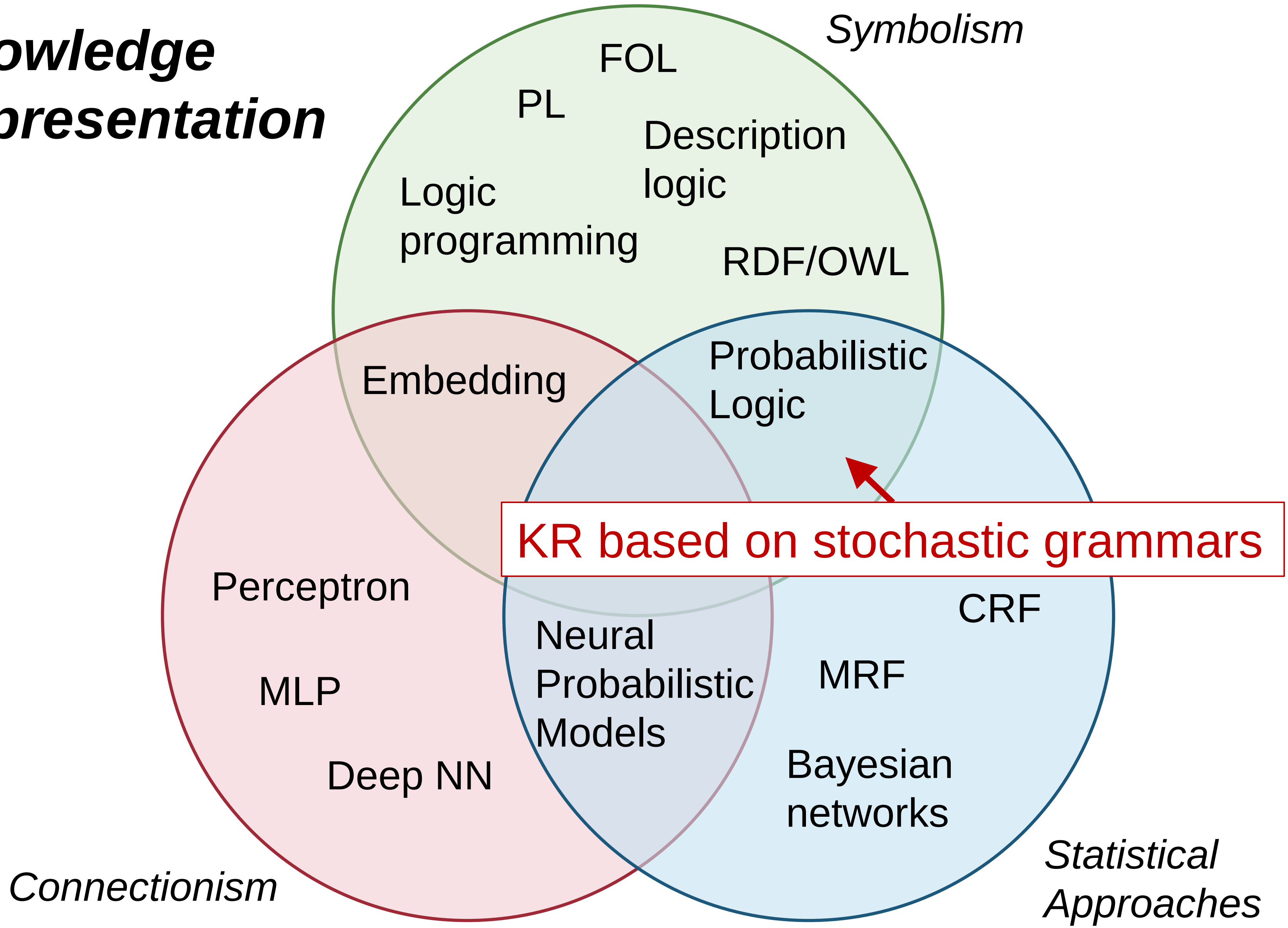
Stochastic And-Or Grammars: Representation & Learning

Kewei Tu

ShanghaiTech University

上海科技大学
ShanghaiTech University

Knowledge Representation



Grammars of language

- A formal grammar has four components
  - A set Σ of terminals (words)
  - A set N of nonterminals (phrases)
  - A start symbol $S \in N$
  - A set R of production rules
    - Specifies how a string of terminals and/or nonterminals can be rewritten to another
    - Context-free grammar (CFG): $A \to \gamma$
- Stochastic grammars
  - Each production rule is associated with a probability

Example grammar

$S \rightarrow NP\ VP$	$Det \rightarrow that \mid this \mid a$
$S \rightarrow Aux\ NP\ VP$	$Noun \rightarrow book \mid flight \mid meal \mid money$
$S \rightarrow VP$	$Verb \rightarrow book \mid include \mid prefer$
$NP \rightarrow Pronoun$	$Pronoun \rightarrow I \mid she \mid me$
$NP \rightarrow Proper\text{-}Noun$	$Proper\text{-}Noun \rightarrow Houston \mid NWA$
$NP \rightarrow Det\ Nominal$	$Aux \rightarrow does$
$Nominal \rightarrow Noun$	$Preposition \rightarrow from \mid to \mid on \mid near \mid through$
$Nominal \rightarrow Nominal\ Noun$	
$Nominal \rightarrow Nominal\ PP$	
$VP \rightarrow Verb$	
$VP \rightarrow Verb\ NP$	
$VP \rightarrow Verb\ NP\ PP$	
$VP \rightarrow Verb\ PP$	
$VP \rightarrow VP\ PP$	
$PP \rightarrow Preposition\ NP$	

Example stochastic grammar

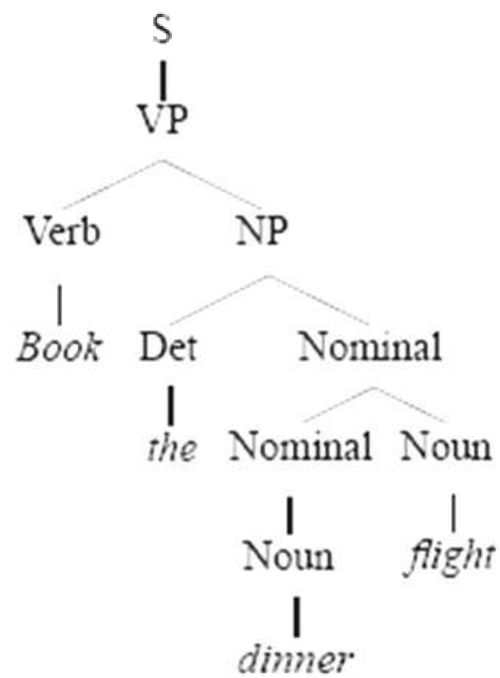
$S \rightarrow NP\ VP$	[.80]		
$S \rightarrow Aux\ NP\ VP$	[.15]		
$S \rightarrow VP$	[.05]		
$NP \rightarrow Pronoun$	[.35]		
$NP \rightarrow Proper\text{-}Noun$	[.30]		
$NP \rightarrow Det\ Nominal$	[.20]		
$NP \rightarrow Nominal$	[.15]		
$Nominal \rightarrow Noun$	[.75]		
$Nominal \rightarrow Nominal\ Noun$	[.20]		
$Nominal \rightarrow Nominal\ PP$	[.05]		
$VP \rightarrow Verb$	[.35]		
$VP \rightarrow Verb\ NP$	[.20]		
$VP \rightarrow Verb\ NP\ PP$	[.10]		
$VP \rightarrow Verb\ PP$	[.15]		
$VP \rightarrow Verb\ NP\ NP$	[.05]		
$VP \rightarrow VP\ PP$	[.15]		
$PP \rightarrow Preposition\ NP$	[1.0]		

$Det \rightarrow$	*that* [.10] | *a* [.30] | *the* [.60]
$Noun \rightarrow$	*book* [.10] | *flight* [.30] | *meal* [.15] | *money* [.05] | *flights* [.40] | *dinner* [.10]
$Verb \rightarrow$	*book* [.30] | *include* [.30] | *prefer;* [.40]
$Pronoun \rightarrow$	*I* [.40] | *she* [.05] | *me* [.15] | *you* [.40]
$Proper\text{-}Noun \rightarrow$	*Houston* [.60] | *NWA* [.40]
$Aux \rightarrow$	*does* [.60] | *can* [40]
$Preposition \rightarrow$	*from* [.30] | *to* [.30] | *on* [.20] | *near* [.15] | *through* [.05]

Example

$S \rightarrow NP\ VP$	[.80]
$S \rightarrow Aux\ NP\ VP$	[.15]
$S \rightarrow VP$	[.05]
$NP \rightarrow Pronoun$	[.35]
$NP \rightarrow Proper\text{-}Noun$	[.30]
$NP \rightarrow Det\ Nominal$	[.20]
$NP \rightarrow Nominal$	[.15]
$Nominal \rightarrow Noun$	[.75]
$Nominal \rightarrow Nominal\ Noun$	[.20]
$Nominal \rightarrow Nominal\ PP$	[.05]
$VP \rightarrow Verb$	[.35]
$VP \rightarrow Verb\ NP$	[.20]
$VP \rightarrow Verb\ NP\ PP$	[.10]
$VP \rightarrow Verb\ PP$	[.15]
$VP \rightarrow Verb\ NP\ NP$	[.05]
$VP \rightarrow VP\ PP$	[.15]
$PP \rightarrow Preposition\ NP$	[1.0]

……



Book the dinner flight

$$P(T) = .05 \times .20 \times .20 \times .20 \times .75 \times .30 \times .60 \times .10 \times .40 = 2.2 \times 10^{-6}$$

The And-Or Normal Form of SCFG

- Two types of non-terminals
  - And-nodes, Or-nodes
- Two types of production rules
  - And-rule (composition): A → N1 N2 …
  - Or-rule (alternative configurations): O → N1 | N2 | t1 | …

SCFG

$S \rightarrow a\ (0.4) \mid AB\ (0.6)$

$A \rightarrow a\ (1.0)$

$B \rightarrow b_1\ (0.2) \mid b_2\ (0.5) \mid b_3\ (0.3)$

The AND-OR Form

$\mathrm{OR}_S \rightarrow a\ (0.4) \mid \mathrm{AND}_{AB}\ (0.6)$

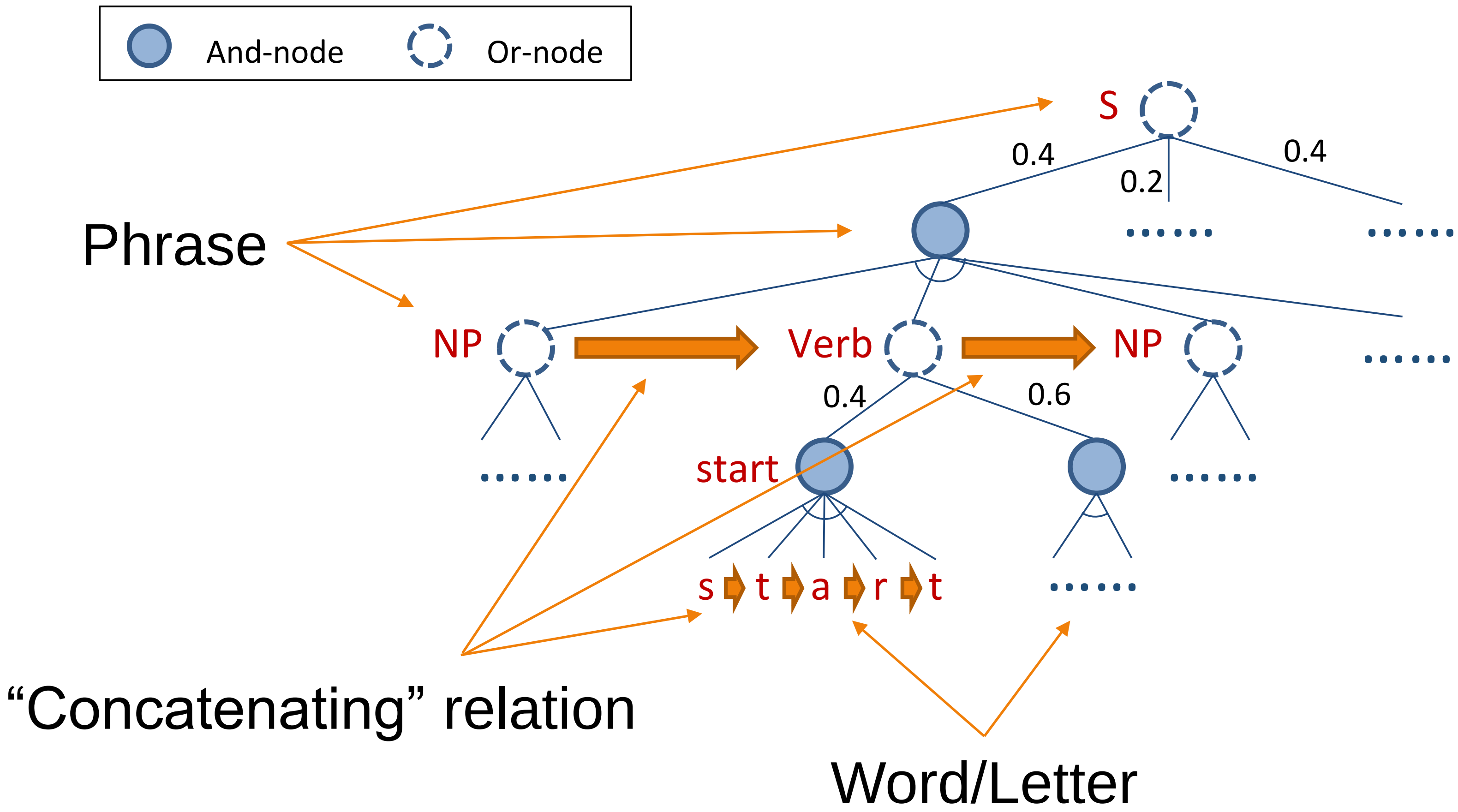
$\mathrm{AND}_{AB} \rightarrow \mathrm{OR}_A \mathrm{OR}_B$

$\mathrm{OR}_A \rightarrow a\ (1.0)$

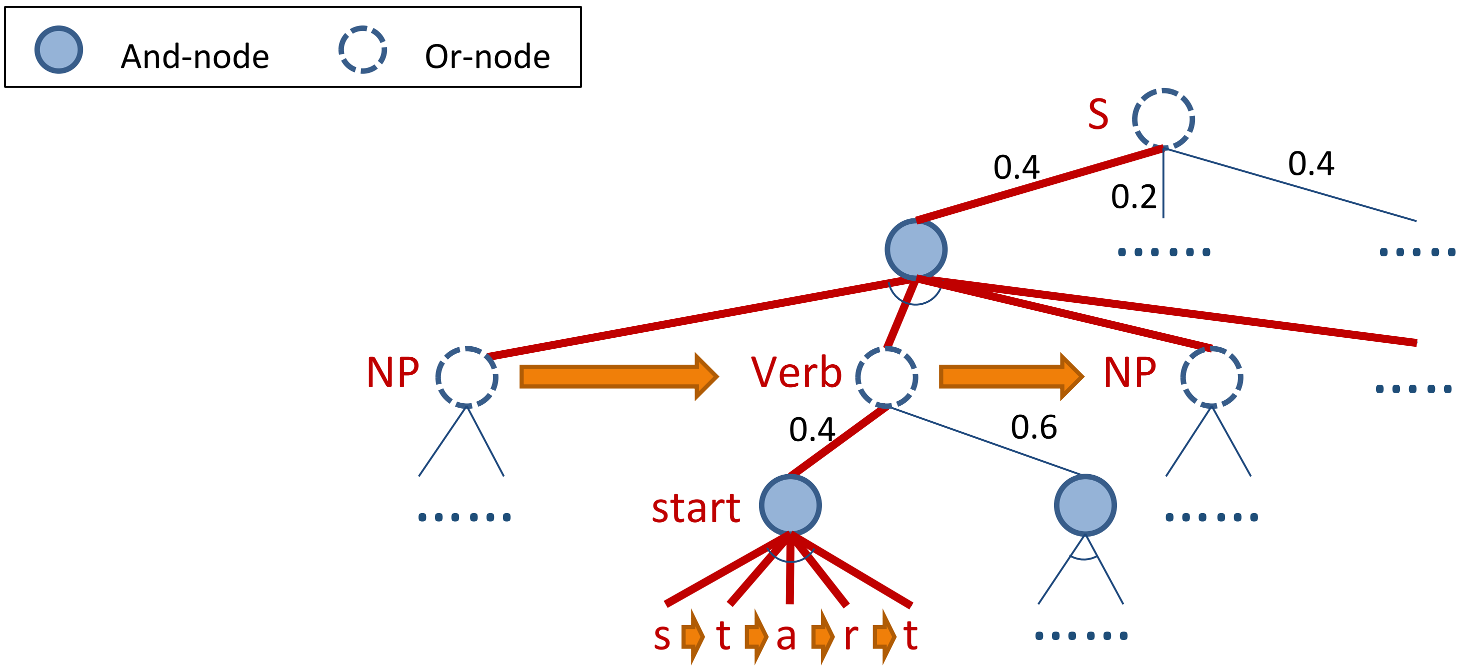
$\mathrm{OR}_B \rightarrow b_1\ (0.2) \mid b_2\ (0.5) \mid b_3\ (0.3)$

[Tu & Honavar, ICGI 2008]

The And-Or Normal Form of SCFG

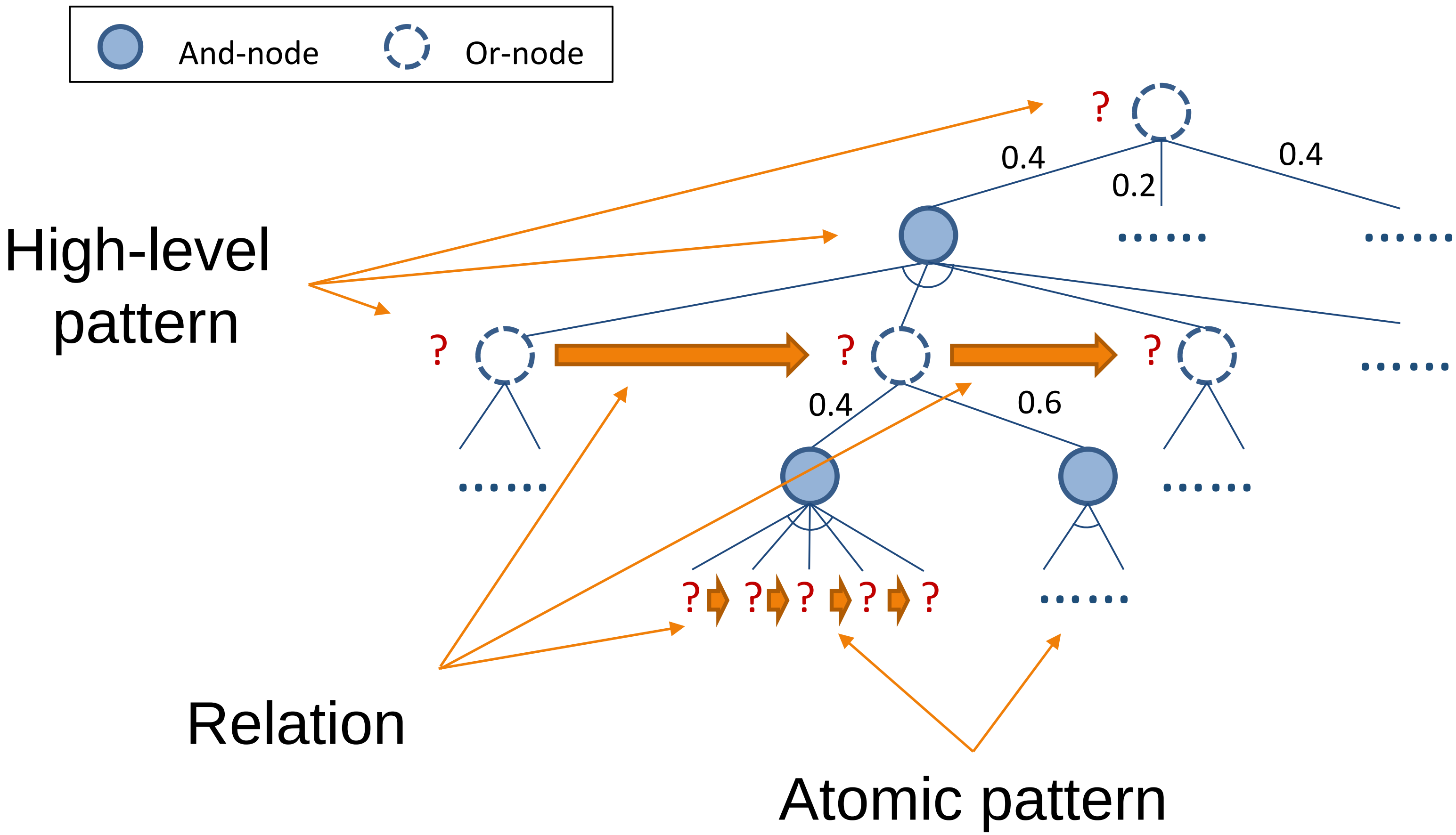


The And-Or Normal Form of SCFG

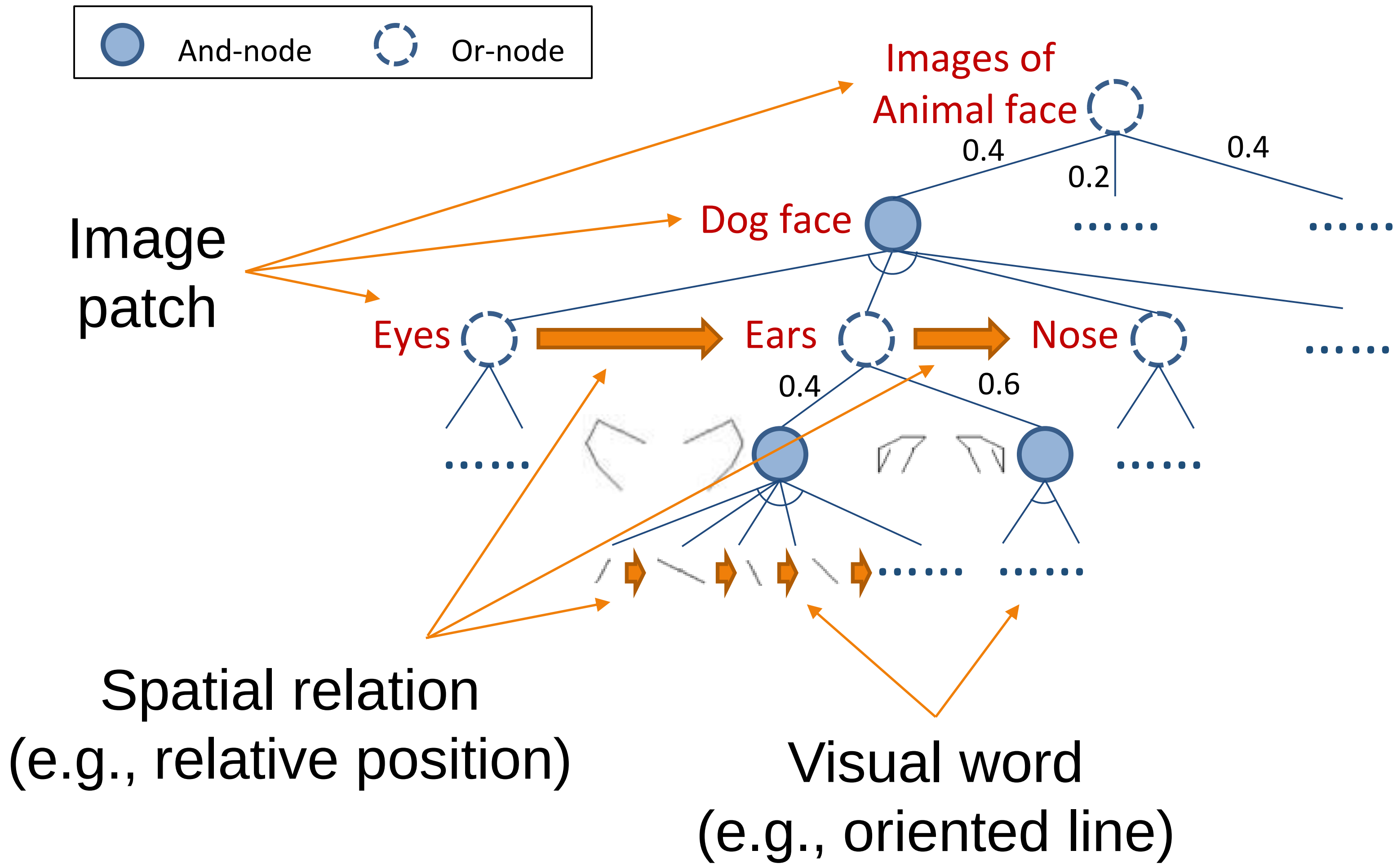


"They start the book today."

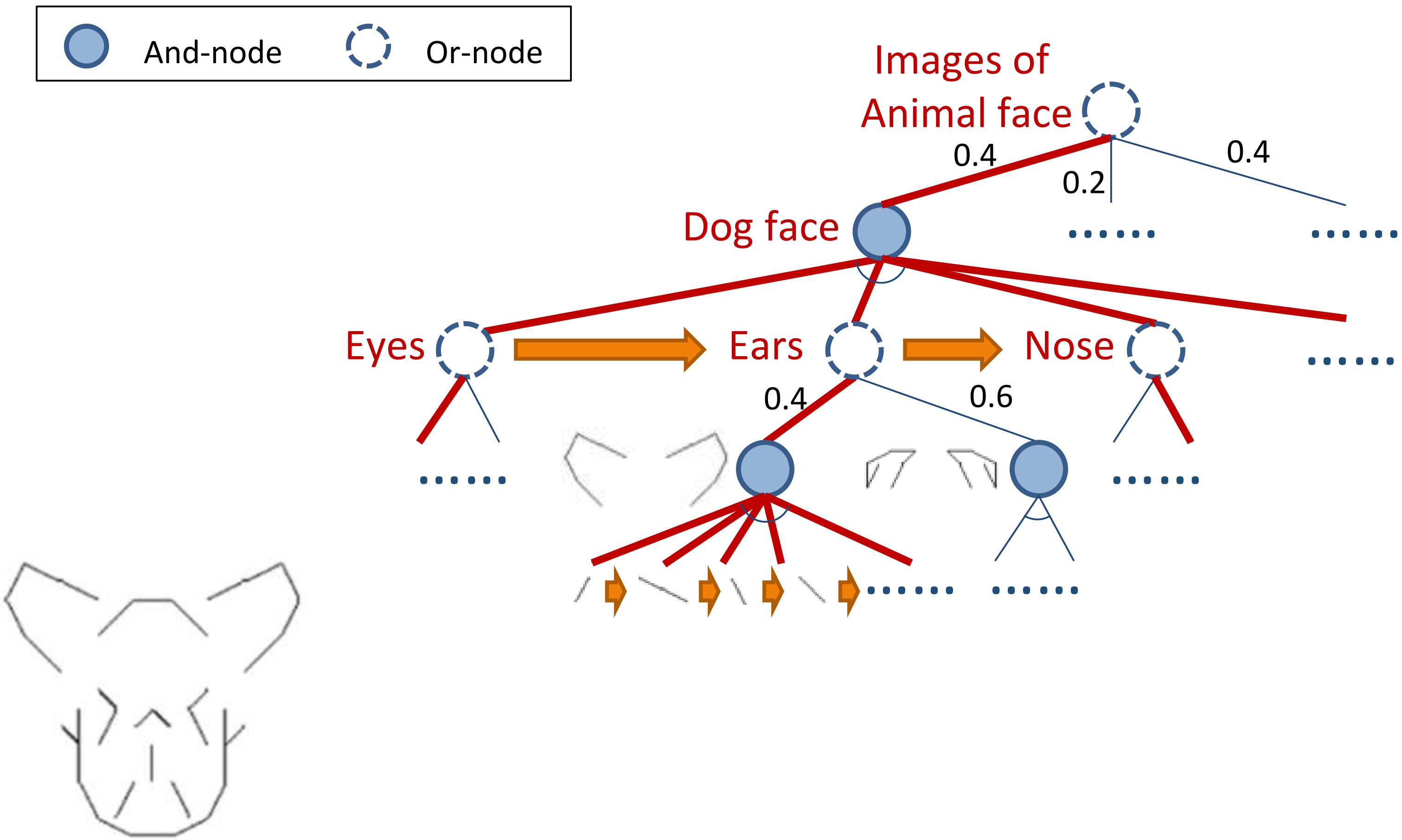
Generalization of SCFG



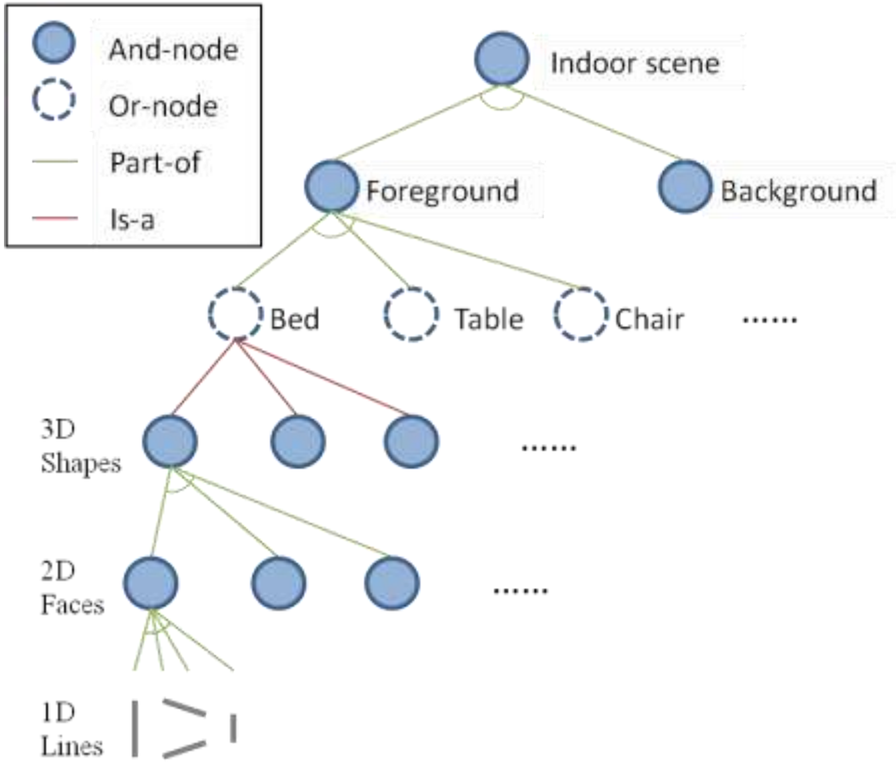
Generalized SCFG of Images



Generalized SCFG of Images

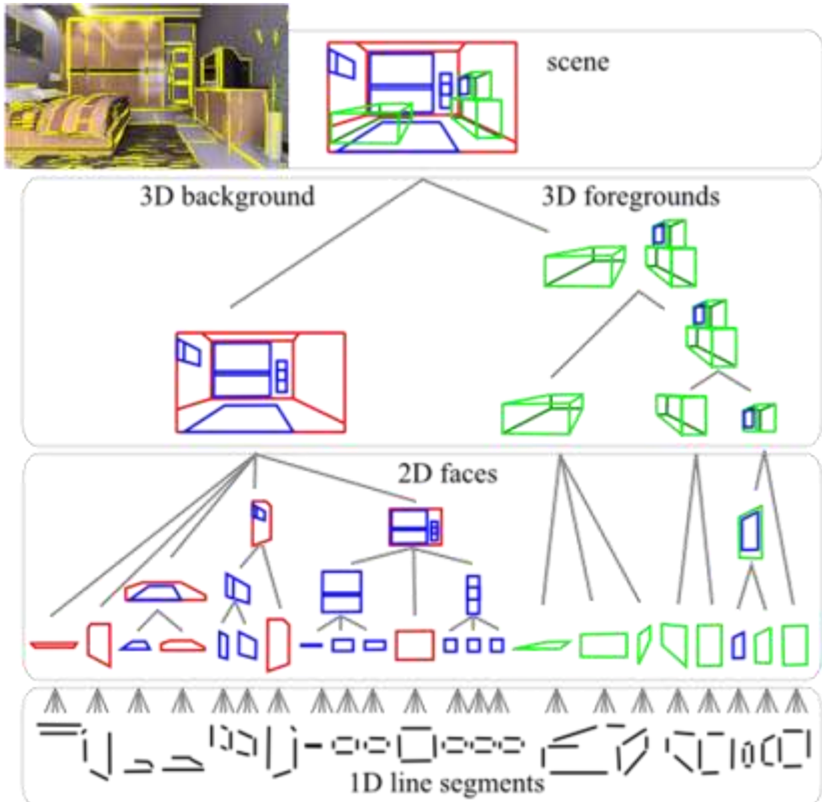


Applications – indoor scene parsing



(a)

Grammar

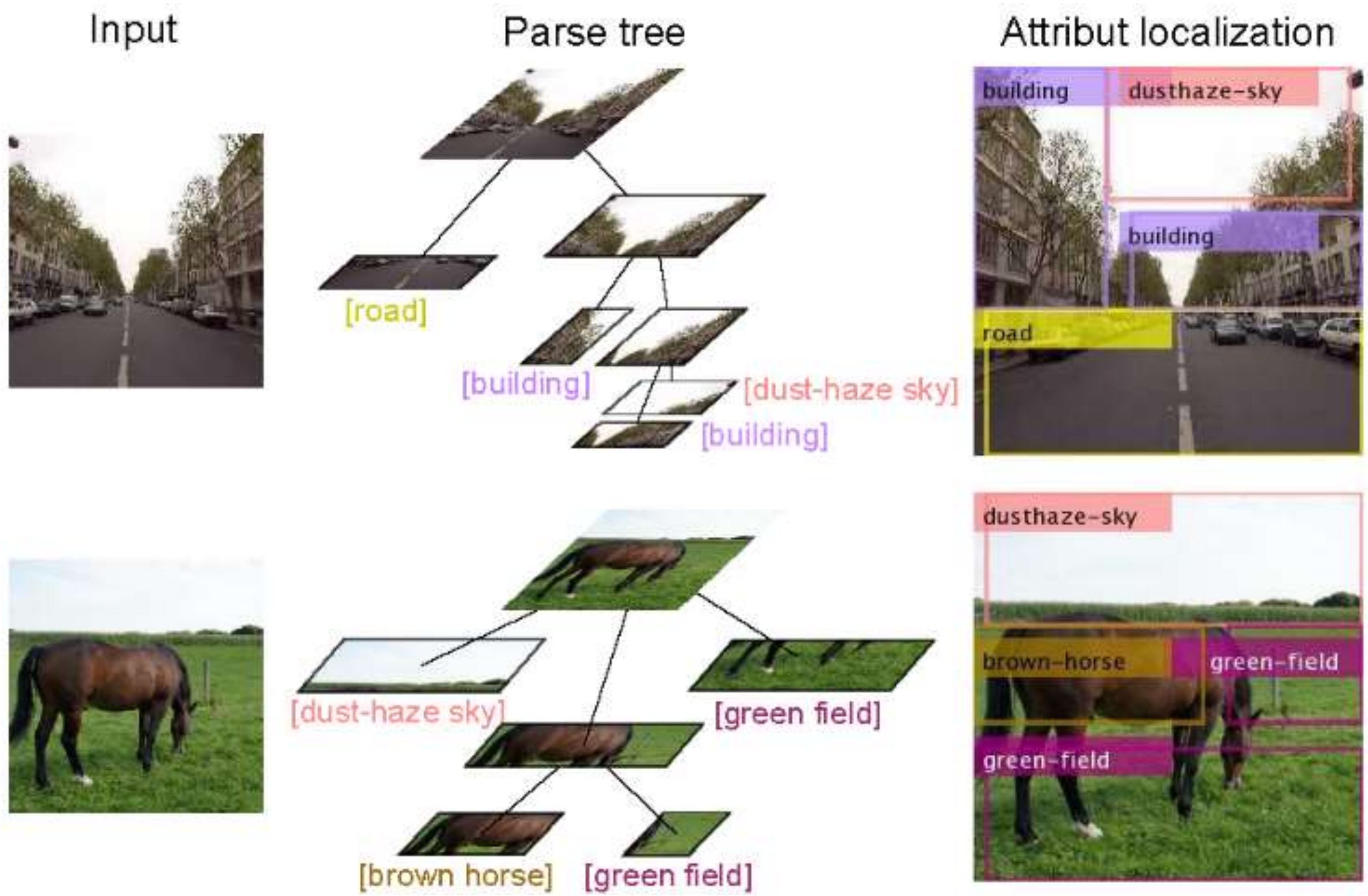


(b)

Parse Tree

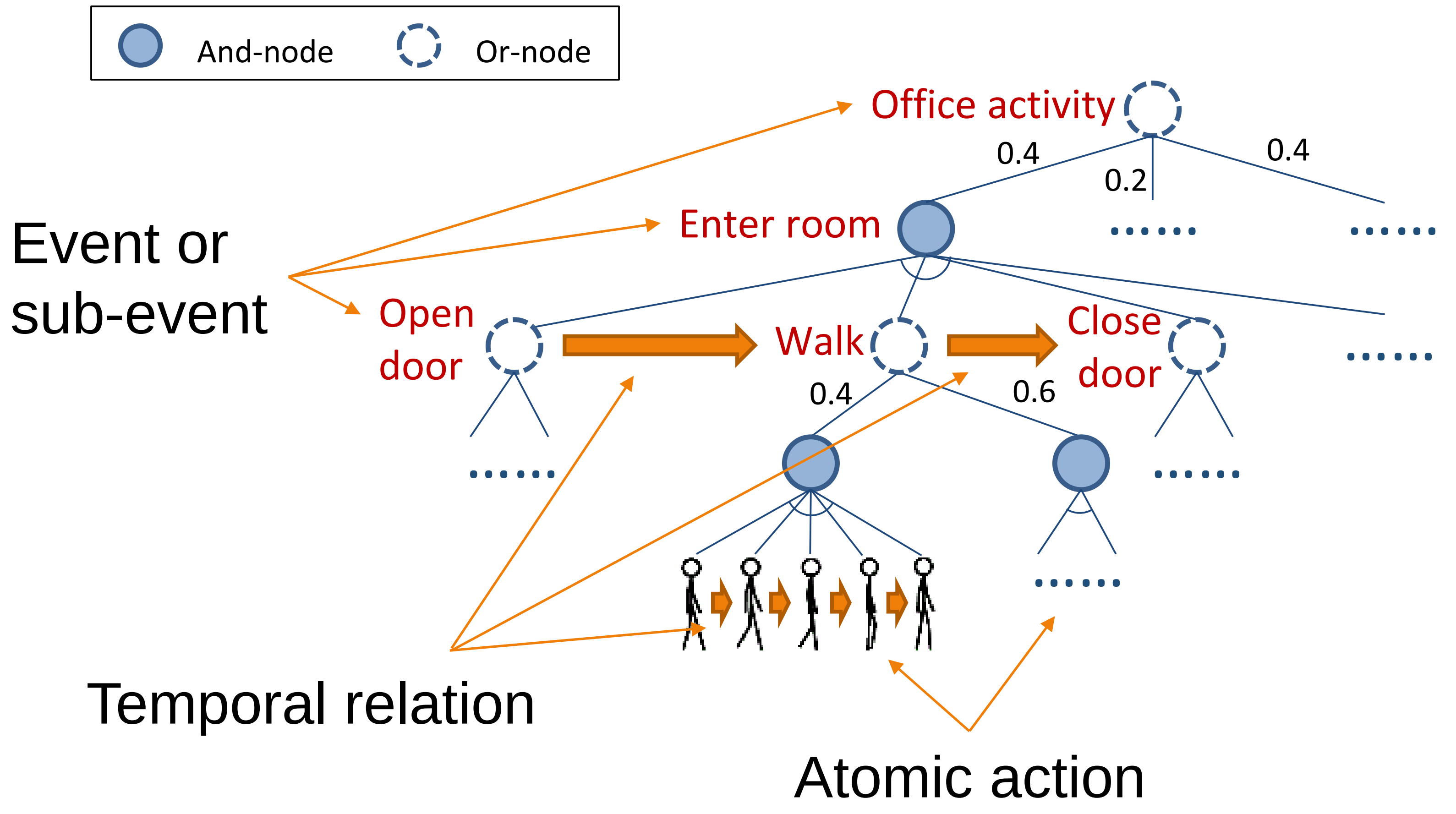
[Zhao & Zhu, NIPS 2011]

Applications – outdoor scene parsing

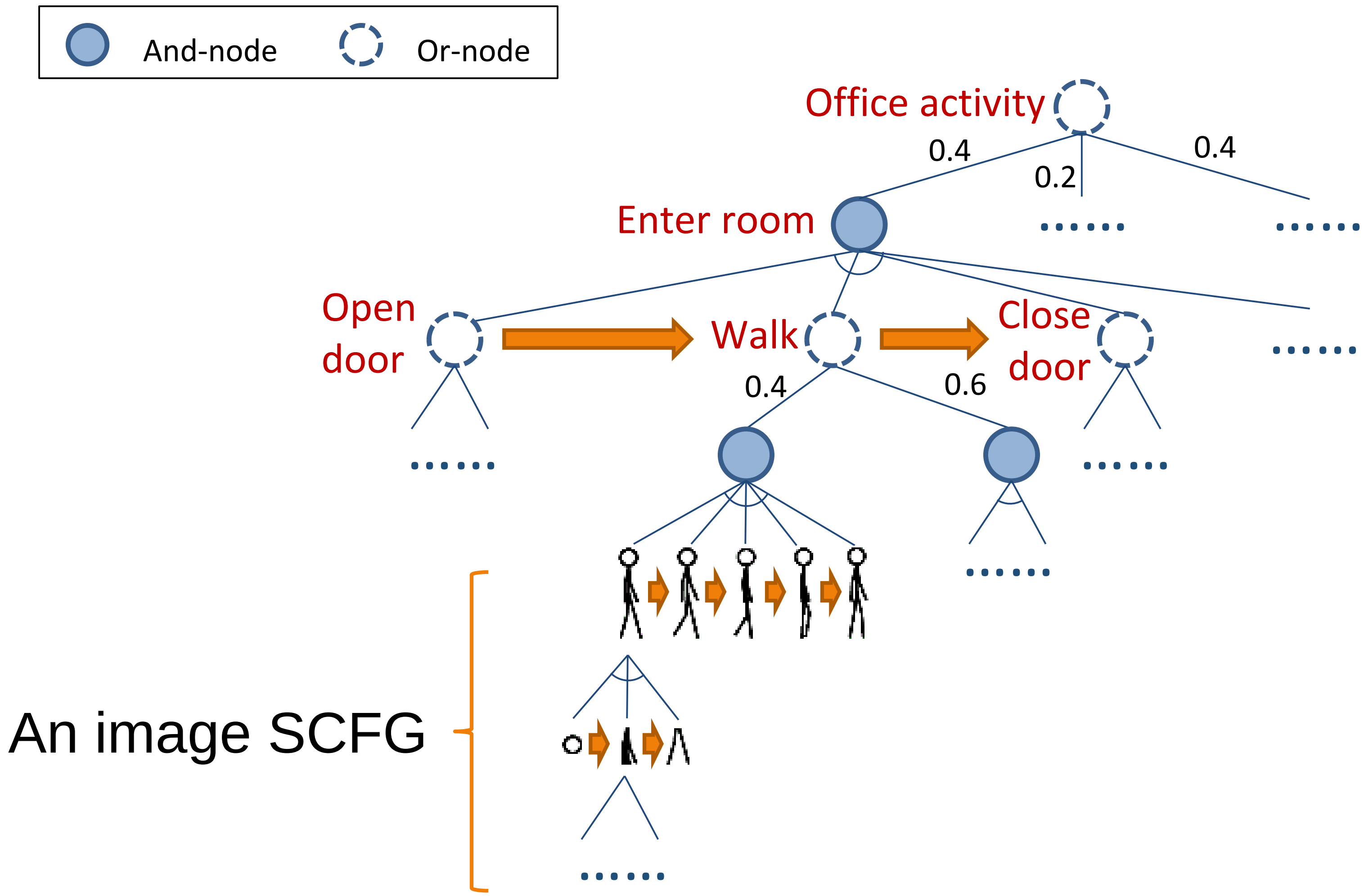


[Wang, Wang & Zhu, PAMI 2015]

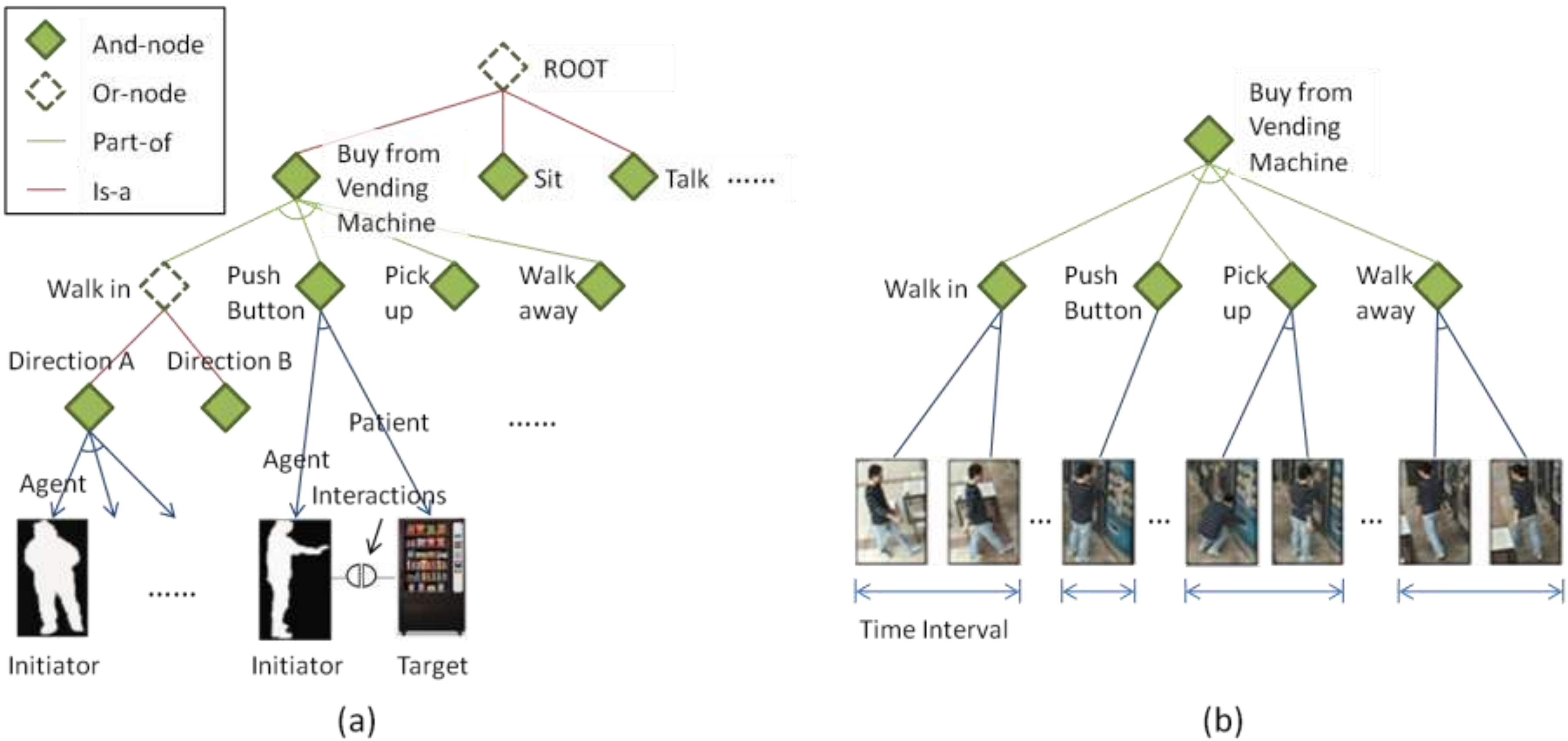
Generalized SCFG of Events



Generalized SCFG of Events in Videos



Applications – video event parsing



Grammar

Parse Tree

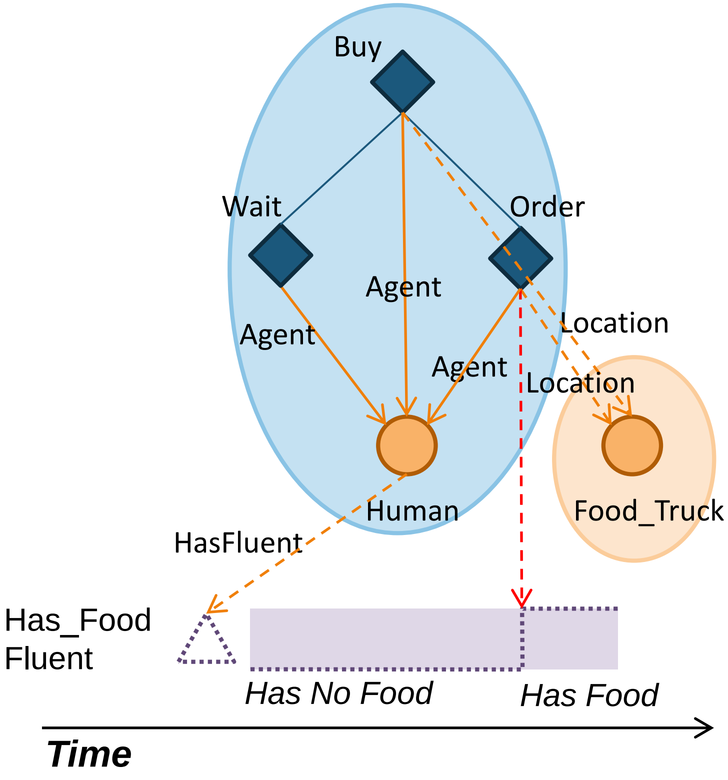
[Pei, Jia & Zhu, ICCV 2011]

Applications – joint event parsing from video&text

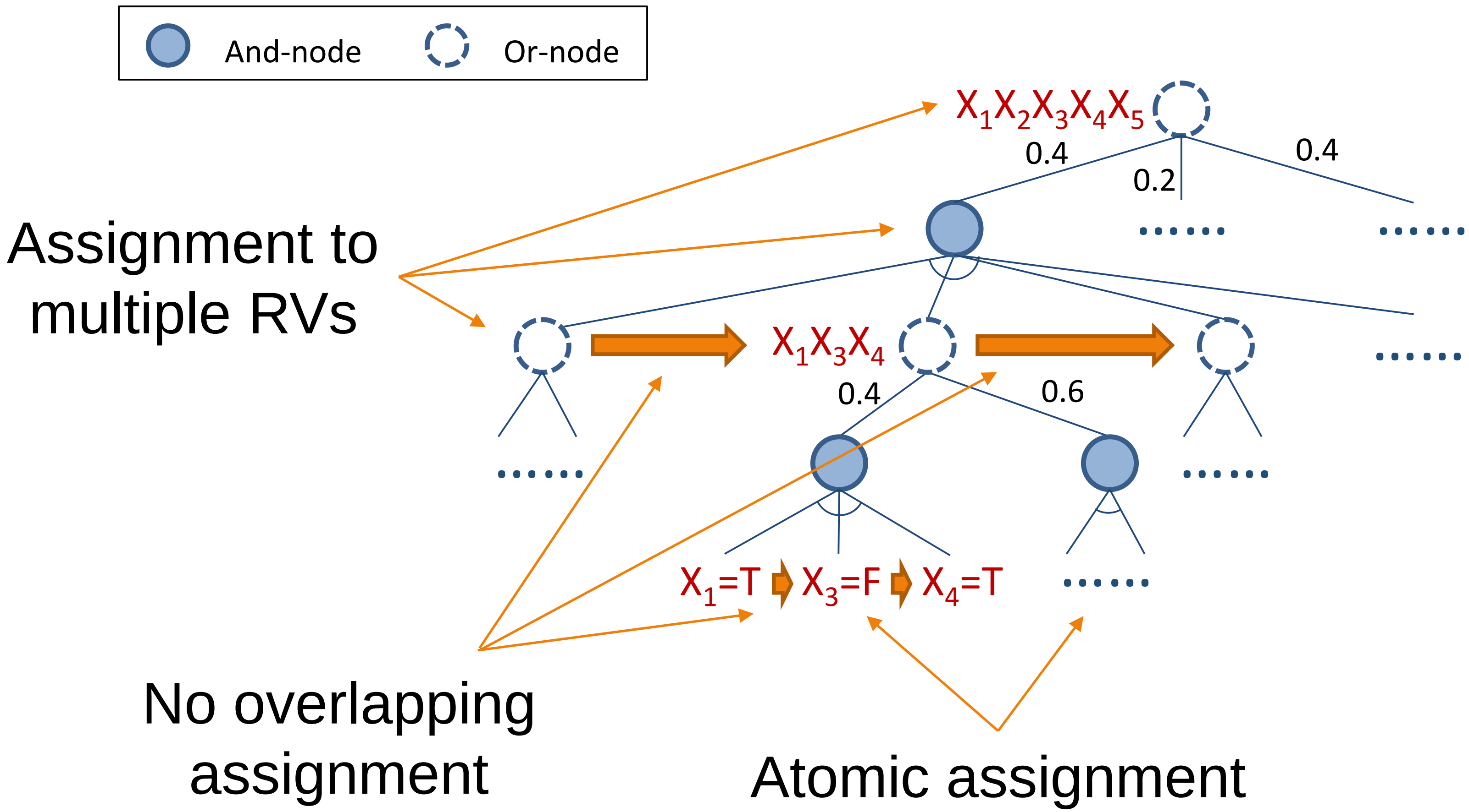
Input Video

Input Text

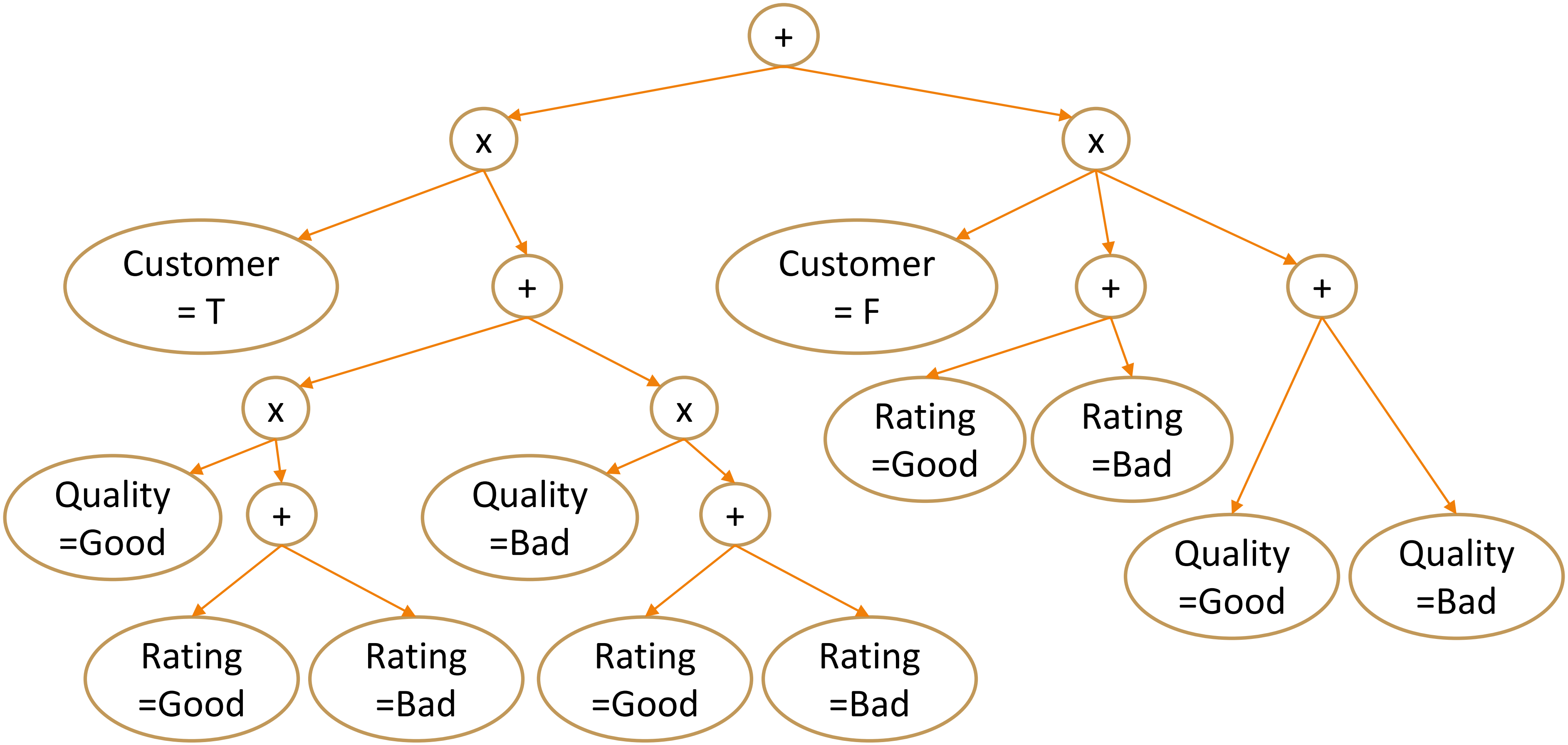
“There is a food truck.”



Generalized SCFG of Vector Data



Sum-Product Network



[Poon & Domingos, UAI 2011]

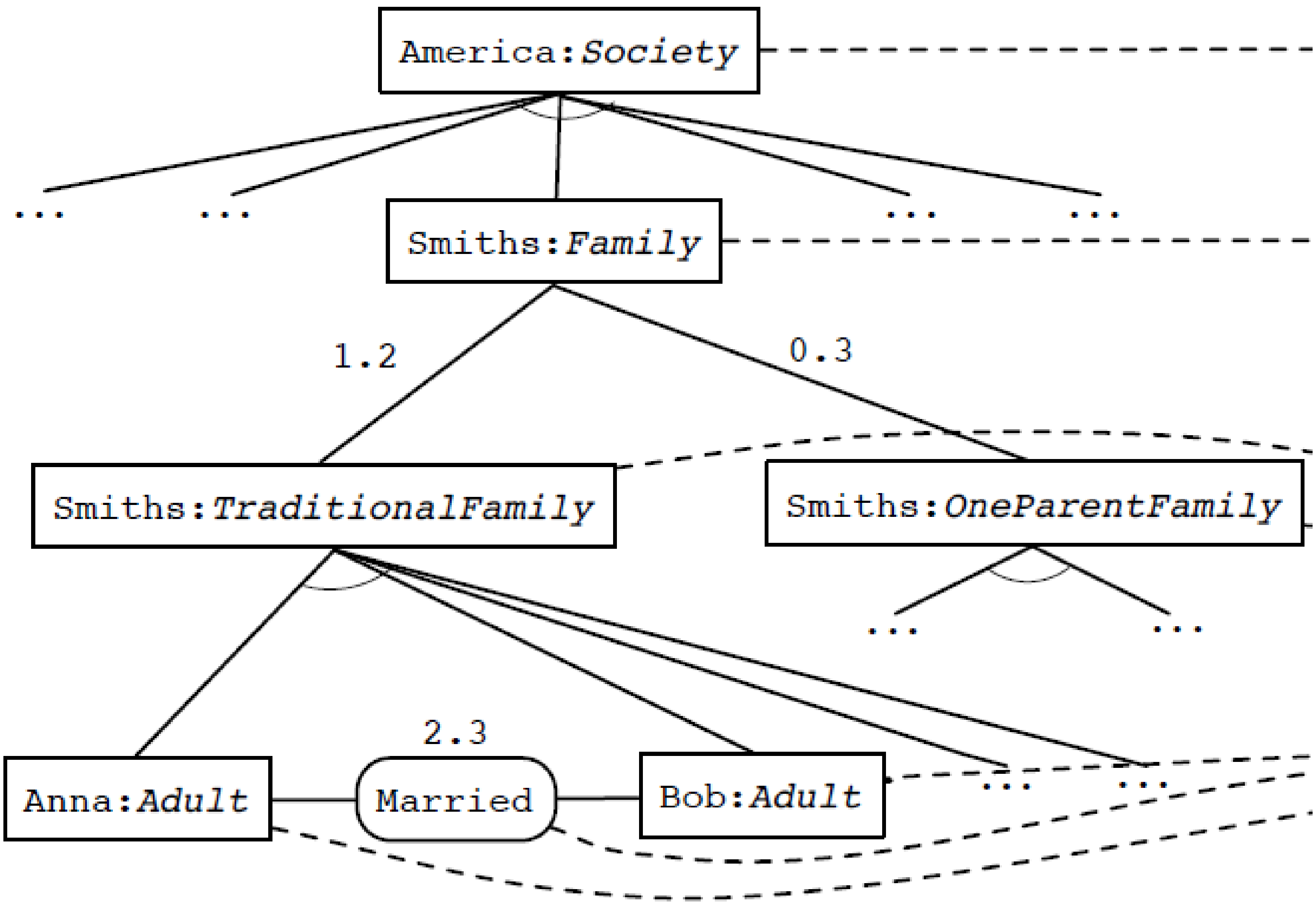
Stochastic And-Or Grammar (AOG)

- A general definition
  - A set Σ of atomic patterns
  - A set N of nonterminal patterns
    - Two disjoint subsets: And-nodes, Or-nodes
  - A start symbol $S \in N$ (a complete entity)
  - A set R of production rules
    - And-rule (composition): A → N1 N2 …
      - A set of parameterized relations between {A, N1, N2, …}
    - Or-rule (alternative configurations):
      O → N1 | N2 | t1 | … (with a probability for each config.)

Special cases of AOG

Natural Language Processing	Computer Vision	Machine Learning
• SCFG • Hidden Markov Models • Linear Context-free Rewriting System • Constraint-based Grammar Formalisms • etc.	• Pictorial Structures Models • Deformable Part Models • Flexible Mixture-of-Parts Models • etc.	• Sum-Product Networks • Naïve Bayes • Biclustering • Mixtures of Trees • Thin Junction Trees • Latent Tree Models • etc.

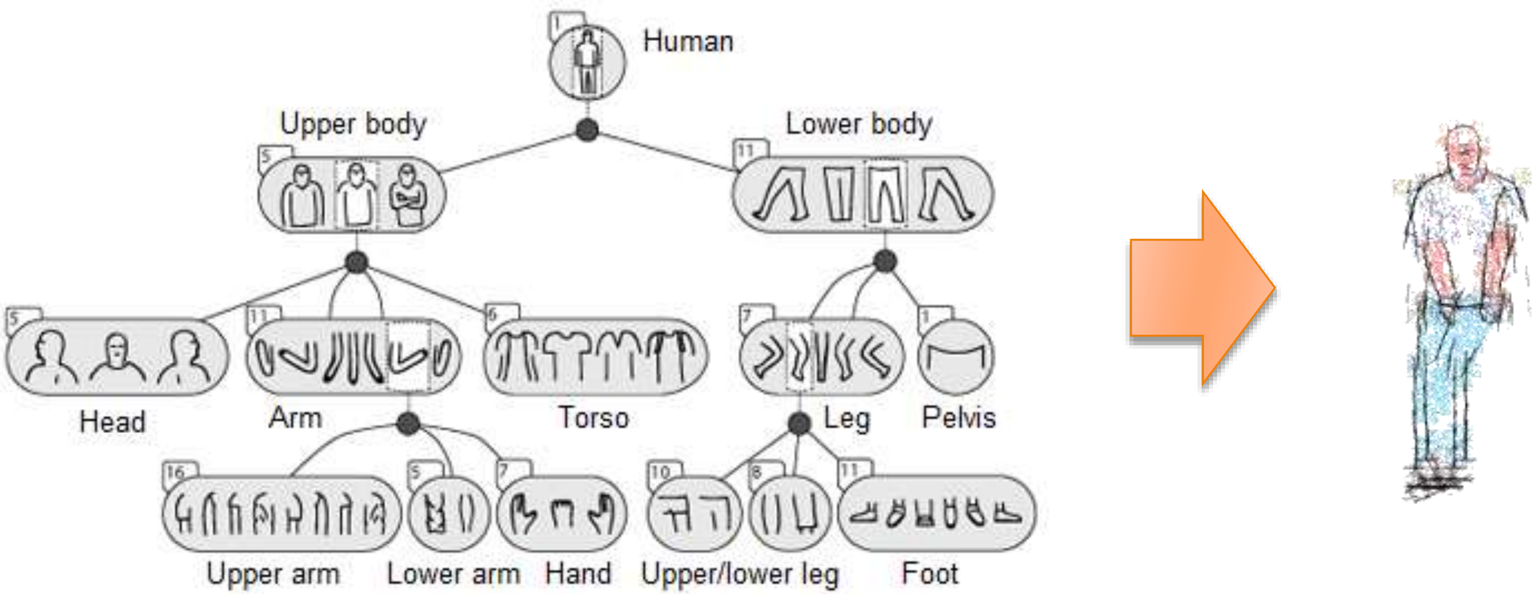
Related Work: Tractable Markov Logic



[Domingos & Webb, AAAI 2012]

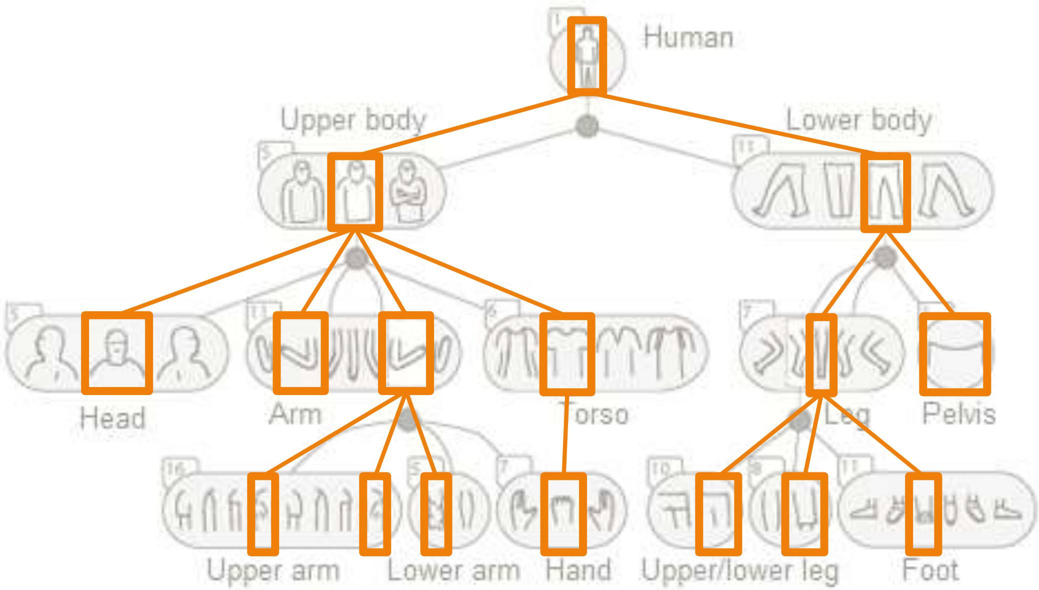
Why Stochastic AOG?

- Representational power
  - Simultaneous modeling of hierarchical compositions and alternatives
  - Representing a large number of patterns in a compact way



Why Stochastic AOG?

- Utility in applications
  - Efficient inference of hidden structures (parsing)
  - Helping solve multiple tasks in a unified way
    - classification, annotation, segmentation, feature generation, etc.



Learning Stochastic AOGs

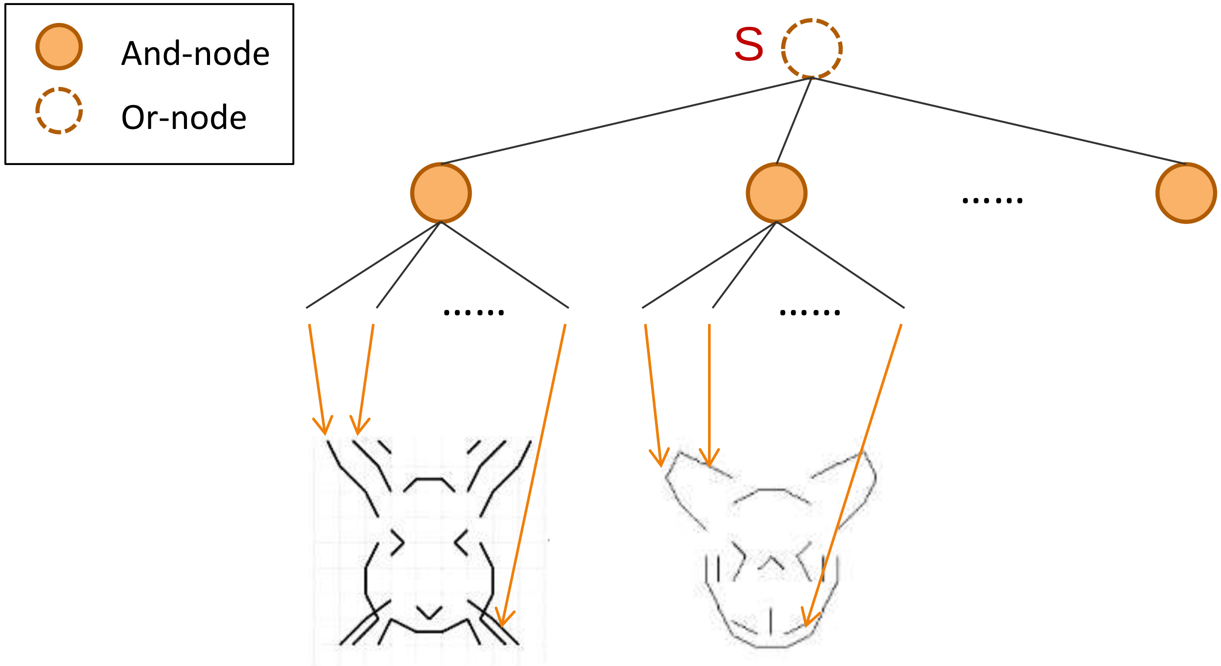
- Motivation
  - Very difficult and sometimes impossible to manually construct an accurate stochastic grammar
- Two types of learning:
  - Supervised learning
    - Learn from data annotated with parses
    - Problem: parse data is usually not available, manual annotation of parses is expensive and error-prone
  - Unsupervised learning
    - Learn from unannotated data

Structure Search

- Try to find an optimal set of grammar rules
- Our approach
  - Bottom-up iterative induction of grammar fragments

Algorithm Framework

- Start with the maximum-likelihood grammar
  - simply the union of all the training samples

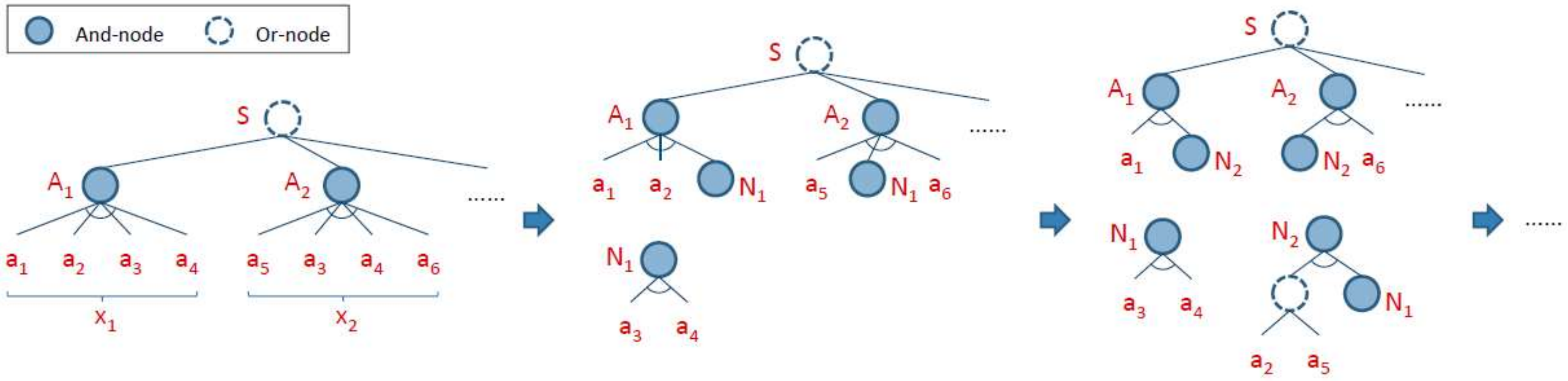


Training set X = { x1, x2, … … }

Algorithm Framework

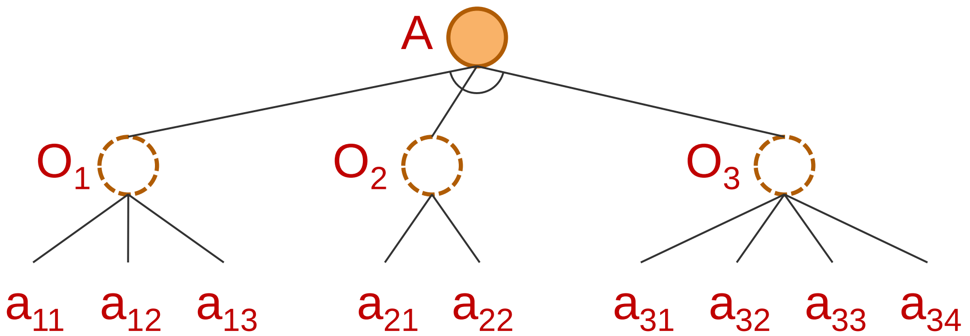
- Start with the maximum-likelihood grammar
  - simply the union of all the training samples
- Repeat:
  - Add a new grammar fragment s.t. the posterior is maximally increased

Until no more fragment can be learned



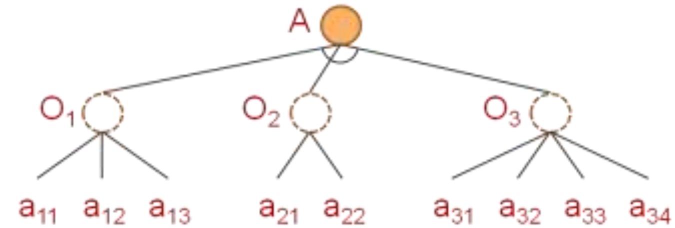
Grammar Fragment

- We propose to learn AND-OR fragments
  - A new AND rule
  - A new set of OR rules

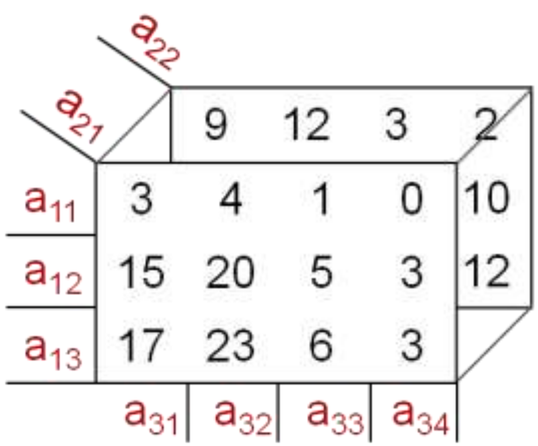


- More coherent, robust and efficient than alternatives

Posterior gain formulation



- Efficient computing based on a set of sufficient statistics
  - Likelihood gain
    - The product of the coherence of the n-gram tensor and the coherence of the context matrix
  - Prior gain
    - Determined by the size of the And-Or fragment and the sum of the n-gram tensor

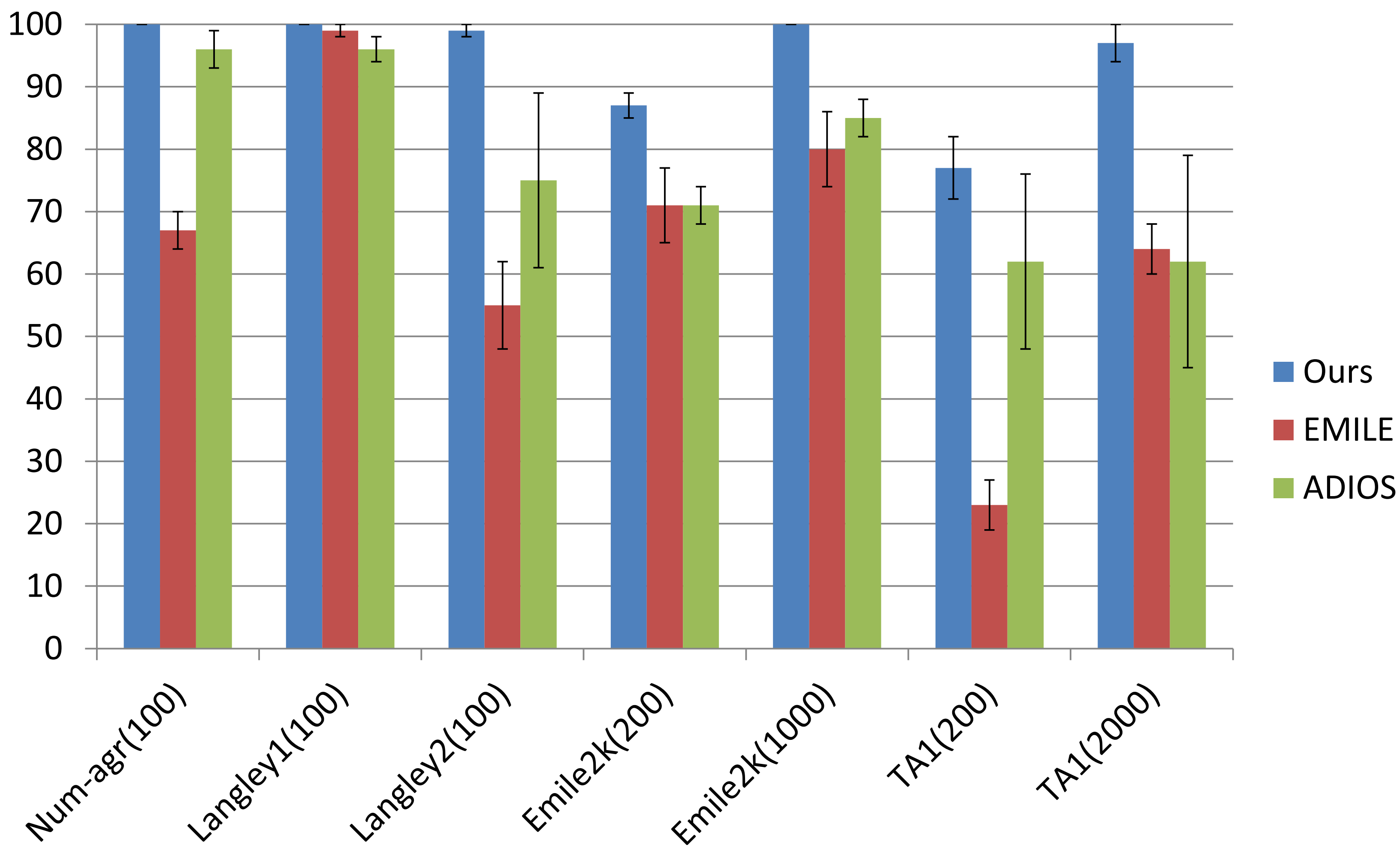


n-gram tensor

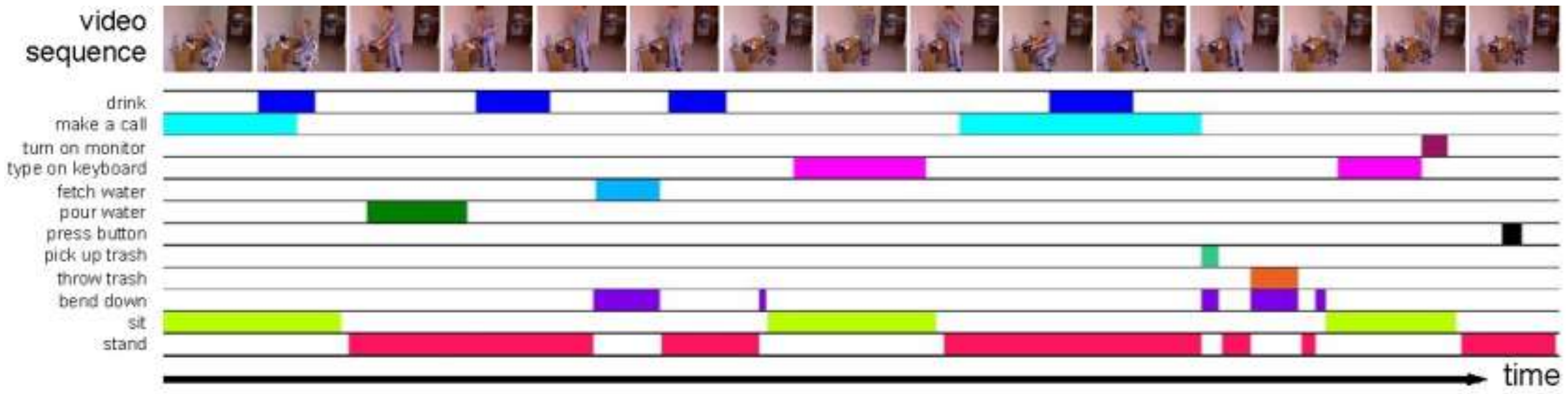
	$context_1$	$context_2$	$context_3$	… …
$a_{11}a_{21}a_{31}$	1	0	0	…
$a_{12}a_{21}a_{31}$	5	1	2	…
… …	…	…	…	…
$a_{13}a_{22}a_{34}$	4	1	1	…

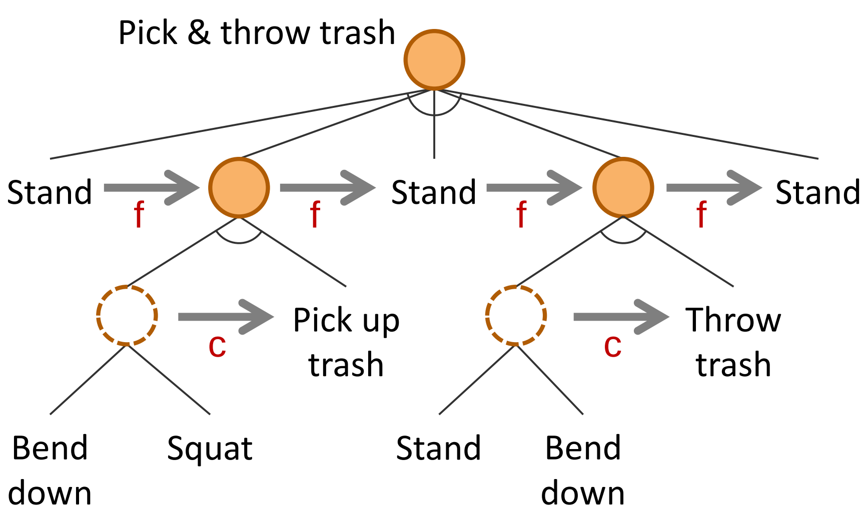
context matrix

Experiments on text data



Experiments on human activity data

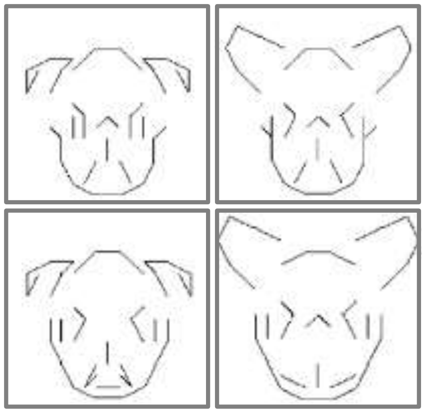




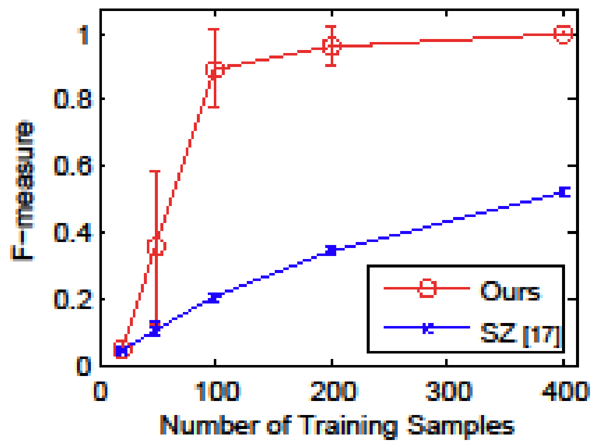
	Data 1	Data 2
ADIOS [15]	0.810	0.204
SPYZ [18]	0.756	0.582
Ours (f)	**0.831**	0.702
Ours (c+f)	0.768	0.624
Ours (cf)	0.767	**0.813**

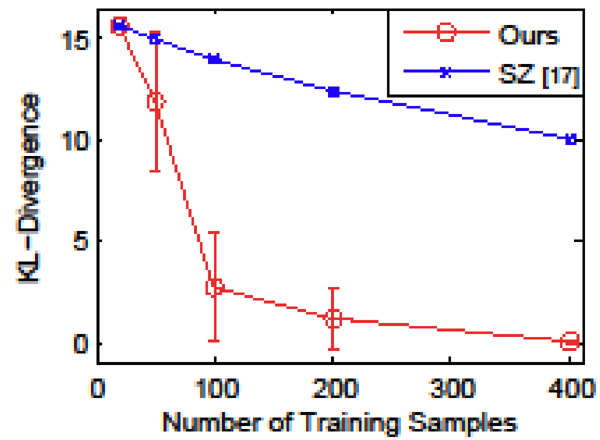
Experiments on images

Synthetic Data



Results



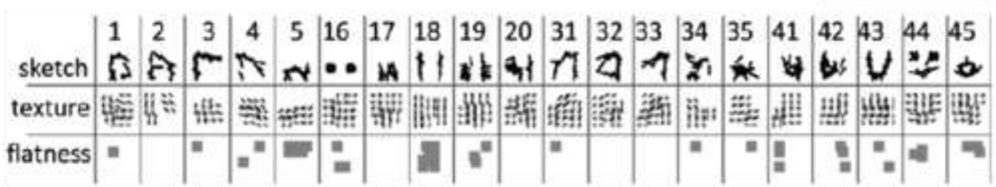


Real Data

Example images

Example quantized images

Atomic patterns (terminal nodes)



Results

	Perplexity
Ours	67.5
SZ [17]	129.4

Summary

- Stochastic And-Or Grammars
  - Compact representation via hierarchical compositions and alternatives
  - Applicable to various types of data
    - Images
    - Events
    - Vector data
    - Relational data in general
- Unsupervised learning of stochastic And-Or grammars
  - Structure search by iterative And-Or fragment induction

Papers

- Kewei Tu, "Stochastic And-Or Grammars: A Unified Framework and Logic Perspective". *Technical report*, 2015 (arXiv:1506.00858).
- Kewei Tu, Maria Pavlovskaia and Song-Chun Zhu, "Unsupervised Structure Learning of Stochastic And-Or Grammars". In *Advances in Neural Information Processing Systems 26* (NIPS 2013), Lake Tahoe, Nevada, USA, December 5-10, 2013.

Thank you!

Q&A