IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 13, 2021, accepted September 5, 2021, date of publication September 23, 2021, date of current version October 11, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3115445

A Theoretical Foundation for Syntactico-Semantic Pattern Recognition

SHRINIVASAN PATNAIKUNI AND SACHIN GENGAJE
Department of Computer Science and Engineering, Walchand Institute of Technology, Solapur 413006, India
Corresponding author: Shrinivasan Patnaikuni (psrpatnaik@gmail.com)

ABSTRACT Conventionally syntactic pattern recognition tasks have been driven by grammars defining a syntactic structure. Syntactic Pattern recognition tasks were primarily relying on the ability of parsing algorithms to recognize the patterns in the input data. These algorithms essentially were dependent on the syntactic grammars defining the patterns. Context free grammars, a type of grammars have been particularly well studied for pattern recognition tasks to be solved by computer efficiently. Some of the key pattern recognition tasks had applications in Natural Language Processing (NLP). Though context free grammars are well suited for capturing rigid patterns and unambiguous patterns, there was a need to encapsulate the uncertainty aspects involved in some pattern recognition processes. Probabilistic context free grammars can well handle the need to capture uncertainty in the processes but not in a true sense they are able to capture the uncertainty associated with the semantic context governing the domain in which the pattern recognition processes are being attempted at. The paper formally puts forth an approach for syntactico-semantic pattern recognition. The syntactico-semantic pattern recognition attempts to capture the semantic context and the uncertainties involved thereof along with probabilistic reasoning. The approach consists of integration mapping between probabilistic context free grammar (PCFG) and Multi Entity Bayesian network (MEBN), a first-order logic for modeling probabilistic knowledge bases. Additionally, the paper outlines a modified version of the CYK parser algorithm for the defined mapping between PCFG and MEBN with a method to ensure the properness and consistency of such PCFG along with its key application, disambiguation of PP (Prepositional Phrase) attachment. The theoretical foundation proposed has been validated by a proof-of-concept implementation of the modified CYK algorithm for syntactico-semantic reasoning in Java with promising ability to disambiguate PP attachment uses cases of New York Times and Wikipedia corpus dataset samples.

INDEX TERMS Syntactic pattern recognition, probabilistic reasoning, MEBN.

I. INTRODUCTION

Grammars form the core of syntactic pattern recognition tasks, of specific importance, are context free grammars (CFG). Often the statistical and syntactic patterns find their best descriptions by a CFG than a regular expression. Probabilistic graphical models are being used for modeling uncertainty reasoning problems for several decades. A well-defined and practical example of probabilistic graphical models is the Bayesian network. The work reported in this paper fundamentally relies on the CFG and Bayesian Network.

The associate editor coordinating the review of this manuscript and approving it for publication was Joewono Widjaja.

A. PROBABILISTIC CONTEXT FREE GRAMMARS (PCFG)

Probabilistic Context Free Grammar [1] 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
- $R_{PCFG} = \{N^i \rightarrow \zeta^j : \zeta^j \in (M_{PCFG} \bigcup T_{PCFG})^*\}$ 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)
  - $R_{PCFG} = \{N^i \rightarrow N^r N^s, N^i \rightarrow w^k\}$
  - $\forall i \sum_{r,s} P(N^i \rightarrow N^r N^s) + \sum_k P(N^i \rightarrow w^k) = 1$

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

$N^1 \rightarrow N^2 N^3 1.0 \quad N^2 \rightarrow w^2 0.1$

$N^2 \rightarrow N^2 N^4 0.4 \quad N^2 \rightarrow w^3 0.04$

$N^4 \rightarrow N^5 N^2 1.0 \quad N^2 \rightarrow w^4 0.18$

$N^3 \rightarrow N^6 N^2 0.7 \quad N^2 \rightarrow w^5 0.1$

$N^3 \rightarrow N^3 N^4 0.3 \quad N^6 \rightarrow w^6 1.0$

$N^2 \rightarrow w^1 0.18 \quad N^5 \rightarrow w^7 1.0$

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

B. MULTI ENTITY BAYESIAN NETWORK (MEBN)

A MEBN (Multi Entity Bayesian Network) theory [2] 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. The arguments used in the resident random variable representation are used to represent variables, constant symbols, or the other resident random variables.
- G^i_{MEBN} is a directed acyclic graph representing the dependency between input random variables and resident random variables conditional on context random variables in one to one correspondence similar to the Bayesian network.
- 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 MEBN theory T_{MEBN}, a set of MFrags $\{F_1, F_2, F_3, \ldots F_n\}$, 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 the MFrag [2].

Collectively MFrags represent a knowledge base for a specific scenario requiring probabilistic reasoning. Once an MFrag is defined it can be applied as a repeatable pattern of reasoning queries, this is achieved by allowing the random variables to accept arguments. The arguments passed to a random variable within an MFrag are called ordinary variables and are distinct from the concept of random variables. In addition to random and ordinary variables, MEBN defines the notion of context nodes, resident nodes, and input nodes. The context nodes are represented as green pentagons and represent a condition to be satisfied for the probability distributions to be applied to the random variables within the MFrag. Input nodes are represented as trapezoids and are the random variables having an effect on resident random variables. Resident random variables are represented as yellow ovals.

Since the Bayesian networks are fixed and rigid, when used as a tool for knowledge representation put the systems relying on it for uncertainty reasoning at disadvantage. Often systems catering to real-world situations demand flexibility in modeling the knowledge representation networks dynamically and MEBN being a fusion of Bayesian Networks with First-Order Logic expressivity stands as the most desirable probabilistic knowledge reasoning tool.

In an MFrag, resident random variables are conditioned on input random variables, these input random variables are the resident random variables in some other MFrag of the same MEBN, overall, the MEBN is a large Bayesian network with MFrags being smaller subnetworks. The arguments in the form of logical variables or even as a function passed to the resident random variable are responsible for the instantiation of situation-specific Bayesian network (SSBN) specific to the values of the arguments, thereby accommodating flexibility.

MEBN Theory is queried using first-order logic constructs, connectives, and operators. Every query on MEBN involves the construction of a situation-specific Bayesian network (SSBN) from the set of MFrags belonging to the concerned MTheory.

Given the probabilistic approach of formally defined systems like PCFG and MEBN theory for uncertainty reasoning in syntactic and semantic aspects respectively, a mapping between these two formal systems lays the foundation for syntactico-semantic reasoning. MEBN theory has been widely adopted and used in various fields [3]. MEBN theory has been mapped with the Relational Model of Relational Databases [4] in a similar attempt as this paper puts forth.

II. BACKGROUND

The background of the research primarily encompasses the following aspects.

1. Augmenting context-free grammars with semantic information.
2. Works based on semantic compositionality.
3. Uncertainty representation in Semantic web.
4. Use of MEBN as a probabilistic ontology-based reasoning system.

Context-free grammar as its definition implies, cannot capture, represent, or encode any information representational of the contextual information. In other words, this simply means their primary role is in validating the sentences for syntactic validity, restricting it from analyzing the sentence in semantic form. There have been some pioneering works by [28] and [29] namely, Affix grammars and Attribute grammars to address the need to represent contextual semantic information in the grammar's rule productions. Especially, production rules in attribute grammars with the help of additional attributes capture semantic information. The notion of Affix and Attribute grammars develop the mechanism of encapsulating the semantic information in the rule productions of the grammar. Reference [30] used linear constraints to represent semantic properties within context-free grammar and was one of the early attempts to augment context-free grammar with semantic information. Though the probabilistic/stochastic grammar uses probability distribution over the rules of the grammar they are purely based on the syntactic and statistical properties of the strings in the grammar [7], [8]. Moreover, [31] and [32] have shown that semantic information in the form of distributional semantics will be challenging and not sufficient enough for identifying logical relatedness between the constructs of the grammar. Frege's Principle, also known as the principle of semantic compositionality [12], states that the meaning of a complex sentence is dependent on the meaning of individual subparts of the sentence and the way the subparts form the complex sentence. The principle of semantic compositionality in an interesting way suggests that for understanding a complex expression it is pertinent to know the way the complex expression is formed from sub expressions and their meaning in semantic context [33], [34]. Nevertheless, probabilistic context-free grammar is the candidate henceforth studied for augmentation with semantic information representation systems due to their ability to represent the uncertainty of the rule production using a probability distribution. In the context of parse trees generated by the context-free grammar works by [35] and [36] suggest an approach based on attention mechanism during the parse tree generation thereby paving a way to realize a form of semantic compositionality. Reference [37]'s work on Affix grammars is the one inspired by [28], attempting direct integration with ontology for syntactic assisted semantic parsing of weighted affix grammar strings. Reference [9]'s work attempted augmenting probabilistic grammars with Markov Random fields in an unsupervised learning model for the structure induction to identify a set of images with image features modeled as structural aspects of PCFG. However, the notion of (PGMM) Probabilistic Grammar-Markov Model does not consider the role of semantics played in image classification.

For knowledge-based systems, the information is represented in the form of collections of symbolic structures which is often referred to as knowledge base. Reasoning over the knowledge base is primarily a process of understanding, analysis, and processing the collection of symbolic structures to produce useful insights. The work by [38] and [39] treats the knowledge-based system as a problem-solving activity. The work by [40] describes the concept of the semantic web as a web of data that can be processed by computers paving away towards knowledge-driven systems from data-driven systems. In the world of knowledge representation, an ontology is a concept which encapsulates some common terms in a domain that represents the domain knowledge in a fashion that allows the knowledge to be processed by computers. The Web ontology language (OWL) assists the semantic web by providing a well-defined semantics and reasoning capability [41]. Despite well-defined semantics, OWL fails to capture the inherent uncertainties in the knowledge bases thereby making the semantic web unable to capture uncertainty. To address this issue, the consortium of the world wide web (W3C) put forth the project "Uncertainty Reasoning for the World Wide Web" [42]. The notion of probabilistic ontologies with capabilities like expressiveness, probabilistic reasoning to model the uncertainty was proposed by [43], [44]. This notion of probabilistic ontology is called PROWL and is developed as an upper ontology over the Bayesian network-based probabilistic graphical reasoning model called MEBN (Multi Entity Bayesian Network). PR-OWL inherits the capability of probabilistic reasoning with a First-order logic expressivity using the underlying MEBN. Reference [10] has further refined the PR-OWL overcoming some of the issues in it. Due to the well-defined semantics and ability to probabilistically capture the uncertainty and reasoning capabilities thereof this research builds its work based on PR-OWL and MEBN. MEBN theory has been mapped with the Relational Model of Relational Databases [4] in a similar attempt as this paper puts forth. Moreover, the MEBNs are being used across various systems for their ability of ontology-based probabilistic reasoning. The primary objective of using MEBN is to model the uncertainty in the ontologies. The uncertainty modeling in ontology leads to a notion called probabilistic ontologies. From modeling situational and contextual information in real-world perspective to building semantically enriched sophisticated intelligent systems, probabilistic ontologies and MEBN are at the core of such systems. Reference [45] modeled user behavior and activity pattern into a MEBN. A spiral model like processes were used to model MEBN for user's behavior, the spiral nature of processes meant the involvement of several iterations, experiments, and improvements over a period. Using MEBN resulted in the achievement of high reliability and low rate of false alarms, the user modeling was in the context of internet security through simulated user actions by the researchers. The researchers did point in their work that additional inclusion of several affecting factors and properties could further improve the reliability and robustness of their system.

Reference [46] highlighted the need for first-order probabilistic reasoning methods in the field of robotics to cater to the requirement of autonomously operating robotic systems. The concerned researcher proposed a method named Bayesian Logic Networks (BLN) which resembles the concept of MEBN. Primarily MEBN is being used as a probabilistic knowledge representation system with first-order logic expressivity. Reference [47]'s work mainly focused on merging the ontologies using PR-OWL and MEBN based on temporal aspects of the events. Reference [48] proposed a method of identification of most influencing classes for entities using multivariate analysis. The methods, kind of resemble pre MEBN modeling methodology.

Research Gap:

1. Lack of a well-defined formal method to integrate stochastic grammar with ontology-based probabilistic reasoning.

2. No attempts were made for the integration of stochastic grammar with well-established MEBN's ontology-based probabilistic reasoning capability

III. MAPPING BETWEEN PCFG AND MEBN

To establish a connecting link between PCFG and MEBN, there needs to be a mapping between the members of quintuples of PCFG and MEBN. The process is outlined in two steps.

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.

A. STEP 1

For a mapping to exist between a PCFG and a MEBN primarily there must be a relation between constructs of PCFG and MEBN. This implies the non-terminals and terminal symbols of PCFG shall have a kind of mapping with MEBN's constructs, specifically, the sets of input, context, and resident variables of each MFrags. The relation is established as follows,

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.

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

There shall be a unique resident random variable *hasProbability* $(\theta_{N^i}, \theta_{N^r N^s})$ *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,

$$\forall (F_i \in T_{MEBN}) \; has \; Probability \; (\theta_{N^i}, \theta_{N^r N^s}) \in R^{i}_{MEBN}$$

An example MFrag for the nonterminal N^{i} is represented in Fig. 1.

B. STEP 2

The mapping between probability distributions of PCFG and MEBN is achieved by combined probability distribution across PCFG and MEBN theory using a method of combining probabilities with least Shannon's information loss, the method is known as the conflation of probabilities [5] and is represented using symbol &(). The reason to use conflation of probability is that the conflation method gives more priority to distribution based on smaller standard deviation and

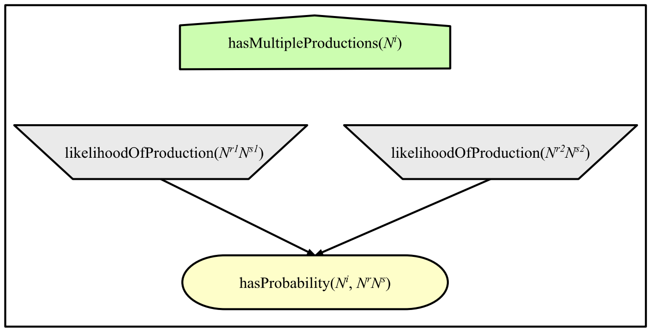

FIGURE 1. **MFrag for a nonterminal.**

avoids the issues which arise if a simple average of probability would have been considered instead. The method of using conflation of probability distribution has been successfully adopted by various other works [18]–[20]. Since the probability distribution of PCFG and MEBN is defined discrete sets, the conflation method to consolidate the probabilities in the mapping process is more effective. In general, if we have probability mass functions, pmf_1, pmf_2, and pmf_3 the conflation of these functions is commutative, associative, and iterative.

$$\begin{aligned} \&(pmf_1, pmf_2) &= \&(pmf_2, pmf_1), \\ \&(\&(pmf_1, pmf_2), pmf_3) &= \&(pmf_1, \&(pmf_2, pmf_3)), \\ \&(pmf_1, pmf_2, pmf_3) &= \&(\&(pmf_1, pmf_2), pmf_3) \end{aligned}$$

The conflation of probabilities in addition to being commutative, associative, and iterative also holds for the following lemmas.

Lemma 1 (Equality of Sum of Conflated Probabilities): Let, $\mathrm{P} = \{p_1, p_2, p_3, \ldots p_n\}$ and $\mathrm{P}' = \{p'_1, p'_2, p'_3, \ldots p'_n\}$ be a set of probability values such that, $|\mathrm{P}| = |\mathrm{P}'|$ 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 $p_i \in \mathrm{P}$, $p_j \in \mathrm{P}$ and $\mathrm{M}(p_i) = p'_i \in \mathrm{P}'$, $\mathrm{M}(p_j) = p'_j \in \mathrm{P}'$ if $p_i + p_j = S$ and given that

$$\begin{aligned} p_{i\,conflation} &= (p_i \times p'_i) \div ((p_i \times p'_i) + ((1-p_i) \times (1-p'_i))) \text{ and} \\ p_{j\,conflation} &= (p_j \times p'_j) \div ((p_j \times p'_j) + ((1-p_j) \times (1-p'_j))) \end{aligned}$$

then for,

$$\begin{aligned} p_{ii} &= (p_{i\,conflation} \div (p_{i\,conflation} + p_{j\,conflation})) \times (p_i + p_j) \text{ and} \\ p_{jj} &= (p_{j\,conflation} \div (p_{i\,conflation} + p_{j\,conflation})) \times (p_i + p_j), \end{aligned}$$

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 $p_l \in \mathrm{P}$, $p_m \in \mathrm{P}$. If

$$\begin{aligned} \mathrm{P}_{t1} &= p_i \times p_l \times p_m \\ \mathrm{P}_{t2} &= p_j \times p_l \times p_m \quad \text{and} \\ \mathrm{P}'_{t1} &= p_{ii} \times p_l \times p_m \\ \mathrm{P}'_{t2} &= p_{jj} \times p_l \times p_m \end{aligned}$$

then for,

$$\mathrm{P}_{t1} + \mathrm{P}_{t2} = p_i \times p_l \times p_m + p_j \times p_l \times p_m = p_l \times p_m \times (p_i + p_j)$$

and

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

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

In MEBN theory, for a non-empty finite set of entity identifier symbols $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$ there exists a partial world W of a resident random variable $RV(\theta_i \ldots \theta_n)$ which is the set of all instances of the parents of random variable $RV(\theta_i \ldots \theta_n)$ and the context variables of the MFrag F_i that can be obtained by substituting ε_i for ordinary variables $\{\theta_i \ldots \theta_n\}$ of F_i. A partial world state S_W for partial world 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 $\pi_{RV(\varepsilon)}$ 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 a finite subset which ranges over the set $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\} \bigcup \{T, F\}$ where "T", "F" denotes 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,

$$\forall (N^i \to N^r N^s \in R_{PCFG} \text{ and } N^i \to w^k \in R_{PCFG})$$
$$F_i \in T_{MEBN} \textit{ and has Probability } (\theta_{N^i}, \theta_{N^r N^s}) \in R^i_{MEBN\ldots}$$

The mapping from PCFG and MEBN theory ensures the following.

$$\begin{aligned} &P_{PCFG-MEBN}(N^i \to N^r N^s) \\ &\quad = \&(P(N^i \to N^r N^s), \pi_{RV(\varepsilon)}(\gamma | S_W)) \\ &\qquad \textit{if } N^i \to N^r N^s \in R_{PCFG} \quad \text{otherwise } 0, \end{aligned}$$

where

$$N^r N^s \in \gamma,$$

and

$$\begin{aligned} P_{PCFG-MEBN}(N^i \to w^k) &= \&(P(N^i \to w^k), \pi_{RV(\varepsilon)}(\gamma | S_W)) \\ &\textit{if } N^i \to w^k \in R_{PCFG} \quad \text{otherwise } 0 \end{aligned}$$

where

$$w^k \in \gamma$$

and.

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

$$\begin{aligned} &P(N^i \to N^r N^s) \times \pi_{RV(\varepsilon)}(\gamma | S_W) \div (P(N^i \to N^r N^s) \\ &\quad \times \pi_{RV(\varepsilon)}(\gamma | S_W) + (1 - P(N^i \to N^r N^s)) \\ &\quad \times (1 - \pi_{RV(\varepsilon)}(\gamma | S_W))). \end{aligned}$$

IV. MODIFIED CYK ALGORITHM, PROPERNESS AND CONSISTENCY OF THE PCFG MAPPED WITH MEBN

A. MODIFIED CYK ALGORITHM FOR PCFG MAPPED WITH MEBN

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 $s = w_1^k \ldots w_n^k$ 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.

$\mathrm{P}(N^i \to N^r N^s)$ and $\mathrm{P}(N^i \to w^k)$ denote the probability associated with $N^i \to N^r N^s$ and $N^i \to w^k$ respectively such that.

$$\forall i \sum\nolimits_{r,s} \mathrm{P}(N^i \to N^r N^s) + \sum\nolimits_k \mathrm{P}(N^i \to w^k) = 1.$$

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

$$p(t) = \forall x \prod \mathrm{P}(\alpha x \to \beta x).$$

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

$$arg\ maxp(t)$$

A CYK [21] algorithm to parse the string $s = w_1^k \ldots w_n^k$ of the PCFG takes G_{PCFG} and s as inputs and outputs *arg max* $p(t)$ for $F_{PCFG}(s)$.

Since the algorithm is recursive, $\Delta(l,o,N^i)$ is defined as for the string $s = w_1^k \ldots w_n^k$

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

where $F_{PCFG}(l, o, N^i)$ is a set of all parse trees for $w_1^k \ldots w_o^k$ for any l and o such that $1 \le l \le o \le n$ and N^i is a root node of the parse tree. It is to be noted that $\Delta(l,o,N^i) = 0$ if $F_{PCFG}(l,o,N^i)$ is an empty set.

Also,

$$\Delta(l,l,N^i) = \mathrm{P}(N^i \to w_l^k) \quad \text{if } N^i \to w_l^k \in R_{PCFG} \quad \text{otherwise } 0.$$

The mapping of PCFG with MEBN as defined in section III requires the parsing algorithm CYK to additionally lookup for probability value from MEBN query. The CYK algorithm needs to be modified to the only lookup for probability value from MEBN query when a specific non-terminal symbol is encountered in the parsing process for which Syntactico-semantic reasoning is expected. This is done by creating a set of non-terminal symbols $M'_{PCFG} \subset M_{PCFG}$ for which a MEBN query lookup is performed. Whenever, CYK algorithm matches a non-terminal $N^i \in M'_{PCFG,} \Delta(y, z, N^i)$ is calculated as

$$\begin{aligned} \Delta(y,z,N^i) = max(&P_{PCFG-MEBN}(N^i \to N^r N^s) \times \Delta(y,z',N^r) \\ &\times \Delta(z'+1,z,N^s)) \quad \forall z' \in \{y, \ldots (z-1)\} \end{aligned}$$

where, $\Delta(l,l,N^i)$ is calculated as

$$\Delta(l, l, N^i) = P_{PCFG-MEBN}(N^i \to w^k)$$

Algorithm CYK

Input:
$s = w_1^k \ldots w_n^k$ 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. $\Delta(y,z,N^i) = max(\mathrm{P}\ (N^i \rightarrow N^r N^s) \times \Delta(y,z',N^r) \times \Delta(z'+1,z,N^s))$ *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* $(\mathrm{P}(N^i \rightarrow N^r N^s) \times \Delta(y,z',N^r) \times \Delta(z'+1,z,N^s))$ *for* $z' \in \{y, \ldots (z-1)\}$

END

Otherwise in the case of, $N^i \notin M'_{PCFG}$

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

The probabilities of rules of PCFG are often learned by training and analysis of linguistic corpus [7], [8], usually, these processes are very time-consuming. To preserve the properness and consistency of PCFG driven by MEBN is very crucial to ensure that the semantics of parsing a string by a PCFG are unaffected. T An approach based on reducing the Kullback–Leibler (KL) divergence between the probability distributions is discussed in [25]. However, recalculating the conflated probabilities and further normalizing them preserves the property of properness and consistency without the need for a complex process of applying the Lagrange multiplier method as proved in proposition 9 of [6].

A PCFG is proper if,

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

A PCFG is consistent if,

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

For any given i,

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

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

$$\mathrm{P}'\left(N^i \rightarrow \zeta^j\right) = \frac{\&\left(P\left(N^i \rightarrow \zeta^j\right), \pi_{RV(\varepsilon)}\left(\gamma \mid S_W\right)\right)}{\sum_j \&\left(P\left(N^i \rightarrow \zeta^j\right), \pi_{RV(\varepsilon)}\left(\gamma \mid S_W\right)\right)} * \sum_j P\left(N^i \rightarrow \zeta^j\right) \quad \forall j \in \{1, 2, \ldots, n\}$$

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

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

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

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

And the probability distribution $\pi_{RV(\varepsilon)}(\gamma|S_W)$ will have the probabilities defined or inferred for each state of the random variable corresponding to each production rule for the non-terminal N^i.

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

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

This is ensured as the MTheory defined shall have MFrag for each of the nonterminal belonging to PCFG.

$$|\mathrm{T}_{MEBN}| \geq |M_{PCFG}|$$

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

$$\mathrm{P}'\left(N^i \rightarrow \zeta^j\right) = \frac{\&\left(P\left(N^i \rightarrow \zeta^j\right), \pi_{RV(\varepsilon)}\left(\alpha \mid S_W\right)\right)}{\sum_j \&\left(P\left(N^i \rightarrow \zeta^j\right), \pi_{RV(\varepsilon)}\left(\alpha \mid S_W\right)\right)} * \sum_j P\left(N^i \rightarrow \zeta^j\right) \quad \forall j \in \{1, 2, \ldots, n\}$$

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

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

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.

Also, it is noted that Theorem 1 from [2] specifically states that for a MEBN theory "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 distributions assigned by the MFrags".

The modified CYK Algorithm, $\mathrm{CYK}_{\mathrm{PCFG-MEBN}}$ follows as below.

V. EXPERIMENT

One of the well-suited applications of syntactico-semantic reasoning is disambiguation of the PP attachment [22]. Also, the application of syntactico semantic reasoning for the disambiguation of PP attachment is discussed in [23]. Based on the mapping defined a MEBN is defined for the following PCFG with specific consideration to the semantics of preposition "with" in linguistic terminologies.

S → NP VP 1   *NN → man 0.1*

Algorithm $\text{CYK}_{\mathbf{PCFG-MEBN}}$

Input:

$s = w_1^k \ldots w_n^k$ 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* $argmax(P'(N^i \rightarrow N^r N^s) \times \Delta(y, z', N^r) \times \Delta(z'+1, z, N^s))$ *for* $z' \in \{y, \ldots (z-1)\}$

END

$VP \rightarrow V\ NP\ 0.7$	$NN \rightarrow woman\ 0.1$
$VP \rightarrow VP\ PP\ 0.3$	$NN \rightarrow telescope\ 0.3$
$NP \rightarrow DT\ NN\ 0.8$	$NN \rightarrow dog\ 0.5$
$NP \rightarrow NP\ PP\ 0.2$	$DT \rightarrow the\ 1.0$
$PP \rightarrow IN\ NP\ 1.0$	$IN \rightarrow with\ 0.6$
$V \rightarrow sleeps\ 0.5$	$IN \rightarrow in\ 0.4$
$V \rightarrow saw\ 0.5$	

The sentence to be parsed is *"the dog saw the man with the telescope."*

The defined MEBN has three MFrags that model the ontology to capture the process of resolving the PP attachment resolution, the MFrags defined using a tool UnBBayes [24], and the approach recommended in [11] are shown in Fig. 2. The OWL modeled ontology has four classes and associated object and data properties.

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

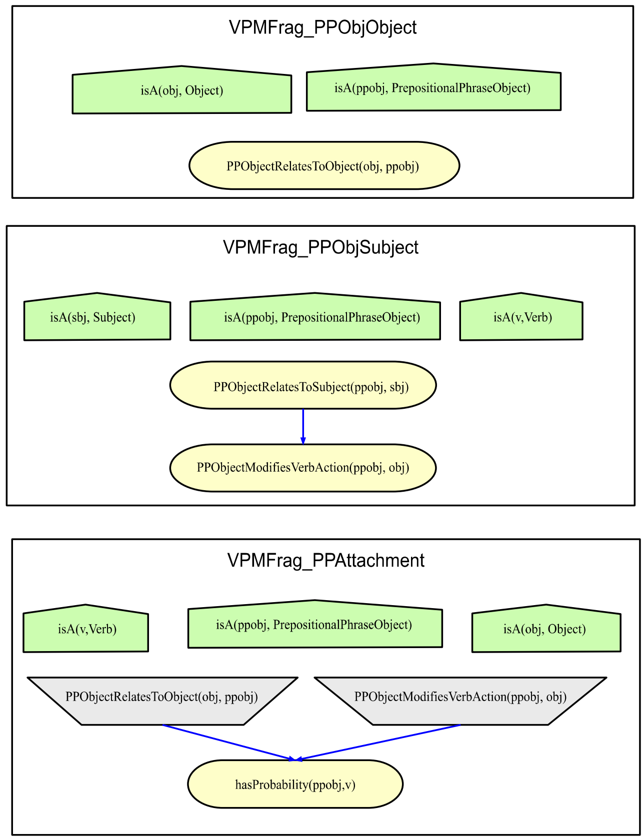

FIGURE 2. MEBN defined for disambiguation of PP attachment.

- PPObjectRelatesToObject
  - Domain: PrepositionalPhraseObject, Object
  - Range: Boolean
- PPObjectRelatesToSubjectAct
  - Domain:PrepositionalPhraseObject, Subject
  - Range: Boolean

For the sentence *"the dog saw the man with the telescope."* the instance of class Verb represents *"saw"*, the instance of class Object represents *"dog"*, the instance of class Subject represents *"man"* and the instance of class PrepositionalPhraseObject represents *"telescope"*.

Though discussing the semantics of linguistics of the English language is beyond the scope of this paper, a light introduction to semantic roles helps here in understanding the background behind the experiment setup more clearly in terms of the prerequisite linguistic understanding needed. Semantic roles in linguistics label the semantic meaning arising out of the verb and associated noun phrases in a sentence. Some of the prominent semantic roles centered around verb and noun phrases and their general descriptions in the context of linguistics are as follows.

- Agent: the entity which is the 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 the 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 is executed

Often these roles will be the semantic information needed to answer the questions based on who, what, how, etc. The experiment in this paper, 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 any of the following WH questions. This approach is based on the findings based on the arguments put forth in [27].

- "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 the source, destination, and place of the action.
- "Whom" type of questions pointing to the benefactive entity.

The labeled samples from the New York Times Corpus and Wikipedia Corpus based dataset are selected. This sample dataset contains labeled PP attachment sentences containing the preposition "with". The probability distributions defined for the MFrags are based on the probabilities calculated from the observed semantic roles in the sentences.

The parse tree obtained for the sentence *"the dog saw the man with the telescope."* using Java-based implementation of algorithm CYK is, see Fig. 3. The parse generated implies that the dog can see through the telescope which is highly unlikely in the real world.

Whereas the parse tree obtained for algorithm $CYK_{PCFG-MEBN}$ is, see Fig. 4. Here, the parse generated has the PP attached to the NP indicating the dog is seeing a man carrying a telescope, which makes much more sense.

The query to MEBN on the resident random variable hasProbability(v, ppobj) generated an SSBN as shown in Fig. 5. Upon evidence propagation across the SSBN, the probability for the rule VP $\rightarrow$ VP PP from the defined MFrag is inferred as 0.34. The parse tree obtained from algorithm $CYK_{PCFG-MEBN}$ shows the effect of the normalized conflated probability of the rule VP $\rightarrow$ VP PP for the sentence under consideration for the parsing to be semantically and syntactically meaningful.

Further, 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.

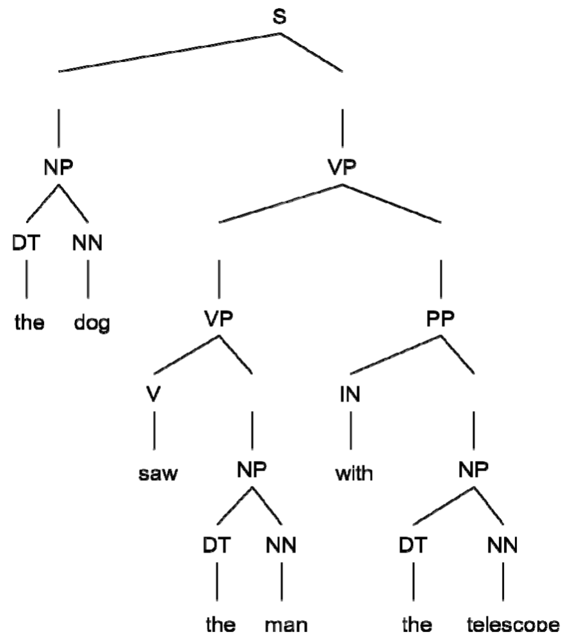

FIGURE 3. **Parse tree obtained using algorithm CYK.**

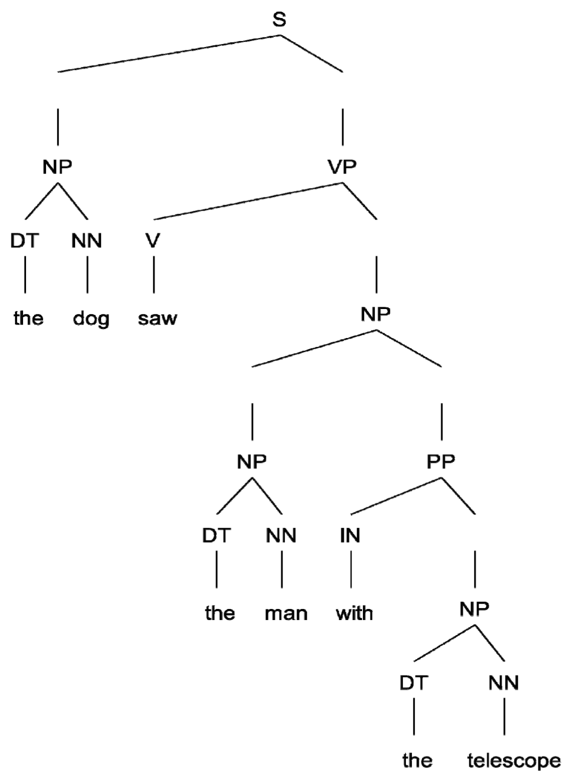

FIGURE 4. **Parse tree obtained using algorithm CYKPCFG-MEBN.**

The experiment further identifies sentences from the sample dataset such that their top two best parse trees obtained from Stanford CoreNLP's parser differ on the

TABLE 1. Sample sentences for disambiguation of PP 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 syntactico-semantic reasoning	PP attached to NP after syntactico-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

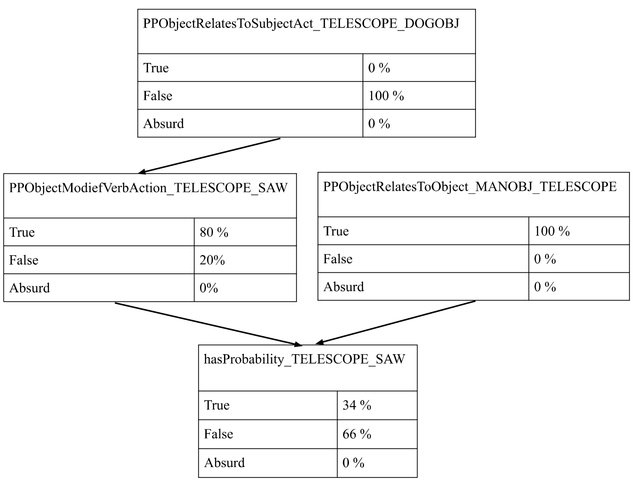

FIGURE 5. SSBN generated for a MEBN query on PP attachment.

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" and "@NodeSet-610143252" -> "@VP^S-VBF-v| VBD^VP_" "NP^VP-R" respectively. Each parse tree has a score, a log probability. It is observed from table 1 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 in section III and IV it is observed that there is a significant additive change in the log probability of parse trees indicating the correct attachment. Though the change is not to the extent to influence the outcome of the Stanford CoreNLP parser but proves the case that syntactico-semantic reasoning is a promising approach to incorporate semantic reasoning capabilities in syntactic reasoning processes. Further analysis of the experiment points out the fact that to achieve a change in log probabilities to an extent to change the outcome of the parser output, especially in the case of Stanford CoreNLP, the MEBNs have to be modeled more closely to resemble domain knowledge capturing every minute uncertainties. Also, various other factors like the depth of the parse tree, the magnitude of the probabilities of the PCFG rules influence the outcome, hence the syntactico-semantic reasoning.

VI. CONCLUSION

Syntactic pattern recognition tasks relying on probabilistic context free grammars have been very popular and widely adopted by several works. In the modern world where the data constantly changes not only statistically but also semantically, relying just on the syntactic or statistical aspects for syntactic pattern recognition is not enough to keep pattern recognition systems to the current data driven times. 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. The work [26] 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 [10], elegantly captures the semantic knowledge bases along with probabilistic reasoning capabilities.

Inspired from the [12]–[14] principle of semantic compositionality and its applications the paper is an attempt to formally define a method to map a probabilistic syntactic pattern recognition process with probabilistic ontology-based graphical knowledge representation and reasoning system. PCFGs have been used with Markov Random fields [9]. The wide adoption of MEBN and its development system 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 [15]–[17]. The approach for PP attachment disambiguation discussed in this paper is one such promising approach not requiring the disambiguation method to frequently train on the text corpus.

REFERENCES

[1] F. Jelinek, J. D. Lafferty, and R. L. Mercer, ''Basic methods of probabilistic context free grammars,'' in *Speech Recognition and Understanding*. Berlin, Germany: Springer, 1992, pp. 345–360.
[2] K. B. Laskey, ''MEBN: A language for first-order Bayesian knowledge bases,'' *Artif. Intell.*, vol. 172, nos. 2–3, pp. 140–178, Feb. 2008.
[3] S. R. P. Patnaikuni and S. R. Gengaje, ''Survey of multi entity Bayesian networks (MEBN) and its applications in probabilistic reasoning,'' *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 5, pp. 2425–2429, Jun. 2017.
[4] C. Y. Park and K. B. Laskey, ''MEBN-RM: A mapping between multi-entity Bayesian network and relational model,'' *Appl. Sci.*, vol. 9, no. 9, p. 1743, Apr. 2019.
[5] T. P. Hill, ''Conflations of probability distributions,'' *Trans. Amer. Math. Soc.*, vol. 363, no. 6, p. 3351, Jun. 2011.
[6] Z. Chi, ''Statistical properties of probabilistic context-free grammars,'' *Comput. Linguist.*, vol. 25, no. 1, pp. 131–160, Mar. 1999.
[7] Z. Chi and S. Geman, ''Estimation of probabilistic context-free grammars,'' *Comput. Linguist.*, vol. 24, no. 2, pp. 299–305, Jun. 1998.
[8] D. Klein and C. D. Manning, ''Accurate unlexicalized parsing,'' in *Proc. 41st Annu. Meeting Assoc. Comput. Linguistics (ACL)*, 2003, pp. 423–430.
[9] L. Zhu, Y. Chen, and A. Yuille, ''Unsupervised learning of probabilistic grammar-Markov models for object categories,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 114–128, Jan. 2009.
[10] R. N. Carvalho, K. B. Laskey, and P. C. G. Costa, ''PR-OWL—A language for defining probabilistic ontologies,'' *Int. J. Approx. Reasoning*, vol. 91, pp. 56–79, Dec. 2017.
[11] L. L. Santos, R. N. Carvalho, M. Ladeira, and L. Weigang, ''Probabilistic ontologies incremental modeling using UnBBayes,'' in *Proc. Encontro Nacional Inteligéncia Artif. Computacional, (ENIAC)*, 2014, pp. 266–271.
[12] F. J. Pelletier, ''The principle of semantic compositionality,'' *Topoi*, vol. 13, no. 1, pp. 11–24, Mar. 1994.
[13] F. J. Pelletier, ''Context, compositionality, and brevity,'' in *Brevity*, 1st ed. Oxford, U.K.: Oxford Univ. Press, 2013, pp. 178–197.
[14] Z. Szabó, ''The case for compositionality,'' *Oxford Handb. Compos.*, vol. 64, no. 1, p. 80, 2012.
[15] E. Agirre, T. Baldwin, and D. Martinez, ''Improving parsing and PP attachment performance with sense information,'' in *Proc. ACL HLT*, 2008, pp. 317–325.
[16] P. Dasigi, W. Ammar, C. Dyer, and E. Hovy, ''Ontology-aware token embeddings for prepositional phrase attachment,'' in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, 2017, pp. 2089–2098.
[17] N. Kim, K. Rawlins, B. van Durme, and P. Smolensky, ''Predicting the argumenthood of English prepositional phrases,'' in *Proc. 33rd AAAI Conf. Artif. Intell. (AAAI)*, vol. 33, no. 1, 2019, pp. 6578–6585.
[18] Y. Oren, O. Weisse, and A. Wool, ''A new framework for constraint-based probabilistic template side channel attacks,'' in *Cryptographic Hardware and Embedded Systems* (Lecture Notes in Computer Science), vol. 8731. Springer, 2014, pp. 17–34.
[19] T. P. Hill and J. Miller, ''How to combine independent data sets for the same quantity,'' *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 21, no. 3, Sep. 2011, Art. no. 033102.
[20] V. Hassija, V. Gupta, S. Garg, and V. Chamola, ''Traffic jam probability estimation based on blockchain and deep neural networks,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 1–10, Jun. 2020.
[21] J. C. Chappelier and M. Rajman, ''A generalized CYK algorithm for parsing stochastic CFG,'' in *Proc. 1st Workshop Tabulation Parsing Deduction (TAPD)*, 1998, pp. 133–137.
[22] M. H. Hamdan and I. H. Khan, ''An analysis of prepositional-phrase attachment disambiguation,'' *Int. J. Comput. Linguistics Res.*, vol. 9, no. 2, p. 60, Jan. 2018.
[23] S. Patnaikuni and S. Gengaje, ''Syntactico-semantic reasoning using PCFG, MEBN & PP attachment ambiguity,'' in *Proc. Int. Conf. Issues Challenges Intell. Comput. Techn. (ICICT)*, Sep. 2019, pp. 1–7.
[24] S. Matsumoto, R. N. Carvalho, M. Ladeira, P. C. G. Costa, L. L. Santos, D. Silva, M. Onishi, E. Machado, and K. Cai, ''UnBBayes: A Java framework for probabilistic models in AI,'' *Java Acad. Res.*, vol. 1, p. 34, Nov. 2011. [Online]. Available: https://www.bookdepository.com/Java-Academia-Research-Ke-Cai/9781463789459
[25] S. Patnaikuni and S. Gengaje, ''Properness and consistency of syntactico-semantic reasoning using PCFG and MEBN,'' in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Jul. 2020, pp. 554–557.
[26] D. N. Popa, J. Perez, J. Henderson, and E. Gaussier, ''Towards syntax-aware token embeddings,'' *Natural Lang. Eng.*, vol. 27, pp. 1–30, Jul. 2020.
[27] C. T. Schütze, ''PP attachment and argumenthood,'' *MIT Work. Paper Linguist.*, vol. 26, no. 95, p. 151, 1995.
[28] C. H. A. Koster, ''Affix-grammars,'' in *Algol 68 Implementation*. Springer, 1970, pp. 95–109.
[29] D. E. Knuth, ''The genesis of attribute grammars,'' in *Attribute Grammars and their Applications* (Lecture Notes in Computer Science), vol. 461. Springer, 1990, pp. 1–12.
[30] P. Bruhn and A. Geyer-Schulz, ''Genetic programming over context-free languages with linear constraints for the knapsack problem: First results,'' *Evol. Comput.*, vol. 10, no. 1, pp. 51–74, Mar. 2002.
[31] O. Levy, S. Remus, C. Biemann, and I. Dagan, ''Do supervised distributional methods really learn lexical inference relations?'' in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2015, pp. 970–976.
[32] G. Kruszewski, D. Paperno, and M. Baroni, ''Deriving Boolean structures from distributional vectors,'' *Trans. Assoc. Comput. Linguistics*, vol. 3, pp. 375–388, Dec. 2015.
[33] T. K. Landauer, D. Laham, B. Rehder, and M. E. Schreiner, ''How well can passage meaning be derived without using word order? A comparison of latent semantic analysis and humans,'' in *Proc. 19th Annu. Meeting Cognit. Sci. Soc.*, 1997, pp. 412–417.
[34] J. Mitchell and M. Lapata, ''Composition in distributional models of semantics,'' *Cognit. Sci.*, vol. 34, no. 8, pp. 1388–1429, Nov. 2010.

[35] Y. Zhou, C. Liu, and Y. Pan, ‘‘Modelling sentence pairs with tree-structured attentive encoder,’’ in *Proc. COLING 26th Int. Conf. Comput. Linguistics*, 2016, pp. 2912–2922.
[36] F. Kokkinos and A. Potamianos, ‘‘Structural attention neural networks for improved sentiment analysis,’’ in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguisticss*, 2017, pp. 586–591.
[37] M. Davydov, O. Lozynska, and V. Pasichnyk, ‘‘Partial semantic parsing of sentences by means of grammatically augmented ontology and weighted affix context-free grammar,’’ *ECONTECHMOD, Int. Q. J.*, vol. 6, no. 2, pp. 27–32, 2017.
[38] H. J. Levesque and G. Lakemeyer, *The Logic of Knowledge Bases*. Cambridge, MA, USA: MIT Press, 2001.
[39] R. J. Brachman and H. J. Levesque, *Knowledge Representation and Reasoning*. Amsterdam, The Netherlands: Elsevier, 2004.
[40] T. Berners-Lee, *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*. San Francisco, CA, USA: Harper, 1999.
[41] I. Horrocks, P. F. Patel-Schneider, and F. van Harmelen, ‘‘From SHIQ and RDF to OWL: The making of a web ontology language,’’ *J. Web Semantics*, vol. 1, no. 1, pp. 7–26, Dec. 2003.
[42] K. J. Laskey and K. B. Laskey, ‘‘Uncertainty reasoning for the world wide web: Report on the URW3-XG incubator group,’’ in *Proc. URSW*, 2008, pp. 115–123.
[43] C. G. C. Paulo, ‘‘Bayesian semantics for the semantic web,’’ in *Proc. 2nd ISWC Workshop Uncertainty Reasoning Semantic Web*, Athens, GA, USA, vol. 218, Nov. 2006. [Online]. Available: http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS//Vol-218/
[44] P. C. G. Costa, K. B. Laskey, and K. J. Laskey, ‘‘Probabilistic ontologies for efficient resource sharing in semantic web services,’’ Ph.D. dissertation, George Mason Univ., 2006. [Online]. Available: http://mars.gmu.edu/handle/1920/455
[45] G. A. AlGhamdi, K. B. Laskey, E. J. Wright, D. Barbará, and K.-C. Chang, ‘‘Modeling insider user behavior using multi-entity Bayesian network,’’ in *Proc. 10th Int. Command Control Res. Technol. Symp.*, vol. 4444, no. 703, 2008, Paper 271. [Online]. Available: http://mars.gmu.edu/handle/1920/539
[46] D. Jain, L. Mosenlechner, and M. Beetz, ‘‘Equipping robot control programs with first-order probabilistic reasoning capabilities,’’ in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2009, pp. 3626–3631.
[47] M. D. Mas, ‘‘Ontology temporal evolution for multi-entity Bayesian networks under exogenous and endogenous semantic updating,’’ Cornell Univ., Tech. Rep., Sep. 2010. [Online]. Available: https://arxiv.org/abs/1009.2084
[48] H. Bouhamed, A. Rebai, T. Lecroq, and M. Jaoua, ‘‘Dataorganization before learning multi-entity Bayesian networks structure,’’ in *Proceeding of World Academy of Science, Engineering and Technology*, vol. 78. World Academy of Science, Engineering and Technology, 2011, pp. 305–308.

SHRINIVASAN PATNAIKUNI is currently an Independent Technology Consultant and a Microsoft Certified Azure AI Engineer. He is also an AWS, Microsoft, and Google Certified Cloud Solutions Architect. He is also working as a Faculty Member with the Department of Computer Science and Engineering, Walchand Institute of Technology (WIT), Solapur. His research interests include AI/ML, cloud computing, and big data.

SACHIN GENGAJE is currently working as a Professor and the Head of the Electronics Department, Walchand Institute of Technology (WIT), Solapur. He has industrial experience with Tata Elxsi Ltd., Bangalore, as the Manager of product design services. His active research interests include artificial neural networks, fuzzy systems, image processing, embedded systems, communication systems, and speech processing.