Ocupado: Visual Analytics for Occupancy Applications

Tamara Munzner
Department of Computer Science
University of British Columbia
Cisco Toronto
26 Jun 2018
@tamaramunzner
www.cs.ubc.ca/~tmm/talks.html#cisco18

Project threads: Completed to date
• visualization research
  – requirements analysis
  – visualization prototyping in Sandbox environment
  • experiment w/ static data
  • integrate with live data
• machine learning research
  – basic prediction: short & long-term forecasting

SBS Bridge2 product
• develop & deploy
• integrate basic ML

Data architecture

- Requirements analysis
- Visualization prototyping in Sandbox environment
- Experiment with static data
- Integrate with live data
- Machine learning research
- Basic prediction: short & long-term forecasting

Visual analytics
- Collects wireless network signals and infers locations of mobile devices via triangulation
- Already deployed, independently of our project

Task analysis: Facilities management stakeholders
- Known in advance
  • SBS first product: occupancy for HVAC control
- Identified as high priority
  • space planning
  • informal learning spaces
  • classroom services
  • custodial services
  • building managers
- Investigated and considered lower priority
  • risk management
  • security and parking
  • transportation

Task analysis: Example stakeholder questions
- Only basic query handled by previous SBS interface
- What is the current activity level of a specific region?
- Many stakeholder questions require bigger picture
  • Which regions are busy/quiet now?
  • Which regions were heavily used and are empty now?
  • What does the long-term activity profile of region X look like?
  • What is the typical usage pattern of a specific region?
  • Weekdays vs weekends/evening/holidays, according to shift boundaries
  • How does the utilization differ between regions?
  • For subset based on size, space type or other attributes
  • What is the predicted activity for a region in the next X hours?
  • Which regions are normally heavy used but quiet now (or vice versa)
  • Detecting current anomalies vs. average patterns

Ocupado collaboration: Partners
• visualization and data analysis: UBC Computer Science
  • led by Tamara Munzner
  • students: Michael Oppermann, Yann Dubois
• building management systems and data recording: Sensible Building Science
  • UBC Sustainability spinoff
  • led by Stefan Storey
• machine learning: UBC Statistics
  • led by Jeff Andrews
• networking infrastructure: Cisco
  • liaison: Rob Barton

Ocupado collaboration: Funding
• kickstarted by Cisco funding ($25K)
  • Locational Service Analytics: Machine Learning and Data Visualization for CMX Data Applications
• matched 3.5x
  • UBC Campus as a Living Lab ($41K)
  • NSERC Engage ($25K)
• substantial increase in project scope & duration
  • 40% spent over 12 months from May 2017 - Sep 2018
  • we’re now at month 13
  • 60% to spend in 24 months from Oct 2018 - Aug 2020

Ocupado Sandbox
alpha 0.1
• Dynamic filtering, slicing, and sorting of regions

Static data: Test deployment, obvious gaps

Ocupado Sandbox
alpha 0.2
• Integrate static data with other data sources: course schedules, predictions

Scheduling data: Actual vs enrolled in courses

Data: Wi-Fi as proxy for human occupancy
• Wi-Fi device activity strongly correlated with occupancy
  • rough proxy for headcounts in rooms
  • device counts every 5 minutes, per zone
  • good spatial precision if zone large enough
  • rooms with multiple people, not single-person offices
  • excellent temporal resolution
• Privacy preserving architecture
  • keep only counts per zone per time slice
  • no tracking of individuals or trajectories
  • privacy built in to SBS Bridge infrastructure at fundamental level
  • MAC addresses thrown away once stored
    • we’d love CMX protocol change so they’re not sent out!
  • (3 month delay in data gathering due to UBC Legal concerns)

Data: WiFi as proxy for human occupancy

Data abstraction

- Requirements analysis
- Visualization prototyping in Sandbox environment
- Experiment with static data
- Integrate with live data
- Machine learning research
- Basic prediction: short & long-term forecasting

Visual analytics
- Collects wireless network signals and infers locations of mobile devices via triangulation
- Already deployed, independently of our project

Task analysis: Facilities management stakeholders
- Known in advance
  • SBS first product: occupancy for HVAC control
- Identified as high priority
  • Space planning
  • Informal learning spaces
  • Classroom services
  • Custodial services
  • Building managers
- Investigated and considered lower priority
  • Risk management
  • Security and parking
  • Transportation

Task analysis: Example stakeholder questions
- Only basic query handled by previous SBS interface
  • What is the current activity level of a specific region?
  • Many stakeholder questions require bigger picture
    • Which regions are busy/quiet now?
    • Which regions were heavily used and are empty now?
    • What does the long-term activity profile of region X look like?
    • What is the typical usage pattern of a specific region?
    • Weekdays vs weekends/evening/holidays, according to shift boundaries
    • How does the utilization differ between regions?
    • For subset based on size, space type or other attributes
    • What is the predicted activity for a region in the next X hours?
    • Which regions are normally heavy used but quiet now (or vice versa)
    • Detecting current anomalies vs. average patterns

Ocupado project
• Occupancy data for facilities management
  • estimate human occupancy of buildings using mobile device connections as common denominator
  • Innovative uses for CMX data
  • Create visual and predictive decision-support tools
  • Visual analytics interface to make data actionable by people
  • Investigate multiple stakeholder contexts of use
  • From energy management to space planning and beyond

Ocupado collaboration: Partners
• Visualization and data analysis: UBC Computer Science
  • Led by Tamara Munzner
  • Students: Michael Oppermann, Yann Dubois
• Building management systems and data recording: Sensible Building Science
  • UBC Sustainability spinoff
  • Led by Stefan Storey
• Machine learning: UBC Statistics
  • Led by Jeff Andrews
• Networking infrastructure: Cisco
  • Liaison: Rob Barton

Ocupado collaboration: Funding
• Kickstarted by Cisco funding ($25K)
  • Locational Service Analytics: Machine Learning and Data Visualization for CMX Data Applications
• Matched 3.5x
  • UBC Campus as a Living Lab ($41K)
  • NSERC Engage ($25K)
• Substantial increase in project scope & duration
  • 40% spent over 12 months from May 2017 - Sep 2018
  • We’re now at month 13
  • 60% to spend in 24 months from Oct 2018 - Aug 2020

Ocupado Sandbox
alpha 0.1
• Dynamic filtering, slicing, and sorting of regions

Static data: Test deployment, obvious gaps

Ocupado Sandbox
alpha 0.2
• Integrate static data with other data sources: course schedules, predictions

Scheduling data: Actual vs enrolled in courses

Data architecture

- Requirements analysis
- Visualization prototyping in Sandbox environment
- Experiment with static data
- Integrate with live data
- Machine learning research
- Basic prediction: short & long-term forecasting

Visual analytics
- Collects wireless network signals and infers locations of mobile devices via triangulation
- Already deployed, independently of our project

Task analysis: Facilities management stakeholders
- Known in advance
  • SBS first product: occupancy for HVAC control
- Identified as high priority
  • Space planning
  • Informal learning spaces
  • Classroom services
  • Custodial services
  • Building managers
- Investigated and considered lower priority
  • Risk management
  • Security and parking
  • Transportation

Task analysis: Example stakeholder questions
- Only basic query handled by previous SBS interface
  • What is the current activity level of a specific region?
  • Many stakeholder questions require bigger picture
    • Which regions are busy/quiet now?
    • Which regions were heavily used and are empty now?
    • What does the long-term activity profile of region X look like?
    • What is the typical usage pattern of a specific region?
    • Weekdays vs weekends/evening/holidays, according to shift boundaries
    • How does the utilization differ between regions?
    • For subset based on size, space type or other attributes
    • What is the predicted activity for a region in the next X hours?
    • Which regions are normally heavy used but quiet now (or vice versa)
    • Detecting current anomalies vs. average patterns
Integrate with ML prediction data

**Ocupado Sandbox**

*alpha 0.3*

- Flexible visual exploration interface between the user and the Bridge API
- Integration of live activity data
- Presets for quickly answering common domain questions
- URL bookmarks for replicating and sharing a certain application state

**Cisco office: Live data testbed (real vs synthetic)**

*alpha 0.4*

- Live data flowing from UBC
- Continued development of visual interface

**Demo**

Overview: Busiest zones, on average

Building view

Busiest buildings, by floor, with floor plans

Browsing patterns within building: Room by room

... Scrolling down

Investigating anomalous zone

Zones in one building, evening custodial shift

Details for one zone

Zooming in

**Ocupado timeline: Milestones to Sep 2018**

May 2017 - Sep 2018

- **UBC: Machine learning (prediction), Task/requirements analysis**
  - Yann Dubois (BSc/Intern), Munzner, Andrews
  - May - Sep 2017

- **UBC: Visualization prototypes, Task/requirement analysis**
  - Michael Oppermann (PhD), Munzner
  - Aug 2017 - Sep 2018

- **SBS: Bridge2 (data infrastructure), Machine learning (integration)**
  - Felipe Deo, Nick Bradley (MSc/Intern)
  - May 2017 - Apr 2018

- **Ocupado timeline: Milestones to Sep 2018**

First 1.3 yrs: 40% funds used

- May - Dec 2017
  - 18K (Cisco/CLL)

- Jan - Sep 2018
  - 25K (NSERC Engage)
  - 48K (Cisco/CLL)
  - 25K (planned MITACS)

Final 2 yrs: 60% funds left

- Sep 2018 - Aug 2020
  - 48K (Cisco/CLL)
  - 25K (planned MITACS)

Project threads

- visualization research
  - requirements analysis
  - visualization prototyping in Sandbox environment
    - experiment w/ static data
    - integrate with live data
  - customized visualization Skins for stakeholders
    - initial development (Jun 2018)
    - deployment & testing (Sep 2018)
    - exploit advanced ML in visualization (Sep 2019)

- machine learning research
  - basic prediction: short & long-term forecasting
  - semi-supervised asset tagging (Sep 2018)
  - advanced prediction: gaps & assets (Apr 2019)

- SBS Bridge2 product
  - develop & deploy
  - integrate basic ML
  - integrate advanced ML (Jan 2020)
Intellectual property
- open-source everything created at UBC
  - after moderate delay, under commercializable license
- unified whole
  - each part builds on and depends on others
  - impossible to disentangle IP into multiple buckets based on chronology/source
- benefits to partner companies
  - open-source Ocupado specifically designed as front end that interoperates with proprietary Bridge infrastructure from SBS
  - Ocupado+Bridge combination showcases benefits of occupancy tracking via Cisco CMX product
  - intellectual contribution of task analysis of stakeholder needs for different verticals is crucial but not patentable

Research agenda: Interleaved angles of attack

Problem-driven work
- design studies
  - in collaboration with target users
  - real data, real tasks
  - intensive requirements analysis
  - iterative refinement
  - deploy tool/systems
- typical evaluation: field studies
- my strategy: opportunistic collaboration
  - many domains
  - both industrial and academic partners

Study-driven work
- scalable algorithms & systems
  - typical evaluation: computational benchmarks
- new input & interaction techniques
  - typical evaluation: controlled experiments on human subjects

Evaluation in the field: Dim. reduction
- guided by domain experts
- theoretical foundations
- evaluation
- network and social analysis
- visual displays
- typical evaluation: field studies
- my strategy: opportunistic collaboration
  - many domains
  - both industrial and academic partners

Problem-driven: Building mgmt, fisheries
- current work: Ocupado (BC Public Health & BC CDC)
  - current work: genomics epidemiology
- current work: genomics (UBC Zoology)
- current work: e-commerce
- current work: facilities management
- current work: clickstreams
- current work: building occupancy
- current work: real-time building occupancy
- current work: building management
- current work: interactive visualization of building occupancy

Problem-driven: Automotive, journalism
- current work: RelEx (BMW)

Problem-driven: Genomics
- current work: genomics epidemiology
- current work: genomics (BC Cancer)

Problem-driven: Current data science
- current work: Scalable Building Science (UBC)
- current work: Build tools for interactive high-dimensional data analysis

Technique-driven work
- scalable algorithms & systems
  - typical evaluation: computational benchmarks
- new layout & interaction techniques
  - typical evaluation: controlled experiments on human subjects

Technique-driven: Dimensionality reduction
- guided by domain experts
- theoretical foundations
- evaluation
- network and social analysis
- visual displays
- typical evaluation: field studies
- my strategy: opportunistic collaboration
  - many domains
  - both industrial and academic partners

Evaluation in the field: Dim. reduction
- guided by domain experts
- theoretical foundations
- evaluation
- network and social analysis
- visual displays
- typical evaluation: field studies
- my strategy: opportunistic collaboration
  - many domains
  - both industrial and academic partners

Problem-driven: Tech industry
- current work: Glint (UBC)
  - current work: visual data analysis
- current work: human-in-the-loop
- current work: building occupancy
- current work: interactive visualization of building occupancy

Problem-driven: Genomics
- current work: genomics epidemiology
- current work: genomics (BC Cancer)

Problem-driven: Current data science
- current work: Scalable Building Science (UBC)
- current work: Build tools for interactive high-dimensional data analysis

Technique-driven work
- scalable algorithms & systems
  - typical evaluation: computational benchmarks
- new layout & interaction techniques
  - typical evaluation: controlled experiments on human subjects

Technique-driven: Dimensionality reduction
- guided by domain experts
- theoretical foundations
- evaluation
- network and social analysis
- visual displays
- typical evaluation: field studies
- my strategy: opportunistic collaboration
  - many domains
  - both industrial and academic partners

Evaluation in the field: Dim. reduction
- guided by domain experts
- theoretical foundations
- evaluation
- network and social analysis
- visual displays
- typical evaluation: field studies
- my strategy: opportunistic collaboration
  - many domains
  - both industrial and academic partners
Theoretical foundations

Sara Ackerman (Microsoft)
Eisenstein (Microsoft)
Nedalia Henry-Riche (Microsoft)
Bongshin Lee (Microsoft)
Benjamin Bach (Microsoft)
Matt Brehmer

Visualization Analysis & Design

Abstract Tasks

Nested Model

algorithm
idiom
abstraction
domain

Theoretical foundations

Papers Process & Pitfalls

Design Study Methodology

Abstract Tasks

More information

• papers, videos, open source software, talks, courses
http://www.cs.ubc.ca/group/infovis
http://www.cs.ubc.ca/~tmm
@tamarasmunzer

Curation & Presentation: Timelines

Johanna Fulda (Sud. Zeitung)

TimeLineCurator

www.cs.ubc.ca/~tmm/talks.html#cisco18

Timeline Revisited
timelinesrevisited.github.io/

Michael Sedlmair
Miriah Meyer
Matt Brehmer