Ocupado: Visual Analytics for Occupancy Applications

Tamara Munzner
Department of Computer Science
University of British Columbia

Cisco Toronto
26 Jun 2018
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

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

www.cs.ubc.ca/~tmm/talks.html#cisco18
Oocupado 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)
  – planned: MITACS ($25K)

• substantial increase in project scope & duration
  – 40% spent over 16 months from May 2017 - Sep 2018
    • we’re now at month 13
  – 60% to spend in 24 months from Oct 2018 - Aug 2020

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

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

Data sources

• Collects wireless network signals and infers locations of mobile devices via triangulation

• Already deployed, independently of our project

Data preprocessing and aggregation

UBC Cisco CMX

Sensible’s Bridge API

Visual analytics

Ocupado

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

- Requests data from UBC Cisco CMX every 5 min
- Aggregates device coordinates by pre-defined zones (a zone can be a research lab, hallway, composite of multiple offices, ...)
- Provides a REST API for external user interfaces

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

UBC Cisco CMX → Sensible’s Bridge API → Ocupado

- Tool for visual exploration of W-Fi activity data (estimated occupancy)
- Support different stakeholders in decision-making process

www.cs.ubc.ca/~tmm/talks.html#cisco18
Data: Wifi as proxy for human occupancy

• wifi 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, not stored
      – we’d love CMX protocol change so they’re not sent out!
      – (3 month delay in data gathering due to UBC Legal concerns)

www.cs.ubc.ca/~tm/talks.html#cisco18
Data abstraction

SPACE (regions of interest)

- Campus
- Building
- Floor
- Zone
- Room

TIME (periods of interest)

- Continuous vs. discrete time series
  - custodial shifts
  - course sessions
  - weekdays
  - past/next X hours
  - weekends and holidays
  - term vs exams

www.cs.ubc.ca/~tmm/talks.html#cisco18
Task analysis: Facilities management stakeholders

- known in advance
  - energy systems
    - 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

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

*alpha 0.1*

- Dynamic filtering, slicing, and sorting of regions

[www.cs.ubc.ca/~tmm/talks.html#cisco18](http://www.cs.ubc.ca/~tmm/talks.html#cisco18)
Static data: Test deployment, obvious gaps

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

*alpha 0.2*

- Integrate static data with other data sources: course schedules, predictions

[www.cs.ubc.ca/~tmm/talks.html#cisco18](www.cs.ubc.ca/~tmm/talks.html#cisco18)
Scheduling data: Actual vs enrolled in courses
Integrate with ML prediction data

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

alpha 0.4

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

www.cs.ubc.ca/~tmm/talks.html#cisco18
Demo
Overview: Busiest zones, on average

www.cs.ubc.ca/~tmm/talks.html#cisco18
Building view
Busiest buildings, by floor, with floor plans

www.cs.ubc.ca/~tmm/talks.html#cisco18
Browsing patterns within building: Room by room

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

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

www.cs.ubc.ca/~tmm/talks.html#cisco18
Zones in one building, evening custodial shift

www.cs.ubc.ca/~tmm/talks.html#cisco18
Details for one zone

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

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

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

First 1.3 yrs: 40% funds used

May - Dec 2017
18K (Cisco/CLL)

Jan - Sep 2018
25K (NSERC Engage)

Final 2 yrs: 60% funds left

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

www.cs.ubc.ca/~tmmtalks.html#cisco18
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)

www.cs.ubc.ca/~tmm/talks.html#cisco18
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
UBC InfoVis Group: Research Approach
Research agenda: Interleaved angles of attack

- Problem-driven work
- Technique-driven work
- Theoretical foundations
- Evaluation

www.cs.ubc.ca/~tmm/talks.html#cisco18
Problem-driven work

• **design studies**
  – in collaboration with target users
    • real data, real tasks
    • intensive requirements analysis
  – iterative refinement
    • deploy tools/systems
  – typical evaluation: field studies

• **my strategy: opportunistic collaboration**
  – many domains
  – both industrial and academic partners

www.cs.ubc.ca/~tmm/talks.html#cisco18
Problem-driven: Tech industry

SessionViewer: web log analysis
https://youtu.be/T4MaTZd56G4

Peter McLachlan
(AT&T Research)

Stephen North

Heidi Lam

Diane Tang
(Google)

LiveRAC: systems time-series logs
https://youtu.be/ld0c3H0VSkw

www.cs.ubc.ca/~tmm/talks.html#cisco18
Problem-driven: Genomics

Aaron Barsky  
Jenn Gardy  
Robert Kincaid  
Miriah Meyer  
Hanspeter Pfister

Cerebral
https://youtu.be/76HhG1FQngl

MizBee
https://youtu.be/86p7brwuz2g

www.cs.ubc.ca/~tmm/talks.html#cisco18
Problem-driven: Genomics

Joel Ferstay
(BC Cancer)

Cydney Nielsen

https://youtu.be/AHDnv_qMXxQ

Variant View

Ana Crisan
(UBC Public Health & BC CDC)

current work:
genomic epidemiology

Jenn Gardy

Zipeng Liu
(UBC Zoology)

current work:
gene trees

www.cs.ubc.ca/~tmm/talks.html#cisco18
Problem-driven: Automotive, journalism

RelEx (BMW)
https://youtu.be/89IsQXc6Ao4

Overview
https://vimeo.com/71483614

Jonathan Stray (Assoc Press)

Michael Sedlmair

www.cs.ubc.ca/~tmm/talks.html#cisco18
Problem-driven: Building mgmt, fisheries

Vismon  https://youtu.be/h0kHoS4VYmk

Matt Brehmer  Kevin Tate
(Pulse/EnerNOC)

Maryam Booshehrian  Torsten Moeller (SFU)

www.cs.ubc.ca/~tmm/talks.html#cisco18
Problem-driven: Current data science

Kimberly Dextras-Romagnino

- current work: Segmentifier (Mobify)
- e-commerce clickstreams
- build tools for human-in-the-loop visual data analysis

Michael Oppermann

- current work: Ocupado (Sensible Building Science, Cisco)
- wifi proxy for real-time building occupancy
- integrate visual analytics and predictive ML for facilities management

www.cs.ubc.ca/~tmm/talks.html#cisco18
Technique-driven work

• scalable algorithms & systems
  – typical evaluation: computational benchmarks

• new layout & interaction techniques
  – typical evaluation: controlled experiments on human subjects
Technique-driven: Graph drawing

Daniel Archambault (Bordeaux)

David Auber (Bordeaux)

TopoLayout
SPF
Grouse
GrouseFlocks
TugGraph

Benjamin Renoust

Detangler  https://youtu.be/QOtnHSsUV6k

Guy Melançon (Bordeaux)

TreeJuxtaposer  https://youtu.be/GdaPj8a9QEo
Evaluation experiments: Graph drawing

Dmitry Nekrasovski  Adam Bodnar  Joanna McGrenere

www.cs.ubc.ca/~tmm/talks.html#cisco18
Technique: Dimensionality reduction

Stephen Ingram

Glimmer

DimStiller

Glint

QSNE

www.cs.ubc.ca/~tmm/talks.html#cisco18
Evaluation experiments: Dim. reduction

Melanie Tory

Points vs landscapes for dimensionally reduced data

Michael Sedlmair

Guidance on DR & scatterplot choices

Taxonomy of cluster separation factors

www.cs.ubc.ca/~tmm/talks.html#cisco18
Evaluation in the field: Dim. reduction

DR in the Wild

Matt Brehmer  Michael Sedlmair  Melanie Tory  Stephen Ingram

www.cs.ubc.ca/~tmm/talks.html#cisco18
Curation & Presentation: Timelines

TimeLineCurator
https://vimeo.com/123246662

Johanna Fulda
(Sud. Zeitung)

Matt Brehmer

Bongshin Lee
(Microsoft)

Benjamin Bach
(Microsoft)

Nathalie Henry-Riche
(Microsoft)

Timelines Revisited
timelinesrevisited.github.io/

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

Papers Process & Pitfalls

Design Study Methodology

Michael Sedlmair  Miriah Meyer

Abstract Tasks

Matt Brehmer

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

Visualization Analysis & Design

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

- papers, videos, open source software, talks, courses

http://www.cs.ubc.ca/group/infovis
http://www.cs.ubc.ca/~tmm