Script Induction as Association Rule Mining

Anton Belyy

Benjamin Van Durme

JOHNS HOPKINS UNIVERSITY

Schematic / Script Knowledge

Understanding the types of events that go together is essential to learn scripts

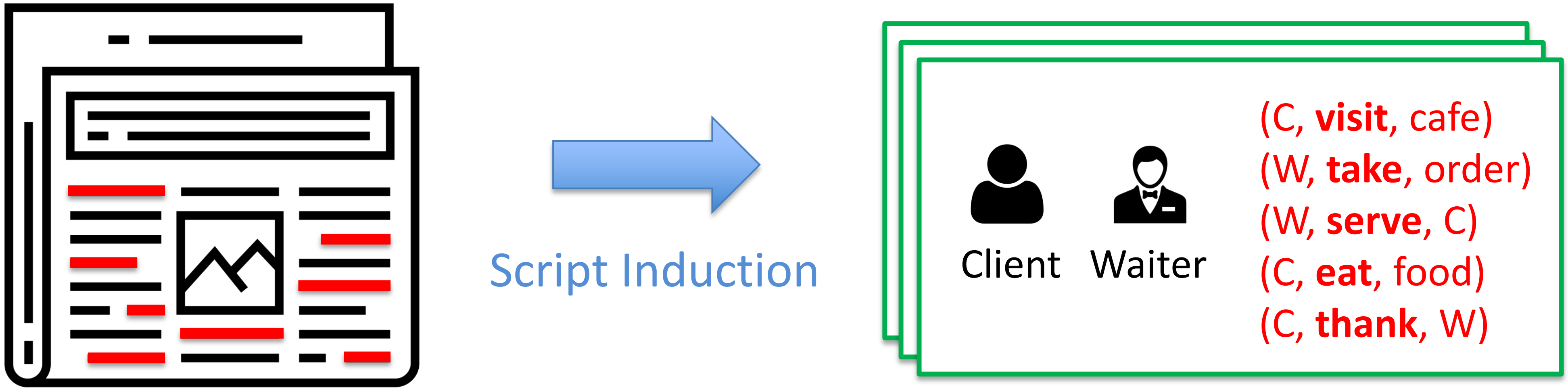
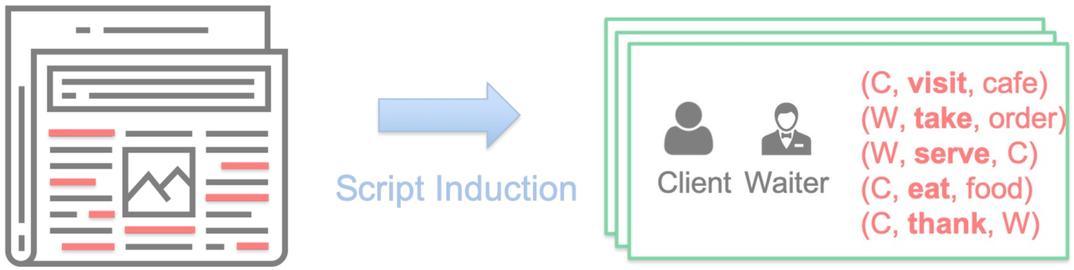


Image credit: ProSymbols, DvM Design (The Noun Project)

Schematic / Script Knowledge

Understanding the types of events that go together is essential to learn scripts



Popular approaches:

- Knowledge-based (Minsky, 1974; Schank and Abelson, 1977; Mooney and DeJong, 1985)
- **Count-based (Chambers and Jurafsky, 2008, 2009; Jans et al., 2012; Pichotta and Mooney, 2014)**
- LM-based (Rudinger et al, 2015; Pichotta and Mooney, 2016; Weber et al., 2018)

Primary focus of this work

Image credit: ProSymbols, DvM Design (The Noun Project)

Looking at things differently

Different views of popular approaches is also important

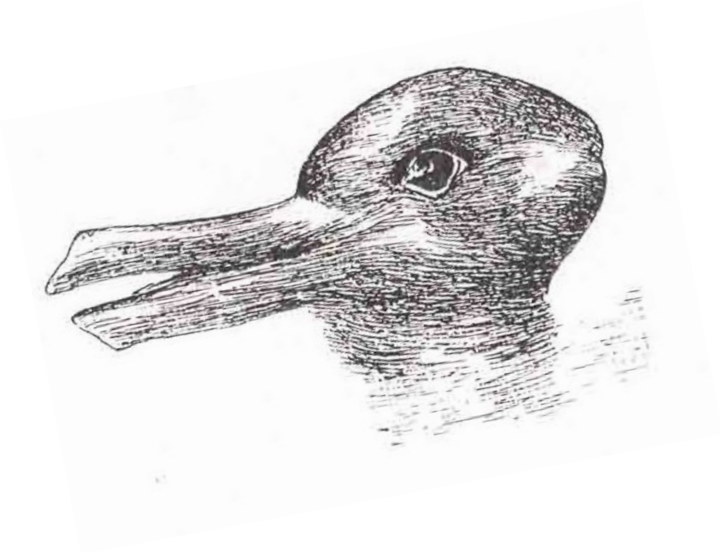
Brings in fresh perspective

More literature to borrow from

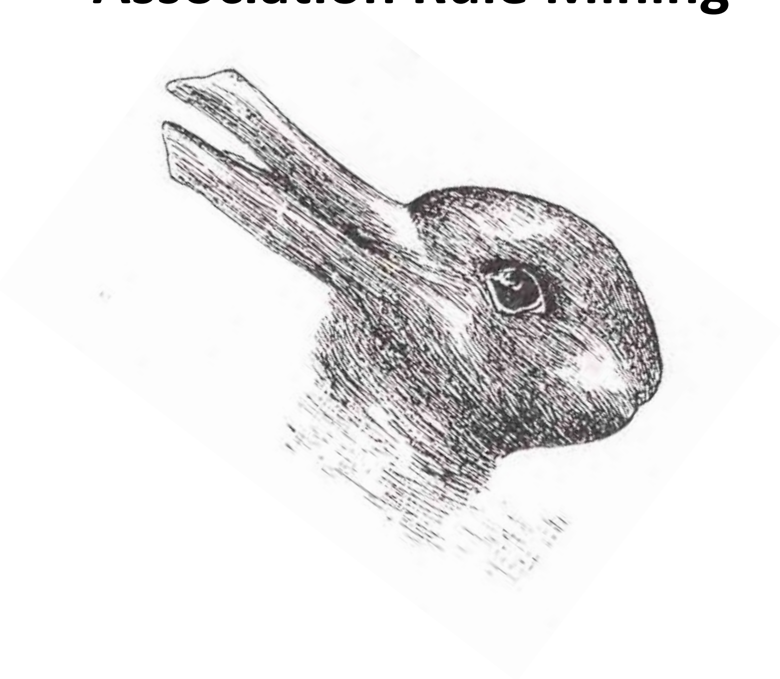
i.e. tomorrow's advances were maybe considered 30 years ago…

In this work

(Count-based) Script Induction



Association Rule Mining



Narrative Cloze Test (Chambers and Jurafsky, 2008)

event_1 event_2 __________ event_4 … event_L

narrative chain

- **partially ordered** set of events
- shares a common actor (**protagonist**)

Narrative Cloze Test (Chambers and Jurafsky, 2008)

event$_1$ event$_2$ __________ event$_4$ … event$_L$

narrative chain

Client Waiter

(C, **visit**, cafe)
(W, **take**, order)
(W, **serve**, C)
(C, **eat**, food)
(C, **thank**, W)

Protagonist (Waiter)

~~(C, **visit**, cafe)~~
(**take**, nsubj)
(**serve**, nsubj)
~~(C, **eat**, food)~~
(**thank**, dobj)

Protagonist (Client)

(**visit**, nsubj)
~~(W, **take**, order)~~
(**serve**, dobj)
(**eat**, nsubj)
(**thank**, nsubj)

Narrative Cloze Test (Chambers and Jurafsky, 2008)

(visit, nsubj) (serve, dobj) ________ (thank, nsubj)

Which event should fill in the gap?

(take, nsubj)
(tell, nsubj)
(eat, nsubj)
...
(persuade, dobj)
(sail, nsubj)
(regard, dobj)

Event vocabulary

Protagonist (Client)

(**visit**, nsubj)
~~(W, **take**, order)~~
(**serve**, dobj)
(**eat**, nsubj)
(**thank**, nsubj)

JOHNS HOPKINS UNIVERSITY

Unordered PMI Model (Chambers and Jurafsky, 2008)

event_1 event_2 ________ event_4 … event_L

candidate event ***e***

UOP model:

$$S(\boldsymbol{e}) = \text{pmi}(event_1, \boldsymbol{e}) + \text{pmi}(event_2, \boldsymbol{e}) + \text{pmi}(event_4, \boldsymbol{e}) + \cdots + \text{pmi}(event_L, \boldsymbol{e})$$

Choose candidate event ***e*** with the **highest score** S(***e***)

$$\text{pmi}(\boldsymbol{e}_1, \boldsymbol{e}_2) \propto \log \frac{C(\boldsymbol{e_1}, \boldsymbol{e_2})}{C(\boldsymbol{e_1}, *)C(*, \boldsymbol{e_2})}$$

Unordered PMI Model: A Closer Look

event$_1$ event$_2$ __________ event$_4$ … event$_L$

candidate event $\boldsymbol{e}$

$$\underset{\boldsymbol{e} \in \boldsymbol{Vocab}}{\operatorname{argmax}} S(\boldsymbol{e}) = \underset{\boldsymbol{e} \in \boldsymbol{Vocab}}{\operatorname{argmax}} \sum_{i=1}^{L} \operatorname{pmi}(event_i, \boldsymbol{e})$$

$$= \underset{\boldsymbol{e} \in \boldsymbol{Vocab}}{\operatorname{argmax}} \sum_{i=1}^{L} \log \frac{C(event_i, \boldsymbol{e})}{C(*, \boldsymbol{e}) \cancel{C(event_i, *)}}$$

$$= \underset{\boldsymbol{e} \in \boldsymbol{Vocab}}{\operatorname{argmax}} \prod_{i=1}^{L} \Pr(\mathrm{event_i} \mid \boldsymbol{e})$$

$$\boxed{= \underset{\boldsymbol{e} \in \boldsymbol{Vocab}}{\operatorname{argmax}} \Pr(\mathrm{event}_1, \mathrm{event}_2, \ldots, \mathrm{event_L} \mid \boldsymbol{e})}$$

Looks like a probabilistic model with full conditional independence!

JOHNS HOPKINS UNIVERSITY

Association Rule Mining (Agrawal et al, 1993)

Given a set of narrative chains (transactions):

Mine all **interesting*** rules X ⇒ Y of form "if we see a set of events **Y**, then we probably see another set of events **X**"

$$int(X \Longrightarrow Y) = P(X \mid Y) = \frac{\sup(X \cup Y)}{\mathrm{wsup}(Y)}$$

*Interesting rules: all tuples (X, Y) s.t.

$$P(X \mid Y) > t_{int}$$

ARM Decoder for Narrative Cloze Test

1

$$S(e) = \Pr(\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6} \mid e)$$

existing chain

candidate event **e**

2 Collect all ARM rules that "predict" parts of existing chain given **e**:

$\boxed{1}\ \boxed{2}\ \boxed{4} \Rightarrow e$ $\quad \boxed{3}\ \boxed{5} \Rightarrow e$ $\quad \boxed{6} \Rightarrow e$ $\quad \boxed{3} \Rightarrow e$

3 **"Cover"** the existing chain with rules s.t. the rules have **minimum total weight**:

$$S(e) = \Pr(\boxed{1}\ \boxed{2}\ \boxed{4} \mid e)\ \Pr(\boxed{3}\ \boxed{5} \mid e)\ \Pr(\boxed{6} \mid e)$$

weight = -log(Pr) $\Rightarrow$ **maximum total probability**

ARM Decoder vs Unordered PMI model

event_1 event_2 ________ event_4 … event_L

candidate event ***e***

$$\text{argmax}\, S(\boldsymbol{e}) = \text{argmax}\, \Pr(\text{event}_1, \text{event}_2, \ldots, \text{event}_L \mid \boldsymbol{e})$$

ARM Decoder: Find most **interesting rules** and **combine** them to explain ***e***

event_{i1} … event_{ik} $\Rightarrow$ $\boldsymbol{e}$

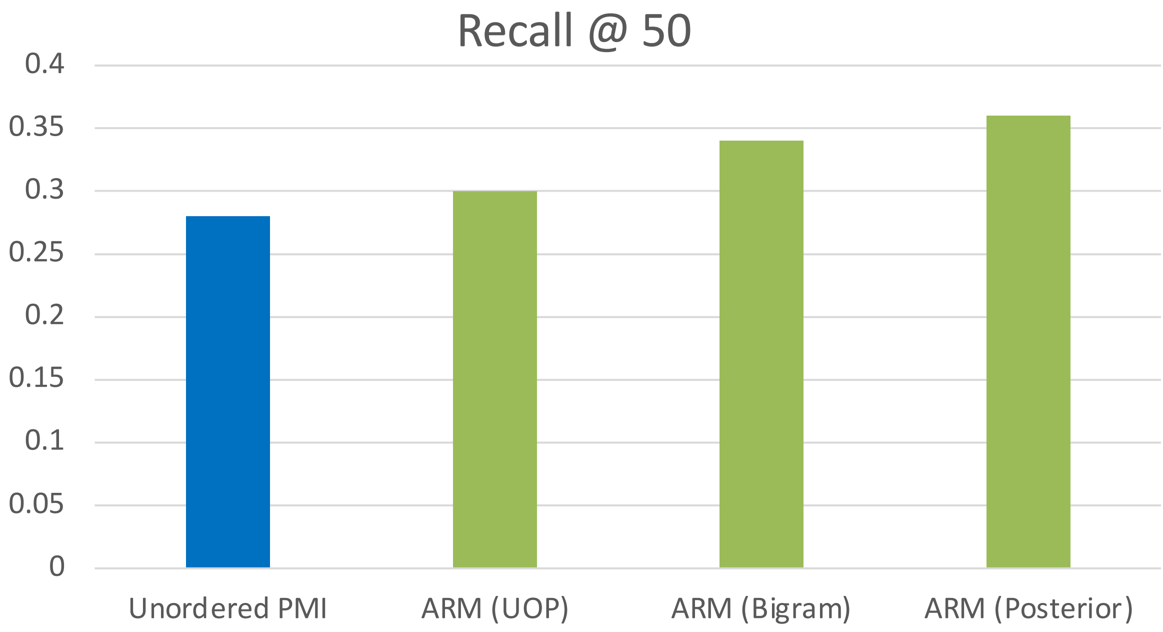
Unordered PMI: Explain ***e*** using PMI between **pairs** of events

event_{i} $\Rightarrow$ $\boldsymbol{e}$

JOHNS HOPKINS UNIVERSITY

Experiments & Results

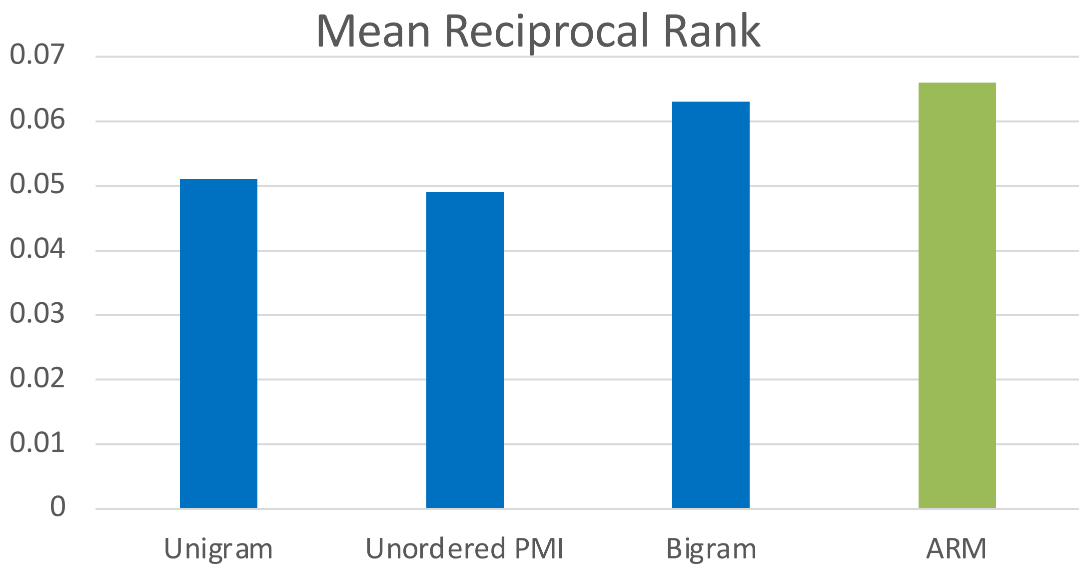
- New York Times part of Annotated Gigaword
- 20K unique events
- Train: 1.3M chains, 8.7M events
- Dev: 10K chains, 62K events
- Test: 5K chains, 31K events

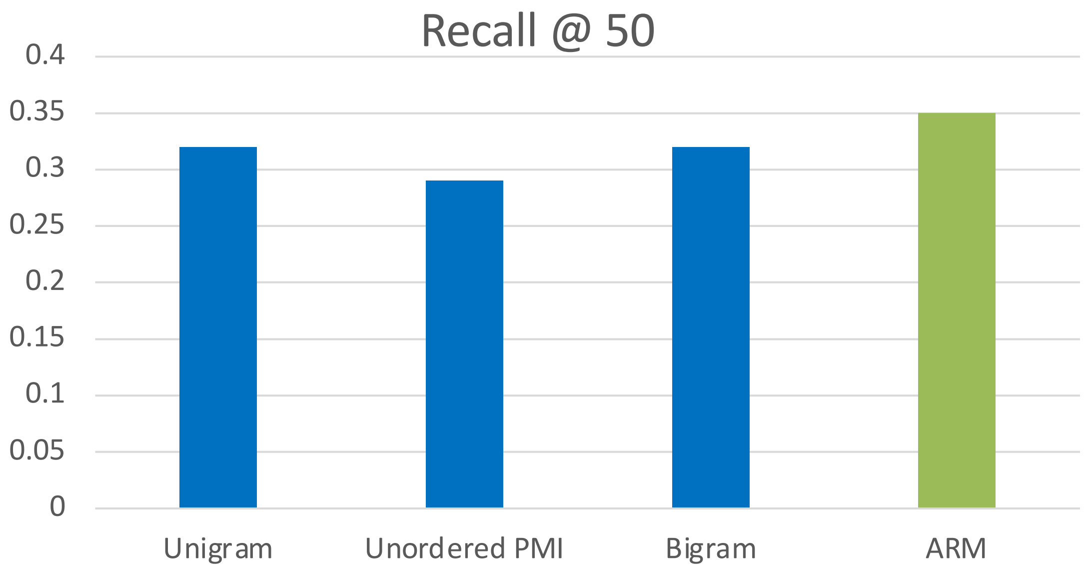


Performance of different ARM models on the narrative cloze test (dev set)

Johns Hopkins University

Experiments & Results





Performance of Count-Based SI models on the narrative cloze test (test set)

Thank you!

Anton Belyy

Benjamin Van Durme

JOHNS HOPKINS UNIVERSITY