A Theoretical Foundation for Probabilistic Ontology Driven Stochastic Grammars for Syntactico-Semantic Pattern Recognition

By

Patnaikuni Shrinivasan R Patnaik

Under supervision of

Dr. S.R. Gengaje

Walchand Institute of Technology, Solapur

(Research Centre, PAH Solapur University, Solapur, MH, India)

Date: 9th May 2022

Publications

1) Patnaikuni, S., & Gengaje, D. S. (2021). A Theoretical Foundation for Syntactico Semantic Pattern Recognition. IEEE Access, 1–1.
[SCI Indexed Q1 Journal] https://ieeexplore.ieee.org/document/9547263

2) Patnaikuni, S., & Gengaje, S. R. (2021). Probabilistic, syntactic, and semantic reasoning using MEBN, OWL, and PCFG in healthcare. In Web Semantics (pp. 87–94). **Elsevier. [Scopus Indexed]** https://www.sciencedirect.com/science/article/pii/B9780128224687000092

3) Patnaikuni, S., & Gengaje, S. (2020). Properness and Consistency of Syntactico Semantic Reasoning using PCFG and MEBN. In Proceedings of the 2020 IEEE International Conference on Communication and Signal Processing, ICCSP 2020 (pp. 554–557). Institute of Electrical and Electronics Engineers Inc. **[Scopus Indexed]** https://ieeexplore.ieee.org/document/9182050

4) Patnaikuni, S., & Gengaje, S. (2019). Syntactico-Semantic Reasoning using PCFG, MEBN PP Attachment Ambiguity. In IEEE International Conference on Issues and Challenges in Intelligent Computing Techniques, ICICT 2019. Institute of Electrical and Electronics Engineers Inc. **[Scopus Indexed]** https://ieeexplore.ieee.org/document/8977712

5) Patnaikuni, S. R. P., & Gengaje, S. R. (2017). Survey of Multi Entity Bayesian Networks (MEBN) and its applications in probabilistic reasoning. International Journal of Advanced Research in Computer Science, 8(5), 2425–2429. Retrieved from www.ijarcs.info [UGC Approved Journal List]
http://www.ijarcs.info/index.php/Ijarcs/article/view/3954

Presentation Outline

- Introduction of PCFG and MEBN
- Background and Research Gap
- Problem Statement, Objectives, and Scope
- Theory of Syntactico-Semantic Reasoning
- Discussions on Experiment and Results
- Conclusion
- Future Scope

Introduction (PCFG)

Probabilistic Context Free Grammar is a quintuple,

$G_{PCFG} = (M_{PCFG}, T_{PCFG}, R_{PCFG}, S_{PCFG}, P_{PCFG})$, where

$M_{PCFG} = \{N^i : i = 1, \ldots, n\}$ is a set of nonterminals

$T_{PCFG} = \{w^k : k = 1, \ldots, v\}$ is a set of terminals

$R_{PCFG} = \{N^i \rightarrow \zeta^j : \zeta^j \in (M_{PCFG} \cup T_{PCFG})^*\}$ is a set of rules

$S_{PCFG} = N^1$ is the start symbol

P_{PCFG} is a corresponding set of probabilities on rules such that.

$$\forall i \sum_j P(N^i \rightarrow \zeta^j) = 1$$

For a PCFG in chomsky normal form (CNF)

$$R_{PCFG} = \{N^i \rightarrow N^r N^s, N^i \rightarrow w^k\}$$

$$\forall i \sum_{r,s} P(N^i \rightarrow N^r N^s) + \sum_k P(N^i \rightarrow w^k) = 1$$

Introduction (PCFG)

An example of probabilistic context free grammar in CNF is represented as,

$N^1 \rightarrow N^2 N^3$ 1.0	$N^2 \rightarrow w^2$ 0.1
$N^2 \rightarrow N^2 N^4$ 0.4	$N^2 \rightarrow w^3$ 0.04
$N^4 \rightarrow N^5 N^2$ 1.0	$N^2 \rightarrow w^4$ 0.18
$N^3 \rightarrow N^6 N^2$ 0.7	$N^2 \rightarrow w^5$ 0.1
$N^3 \rightarrow N^3 N^4$ 0.3	$N^6 \rightarrow w^6$ 1.0
$N^2 \rightarrow w^1$ 0.18	$N^5 \rightarrow w^7$ 1.0

The number beside the rule indicates the probability associated with the rule.

Introduction (MEBN)

A MEBN (Multi Entity Bayesian Network) theory T_{MEBN} is a set of MFrags $\{F_1, F_2, F_3, \ldots., F_n\}$.

An MFrag F_i is a quintuple $F_i = (C^i_{MEBN}, I^i_{MEBN}, R^i_{MEBN}, G^i_{MEBN}, D^i_{MEBN})$ where

C^i_{MEBN} is a finite set of values a context can take form as a value; context serves as constraints under which the variables in MFrag are instantiated.

I^i_{MEBN} is a set of input random variables.

R^i_{MEBN} is a finite set of resident random variables, the term "resident random variable" indicates a random variable symbol with a parenthesized list of arguments separated by commas.

G^i_{MEBN} is a directed acyclic graph representing the dependency between input random variables and resident random variables conditional on context random variables

D^i_{MEBN} is a set of local conditional probability distributions where each member of R^i_{MEBN} has its own conditional probability distribution in set D^i_{MEBN}.

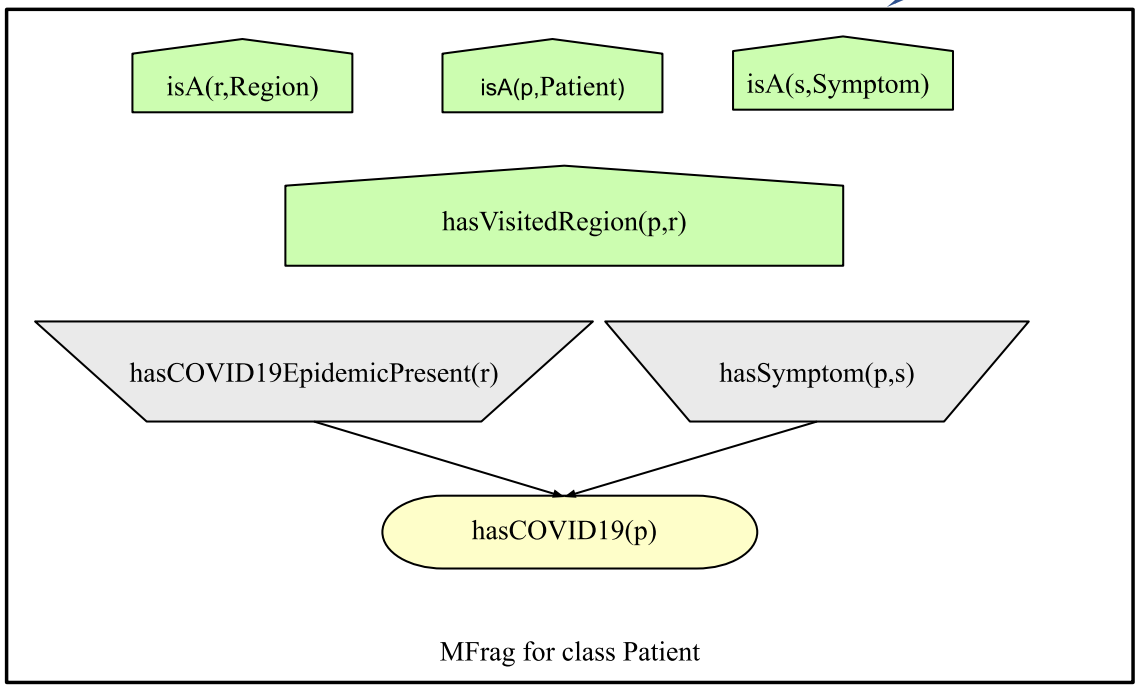
Sets C^i_{MEBN}, I^i_{MEBN}, and R^i_{MEBN} are pairwise disjoint. For a set of MFrags $\{F_1,F_2,F_3,....F_n\}$, there exists a joint unique probability distribution on the set of instances of the random variables of its MFrags that is consistent with the local probability distributions assigned within a MFrag.

Introduction (MEBN)

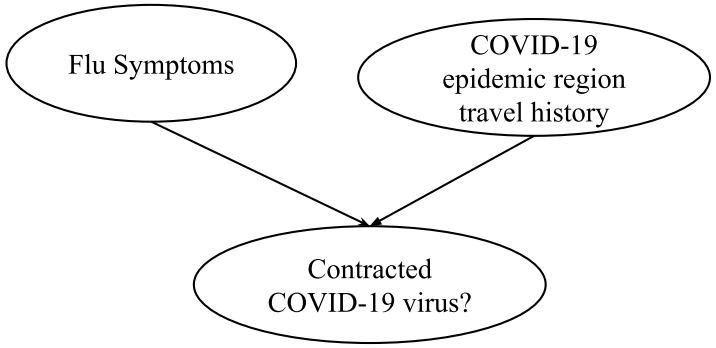
- Classes:
  - Patient
  - Region
  - Diagnosis
- Properties:
  - hasCOVID19 (Patient p)
    - Domain: Patient
    - Range: Diagnosis
  - hasCOVID19EpidemicPresent (Region r) | Range data type: Boolean
    - Domain: Region
    - Range: Boolean
  - hasVisitedRegion (Patient p, Region r) | Range data type: Boolean
    - Domain: Patient X Region
    - Range: Boolean

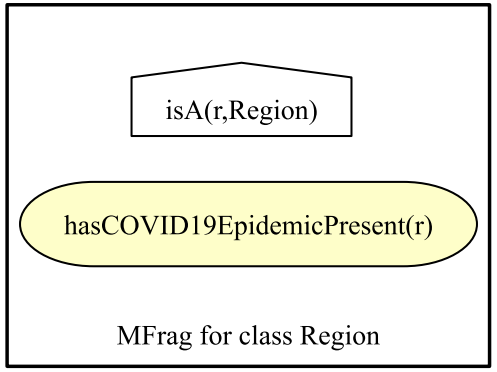
Ontology

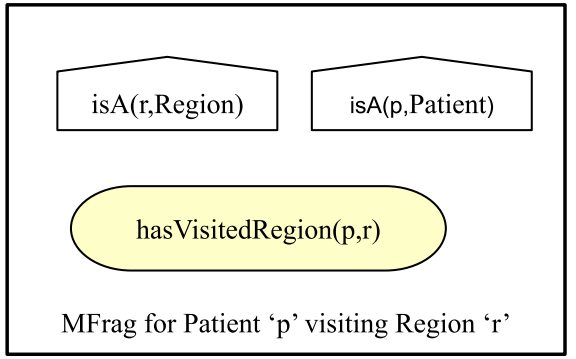
MEBN

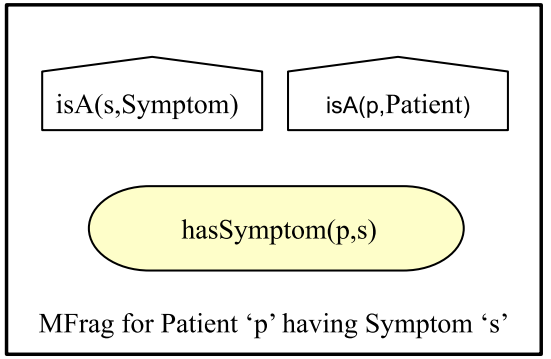


Bayesian PGM









Literature Review

The literature survey (1970- 2021) of the research primarily focuses on the following aspects.

I. Extending the context free grammars by augmenting with semantic information.

II. Semantic compositionality and works on the applicability of the principle.

III. Probabilistic ontologies, especially using multi entity Bayesian network as a reasoning system based on probabilistic ontology.

IV. Representing uncertainty, the domain of Semantic web.

Research Gap

1. Lack of a well-defined formal method to integrate stochastic grammar with ontology based probabilistic reasoning.
2. Research gap in integration methods for PCFG with probabilistic semantic reasoning system like MEBN.

This research is an attempt to bridge the research gap identified and pave the way for probabilistic reasoning in syntactic processes with semantic uncertainty reasoning, also called as syntactico-semantic reasoning.

Problem Statement

To develop a theoretical foundation for probabilistic ontology driven stochastic grammars for syntactico-semantic pattern recognition.

Objectives

1. To define a formal method for using semantic information in the form of probabilistic ontologies, in the process of syntactic pattern recognition.

2. To propose a theoretical foundation for using probabilistic ontologies in learning the rule probabilities of stochastic grammars in the context of syntactic pattern recognition.

Scope

1. The research will be mainly be focusing on laying out a theoretical foundation for using semantic information in form of ontologies to drive the probabilistic reasoning of the stochastic grammars

2. The research aims to make use of stochastic grammars in the area of syntactic pattern recognition.

3. The validation of proposed work will be done by developing an application framework which subsumes the processes of extracting semantic information or ontology construction methods as preprocessing part of the input dataset.

The theoretical foundation for syntactico-semantic reasoning using PCFG and MEBN

Mapping (formal specification)

To an integration method to exist between PCFG and MEBN, there is a need for mapping between the entities of PCFG and MEBN. This mapping is a two step process as outlined.

Step 1: Mapping of Non-terminals and Terminals in PCFG to the sets of Context, Input, and Random variables in the MEBN theory.

Step 2: Mapping between probability distributions of PCFG and MEBN.

Mapping Step 1

Every non-terminal shall have a corresponding input variable of an MFrag, which implies.

$$M_{PCFG} \subset \bigcup I^{i}_{MEBN}$$

Every derivation of grammar rule shall be part of an infinite set ε of entity identifier symbols across all MFrags of the MEBN theory, which implies.

$$\forall (N^{i} \rightarrow N^{r}N^{s} \in R_{PCFG} \text{ and } N^{i} \rightarrow w^{k} \in R_{PCFG})\ N^{r}N^{s} \in \varepsilon \text{ and } w^{k} \in \varepsilon$$

There shall be a unique resident random variable $hasProbability(\theta_{Ni}, \theta_{NrNs})$ *for each MFrag* F_i where θ_i is an ordinary variable belonging to the set of constant and identifier symbols of MEBN theory $T_{MEBN,}$ implying,

$$\forall\ (F_i \in T_{MEBN})\ hasProbability(\theta_{Ni}, \theta_{NrNs}) \in R^{i}_{MEBN}$$

Mapping Step 2

- The mapping between probability distributions of PCFG and MEBN is achieved by combining probability distributions of PCFG and MEBN theory using a method of combining probabilities called as conflation of probabilities, represented using symbol &().
- The special property of conflation of probability is its least Shannon's information loss in combining probability values.
- The reason to use conflation of probability is that the conflation method gives more priority to distribution based on smaller standard deviation and avoids the issues which arise if a simple average of probability would have been considered instead.
- Conflation of probabilities is commutative, associative, and iterative and holds for the Lemma 1 and Lemma 2 proposed under this research.

Lemma 1: Equality of Sum of Conflated Probabilities

Let, P = $\{p_1, p_2, p_3, \ldots. p_n\}$ and P' = $\{p'_1, p'_2, p'_3, \ldots. p'_n\}$ be a set of probability values such that, |P| = |P'| and for $i \in \{1, 2, 3, \ldots, n\}$ $\sum_i p_i = 1$ and $\sum_i p'_i = 1$. Also, there exists a one-to-one injective mapping M between P and P'.

For any $p_i \in$ P, $p_j \in$ P and M(p_i) = $p'_i \in$ P' , M(p_j) = $p'_j \in$ P' **if $p_i + p_j = S$** and given that

$$p_{i\ conflation} = (p_i \times p'_i) \div ((p_i \times p'_i) + ((1 - p_i) \times (1 - p'_i)))$$ and

$$p_{j\ conflation} = (p_j \times p'_j) \div ((p_j \times p'_j) + ((1 - p_j) \times (1 - p'_j)))$$

then for,

$$p_{ii} = (p_{i\ conflation} \div (p_{i\ conflation} + p_{j\ conflation})) \times (p_i + p_j)$$ and

$$p_{jj} = (p_{j\ conflation} \div (p_{i\ conflation} + p_{j\ conflation})) \times (p_i + p_j),$$

The equality $p_{ii} + p_{jj} = S$ holds True.

Lemma 2: Equality of conflated probabilities in product of probabilities

From Lemma 1 given that $p_i + p_j = S$, $p_{ii} + p_{jj} = S$ and for any $p_l \in$ P, $p_m \in$ P. If

$$\mathbf{P_{t1} = p_i \times p_l \times p_m}$$

$$\mathbf{P_{t2} = p_j \times p_l \times p_m}$$

and

$$\mathbf{P'_{t1} = p_{ii} \times p_l \times p_m}$$

$$\mathbf{P'_{t2} = p_{jj} \times p_l \times p_m}$$

then for,

$$\mathbf{P_{t1} + P_{t2} = p_i \times p_l \times p_m + p_j \times p_l \times p_m = p_l \times p_m \times (p_i + p_j)}$$

and

$$\mathbf{P'_{t1} + P'_{t2} = p_{ii} \times p_l \times p_m + p_{jj} \times p_l \times p_m = p_l \times p_m \times (p_{ii} + p_{jj})}$$

The equality $P_{t1} + P_{t2} = P'_{t1} + P'_{t2}$ holds True.

Mapping Step 2

For a partial world state for partial world W, S_W, is the set of assignments of values for each one of the random variables of the MFrag F_i in the partial world.

A local probability distribution $\pi_{RV(\varepsilon)}$ defined for a resident random variable $RV(\theta_i \ldots \theta_n)$ in MFrag F_i in addition to specifying a subset of values for the resident random variable provides a probability distribution function such that,

$\pi_{RV(\varepsilon)}(\gamma|S_W) >= 0$ and $\sum_\gamma \pi_{RV(\varepsilon)}(\gamma|S_W)=1$, where γ is finite subset which ranges over the set $\{\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n\} \cup \{T,F\}$ where "T", "F" denote Boolean truth values TRUE and FALSE respectively.

For a mapping to exists between PCFG and MEBN there shall be a unique MFrag for each of non-terminal having production rules, implying,

$$\forall (N^i \rightarrow N^r N^s \in R_{PCFG} \text{ and } N^i \rightarrow w^k \in R_{PCFG})$$

$$F_i \in T_{MEBN} \text{ AND } hasProbability(\theta_{Ni}, \theta_{NrNs}) \in R^i{}_{MEBN}.$$

Mapping Step 2

The mapping from PCFG and MEBN theory ensures the following.

$$P_{PCFG\text{-}MEBN}(N^i \rightarrow N^r N^s) = \&\ (P\ (N^i \rightarrow N^r N^s),\ \pi_{RV(\varepsilon)}(\gamma | S_W))$$

$$\text{if } N^i \rightarrow N^r N^s \in R_{PCFG} \text{ otherwise 0, where } N^r N^s \in \gamma,$$

and

$$P_{PCFG\text{-}MEBN}(N^i \rightarrow w^k) = \&\ (P\ (N^i \rightarrow w^k),\ \pi_{RV(\varepsilon)}(\gamma | S_W))$$

$$\text{if } N^i \rightarrow w^k \in R_{PCFG} \text{ otherwise 0 where } w^k \in \gamma.$$

The conflation, $\&\ (P\ (N^i \rightarrow N^r N^s),\ \pi_{RV(\varepsilon)}(\gamma | S_W))$ is defined as

$$= \frac{P\ (N^i \rightarrow N^r N^s) \times \pi_{RV(\varepsilon)}(\gamma | S_W)}{(P\ (N^i \rightarrow N^r N^s) \times \pi_{RV(\varepsilon)}(\gamma | S_W) + (1 - P\ (N^i \rightarrow N^r N^s)) \times (1 - \pi_{RV(\varepsilon)}(\gamma | S_W)))}$$

CYK Parsing for the mapping

Given a context-free grammar G_{PCFG}, F_{PCFG} is the set of all derivations of the grammar G_{PCFG}. Let *gen(t)* denote the string $s=w^k_1 \ldots w^k_n$ where $s \in T^*_{PCFG}$ and $F_{PCFG}(s)= \{t: t \in F_{PCFG}, gen(t) = s\}$ is a set of all possible parse trees for string *s*.

$P(N^i \rightarrow N^r N^s)$ and $P(N^i \rightarrow w^k)$ denote the probability associated with $N^i \rightarrow N^r N^s$ and $N^i \rightarrow w^k$ respectively such that

$$\forall i \sum_{r,s} P(N^i \rightarrow N^r N^s) + \sum_k P(N^i \rightarrow w^k) = 1.$$

For a given parse tree $t \in F_{PCFG}$ derived using set of rules $\alpha 1 \rightarrow \beta 1 \in R_{PCFG}$, $\alpha 2 \rightarrow \beta 2 \in R_{PCFG}$, . . ., $\alpha n \rightarrow \beta n \in R_{PCFG}$, *p(t)* is defined as,

$$p(t)= \forall x \prod P(\alpha x \rightarrow \beta x).$$

For $F_{PCFG}(s)$ the highest scoring parse tree is

$$arg\ max\ p(t)$$

A CYK algorithm to parse the string $s=w^k_1 \ldots w^k_n$ of the PCFG takes G_{PCFG} and *s* as inputs and outputs *arg max p(t)* for $F_{PCFG}(s)$. Since the algorithm being a recursive, $\Delta(l, o, N^i)$ is defined as, for the string $s=w^k_1 \ldots w^k_n$

$$\Delta(l, o, N^i) = max\ p(t) \text{ for } t \in F_{PCFG}(l, o, N^i).$$

CYK Parsing for the mapping

Algorithm: CYK

Input:

$s=w^k_1 \ldots w^k_n$ and $G_{PCFG} = (M_{PCFG}, T_{PCFG}, R_{PCFG}, S_{PCFG}, P_{PCFG})$

START:

1. ***For*** $x = 1 \ldots (n-1)$
2. ***For*** $y = 1 \ldots (n-1)$
3. $z = y + x$
4. $\forall N^i \in R_{PCFG}$ *calculate*
5. $\Delta\ (y, z, N^i) = max\ (P\ (N^i \rightarrow N^r N^s) \times \Delta\ (y, z', N^r) \times \Delta\ (z'+1, z, N^s))$ *for* $z' \in \{y, \ldots (z-1)\}$

   // *pointers are to be stored for retrieval of*

   //*the highest scoring parse tree*
6. *store the pointers to y, z and* N^i *for*

   $arg\ max\ (P\ (N^i \rightarrow N^r N^s) \times \Delta\ (y, z', N^r) \times \Delta\ (z'+1, z, N^s))$ *for* $z' \in \{y, \ldots (z-1)\}$

END

Properness and Consistency

A PCFG is proper if,

$$\forall i \sum_j P(N^i \rightarrow \zeta^j) = 1$$

A PCFG is consistent if,

$$\forall s \in F_{PCFG} \sum p(t)\ \forall t \in F_{PCFG}(s) = 1$$

For any given i,

$$\sum_j P(N^i \rightarrow \zeta^j) = 1$$

And let $P'(N^i \rightarrow \zeta^j)$ indicate the normalized probability after conflation operation with the probability distribution function, $\pi_{RV(\varepsilon)}(\gamma|S_W)$, of a resident random variable $RV(\varepsilon)$ from MTheory's MFrag F_i.

$$P'(N^i \rightarrow \zeta^j) = \frac{\&(P(N^i \rightarrow \zeta^j), \pi_{RV(\varepsilon)}(\gamma|S_W))}{\sum_j \&(P(N^i \rightarrow \zeta^j), \pi_{RV(\varepsilon)}(\gamma|S_W))} * \sum_j P(N^i \rightarrow \zeta^j),$$

$$\forall j \in \{1, 2, \ldots, n\}.$$

However, in case of $N^i \notin M'_{PCFG,}$ or $j \in \{1\}$,

$$P'(N^i \rightarrow \zeta^j) = P(N^i \rightarrow \zeta^j).$$

By definition of $P'(N^i \rightarrow \zeta^j)$ and given that $\sum_\gamma \pi_{RV(\varepsilon)}(\gamma|S_W)=1$, it is evident considering the Lemma 1 and 2 that

$$\sum_j P'(N^i \rightarrow \zeta^j) = 1.$$

Properness and Consistency

And the probability distribution $\pi_{RV(\varepsilon)}(\gamma|S_W)$ will have the probabilities defined or inferred for each state of random variable corresponding to each production rule for the non-terminal N^i. The equality, $\sum_j P'(N^i \rightarrow \zeta^j) = 1$, implies that the property of properness of the PCFG driven by MEBN shall be preserved.

The consistency requirement of the PCFG driven by MBEN requires,

$$\forall s \in F_{PCFG} \sum p(t)\ \forall t \in F_{PCFG}(s) = 1.$$

This is ensured as the MTheory defined shall have MFrag for each of the nonterminal belonging to PCFG, such that ,$|T_{MEBN}| \geq |M_{PCFG}|$.

Given that,

$$\sum_\gamma \pi_{RV(\varepsilon)}(\gamma|S_W)=1$$

and

$$P'(N^i \rightarrow \zeta^j) = \frac{\&(P(N^i \rightarrow \zeta^j), \pi_{RV(\varepsilon)}(\alpha|S_W))}{\sum_j \&(P(N^i \rightarrow \zeta^j), \pi_{RV(\varepsilon)}(\alpha|S_W))} * \sum_j P(N^i \rightarrow \zeta^j),$$

$$\forall j \in \{1, 2, \ldots, n\}.$$

for any $p'(t)$ defined as,

$$p'(t) = \forall x \prod P'(\alpha x \rightarrow \beta x)$$

and considering Lemma 1 and 2, implies that $\forall s \in F_{PCFG} \sum p'(t)\ \forall t \in F_{PCFG}(s) = 1$, preserving the consistency of the grammar.

Modified CYK Parsing for the mapping

Algorithm: $CYK_{PCFG\text{-}MEBN}$

Input:

$s=w^k_1 \ldots w^k_n$ and $G_{PCFG} = (M_{PCFG}, T_{PCFG}, R_{PCFG}, S_{PCFG}, P_{PCFG})$

START:

1. ***For*** $x = 1 \ldots (n-1)$
2. ***For*** $y = 1 \ldots (n-1)$
3. $z = y + x$
4. $\forall N^i \in R_{PCFG}$ *calculate*
5. $\Delta\ (y, z, N^i) = max\ (P'(N^i \rightarrow N^r N^s) \times \Delta\ (y, z', N^r) \times \Delta\ (z'+1, z, N^s))$ *for* $z' \in \{y, \ldots (z-1)\}$

   // pointers are to be stored for retrieval of

   //the highest scoring parse tree
6. *store the pointers to y, z and* N^i *for*

   $arg\ max\ (P'(N^i \rightarrow N^r N^s) \times \Delta\ (y, z', N^r) \times \Delta\ (z'+1, z, N^s))$ *for* $z' \in \{y, \ldots (z-1)\}$

END

Experiments & Results

Syntactic relations are often best handled by computational tasks finding syntactic relations. These conventional computational methods are yet ambiguous and fail to outperform the elegance exhibited by the human brain in processing syntactic and semantic relations without ambiguity. One famous form of ambiguity is prepositional phrase (PP) attachment ambiguity

Considering two sentences S1 and S2, , their parse trees based on the sample PCFG grammar

S1: *"Alex eats fish with fork"*

S2: *"Alex eats fish with eggs"*

S⟶ NP VP 1.0

NP⟶ NP PP 0.5

PP⟶ P NP 1.0

VP⟶ V NP 0.7

VP⟶ VP PP 0.3

NP⟶ fish 0.18

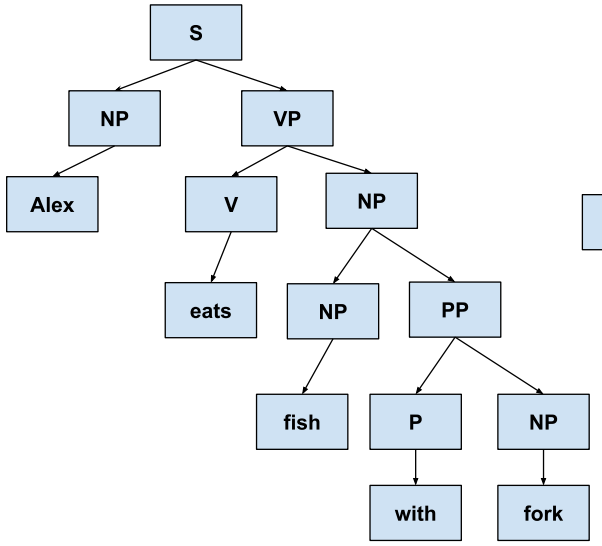
NP⟶ eggs 0.1

NP⟶ fork 0.04

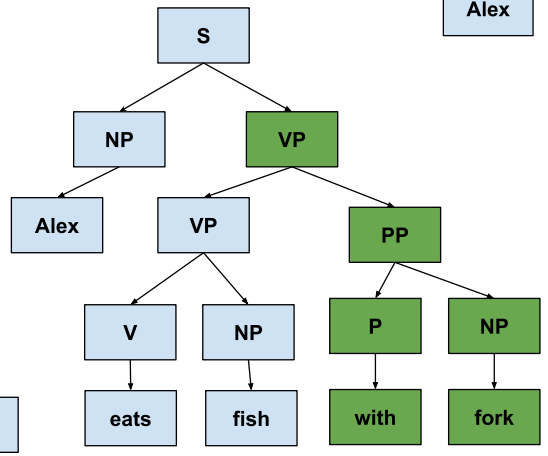
NP⟶ Alex 0.18

V⟶ eats 1.0

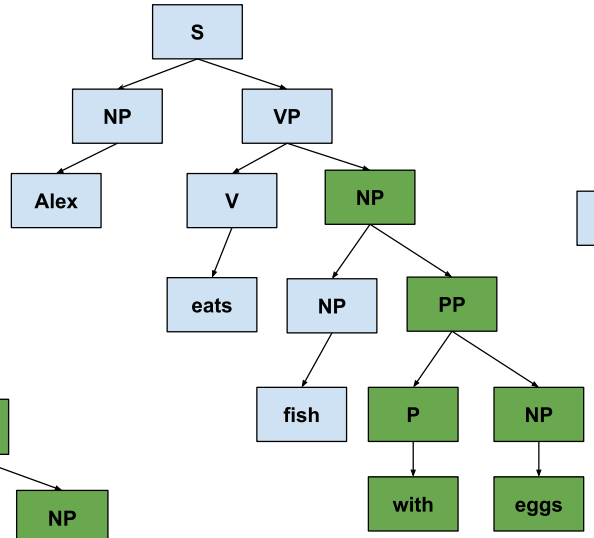
P⟶ with 1.0



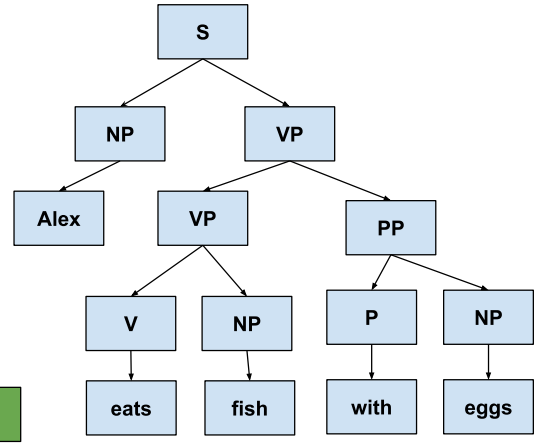
Parse Tree 1 for Sentence S1



Parse Tree 2 for Sentence S1
(semantically correct PP attachment to VP)



Parse Tree 1 for Sentence S2
(semantically correct PP attachment to NP)



Parse Tree 2 for Sentence S2

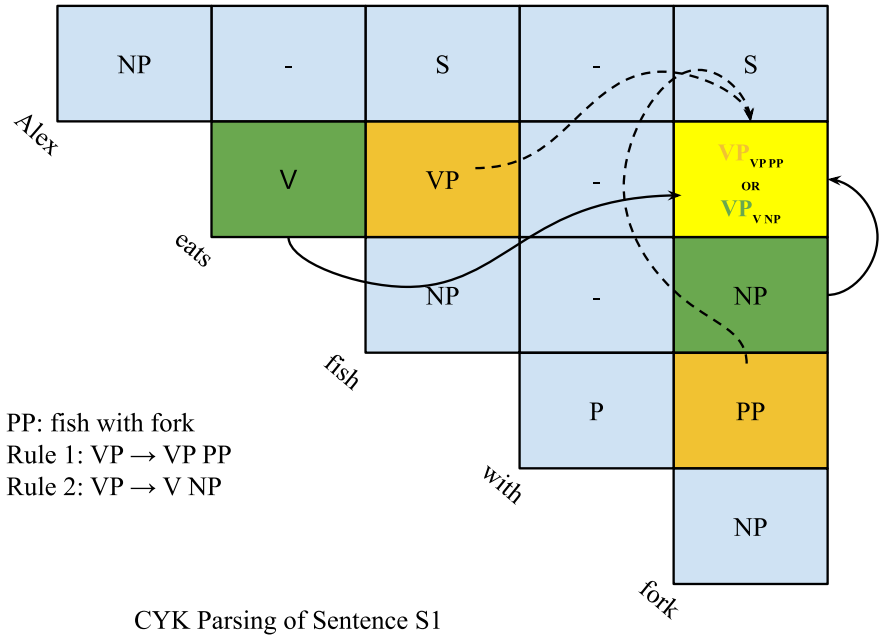
Experiments & Results

The PP attachment disambiguation can be attempted if the CYK parsing algorithm is modified to probabilistically reason to select a production rule rather than only considering the max product of probabilities.

In a step where CYK algorithm decides among two rules to choose from, *Rule 1*, *Rule 2* with probability products $VP_{VP\ PP}$ and $VP_{V\ PP}$ respectively for a derivation string *D*, a MEBN defined specifically to model the semantics of PP attachment for preposition "with" in the context of verb "eat" and prepositional phrase objects "eggs" or "fork" is queried on the resident random variable conditioned to reason about attachment semantics.

The query generates a situation specific Bayesian network (SSBN) and infers a probability value indicating with what likelihood the string *D* is semantically correct derivation from production rules *Rule 1* and *Rule 2*.

The knowledge represented using ontology forms the basis for the MEBN query over a MFrag's resident random variable for PP attachment disambiguation.



CYK Parsing of Sentence S1

Experiments & Results

To validate the proposed syntactico-semantic reasoning using PCFG and MEBN and the discussed PP attachment disambiguation process, the Algorithm CYK PCFG-MEBN has been implemented in Java and PP attachment disambiguation knowledge is modelled using Protégé and UnBBayes tools.

An MEBN is defined for the following PCFG with specific consideration to the semantics of preposition “with” in linguistic terminologies for the use case sentence: *“the dog saw the man with the telescope”*

Experiments & Results

The sentence to be parsed is,

"the dog saw the man with the telescope."

S → NP VP 1	*NN → man 0.1*
VP → V NP 0.7	*NN → woman 0.1*
VP → VP PP 0.3	*NN → telescope 0.3*
NP → DT NN 0.8	*NN → dog 0.5*
NP → NP PP 0.2	*DT → the 1.0*
PP → IN NP 1.0	*IN → with 0.6*
V → sleeps 0.5	*IN → in 0.4*
V → saw 0.5	

PP Attachment disambiguation Ontology

- Classes:
  - Verb
  - Subject
  - Object
  - PrepositionalPhraseObject
- Object properties:
  - hasObject
    - Domain: Verb
    - Range: Object
  - hasSubject
    - Domain: Verb
    - Range: Object
- Data properties:
  - PPObjectModifiesVerbAction
    - Domain: PrepositionalPhraseObject, Verb
    - Range: Boolean
  - PPObjectRelatesToObject
    - Domain: PrepositionalPhraseObject, Object
    - Range: Boolean
  - PPObjectRelatesToSubjectAct
    - Domain: PrepositionalPhraseObject, Subject
    - Range: Boolean

Experiments & Results

The prominent semantic roles centered around verb and noun phrases and their general descriptions in the context of the linguistics are as follows.

- Agent: the entity which is initiator of the action implied in the verb
- Patient: the entity which is affected by the action implied in the verb
- Instrument: the mode or tool with which the action implied in the verb was performed
- Benefactive: the entity recipient of the action implied in the verb.
- Goal: the entity representing the purpose of the action implied in the verb
- Source: the entity representing origin of the action implied in the verb
- Destination: the entity representing the destination of the action implied in the verb.
- Location: the entity representing the place of the action being executed

Experiments & Results

Often these roles will be the semantic information needed to answer the questions based on who, what, how, etc. The experiment to realistically model the prior probabilities for the MEBN designed relies on the semantic roles centered around the WH questions. The relation is deemed established between the entities if there exists a convincing answer for the any of the following WH question.

“Who” and “What” type of questions pointing to the Agent-Patient relation.

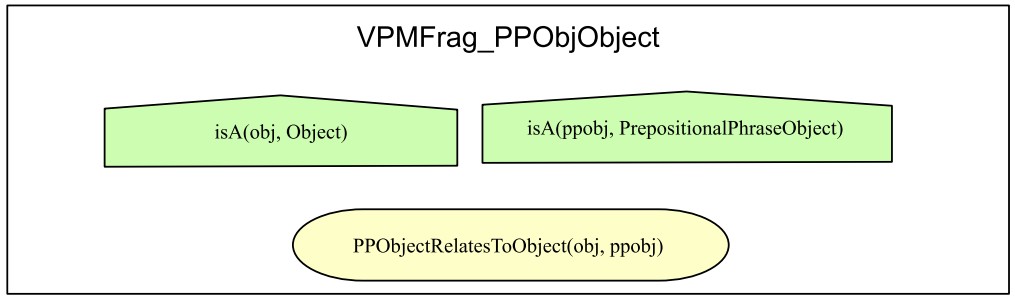
“How” and “with what” type of questions pointing to the use of instrument or mode or method.

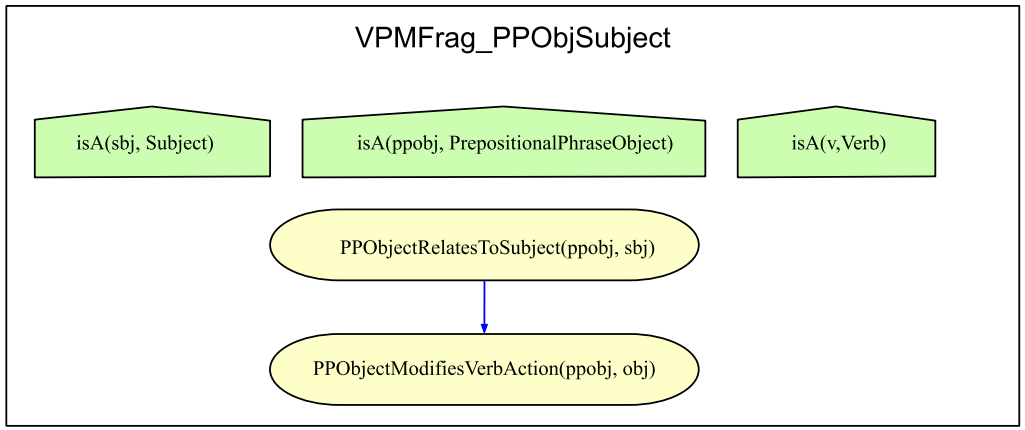
“Where” type of questions pointing to source, destination, and place of the action.

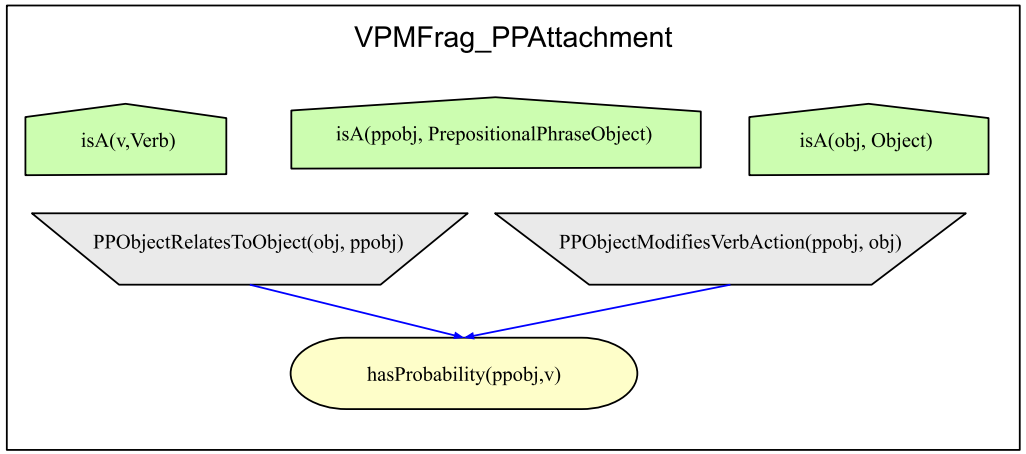
“Whom” type of questions pointing to the benefactive entity.

The labelled sample from the New York Times Corpus and Wikipedia Corpus based dataset samples is selected. This sample dataset contains labelled PP attachment sentences containing preposition “with”. The probability distributions defined for the MFrags are based on the probabilities calculated from the observations made in the sample dataset w.r.t to the semantic role identification.

Experiments & Results







PPObjectRelatesToSubjectAct_TELESCOPE_DOGOBJ	
True	0 %
False	100 %
Absurd	0 %

SSBN generated by MEBN Query

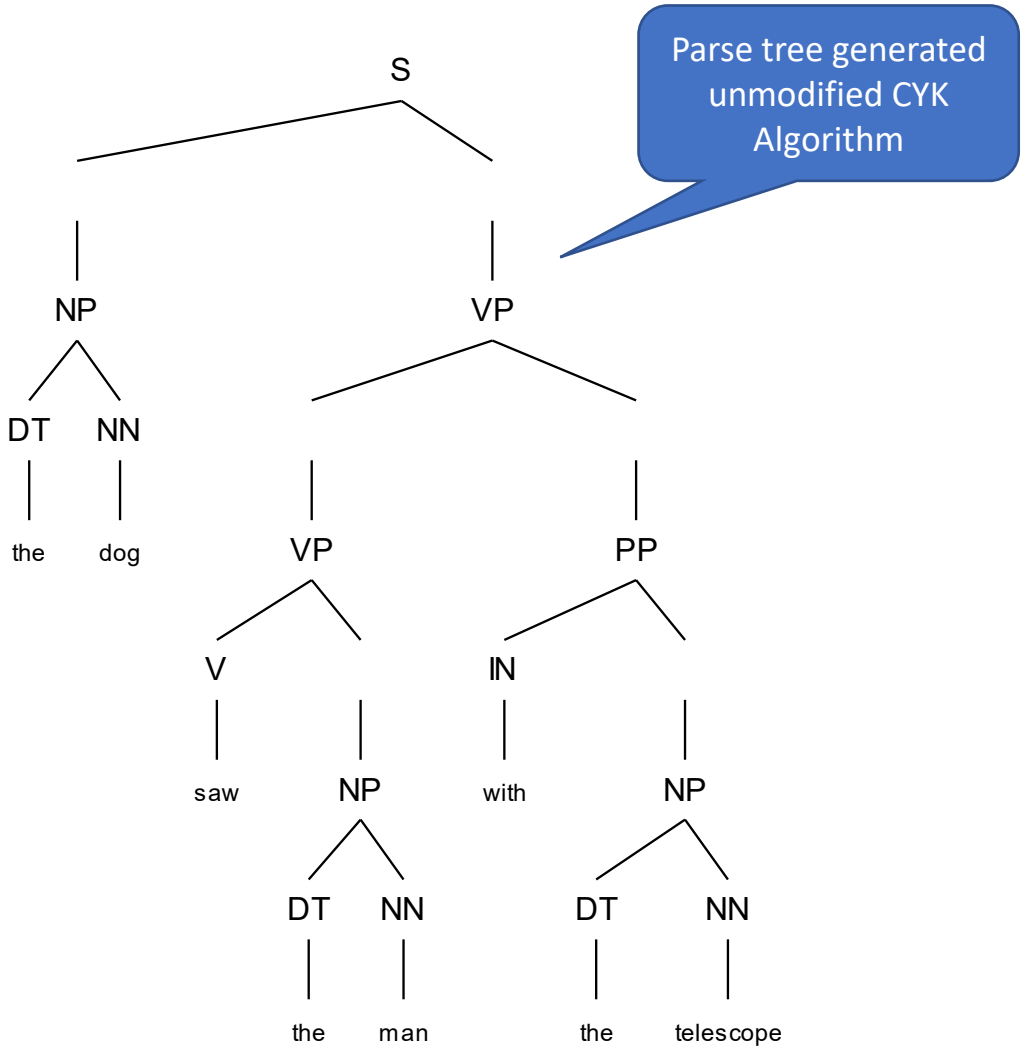
PPObjectModiefVerbAction_TELESCOPE_SAW	
True	80 %
False	20%
Absurd	0%

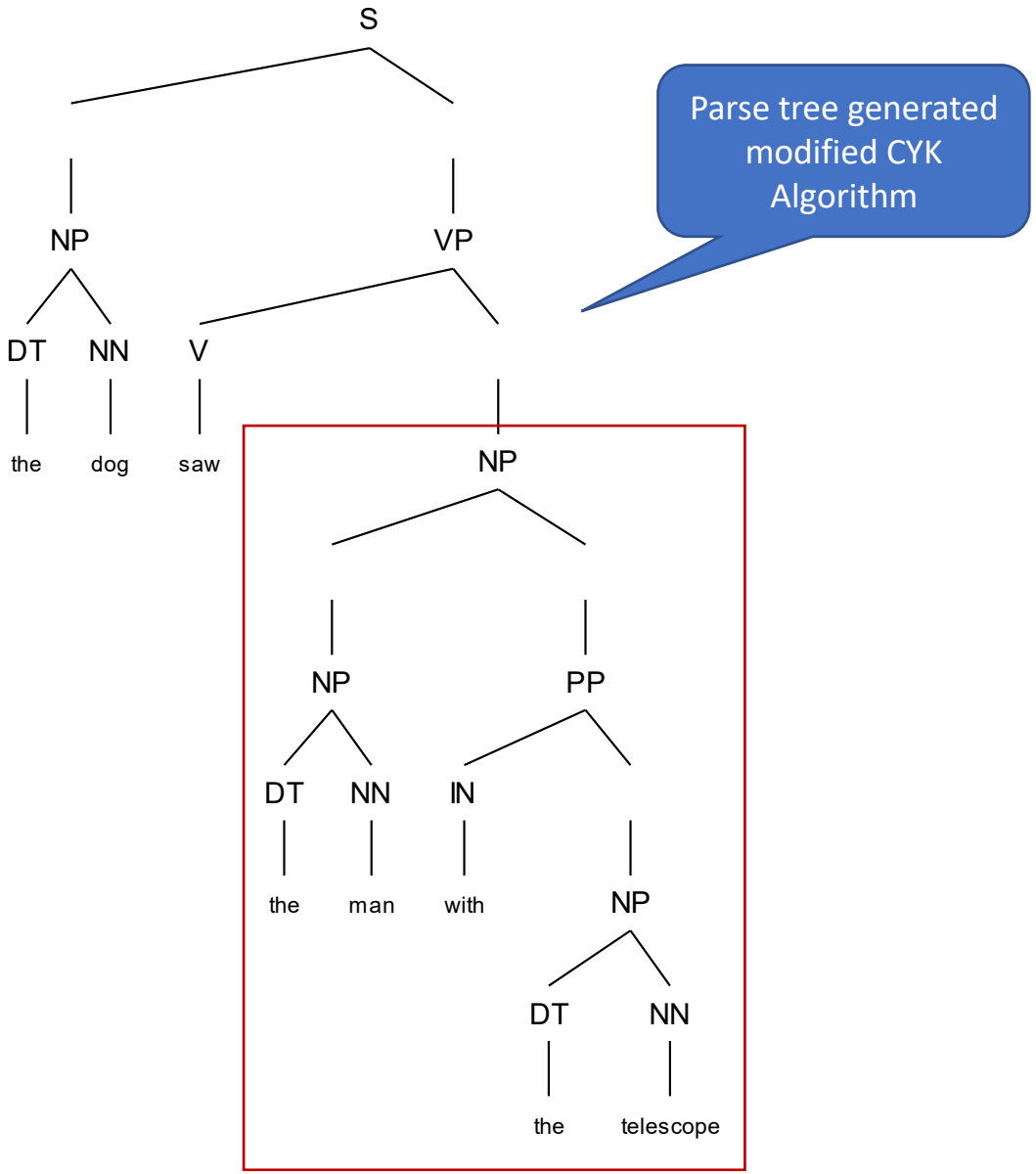
PPObjectRelatesToObject_MANOBJ_TELESCOPE	
True	100 %
False	0 %
Absurd	0 %

MEBN's MTheory for PP Attachment disambiguation

hasProbability_TELESCOPE_SAW	
True	34 %
False	66 %
Absurd	0 %

Experiments & Results





Experiments & Results

The experiments further identify sentences from the sample dataset such that their top two best parse trees obtained from Stanford CoreNLP's parser differ on the prepositional phrase attachment.

The Stanford CoreNLP's, version 4.2.2, trained PCFG's rules deciding the PP's attachment to VP and NP are

"@NodeSet-610143252" -> "@VP^S-VBF-v| VBD^VP_ NP^VP-B>" "PP^VP"

"@NodeSet-610143252" -> "@VP^S-VBF-v| VBD^VP_" "NP^VP-R" respectively.

Each parse tree has a score, a log probability.

It is observed that the state of the art Stanford CoreNLP's PCFG parser still could not disambiguate the PP attachment correctly to imply the real world meaning.

Upon applying the method of syntactico-semantic reasoning discussed it is observed that there is significant additive change in the log probability of parse trees indicating the correct attachment.

sentence #	subject	verb	object	preposition	PP object	attachment	PPObj modifies VbAction	PPO Relates to Obj	PPO Relates to Sbj	Relation to WH Question (Semantic Role)	Log Probability of Parse Trees			
											PP attached to VP	PP attached to NP	PP attached to VP after syntatico-semantic reasoning	PP attached to NP after syntatico-semantic reasoning
1	coleman	reached	base	with	regularity	v	Yes	No	Yes	how	-56.4137	-57.6083	-56.99	-57.608
2	taft	played	golf	with	passion	v	Yes	No	Yes	how	-55.9678	-57.1624	-56.544	-57.162
3	mayor	began	talks	with	unions	v	Yes	No	Yes	whom	-45.8165	-46.8976	-46.393	-46.897
4	mr rowland	shook	hands	with	voters	v	Yes	No	Yes	what	-64.8046	-65.8857	-65.381	-65.885
5	james	told	reporters	with	bluntness	v	Yes	No	Yes	how	-54.1544	-55.2355	-54.731	-55.235
6	mr rowland	played	golf	with	matthews	v	Yes	No	Yes	whom	-71.711	-72.9056	-72.287	-72.905
7	north shore	created	affiliations	with	college	v	Yes	No	Yes	whom	-64.5973	-65.6784	-65.174	-65.678
8	new york city police	charged	mason	with	possession	v	Yes	No	Yes	what	-84.3933	-85.5879	-84.97	-85.587
9	badgers	won	games	with	goals	v	Yes	No	Yes	how, with what	-51.2124	-52.2935	-51.789	-52.293
10	vocal ability	evoked	comparison	with	evans	n	No	No	Yes	whom	-69.0384	-70.233	-69.615	-70.232
11	agent	attended	meeting	with	tranter	v	Yes	No	Yes	whom	-52.3507	-53.5453	-52.927	-53.545
12	forbes	signed	contract	with	league	n	Yes	No	Yes	whom	-51.053	-52.2476	-51.629	-52.247
13	ksol	swapped	letters	with	kemr	v	Yes	No	Yes	whom	-63.8576	-64.3796	-64.434	-64.379
14	cedric sharpe	formed	quartet	with	beckwith	v	Yes	No	Yes	whom	-79.0135	-79.5981	-79.59	-79.598
15	davis	issued	statement	with	delay	v	Yes	No	Yes	how	-51.196	-52.3906	-51.772	-52.39
16	protesters	attacked	forces	with	bombs	v	Yes	No	Yes	how, with what	-48.8452	-49.9263	-49.421	-49.926
17	buzorgi	signed	contract	with	haifa	n	Yes	No	Yes	whom	-56.0198	-57.2144	-56.596	-57.214
18	workman	replaced	mcintosh	with	evans	v	Yes	No	Yes	what	-64.4351	-65.6296	-65.011	-65.629
19	wabdullah bosnevire abdullah bosnevi	studied	biology	with	scholars	v	Yes	No	Yes	whom	-103.578	-104.773	-104.15	-104.77
20	reecen	released	album	with	stream	v	Yes	No	Yes	how, with what	-57.659	-58.8536	-58.235	-58.853
21	delmas	signed	deal	with	delmas	n	No	No	No	---	-55.7816	-56.9762	-56.358	-56.976
22	o'brien	arranged	trade	with	bulls	n	No	No	No	---	-55.4693	-56.6639	-56.046	-56.663
23	clotilde dusoulier	signed	deal	with	books	n	No	No	No	---	-66.82	-67.1142	-67.396	-67.114
24	ponting	received	sponsorship	with	sport	n	No	Yes	No	---	-57.2913	-58.4859	-57.868	-58.485
25	owen	developed	fascination	with	song	n	Yes	Yes	Yes	whom	-59.0739	-60.2685	-59.65	-60.268
26	ali crawford	signed	contract	with	hamilton	n	No	Yes	Yes	whom	-74.1511	-75.0612	-74.727	-75.061

Experiments & Results

t-Test: Paired Two Sample for Means		
	Log Difference (7-8)	*Log Difference (9-10)*
Mean	-58.56075212	-59.26303056
Variance	50.22716257	50.25345428
Observations	86	86
Pearson Correlation	0.998103239	
Hypothesized Mean Difference	0	
df	85	
t Stat	14.91788407	
P(T<=t) one-tail	9.28052E-26	
t Critical one-tail	1.6629785	
P(T<=t) two-tail	1.8561E-25	
t Critical two-tail	1.988267907	

Shows the results obtained are significant.

Table 5.3: Significance test using t-Test (p-value = 0.05)

Repository of all codebases and dataset samples and other ancillary materials to reproduce the results and experiment
https://github.com/psrpatnaik/syntactico-semantic-pcfg-mebn

Experiments & Results

The modified algorithm's time complexity does remain asymptotically as **O(n^3.|G|)** where n is the length of string and |G| is the size of grammar, since the lookup for the non-terminal which requires syntactico-semantic reasoning is based **on O(1) lookup time complexity data structure.**

Consideration which needs discussion here is that the experiment focused specifically on preposition "with" and modelled the MEBN network specific to it. For a different preposition like "in" or other prepositions there is need to tweak the defined MEBN network.

The syntactico-semantic reasoning in its application successfully disambiguates the PP attachment problem and has potential application in other syntactic pattern recognition tasks which demand strong influence of the semantics of the domain the pattern recognition processes.

Experiments & Results

There are some optimizations proposed for the CYK algorithm, of which this research explored the valiant's algorithm and the parallelized versions of CYK algorithm proposed by [Yi et al., 2014]. The optimizations proposed in the valiant's algorithm primarily rely on efficient matrix product algorithm, to be specific, the Strassen's matrix product algorithm.

The valiant's algorithm proposes a mapping between matrix multiplication with transitive closure and sequence of operations a CFG performs to generate a string. Given the working of the valiant's algorithm and the algorithm $CYK_{PCFG\text{-}MEBN}$ can also reap the same optimizations leading to the time-complexity of O(n^2.8 * |G|).

Experiments & Results

Also, the contour surface plot, surface plot, and z-view of the conflation of probabilities represented over the x-axis and y-axis on the z-axis show the behavior of any system dependent on the resultant conflation of the probabilities.

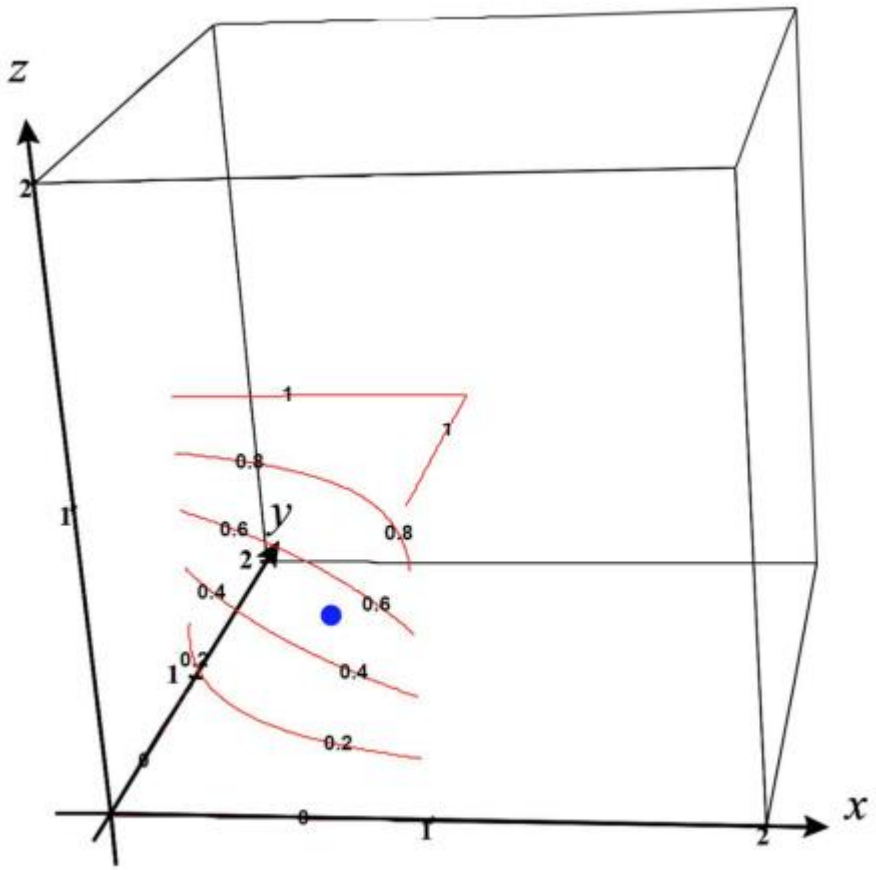


Figure 5.8: Contour surface plot.

Experiments & Results

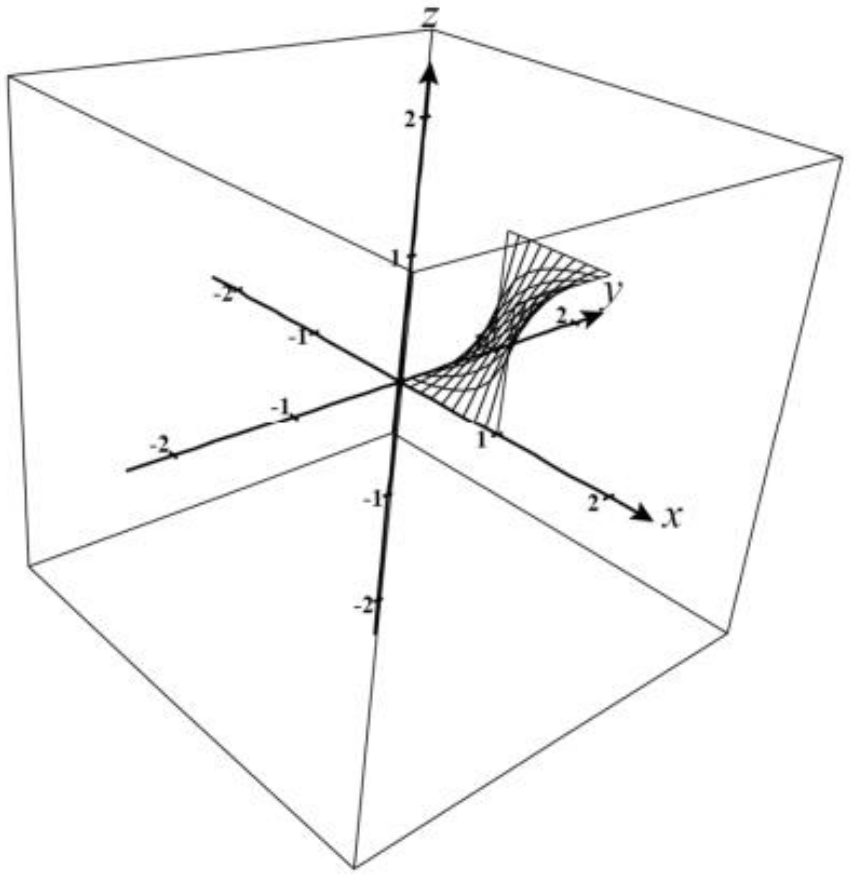


Figure 5.9: Surface plot.

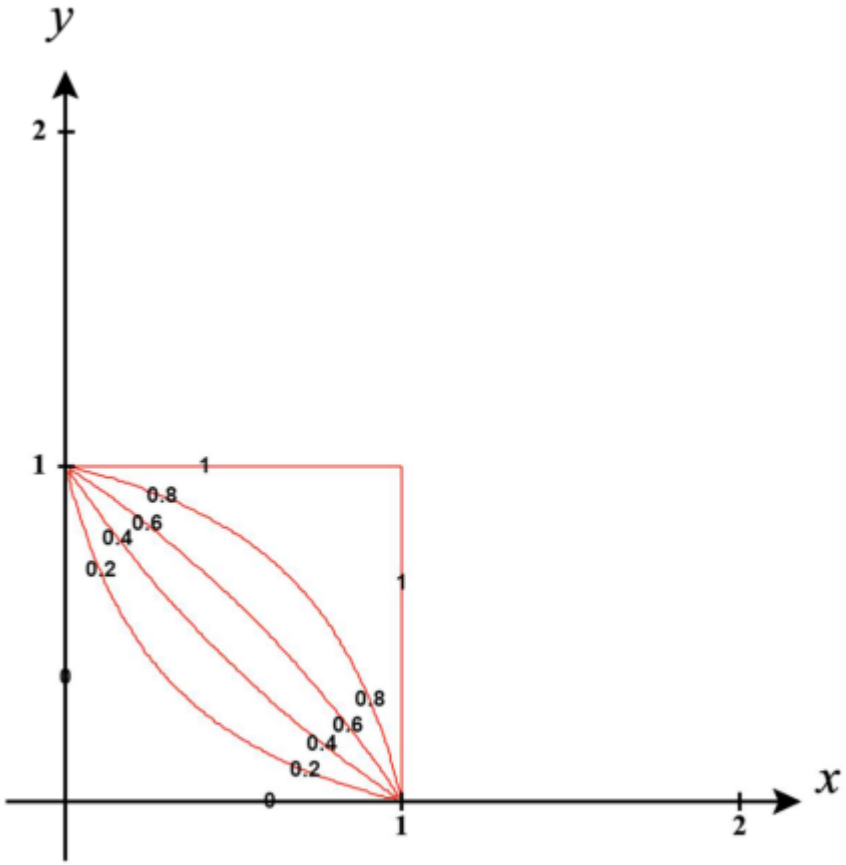


Figure 5.10: z-view of contour plot.

Conclusion

Today a sentence, “Apple’s stocks fell”, could mean something completely different based on what the word “Apple” means, hence the need to consider the contextual and semantic information.

Earlier works have heavily stressed the need to give larger importance to contextual and semantic information for modern NLP tasks. Ontology Web Language (OWL) and upper ontologies built on top of the MEBN for example the PR-OWL, elegantly capture the semantic knowledge bases along with probabilistic reasoning capabilities.

There is a research gap in integration methods for PCFG with Probabilistic semantic reasoning systems which this research successfully attempts to bridge. Inspired by the principle of semantic compositionality and its applications, the research is a successful attempt to formally define a method to map a probabilistic syntactic pattern recognition process with probabilistic ontology based graphical knowledge representation and reasoning systems preserving the PCFG’s consistency and properness properties.

Conclusion

The research successfully attempts to bridge the gap in integration methods for PCFG with probabilistic semantic reasoning system, MEBN.

The research in this thesis is a promising and a successful attempt to define an integration mapping method in formal specifications to map a syntactic & probabilistic pattern recognition process with ontology based probabilistic & graphical knowledge representation and reasoning system.

Using the Lemmas derived in the research the PCFG's properness and consistency properties are ensured in the proposed integration mapping. The formal specification of mapping between PCFG and MEBN and the lemmas proposed in the research are novel contributions leading towards a concept of syntactico-semantic reasoning in syntactic pattern recognition processes in general.

Conclusion

The wide adoption of MEBN and its development ecosystem makes the proposed syntactico-semantic pattern recognition method to be directly used in existing systems.

Attempts for the disambiguation of PP attachment using context aware semantic information have proved to be effective compared to syntactic or text corpus based token embeddings. The approach for PP attachment disambiguation discussed in this work is one such promising approach not requiring the disambiguation method to frequently train on the text corpus.

The proposed syntactico-semantic reasoning also has immense potential in medical applications like protein sequencing and decision support systems in the context of medical texts.

Future Scope

Future scope of this research specifically stresses the additional work and research needed for the following.

1) Semantic reasoning in MEBN is computationally intensive, for a large MEBN network SSBN generation and inference is slower, appropriate parallel or tensor computation based reasoning algorithms needs to be developed to improve scalability of MEBN for enterprise level application integration.

2) Though there exists an uncertainty modelling process for MEBN, there is need to defined custom domain specific probabilistic ontology modelling processes with more focus on domain expert involvement.

3) The proposed mapping being novel in its nature, there is a need to develop gold datasets for performance evaluations and benchmarking further extensions in the field of syntactico-semantic reasoning for various application domains.

Thank you