Stochastic And-Or Grammars: A Unified Framework and Logic Perspective

Kewei Tu

ShanghaiTech University

上海科技大学
ShanghaiTech University

Outline

- Stochastic And-Or Grammars
  - Extension of stochastic grammars of language, originally proposed to model images [Zhu & Mumford, 2006]
- This paper:
  - A unified representation framework agnostic to the type of data being modeled
    - It subsumes many existing models as special cases.
  - A domain-independent inference algorithm
    - Tractable under a reasonable assumption
  - Interpretations as a subset of probabilistic logic
    - Connection to statistical relational learning

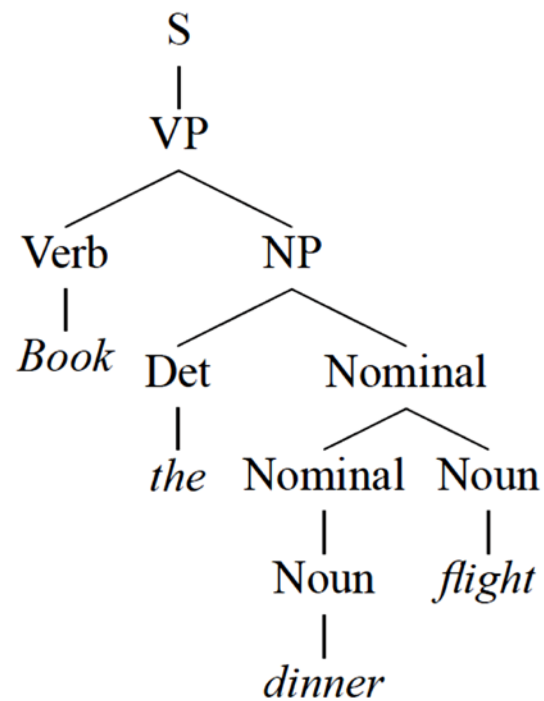
Grammars of language

- A context-free 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 nonterminal can produce a string of terminals and/or nonterminals
- Stochastic context-free grammars
  - Each production rule is associated with a probability

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"

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

[Jurafsky & Martin, 2006]

And-Or Normal Form of Stochastic CFG

- Two types of non-terminals
  - And-nodes, Or-nodes
- Two types of production rules
  - And-rule (composition): $A \rightarrow N_1 N_2 \cdots$
  - Or-rule (alternative configurations): $O \rightarrow N_1 | N_2 | t_1 | t_2 | \cdots$

SCFG

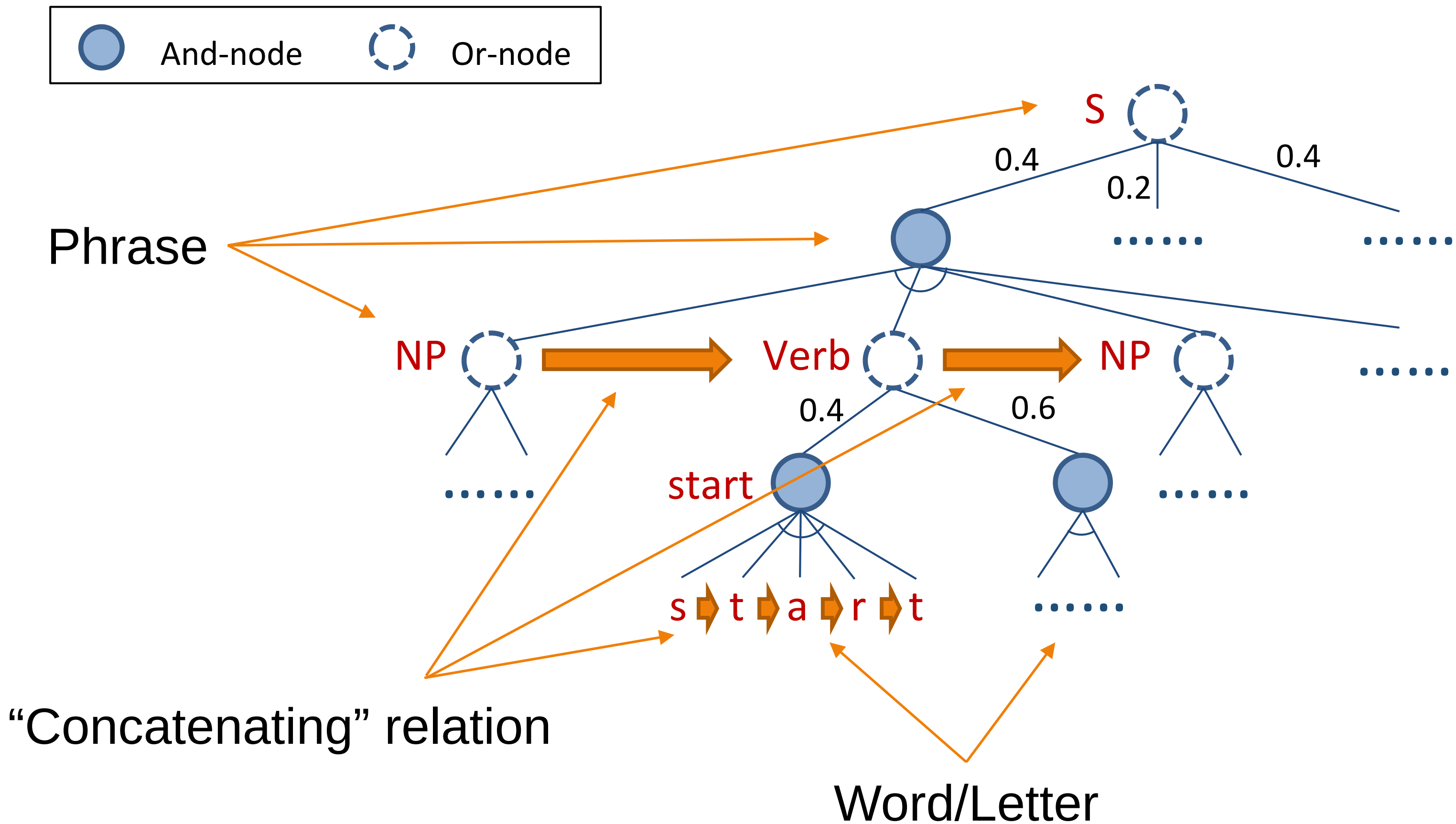
$S \longrightarrow a \ (0.4) \mid AB \ (0.6)$
$A \longrightarrow a \ (1.0)$
$B \longrightarrow b_1 \ (0.2) \mid b_2 \ (0.5) \mid b_3 \ (0.3)$

The AND-OR Form

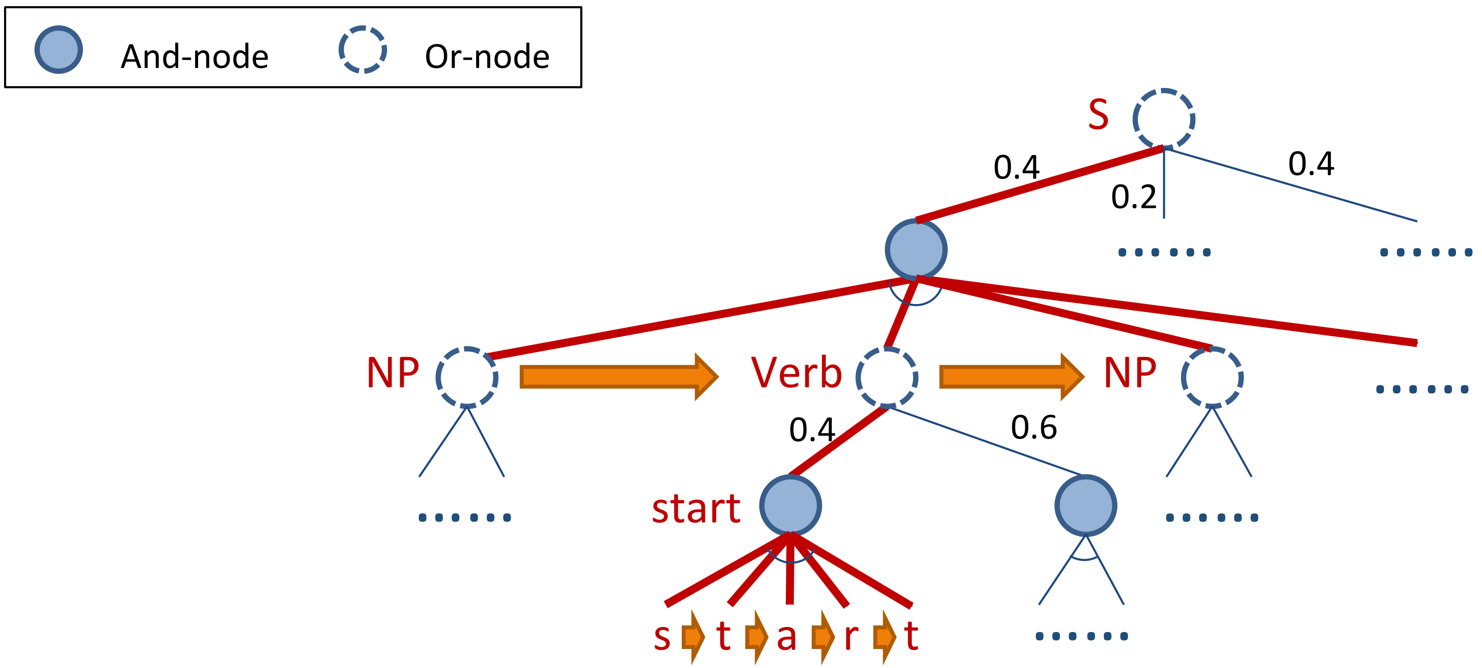
$\mathrm{OR}_S \longrightarrow a \ (0.4) \mid \mathrm{AND}_{AB} \ (0.6)$
$\mathrm{AND}_{AB} \longrightarrow \mathrm{OR}_A \mathrm{OR}_B$
$\mathrm{OR}_A \longrightarrow a \ (1.0)$
$\mathrm{OR}_B \longrightarrow b_1 \ (0.2) \mid b_2 \ (0.5) \mid b_3 \ (0.3)$

[Tu & Honavar, 2008]

And-Or Normal Form of Stochastic CFG

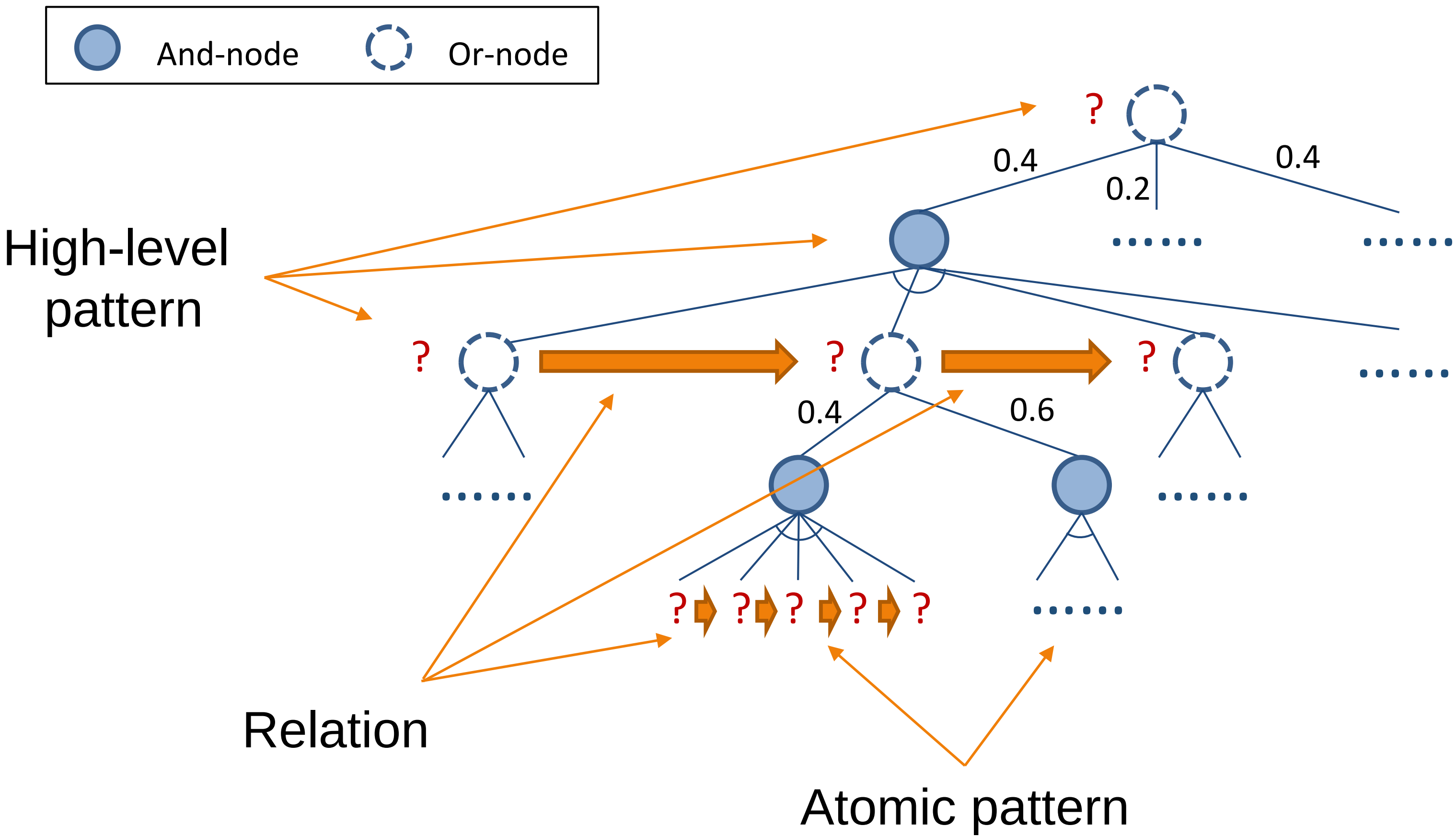


And-Or Normal Form of Stochastic CFG

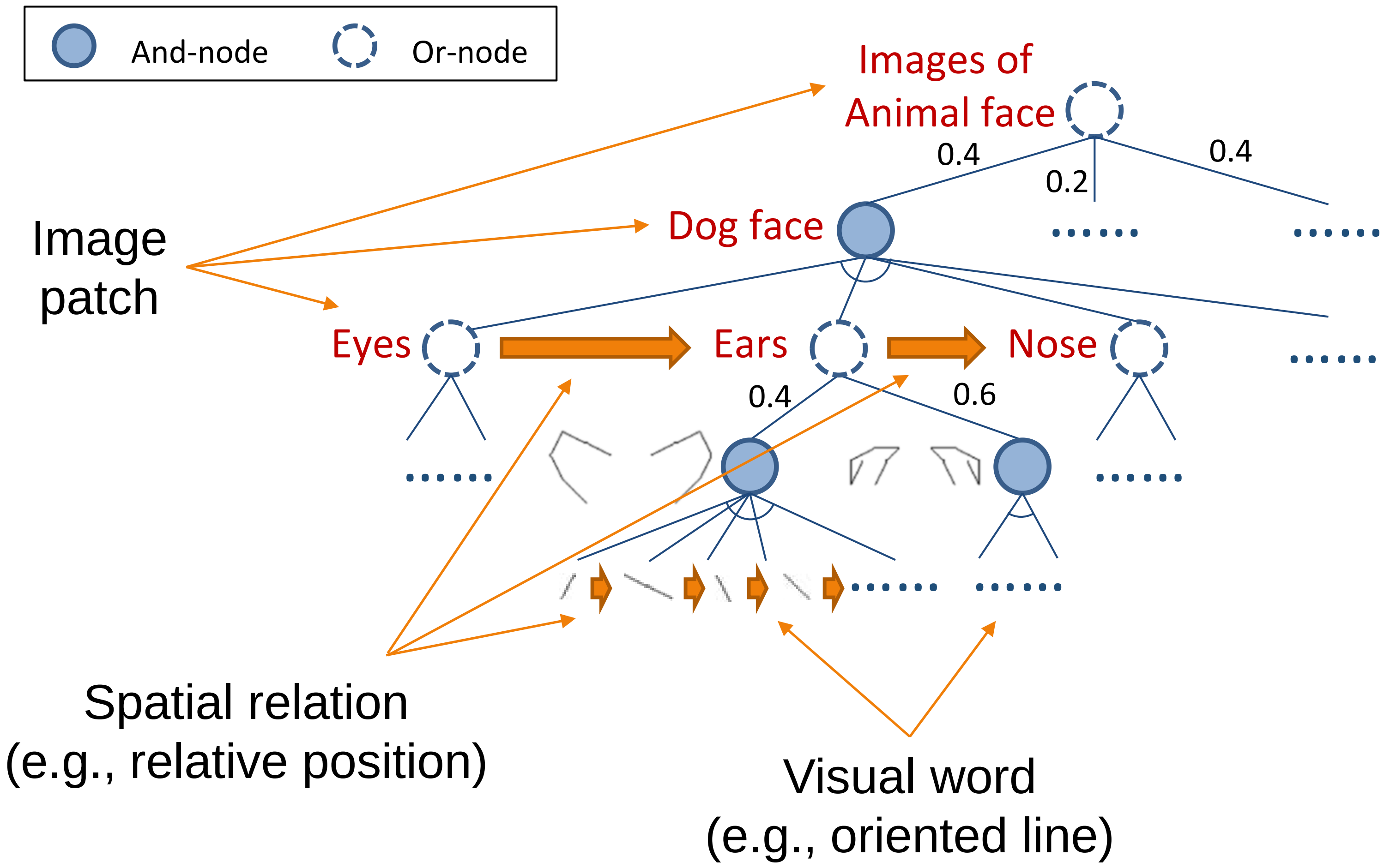


"They start the book today."

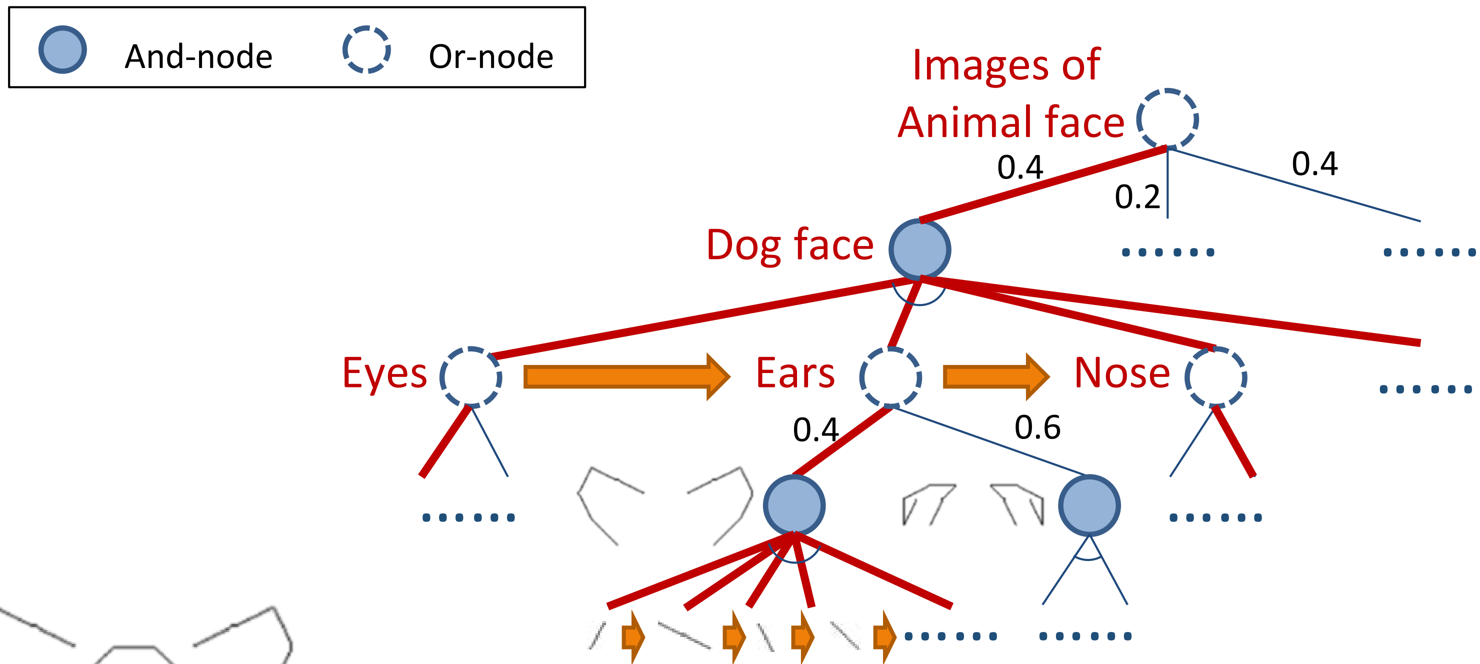
Stochastic Context-free And-Or Grammar (AOG)



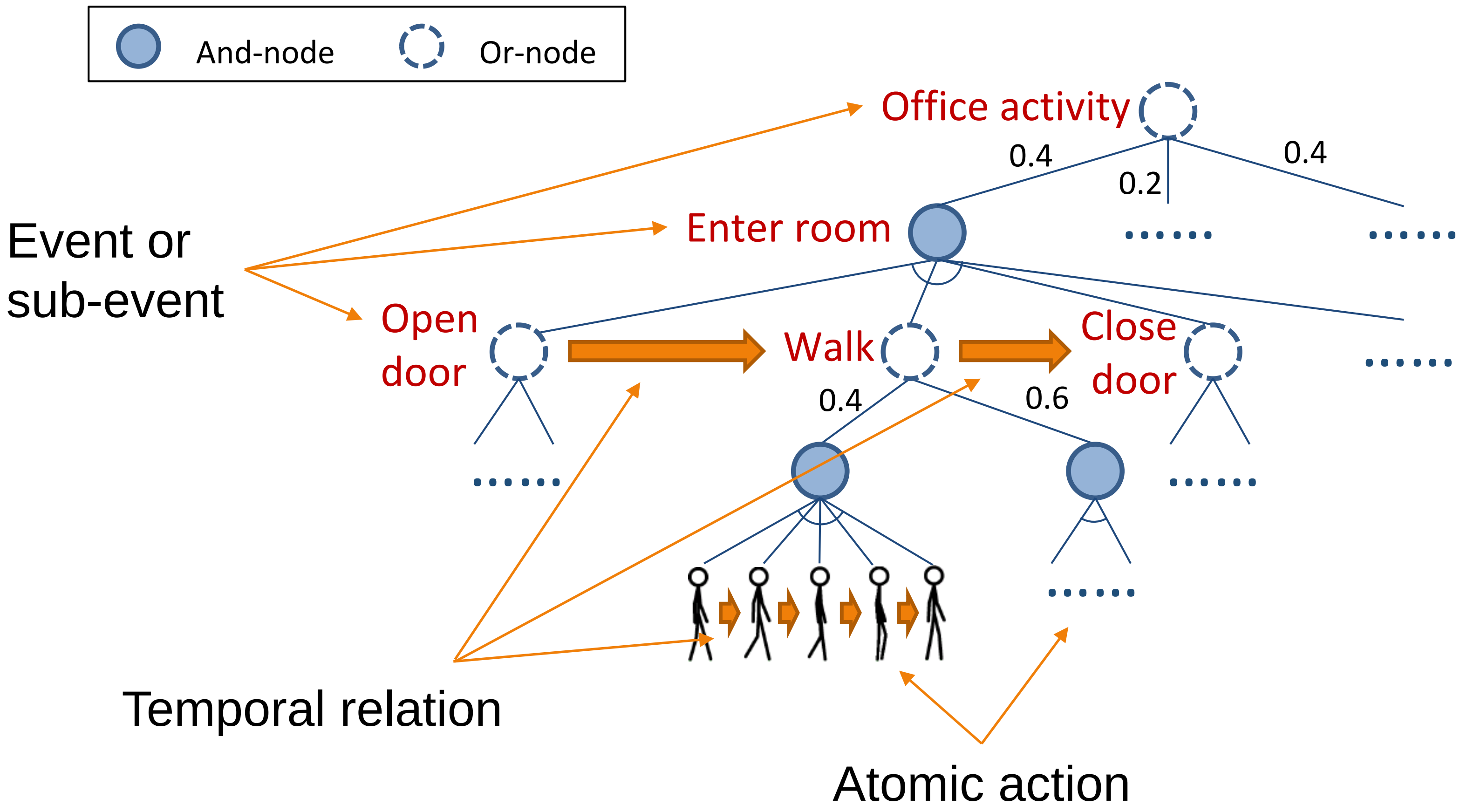
AOG of Images



AOG of Images

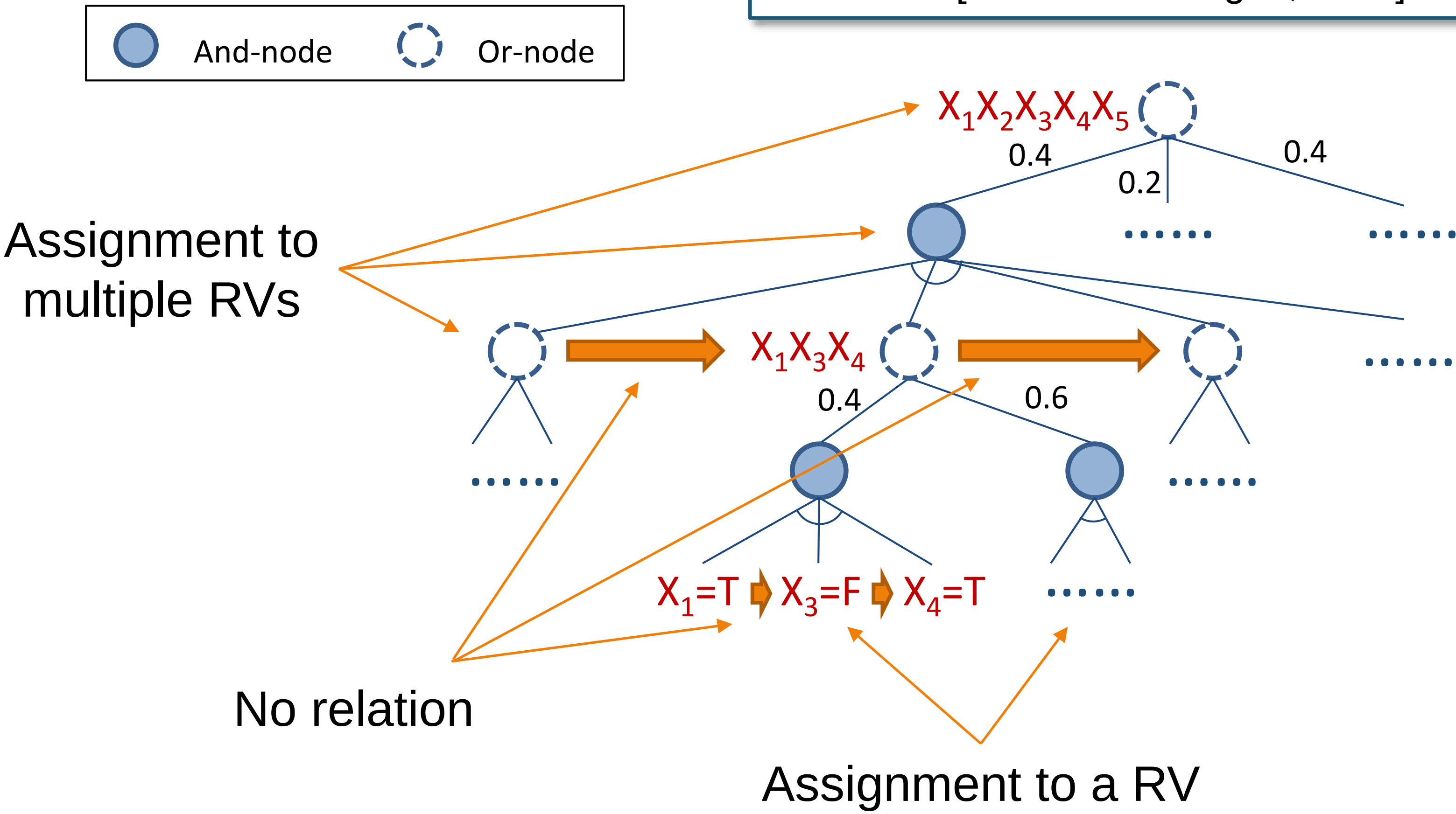


AOG of Events



AOG of Vector Data

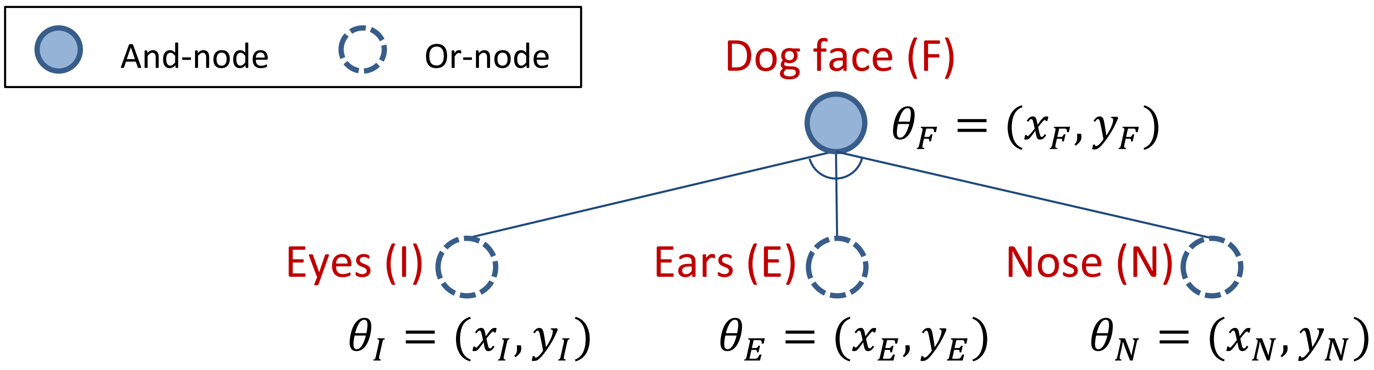
With some additional constraints, we get a decomposable sum-product network [Poon & Domingos, 2011]



Definition

- A stochastic context-free And-Or grammar is a 5-tuple
  - A set Σ of terminals (atomic patterns)
  - A set *N* of nonterminals (high-level patterns)
    - Two disjoint subsets: And-nodes, Or-nodes
  - A start symbol *S*∈*N* (a complete pattern)
  - A function θ that maps an instance of a terminal or nonterminal x to a parameter θ_x
  - A set *R* of production rules
    - Or-rule $\langle r, p \rangle$
      - r: $O \rightarrow x$
      - p is the probability of O producing x
    - And-rule $\langle r, t, f \rangle$
      - $r: A \rightarrow \{x_1, x_2, \dots, x_n\}$ for some $n \geq 2$
      - $t(\theta_{x_1}, \theta_{x_2}, \dots, \theta_{x_n})$ is a parameter relation between child nodes
      - $\theta_A = f(\theta_{x_1}, \theta_{x_2}, \dots, \theta_{x_n})$ is a parameter function

And-rule Example



- Production rule

$$r\colon F \to \{I, E, N\}$$

- Parameter relation

$$t(\theta_I, \theta_E, \theta_N) = T \ \text{ iff. } \ x_I = x_E = x_N \ \wedge \ 0.1 < y_E - y_I < 0.3 \ \wedge \ 0.1 < y_I - y_N < 0.2$$

- Parameter function

$$x_F = x_I, \qquad y_F = \frac{y_I + y_E + y_N}{3}$$

Special cases of AOG

Natural Language Processing

- Stochastic Context-free Grammars
- Hidden Markov Models
- Linear Context-free Rewriting System
- Constraint-based Grammar Formalisms
- etc.

Computer Vision

- And-Or Graphs
- Pictorial Structures Models
- Deformable Part Models
- Flexible Mixture-of-Parts Models
- etc.

Machine Learning

- Sum-Product Networks
- Naïve Bayes
- Biclustering
- Thin Junction Trees
- Mixtures of Trees
- Latent Tree Models
- etc.

Inference

- The main inference problems
  - Infer the most likely compositional structure of a data sample (parsing)
  - Compute the marginal probability of a data sample
  - *Both are NP-hard!*
- Composition Sparsity Assumption
  - *For any data sample X, the number of valid compositions in X is polynomial in |X|.*
  - Reasonable in many scenarios (text, images, etc.)
  - Leads to tractable inference
- Inference by bottom-up dynamic programming
  - An extension of the CYK algorithm for text parsing

Logic Interpretations

- Goal
  - To connect AOGs to statistical relation learning
  - To facilitate exchange of ideas between the two areas
- Two interpretations
  - As a subset of first-order probabilistic logic
    - grammar rules → material implications
    - terminals, nonterminals → unary relations
    - A possible-world semantics

    Resembles tractable Markov logic [Domingos & Webb, 2012]
  - As a stochastic logic program [Muggleton, 1996]
    - grammar rules → SLP clauses
    - start symbol → SLP goal

Summary

- Stochastic And-Or Grammars
  - Representation framework with hierarchical And-nodes (compositions) and Or-nodes (alternatives)
    - Applicable to various types of data: text, images, events, vector data, relational data in general
    - Subsumes many existing models as special cases
  - Tractable inference under the composition sparsity assumption
  - Two interpretations as a subset of probabilistic logic

Thank you!

Q&A