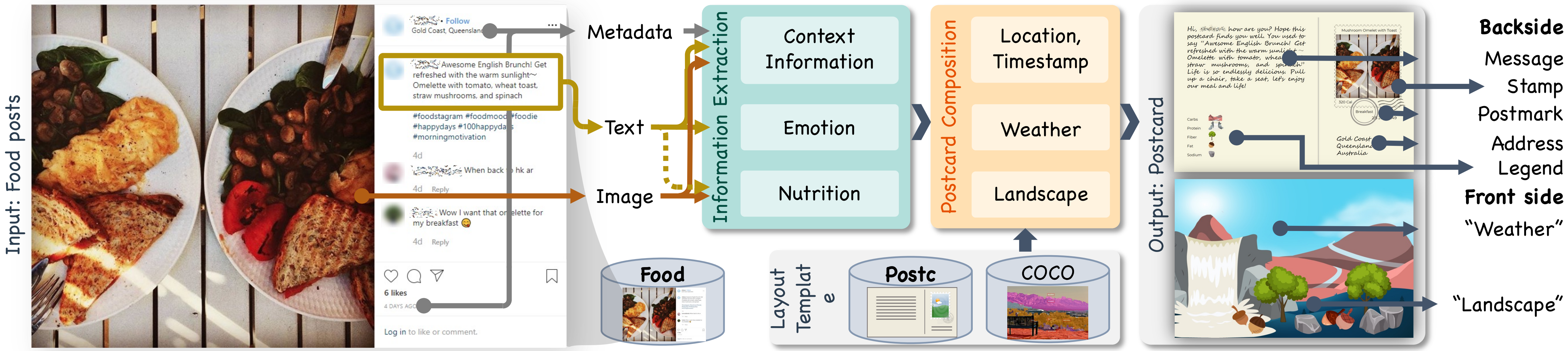
Major

- Existing statistical design in self-tracking tools may evoke negative feelings, judgment, or obsession [Epstein *et al.*, 2016]
- Self-tracking users have a "learned enough" effect on existing tools [Epstein *et al.*, 2016]

Other

- A systematic food logging procedure that tends to be repetitive and tedious [Cordeiro *et al.*, 2015; Konrad *et al.*, 2016]
- Food posts may not be published regularly and are usually not systematically organized

Research Pipeline



上海科技大学
ShanghaiTech University

Information Extraction

- Nutrition
  - Facebook's Inverse Cooking API [Salvador *et al.*, 2019]
  - The official USDA National Nutrient Database
  - **Food image** → carbohydrate (g), protein (g), fat (g), fiber (g), and sodium (mg), and total calories (KCal)
- Emotion
  - A meta-learning approach for text emotion distribution [Zhao *et al.*, 2019]
  - Ekman's six basic emotions
  - **Textual message** → anger, disgust, fear, joy, sadness, and surprise

Postcard Composition: Front side design

- Design encoding scheme
  - Design workshop w/ 6 design background university students
    - 4 females, Age: Mean = 24.67, SD = 1.80
  - 20-min warm-up session + 40-min discussion
  - Design goals
    1. reveal informative and expressive context with diverse elements
    2. enable clear association by leveraging metaphorical encoding
    3. reduce visual clutter and balance visual layout
    4. eliminate ambiguity and cultural bias

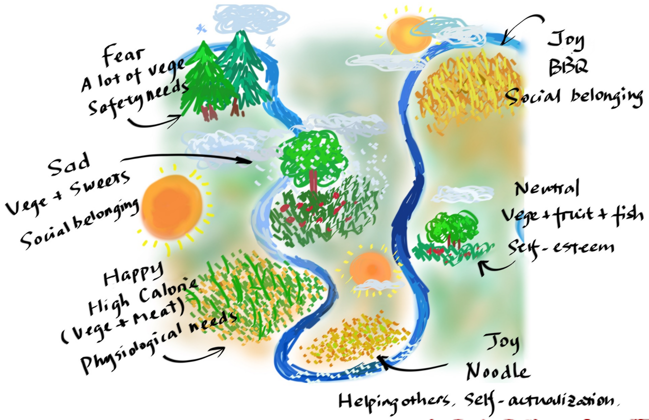
Target

Relationship
emotion + nutrition = life
Emotion
joy, happy, fear, disgust, anger, surprise, sad, neutral
Nutrition
calorie, fat, cholesterol, sodium, carbohydrate, protein, fiber, and vitamin

Source

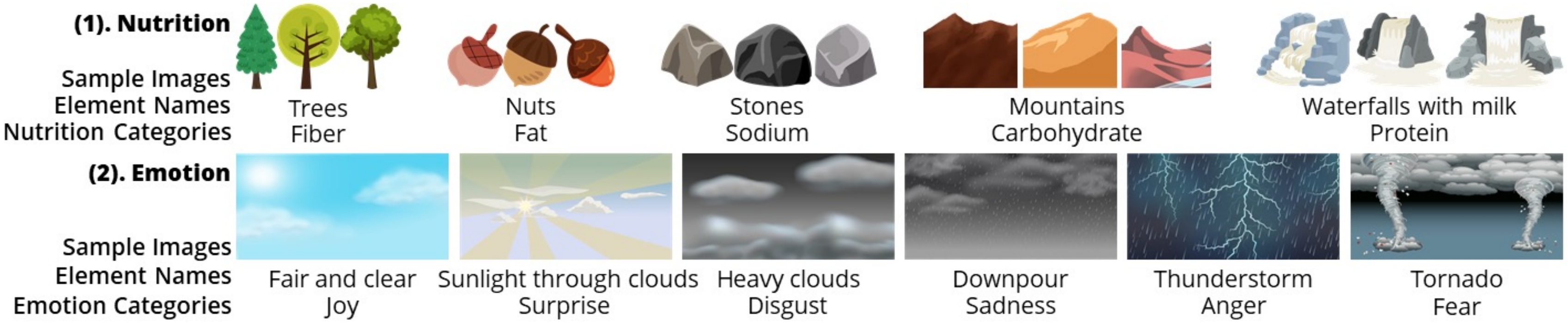
Relationship
weather + plants = ecosystem
Emotion
Joy-sunny, happy-windy, fear-snowy, disgust-foggy, anger-sunburn, surprise-thunder, sad-rainy, neutral-cloudy
Nutrition
Calorie-the area, fat, cholesterol, sodium-red lower, carbohydrate-yellow grass, protein-brown grass, fiber-green trees, and vitamin-green grass

Style



Postcard Composition: Front side design

- Visual design validation
  - Question
    - whether the encoding scheme is intuitive
    - whether users get a consistent interpretation of the trends
  - In-lab user study w/ 10 university students
    - 3 females, Age: Mean = 23.50, SD = 3.01
  - Intuitiveness of Visual Encoding Scheme



  - Interpretation of the Visualization Design
    - Nutrition intake patterns: ICC = .76 ($F(17,153)=4.16$, p <.001)
    - Emotional patterns: ICC = .90 ($F(17,153) = 10.34$, $p < .001$)

上海科技大学
ShanghaiTech University

Postcard Composition: Backside design

- Contextual information layout

Message

1. One greeting sentence
2. The main body of users' posts
3. One closing sentence

Stamp

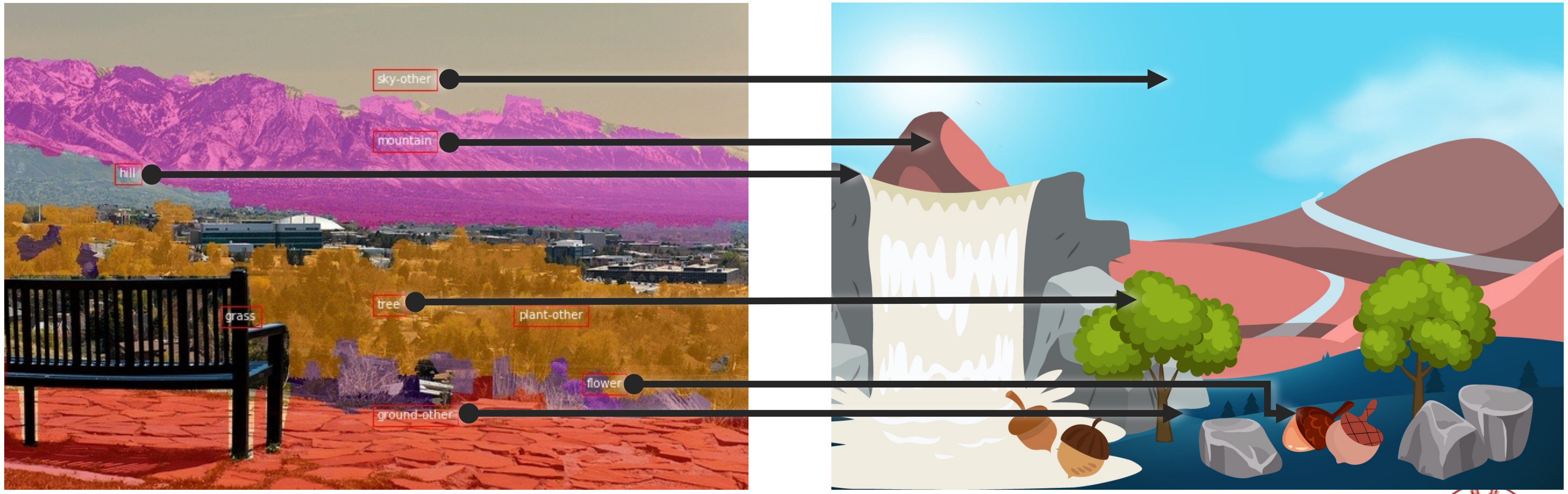
Postmark

Address

Legend

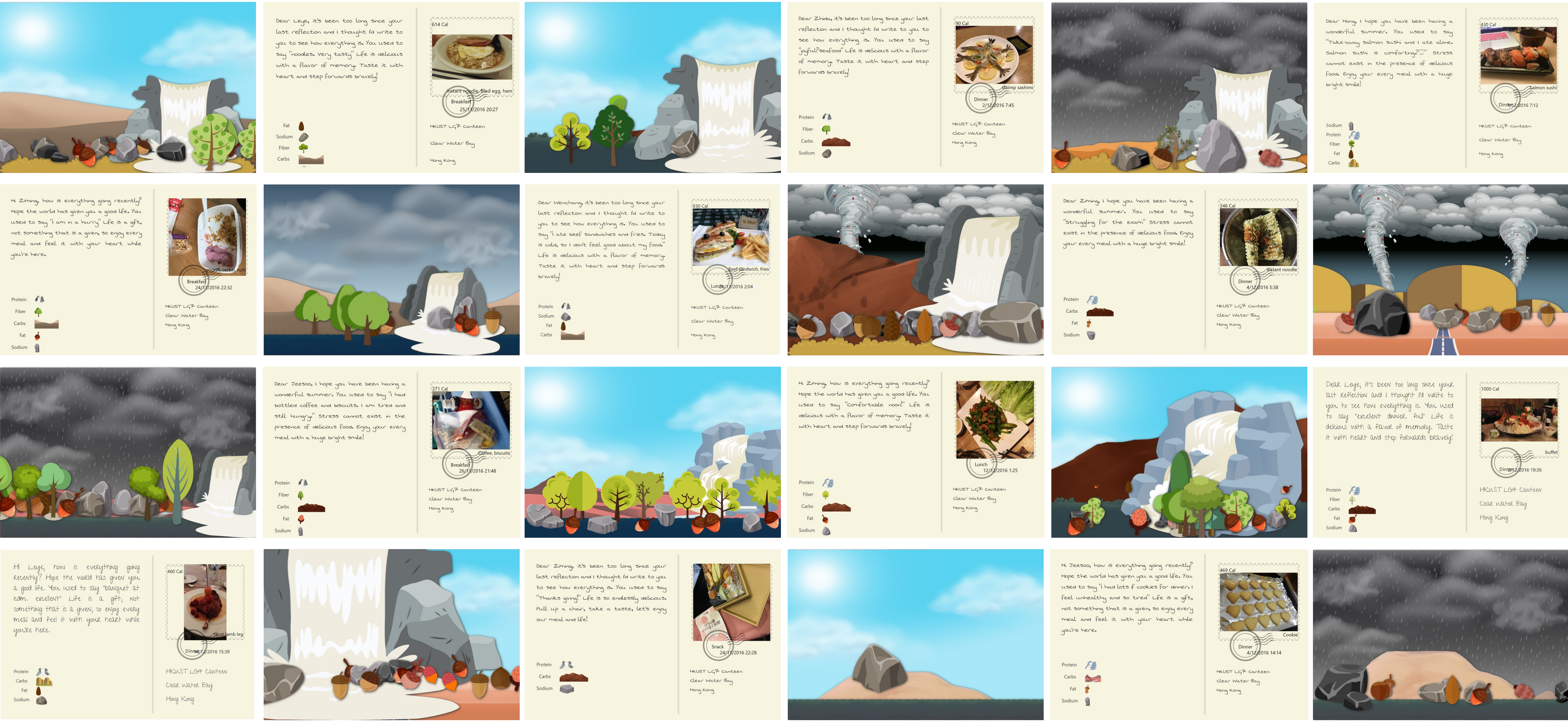
Implementation

- Template-based method
  - The COCO (Common Objects in Context) dataset [Caesar *et al.*, '18]
  - Adjust with two constraints
    - Following real-world common sense
    - Using the geometric perspective



Evaluation

- Goal
  - A descriptive understanding of salient issues by focusing on a smaller sample size
- Participants
  - 20 university students (7 females, Age: Mean =22.25, SD =3.13)
  - Europe (6 participants) & Asia
- Procedure
  - Three weeks: before, during, and after the exam week
  - Semi-structured interview
- Dataset
  - 513 valid entries (151 breakfast; 183 lunch; 169 dinner; 10 snack)



上海科技大学
ShanghaiTech University

Findings

- **Data**: Exploring Food & Emotional Experience

“ *The [front side] readily combines my nutritional and emotional data. – P10 (M, 24)*

- **Context**: Recalling Post-Related Memory

“ *The postcard naturally reminds me of that moment…that was a big union and I was surrounded by a couple of friends. – P9 (M, 27)*

- **Action**: Triggering Future Reaction

“ *I kept a balanced diet during these three weeks. That makes me feel proud as I can take care of my health…it is like a reminder urging me to keep the habit. – P2 (F, 20)*

- **Value**: Involving Personal Pursuit

“ *The landscape design urge me to live a low-carbon life to keep the earth “healthy”. – P9 (M, 27)*

上海科技大学
ShanghaiTech University

Benefits of Visualization for Humans

- Long-term use for end users (e.g., exploratory analysis of scientific data)
- Presentation of known results
- Stepping stone to better understanding of requirements before developing models
- Helps developers of automatic solution refine/debug, determine parameters
- Helps end users of automatic solutions verify and build trust

立志成才报国裕民

TRADITIONAL

- Query for known patterns
- Display results using traditional techniques

Pros:

- Many solutions
- Easier to implement

Cons:

- Can't search for the unexpected

DATA MINING/ML

- Based on statistics
- Black box approach
- Output outliers and correlations
- Human out of the loop

Pros:

- Scalable

Cons:

- Analysts have to make sense of the results
- Makes assumptions on the data

INFOVIS

- Visual interactive interfaces
- Human in the loop

Pros:

- Visual bandwidth is enormous
- Experts decided what to search for
- Identify unknown patterns and errors in the data

Cons

- Scalability can be an issue

Questions?
Thank you ☺

Quan Li

liquan@shanghaitech.edu.cn