Ch.9 Bottom-up/top-down inference in And-Or-Graph

--- Event Recognition and Intent Prediction
Earley-Stolcke parsing algorithm

- $S \rightarrow A B C$
- $S \rightarrow D E F$
- $A \rightarrow a_1$  $A \rightarrow a_2$
- $B \rightarrow b_1$  $B \rightarrow b_2$
- $C \rightarrow c_1$  $C \rightarrow c_2$
- $D \rightarrow a_1$  $D \rightarrow a_2$
- $E \rightarrow b_1$  $E \rightarrow b_2$
- $F \rightarrow c_1$  $F \rightarrow c_2$

- $a_1 b_1 c_1$
- $a_1 b_1 c_2$
- $a_1 b_2 c_1$
- $a_1 b_2 c_2$
- $a_2 b_1 c_1$
- $a_2 b_1 c_2$
- $a_2 b_2 c_1$
- $a_2 b_2 c_2$
### Earley-Stolcke parsing algorithm

The input is \texttt{a1 b1 c1}

<table>
<thead>
<tr>
<th>State set 0</th>
<th>State set 1</th>
<th>State set 2</th>
<th>State set 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-&gt;.S</td>
<td>Scanned</td>
<td>Scanned</td>
<td>Scanned</td>
</tr>
<tr>
<td>Predicted</td>
<td>0A-&gt;a1.</td>
<td>1B-&gt;b1.</td>
<td>2C-&gt;c1.</td>
</tr>
<tr>
<td>0S-&gt;.A B C</td>
<td>0D-&gt;a1.</td>
<td>1E-&gt;b1.</td>
<td>2F-&gt;c1.</td>
</tr>
<tr>
<td>0S-&gt;.D E F</td>
<td>Completed</td>
<td>Completed</td>
<td>Completed</td>
</tr>
<tr>
<td>0A-&gt;.a1</td>
<td>0S-&gt;A B C</td>
<td>0S-&gt;A B C</td>
<td>0S-&gt;A B C</td>
</tr>
<tr>
<td>0A-&gt;.a2</td>
<td>0S-&gt;D E F</td>
<td>0S-&gt;D E F</td>
<td>0S-&gt;D E F</td>
</tr>
<tr>
<td>0D-&gt;.a1</td>
<td>Predicted</td>
<td>Predicted</td>
<td></td>
</tr>
<tr>
<td>0D-&gt;.a2</td>
<td>1B-&gt;.b1</td>
<td>1E-&gt;.b1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1B-&gt;.b2</td>
<td>1E-&gt;.b2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2C-&gt;.c1</td>
<td>2C-&gt;.c2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2F-&gt;.c1</td>
<td>2F-&gt;.c2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0-&gt;S.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Grammars for Events

- UsingWD → ArriveWD UseWD LeaveWD
- ArriveWD → arrivewd
- UseWD → TakeWater
- UseWD → TakeBucket
- TakeWater → benddown1 standup
- TakeBucket → stretchhand drawbackhand
- LeaveWD → leavewd
Grammars for Events

- BendDown $\rightarrow$ Pickup
- BendDown $\rightarrow$ TieString
- Pickup $\rightarrow$ benddown2  standup
- TieString $\rightarrow$ benddown2  standup
AOGs for Events

- UsingWD
  - ArriveWD
  - UseWD
    - TakeWater
      - Bend Down1
      - Stand Up
    - TakeBucket
      - Stretch Hand
      - Drawback Hand
  - LeaveWD

- And Node
- Or Node
- Leaf Node
AOGs for Events

- **And Node**
- **Or Node**
- **Leaf Node**

Diagram:
- **BendDown**
  - **PickUp**
    - **Bend Down2**
    - **Stand Up**
  - **TieString**
    - **Bend Down2**
    - **Stand Up**
    - **Bend down2**
    - **stand up**
The parsing process

- **ArriveWD**
  - **Bend Down1**
  - **Stand Up**
  - **Take Water**
  - **Take Bucket**
  - **Bend Down**
  - **Stand Up**

- **UseWD**
  - **Stretch Hand**
  - **Drawback Hand**

- **LeaveWD**
  - **Bend Down2**
  - **Stand Up**

- **BendDown**
  - **PickUp**
  - **TieString**
The parsing process

UsingWD

ArriveWD

TakeWater

BendDown1

StandUp

Bend down1

stand up

UseWD

TakeBucket

StretchHand

DrawbackHand

Bend Down2

Stretch hand

drawback hand

LeaveWD

Bend

Down2

Stand

Up

Bend down2

stand up

BendDown

PickUp

TieString

Bend Down2

Stand

Up

Bend down2

stand up

arrivewd

leavewd
The parsing process

UsingWD

ArriveWD

TakeWater

BendDown1

Stand Up

Bend Down1

Stand Up

Bend Down2

Stand Up

Bend Down2

Stand Up

UseWD

TakeBucket

Stretch Hand

Drawback Hand

TieString

PickUp

BendDown

LeaveWD

Drawback Hand

Stand Up

Bend Down2

Stand Up

Bend Down2

Stand Up

Drawback Hand

Stand Up

Bend Down2

Stand Up

Bend Down2

Stand Up
The parsing process
The parsing process

UsingWD

UseWD

LeaveWD

ArriveWD

TakeWater

TakeBucket

BendDown

PickUp

TieString

Bend

Down1

Stand

Up

Stretch

Hand

Drawback

Hand

bend
down1

stand
up

stretch
hand
drawback
hand

bend
down2

stand
up

bend
down2

stand
up
The parsing process
The parsing process

UsingWD

ArriveWD

UseWD

TakeWater

TakeBucket

BendDown

PickUp

TieString

BendDown

StandUp

DrawbackHand

StretchHand

BendDown1

StandUp

BendDown2

StandUp
The parsing process
The parsing process
The parsing process
The parsing process

UsingWD

ArriveWD

UseWD

LeaveWD

TakeWater

TakeBucket

Bend

Down1

Stand

Up

Stretch

Hand

Drawback

Hand

BendDown

PickUp

TieString

Bend

Down2

Stand

Up

bend
down1

stand
up

stretch
hand

drawback
hand

bend
down2

stand
up

aw  bd  su  sh
The parsing process

UsingWD

ArriveWD → UseWD

UseWD

TakeWater

TakeBucket

BendDown

BendDown → PickUp

PickUp

TieString

BendDown2

StandUp

bend down1

stand up

stretch hand

drawback hand

bend down2

stand up

bend down2

stand up

aw  bd  su  sh
The parsing process
The parsing process

- **ArriveWD**
  - **UsingWD**
    - **UseWD**
      - **TakeWater**
      - **TakeBucket**
        - **BendDown**
          - **PickUp**
          - **TieString**
        - **Stand Up**
          - **Stretch Hand**
          - **Drawback Hand**
      - **Stand Up**
    - **Bend Down1**
      - **ArriveWD**
      - **Stand Up**
    - **Stretch Hand**
      - **Drawback Hand**
    - **Bend Down2**
      - **LeaveWD**
      - **Stand Up**
  - **LeaveWD**

**Actions:**
- Bend Down1
- Stand Up
- Stretch Hand
- Drawback Hand
- Bend Down2
- Stand Up
- ArriveWD
- LeaveWD

**Terms:**
- aw
- bd
- su
- sh
- dh
The parsing process
The parsing process
The parsing process
Goal: Recognize events in daily scenes

For example, in an office scene

Work by Mingtao Pei, UCLA
Challenges

1. Events happen over an extended time period
   • Variant time-span
   • Could be interrupted
   • Multiple routes
   • Intention and prediction

2. Actions are hard to recognize
   • Subtle and similar
   • No salient motion/pose at most of the time
   • Contextual OBJCTS are key!!
Overview of our approach

Spatial image parsing

Video

Background objects

Refill tea

make phone call

Temporal event parsing

Refill water

Empty the cup

Renew tea leaf

Events

Water collector

Trash can

Water boiler

Phone

Tea canister

Arrive trashcan

Bend & dump

Leaves trashcan

Arrive tea canister

Grab tea canister

Leave tea canister

Arrive boiler

Bend & leave boiler

Arrive boiler

Bend & pour boiler
Scene parsing
Event as temporal And-Or-Graph
Event as temporal And-Or-Graph
Formulation

\[ p(g) = \frac{1}{Z} \exp\{\text{score}(g)\} \]

\[
\text{score}(g) = \sum_{v_t \in T(g)} \lambda_{v_t} \alpha(v_t) + \sum_{v \in V_o(g)} \lambda_v \omega(v) + \sum_{(i,j) \in E(g)} \lambda_{ij} r_{ij}(v_i, v_j)
\]
Formulation

\[ p(g) = \frac{1}{Z} \exp\{\text{score}(g)\} \]

\[
\text{score}(g) = \sum_{v_t \in T(g)} \lambda_{v_t} \alpha(v_t) + \sum_{v \in V_o(g)} \lambda_v \omega(v) + \sum_{(i,j) \in E(g)} \lambda_{ij} r_{ij}(v_i, v_j)
\]

\[
\alpha(v_t) = \sum_{i \in F} \beta_i h_i(v_t) - \text{dist}(P_{\text{person}}, P_{\text{obj}})
\]
Combing action and contextual object

Event 1:
Bend down action + Close to Trash Can = Drop waste

Event 2:
Reach out action + Close to Elevator = Use elevator
Parsing process (Earley Parser [Earley 1970])

And-node

Or-node

Leaf-node

Refill tea

Empty cup

Renew tea leaf

Refill water

Dump old tea

Dump water

Null

Arrive trashcan

Bend down

Leave trashcan

Arrive collector

Bend down

Leave collector

Get close to tea can

Grab tea can

Leave tea-can

Arrive boiler

Bend down

Leave boiler
Parsing process (Earley Parser [Earley 1970])

And-node
Or-node
Leaf-node

Refill tea
Empty cup
Renew tea leaf
Refill water

Dump old tea
Dump water
Null

Arrive trashcan
Bend down
Leave trashcan
Arrive collector
Bend down
Leave collector
Get close to tea can
Grab tea can
Leave tea-can
Arrive boiler
Bend down
Leave boiler
Parsing process (Earley Parser [Earley 1970])

- And-node
- Or-node
- Leaf-node
Parsing process (Earley Parser [Earley 1970])
Duration

Arrive trashcan
Bend down
Leave trashcan
Arrive collector
Bend down
Leave collector
NULL
Get close to tea can
Grab tea can
Leave tea-can
Arrive boiler
Bend down
Leave boiler

Refill tea
Empty cup
Renew tea leaf
Refill water

And-node
Or-node
Leaf-node

Parsing process (Earley Parser [Earley 1970])
**Parsing process (Earley Parser [Earley 1970])**

- **And-node**
- **Or-node**
- **Leaf-node**

- **Refill tea**
- **Empty cup**
- **Renew tea leaf**
- **Refill water**
- **Dump old tea**
- **Dump water**
- **Null**

- **Arrive trashcan**
- **Bend down**
- **Leave trashcan**
- **Arrive collector**
- **Bend down**
- **Leave collector**
- **Get close to tea can**
- **Grab tea can**
- **Leave tea-can**
- **Arrive boiler**
- **Bend down**
- **Leave boiler**

Earley Parser [Earley 1970]
Refill tea

Empty cup

Renew tea leaf

Refill water

Null

Arrive trashcan
Bend down
Leave trashcan
Get close to tea can

Arrive collector
Bend down
Grab tea can

Arrive boiler
Bend down
Leave boiler

And-node
Or-node
Leaf-node
Parsing process (Earley Parser [Earley 1970])

And-node

Or-node

Leaf-node

Arrive trashcan

Bend down

Leave trashcan

Arrive collector

Bend down

Leave collector

Dump old tea

Empty cup

Dump water

Null

Get close to tea can

Grab tea can

Leave tea-can

Arrive boiler

Bend down

Leave boiler

Refill tea

Renew tea leaf

Refill water

Refill water
Refill tea

Empty cup

Renew tea leaf

Refill water

And-node

Or-node

Leaf-node

Arrive trashcan

Bend down

Leave trashcan

Arrive collector

Bend down

Leave collector

NULL

Get close to tea can

Grab tea can

Leave tea-can

Arrive boiler

Bend down

Leave boiler
Parsing process (Earley Parser [Earley 1970])

Refill tea

Empty cup

Renew tea leaf

Refill water

Dump old tea

Dump water

Null

Renew tea leaf

Refill water

Arrive trashcan → Bend down → Leave trashcan → Arrive collector → Bend down → Leave collector → NULL

Get close to tea can → Grab tea can → Leave tea-can → Arrive boiler → Bend down → Leave boiler
Parsing process (Earley Parser [Earley 1970])

And-node
Or-node
Leaf-node

Refill tea
Empty cup
Renew tea leaf
Refill water

Dump old tea
Dump water
Null
Renew tea leaf
Refill water

Arrive trashcan
Bend down
Leave trashcan
Arrive collector
Bend down
Leave collector
NULL
Get close to tea can
Grab tea can
Leave tea-can
Arrive boiler
Bend down
Leave boiler
Parsing process (Earley Parser [Earley 1970])
Parsing process (Earley Parser [Earley 1970])
Parsing process (Earley Parser [Earley 1970])

- And-node
- Or-node
- Leaf-node

Refill tea

Empty cup

Renew tea leaf

Refill water

Dump old tea

Dump water

Null

Get close to tea can

Grab tea can

Leave tea-can

Arrive boiler

Bend down

Leave boiler

Arrive trashcan

Bend down

Leave trashcan

Arrive collector

Bend down

Leave collector

NULL

Arrive boiler

Bend down

Leave boiler
Parsing process (Earley Parser [Earley 1970])

And-node
Or-node
Leaf-node

Refill tea

Empty cup
Renew tea leaf
Refill water

Dump old tea
Dump water
Null
Renew tea leaf
Refill water

Arrive trashcan
Bend down
Leave trashcan
Grab tea can
Leave tea-can
Arrive boiler
Bend down
Leave boiler

Get close to tea can
Leave collector

Earley Parser [Earley 1970]
Parsing process (Earley Parser [Earley 1970])

And-node

Or-node

Leaf-node

Refill tea

Empty cup

Renew tea leaf

Null

Refill water

Dump old tea

Dump water

Renew tea leaf

Get close to tea can

Grab tea can

Leaves tea-can

Arrive boiler

Bend down

Leave boiler
Intention and prediction
Intention and prediction

Time 2

Refill tea

Make phone call

intention

prediction

Refill tea

Make phone call

Grab tea can

Pick up receiver
Handle event interruption

- And-Node
- Or-Node
- Leaf-Node

First Partial parse tree of take water
Parse tree of take a phone
Second Partial parse tree of take water

Observed Data
Comparison with human prediction [Baker, Saxe and Tenenbaum 2009]

![Graph showing comparison between human prediction and algorithm]

- **Refill tea**
- **Phone call**

**Average of participants**

<table>
<thead>
<tr>
<th>Judge points</th>
<th>Goal inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>2</td>
<td>0.6</td>
</tr>
<tr>
<td>3</td>
<td>0.4</td>
</tr>
<tr>
<td>4</td>
<td>0.2</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

**Algorithm**

<table>
<thead>
<tr>
<th>Judge points</th>
<th>Goal inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>2</td>
<td>0.6</td>
</tr>
<tr>
<td>3</td>
<td>0.4</td>
</tr>
<tr>
<td>4</td>
<td>0.2</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>
Synthesize new events by sampling the AoG

Sample from AOG

Synthesized Event
Event synthesis

Sample from AOG

Synthesized Event
Event synthesis

Sample from AOG

Synthesized Event
* We use stored foreground windows and a background frame to synthesize new videos
Computation complexity of parsing

- Initially the number of interpretations above a threshold grows rapidly over time.
- At certain decisive moments, i.e. when informative actions are observed, large number of unlikely interpretation drops below the threshold and hence is pruned.

Pickup phone
Reach water boiler