ConceptNet (Liu and Singh, 2004)

\{wake up\} $\rightarrow$ make coffee
ConceptNet (Liu and Singh, 2004)

\{\text{wake up}\} \rightarrow \text{make coffee}

But what about:

\{\text{go running}\} \rightarrow \text{drink water}

\{\text{cart, broccoli}\} \rightarrow \text{buy food}
No one tweets like this:

“I’m #typing on my #keyboard”

“Now I’m #standing_up”

“Hey, #walking to my #window”
Fiction: “He walked to the bookshelf, picked up his favorite book, and started to read.”

Predict next activity:
pick up book → read

Predict activity from context:
bookshelf → {pick up book, read}
{mountain, tree, backpack} $\rightarrow$ hike
(1) Data
(2) Knowledge Base
(3) Models
(4) Applications
(5) Evaluation
Data: 1.8 billion words of modern fiction from Wattpad, an amateur writing community.
The Augur knowledge base
54,075 activities in Augur:

“He opens the fridge” → open fridge

“She turns off the lights” → turn off lights

“I jumped” → jump
“He opened up Facebook.”

“When we got to the beach, I took off my shirt.”

“We got in the car and drove to the hospital.”
API #1: scene context → activity:
{plate, fork, table} → eating
{car, road} → drive

API #2: activity → activity:
{order, eat} → pay
{shower} → put on clothes
“He drove down the road.”
drove and road co-occur 3590 times

“He drove to the store and parked the car.”
drive and park co-occur 5433 times

“He ate while he drove.”
drove and eat co-occur 102 times
MI(a,b) \sim \log(\ p(a,b) / p(a) \cdot p(b) \ )

p(w) = \text{occurrence of } w
p(w_1, w_2) = \text{co-occurrence of } w_1, w_2

\text{computer and type have high MI}
\text{drive and park have high MI}
\text{tree and eat are have low MI}
Vector spaces allow Augur to leverage multiple examples of scene context.

\[
\text{swim} = [2.1, 1.5, 0.3, \ldots]
\]

\[
\text{swim}_\text{goggles} = 2.1
\]

\[
\text{swim}_\text{pool} = 1.5
\]

\[
\text{swim}_\text{chair} = 0.3
\]
query on “goggles” and “pool”:

\[
\begin{align*}
goggles & \quad \text{pool} \\
0 & \quad 1 & \quad 0 & \quad 0 & \quad 1 & \quad 0 & \quad \ldots
\end{align*}
\]

find vectors (other activities in the space) with the highest cosine similarity
Augur applications

Computer Vision → Augur VSM → Predictions
CV: people, mountain, tree, backpack

Augur: hike, sling, see fire, climb tree, climb, reach top, leap
CV: beach, sand, boy, shoe

Augur: reach beach, lay towel, love beach, take shoe off, swim, lay, dive
CV: boat, sea, sky, ship, ocean, fog

Augur: row, see light, sail, swim, dive
Activity identification: automatic meal photographer
Activity identification: automatic meal photographer

plate + steak + broccoli

fill plate  0.39
put food    0.23
take plate  0.15
**eat food**  0.14
set plate   0.12
cook        0.10
Activity prediction: context aware phone calls
Activity prediction: context aware phone calls

get call + curse

throw phone 0.24
ignore call 0.18
ring 0.18
answer call 0.17
call back 0.17
call number 0.17
leave voicemail 0.17
Spending money wisely

- enter store
- scan 0.19
- ring 0.19
- pay 0.17
- swipe 0.17
- shop 0.13
- buy 0.10
Spending money wisely

call taxi

hail taxi 0.96
pay 0.96
take taxi 0.96
get taxi 0.96
tell address 0.95
get suitcase 0.82
Dynamic music player

stove + pot + spoon

<table>
<thead>
<tr>
<th>Action</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>cook</td>
<td>0.50</td>
</tr>
<tr>
<td>pour</td>
<td>0.39</td>
</tr>
<tr>
<td>place</td>
<td>0.37</td>
</tr>
<tr>
<td>stir</td>
<td>0.37</td>
</tr>
<tr>
<td>eat</td>
<td>0.34</td>
</tr>
</tbody>
</table>
Evaluation
We conducted a two-hour field deployment of our dynamic music player, finding 71% precision and 97% recall over a set of seven common activities.
Computer Vision: road, car, automobile, vehical, blacktop, traffic, people, building, crash action, driver, pavement
Augur detects: Driving
Now playing "The Engine Driver" by "The Decemberists"
Computer Vision: library, book, city, bookshelf, stack, college, indoors, knowledge, research, university, nobody
Augur detects: Reading

Now playing "Bach Concerto" by "Hilary Hahn"
Computer Vision: shopping, urban, city, commerce, store, indoors, path, industry, terminal, beverage, politics
Augur detects: Buying

Now playing "Trojans" by "Atlas Genius"
We tested Augur’s predictions on a dataset of images sampled from the Instagram hashtag #dailylife, and found 94% of predictions were rated as matching the scene.
Imagine a random person is around a chair 100 times. For each action in this list, estimate how many times that action would be taken.

“Imagine a random person is around a chair 100 times. For each action in this list, estimate how many times that action would be taken.”
Mean absolute error: \(12.5\%\) Augur compared to humans
How much more can computers learn from fiction?
In our most compelling visions of human-computer interaction, computers understand the breadth of human life.
How can we teach computers to understand the broad set of things people do?
Thank you to NSF and Toyota for funding support.