Situated Conversational Interaction

LARRY HECK

CONTRIBUTORS:
Dilek Hakkani-Tür, Patrick Pantel

DECEMBER 2013

TWEET COMMENTS/QUESTIONS: #IEEGLOBALSIP
Situated Conversational Interaction

Leverage the situation or context of the user to create a conversational natural user interfaces to the world's knowledge
Modes of Situated Conversations

Human computer interaction

Augmented multi-party interaction

Personal Assistant for Phones

Send email to Larry about the demo

Welcome to home search

Other Screens:

Look for tattooists clip in hand

What about places with outdoor dining?
Conversational Browser

SITUATED INTERACTION WITH THE WEB OF DOCUMENTS AND APPS
Creating a “Web Scale” Conversational Browser

Situated Interactions: Dynamically Adapt to the Page Content

Browsing to a new web page or App affects ASR and SLU

- **ASR**: Dynamic adaptation of LMs to ngrams of page content ➔ 16% ERR in WER
- **SLU**:
  - Add 100s of *click intent* actions to *static SLU*
  - Multi-tiered logic determines final intent

---

User Intention

- **Generic**
  - Greeting
  - Frustration
  - Etc.
- **Page Related**
  - Click on link
  - Click on button
  - Etc.
- **Deep-Down Domains (bottom part of the T)**
  - Restaurants
  - Movies
  - Weather
  - Etc.
- **Navigation**
  - Go back
  - Start over
  - Scroll Down
  - Go to URL
  - Etc.
Creating a “Web Scale” Conversational Browser

Situated Interactions: Multimodal Processing

Multimodal Sensor for Study: Kinect™
- RGB camera
- Depth sensor
- Multi-array microphone running proprietary software

Kinect enables full-body 3D motion capture, facial recognition and voice recognition
Creating a “Web Scale” Conversational Browser

Situated Interactions: Multimodal Processing

Lexical Click Intent

$$\text{TF-IDF}(q, p_k) = \sum_{t \in q} \text{tf-idf}_{t, p_k}$$

Gesture Click Intent

$$\text{Gesture Score} = \exp\left(-\frac{|d|^a}{10^b}\right)$$

Combining Intents

$$S_{M_k} = (1 - \alpha) \text{TF-IDF}(q, p_k) + \alpha \cdot \text{Gesture Score}$$

$$= (1 - \alpha) \sum_{t \in q} \text{tf-idf}_{t, p_k} + \alpha \cdot \exp\left(-\frac{|d|^a}{10^b}\right) \quad (3)$$
Experimental Setup

Living Room: users seated 5-6 feet away from TV

Two Data Collections

- Set #1: 8 speakers over 25 sessions
  - 2,868 user turns
  - 917 (31.9%) with click intent
- Set #2: 7 speakers over 14 sessions
  - 1,101 user turns
  - 284 (25.8%) with click intent
Gesture Intent Model

Results

Studied errors in gesture intent model
- False Accepts (random arm movement triggers intent)
- Miss (system misses pointing intent)

Removed all display control utts (e.g., “scroll up”)

558 remaining turns (ones with/without hand gesture)

<table>
<thead>
<tr>
<th></th>
<th>Gesture Found</th>
<th>No Gesture Found</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Click “that one”</td>
<td>15 (2.7%)</td>
<td>1 (0.1%)</td>
<td>16 (2.8%)</td>
</tr>
<tr>
<td>Click other</td>
<td>25 (4.5%)</td>
<td>102 (18.3%)</td>
<td>127 (22.8%)</td>
</tr>
<tr>
<td>Non-Click</td>
<td>127 (22.8%)</td>
<td>288 (51.6%)</td>
<td>415 (74.4%)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>167 (30.0%)</td>
<td>391 (70.0%)</td>
<td>558</td>
</tr>
</tbody>
</table>

False Accepts  Miss
**Multimodal Click Intent Detection**

*Simulated vs Real Gestures*

- What is the gesture accuracy of humans?  
  \[16.4 \leftrightarrow 28.6\] pixels

- Do humans point every time they speak?  
  Humans only use gesture 30.0% of the time
To dig deeper...

Larry Heck, et al, Multimodal Conversational Search and Browse
IEEE Workshop on Speech, Language and Audio in Multimedia, August 2013
Conversational Insights

SITUATED INTERACTION WITH DOCUMENTS
Fall 2013 Guide to New York City
Frommer's - By Donna Heidemund on September 09, 2013 at 4:23:36 PM EDT

Like the classic Frank Sinatra hit asks, "Autumn in New York... why does it seem so inviting?" New Yorkers like to chalk up the city's third-quarter appeal to its generally sunny weather (with temperatures ranging from pleasantly warm to intermittently nippy), the waning of summer tourist crowds, festive parades, Broadway show openings, gargantuan harvest moons rising over the East River, and, of course, the famed New York City Marathon in early November. Here are some of the things that will keep you busy during a September-to-November visit.

Autumn's Annual Events

Even though it's not officially Fall for a few more weeks, the season kicks off right after Labor Day with Mercedes-Benz Fashion Week (September 5-12) as style-savvy buyers, editors, celebrities, and fashionistas (as well as Project Runway's Heidi Klum and Tim Gunn) descend on the city for nonstop couture shows and late-night parties—it's people-watching at its chicest.

To satisfy something other than your fashion sense, follow your nose to the Feast of San Gennaro (September 12-22), an annual rite of caloric overindulgence featuring favorite Italian foods like sausage and peppers and cannoli, as well as music and entertainment. You can mangia every day from 11:30 am to 11:30 pm on Mulberry and Grand streets in Little Italy.

Autumn plays host to some of NYC's top parades: The festivities along Fifth Ave. begin with oompah music on Sept 21 during the St. Patrick's Day Parade (a celebration of German-American heritage), then march to a polka beat on Oct. 6 with the Polish Day Parade (featuring Polish-Americans), and conclude on Oct. 14 with the Columbus Day Parade (an Italian-American extravaganza). On Oct. 31, head to the West Village for the Village Halloween Parade (6:00-8:30 pm on Sixth Ave. between 14th and 16th streets) to gleefully outrageous costumed marchers. And on the morning of November 28, find a spot along Central Park West or Sixth Ave. above 34th St. to catch the world-famous balloons, floats, and marching bands in the 87th Annual Macy's Thanksgiving Day Parade. Or gather outside the American Museum of Natural History on the evening of Nov. 27 to watch the giant balloons spring to life.

Hot New Eats

Per usual, dozens of restaurants debuted in NYC over the last few months, but these three are getting tons of buzz: Betony, where Chef Bryce Shuman and partner Eamon Rockey (both formerly of Eleven Madison Park) preside over a sophisticated and flavor-rich modern American menu served in a soaring space. In Nolita, New Zealand Chef Matt Lambert (formerly of Public) is helming The Musketeer Room, where the menu is inspired by his homeland and features dishes like mushroom custard, cold smoked scallops, and red New Zealand deer flavored with gin.

And given the late August opening of Juni, you can be among the first to try Australian chef Shaun Hergatt's seasonal tasting menus (his previous restaurant, SHO Shaun Hergatt, had two Michelin stars).
Entity Linking

**Problem:** Annotate mentions of an entity with the knowledge base entry for that entity (linking text to knowledge).

**Solution:** Build sequence models and contextual disambiguation models trained over self-disambiguated data such as Wikipedia pages or query-document click data. Features range from topical information, linguistic patterns, morpho-syntactic clues, web usage logs, anchor texts, etc.

Silviu-Petru Cucerzan (EMNLP’07);
Rakesh Agrawal, Ariel Fuxman, Anitha Kannan, John Shafer (WWW’12)
Larry Heck, Dilek Hakkani-Tur, and Gokhan Tur (Interspeech’13)
Ming-Wei Chang, Emre Kiciman (NAACL 2013)
Chin-Yew Lin, Xiaojiang Huang, Yunbo Cao

Even though it’s not officially Fall for a few more weeks, the season kicks off right after Labor Day with Mercedes-Benz Fashion Week (September 5-12) as style-savvy buyers, editors, celebrities, and fashionistas (as well as Project Runway’s Heidi Klum and Tim Gunn) descend on the city for nonstop couture shows and late-night parties—it’s people-watching at its chicest.
Knowledge: Foundation of Situated Conversations

A vast majority of user interactions are with people, locations, things.

Knowledge refers to these entities/concepts and to how they are interrelated.

The dual-role of knowledge

- People seek to **find information** about things, to **transact** on them, and to **browse for recommendations**.
- Knowledge “joined” with world ➔ **situated conversation**.
Conversational Systems **Challenge**

Scaling Through Situated Knowledge

Scale Strategy: Shallow & Broad

Head Strategy: Deep & Narrow

Semantic Depth

Domain Breadth

Pivot on Knowledge
Breaking Bad

Source page topic

.........................

........Bryan Cranston .........

Aaron Paul ......................

Crime drama ........ protagonist

..........Albuquerque ...........

.......Vince Gillian ............

Destination page topic
Who directed the movie Life is Beautiful?

SELECT ?director
WHERE {"Life is Beautiful" directed_by ?director}.

Sample user utterances:
- "Show me movies by Roberto Benigni"
- "Who directed Life is Beautiful?"

Data & features for training statistical CU models

Web Search

Movie-Director search queries:
- "Life is Beautiful" and "Roberto Benigni"
- "Roberto Benigni directed"
- "James Cameron directed"
- "Titanic and James Cameron"
- "Life is Beautiful and Roberto Benigni"

Web Search Query Click Logs

JSON scheme for the KB, corresponding to the example query "Who directed Life is Beautiful?"

Larry Heck, Dilek Hakkani-Tur, and Gokhan Tur; 2012-2013
Conversational Systems **Approach**

**Scaling Through Situated Knowledge**

Processing Flow of an Interaction with Knowledge

- Scoping
  - Linking the focus entity(ies) to the knowledge base
    - E.g., linking a touch on “marathon” to the Wikipedia ID for “NYC Marathon”; subselecting restaurants on a map after circling a region.

- Intent detection
  - Interpret the intent (what the user wants) in the context of the knowledge (and other context)

- Execution
  - Farm out the request to an appropriate execution engine

Key Point: There are very few intent categories that govern most interactions.
NUI Context: Scope

Entity  Collection  Page  App

Navigate
Browse
Informational
Transactional

#IEEEGLOBALASP
Entity Linking

**Problem:** Annotate mentions of an entity with the knowledge base entry for that entity (linking text to knowledge).

**Solution:** Build sequence models and contextual disambiguation models trained over self-disambiguated data such as Wikipedia pages or query-document click data. Features range from topical information, linguistic patterns, morpho-syntactic clues, web usage logs, anchor texts, etc.

Silviu-Petru Cucerzan (EMNLP'07);
Rakesh Agrawal, Ariel Fuxman, Anitha Kannan, John Shafer (WWW'12)
Larry Heck, Dilek Hakkani-Tur, and Gokhan Tur (Interspeech'13)
Ming-Wei Chang, Emre Kiciman (NAACL 2013)
Chin-Yew Lin, Xiaojiang Huang, Yunbo Cao

Even though it's not officially Fall for a few more weeks, the season kicks off right after Labor Day with Mercedes-Benz Fashion Week (September 5-12) as style-savvy buyers, editors, celebrities, and fashionistas (as well as Project Runway's Heidi Klum and Tim Gunn) descend on the city for nonstop couture shows and late-night parties—it's people-watching at its chicest.

Hot New Eats

The New York City Marathon is an annual marathon that courses through the five boroughs of New York City. It is one of the largest marathons in the world, with 45,103 finishers in 2010. Along with the Boston Marathon and C...
Fall 2013 Guide to New York City

Frommer's - By Donna Hostenfeld on September 09, 2013 at 4:22:36 PM EDT

Like the classic Frank Sinatra hit asks, "Autumn in New York... why does it seem so inviting?" New Yorkers like to chalk up the city's third-quarter appeal to its generally sunny weather (with temperatures ranging from pleasantly warm to invigoratingly nippy), the waning of summer tourist crowds, festive parades, Broadway show openings, gargantuan harvest moons rising over the East River, and, of course, the famed New York City Marathon in early November. Here are some of the things that will keep you busy during a September-to-November visit.

Even though it's not officially Fall for a few more weeks, the season kicks off right after Labor Day with Mercedes-Benz Fashion Week (September 5-12) as style-savvy buyers, editors, celebrities, and fashionistas (as well as Project Runway's Heidi Klum and Tim Gunn) descend on the city for monster couture shows and late-night parties—it's people-watching heaven sense, follow your nose to the Feast of San Gennaro (September 12-22), an annual rite of caloric overindulgence featuring favorite Italian foods like sausage and peppers and sautéed calamari, as well as music and entertainment. You can nosh every day from 11:30 am to 11:30 pm on Mulberry and Grand streets in Little Italy.

Per usual, dozens of restaurants extend their menus, but these three are getting special mention: Bryce Shuman and partner Evan Deger (Sidney Street in Madison Park) preside over a special Italian menu served in a Soviet-inspired restaurant; Matt Lambert (formerly of Passepartout) who has the menu inspired by his hometown, custard, cold smoked scallops, and gin.

And given the late August opening of the new SHO restaurant, first to try Australian chef Shuan Hung's previous restaurant, SHO Shank...
Interestingness

Problem: Predict the k things on the page of most interest to the user.

Solution: Train prediction models on web browsing sessions. Model topic semantics of source and target content.

Michael Gamon, Patrick Pantel, Johnson Apacible, Xinying Song (Microsoft Office)

Even though it's not officially Fall for a few more weeks, the season kicks off right after Labor Day with Mercedes-Benz Fashion Week (September 5-12) as style-savvy buyers, editors, celebrities, and fashionistas (as well as Project Runway's Heidi Klum and Tim Gunn) descend on the city for nonstop couture shows and late-night parties—it's people-watching at its chicest.

To satisfy something other than your fashion sense, follow your nose to the Feast of San Gennaro (September 12-22), an annual rite of caloric indulgence featuring favorite Italian foods like sausage and peppers and cannoli, as well as music and entertainment. You can sample every day from 11:30 am to 11:30 pm on Mulberry and Grand streets in Little Italy.

Some of NYC's top parades: The festivities along Central Park South begin with the Manhattan Hootenanny on Sept 21 during the St. Barts German-American Heritage, then march to a Polonaise with the Pulaski Day Parade (featuring Polish-American costumed marchers) on Oct. 14 with the Columbus Day Parade (an all-American extravaganza). On Oct. 31, head to the West Village for the Greenwich Village Halloween Parade (6:30-9 pm on Sixth Ave. between 16th and 7th Sts.) to see outrageous costumes and marching bands. And November 28, find a spot along Central Park West or 72nd St. to catch the world-famous balloons, floats, and marching bands in the 87th Annual Macy's Thanksgiving Day Parade. Or gather outside the American Museum of Natural History on the evening of Nov. 27 to watch the giant balloons spring to life.

Hot New Eats

Per usual, dozens of restaurants debuted in NYC over the last few months, but these three are getting tons of buzz: Betony, where Chef Bryce Shuman and partner Eamon Rockey (both formerly of Eleven Madison Park) preside over a sophisticated and flavor-rich modern American menu served in a soaring space. In Nolita, New Zealand Chef Matt Lambert (formerly of Public) is helming The Musket Room, where the menu is inspired by his homeland and features dishes like mushroom custard, cold smoked scallops, and red New Zealand deer flavored with gin.

And given the late August opening of Junii, you can be among the first to try Australian chef Shawn Hergatt's seasonal tasting menus (his previous restaurant, SHO Shawn Hergatt, had two Michelin stars).
Conversational Search and Browse

**Problem:** Enable search and browse of content with voice.

**Solution:** Dynamically adapt (at run-time) the speech language models and semantic parsers based on the content of the page; identify actionable elements on the page and add to intent list.

Larry Heck, Dilek Hakkani-Tur, Madhu Chinthakunta, Gokhan Tur, Rukmini Iyer, Partha Parthasarathy, Lisa Stifelman, Elizabeth Shriberg, and Ashley Fidler (IEEE SLAM’13)
Semantic Graphs for Conversational Interaction

Experimental Setup

Scenario: conversational search over movies (Netflix)

Training
- Freebase film (movies) domain, 56 relations with linked Wikipedia articles
- Focused on 4 Netflix properties: movies names, actors, genres, directors

Task: Entity spotting (F-Measures of precision/recall)
- Entity modeling only
- Entity plus relation modeling
Semantic Graphs for Conversational Interaction

Summary Approach and Results

**Knowledge as Priors:** Leverage large KGs (Freebase) to bootstrap web-scale semantic parsers
- ~50M entities
- ~700M relations

**Unsupervised Machine Learning**
- No semantic schema design
- No data collection
- No manual annotations

**Graph Crawling Algorithm for Unsupervised Data Mining**

**Entity and Relation Modeling with Mined Data**
- Netflix Search Entity Spotting
  - Entity modeling: 61.0% and 55.4% F-measure (Manual/ASR transcriptions)
  - Entity modeling plus relation: 84.6% and 80.6% F-measure
  - **Within 5.5% of supervised training**

Situated Conversational Interaction

Leverage the situation or context of the user to create a conversational natural user interfaces to the world's knowledge