Algorithms for Semantic Role Labeling

Abhijith Chandraprabhu

Contents

INTRODUCTION:

In our objective towards building computer systems as intelligent as us, or any step towards it, like natural language understanding applications for instance information extraction, question answering, automatic summarization, machine translation etc then a deep linguistic, mathematical and technical understanding is simply inevitable. The linguists derive motivation to improve the communication between humans using natural language. However as a computational scientist I derive technological motivation to build intelligent computer systems, without worrying too much of the linguistic aspects. In this presentation of my work on the Semantic role labeling I intend to analyze various algorithms for SRL. The task of SRL consists of, given a sentence analyzing the propositions expressed by the target verbs in the sentence. In simpler words analyzing "who" did "what" to "whom", "when" and "where" I.e to represent the full meaning of the sentences that are parsed. In particular recognizing all the constituents of the sentence and assign to them semantic roles.

Natural Language Understanding is the subset of Computational Linguistics, which aims to build a theory for natural language comprehension and productions by computer systems.

SEMANTIC ROLE LABELING:

The field of Natural Language Understanding can be considered to mainly be comprised of four stages namely a) Syntactic processing, b) Semantic Interpretation, c) Context and World Knowledge and d) Response generation.

The syntactic analysis produced by the parsers do not convey any contextual meaning of the sentence, but it is essential to be able to detect the events described in the sentence and relationships with the event participants. The task of SRL takes the output from the syntactic parser as its input and maps the result of the syntactic analysis to the underlying predicate arguments of the sentence. For example consider the processing of the sentence 'Newton fed grains to the pigeons at the park every evening' should result in the system being able to identify feeding as the event involving Newton as the agent or causer of event, grains as the item being fed, pigeons as the recipient of the item, the park as the place and every evening as the time the event occurs. Accurate interpretation of the semantic roles of the verb arguments, i.e., “Who did What to Whom?” is a crucial goal for natural language processing systems. There has been substantial progress recently in our ability to do this automatically, based to a large degree on the availability of annotated corpora. Table1. A set of roles

Role	**Description**	**Examples**
Agent	Initiator of action, capable of volition	The batter smashed the pitch into left field. The pilot landed the plane as lightly as a feather.
Patient	Affected by action, undergoes change of state	David trimmed his beard. John broke the window.
Theme	Entity moving, or being “located”	Paola threw the Frisbee. The picture hangs above the fireplace.
Experiencer	Perceives action but not in control	He tasted the delicate flavor of the baby lettuce. Chris noticed the cat slip through the partially open door.
Beneficiary	For whose benefit action is performed	He sliced me a large chunk of prime rib, and I could hardly wait to sit down to start in on it. The Smiths rented an apartment for their son.
Instrument	Intermediary/means used to perform an action	He shot the wounded buffalo with a rifle. The surgeon performed the incision with a scalpel.
Location	Place of object or action	There are some real monsters hiding in the anxiety closet.The band played on the stage.

LEXICAL RESOURCES:

The most important infrastructure for Natural Language understanding is the lexcial resources from which the systems can train from. Within the English language there are three main lexical resources which provide explicit semantic roles for use in data annotation, FrameNet, PropBank and VerbNet. The most notable difference is in the level of granularity, FrameNet is the most fine grained and the PropBank is the most generic, whereas the VerbNet is mix of the the two though closer to the PropBank.

FrameNet:

This data collection is based on the work of Filmore on Frame Semantics, where each Frame is defined with respect to frame elements, which are fine grained semantic role labels. As a simple example consider the Frame – knowledge, the frame elements could be study, school, teacher, books. It can be observed that the lexical units could not only be verbs but also adjectives and nouns.

The frame elements for a particular frame are classified into three catogories, Core, peripheral and extra-thematic. The core level frame elements are the ones which are syntactically obligatory, they are the ones that are necessary conceptually for the frame.

The peripheral level frame elements are not centered around the frame but they provide additional contextual information like time and place. Lastly the extra-thematic frame elements provide information on a broader context. In FrameNet the items are grouped into frames strictly lexically but not syntactically, hence a frame might have contain sets of verbs with related senses but differing in subcategorization properties. FrameNet mainly aims to provide detailed semantic properties of the lexical units in context.

PropBank:

The primary goal of the PropBank was to develop an annotated corpus which could be used as a training data for machine learning algorithms, unlike the FrameNet and

VerbNet which were mainly built as lexical resources. The PropBank supplies consistent, simple and general purpose labeling of semantic roles for a large quantity of text for training of automatic SRL systems. As mentioned before even though the semantic roles are generic and represented as Arg0, Arg1 they maintain consistency in annotation of same semantic roles across syntactic variations. For example in the sentence 'He broke the bottle' the bottle(syntactic subject) is annotated as Arg1 or patient and in the sentence 'the bottle broke' the same bottle is in the syntactic object position but it still is annotated as Arg1. PropBank also provides a aomprehensive list for each annotated verb all possible arguments in the predicate and their roles in all plausible syntactic realizations.

But generation of a universal set of roles covering all types of predicates is difficult, so PropBank defines roles on a verb by verb manner. For each verb the semantic arguments begin in a numbered manner from Arg0 Arg1 and so on. Generally Arg0 is the Agent (or causer) and Arg1 is patient(or theme). But for verbs with higher number of arguments it is difficult to maintain consistent generalization, PropBank provides more general ArgM – Argument Modifier roles like LOCation, EXTent, ADVerbial, TeMPoral DIRection etc which can be applied to any verb.

MACHINE LEARNING FOR SRL:

Semantic Role Labeling can be considered as classification problem, I.e for a known predicate verb and parse structure, the task is to assign from a pre-defined set correct semantic roles to the constituents with respect to the predicate verb. Like any other classification tasks it involves training the algorithm from a corpus. Basically the design of a SRL system deals with extraction of features(aspects of syntax and lexical semantics) for each constituents of the parse structure, and train the classifier to predict a semantic role to each of the constituents. Evaluation is done by measuring the number of constituents that are labeled correctly.

Identification and Classification:

Most SRL systems divide the task of labeling into identification and classification. The syntactic parsers generate parse trees recursivley, hence the number of constituents are very large. But of these large number of constituents only a small fraction is relevant to the considered predicates. Also another issue is that the constituents can be labelled with more than one role for different predicates(each predicate assigns one semantic role to its corresponding constituent). The SRL system can find all the arguments for a particular predicate only each time the tree is traversed, hence for sentences with more than one predicates the parse tree is to be traversed as many times as there are predicates. In this regard we have positive samples (constituents which are arguments for a predicate), and negative sample(constituents which are not arguments for a predicate). Usually there is a imbalance between the two which causes problems for machine learning algorithms. To work around this problem SRL systems divide the tasks into identification and classification. In identification phase a binary decision is made whether a constituent is a positive sample or a negative sample, and in classification phase the positive samples constituents are assigned a role.

Apart from this some other concerns also exist like choice of parser, impact of test data having different characteristics then training data, the choice of corpus etc.

Features used for classification:

Features are very important since it is what helps to exploit to the maximum the information provided by the syntactic parse trees which are the inputs to SRL systems. In the recent days many researchers have explored new ways of leveraging the syntactic parse trees in order to better analyze the semantic roles. In this section I provide some important features that are used which are Phrase type, Governing Category, Parse tree path, Position, Voice, Head Word, Sub Categorization, Argument set.

Phrase Type: The phrase type feature specifies the syntactic category of the corresponding phrase expressing the semantic role. The phrase type of the constituent which is being classified is an important feature since different syntactic categories tend to produce different arguments(roles). Considering the FrameNet data, the distribution of the phrase types for all the frame elements are – NP(Noun phrases) 47%, PP(prepositional phrases)22%, ADVP(Adverbial phrases) 4%, PRT(Particles) 2% and SBAR(Clauses) 2%.

Governing Category: The syntactic realization could be as a subejct or as an object, the correlation of this with the semantic roles is an important aspect. For example in the sentence 'John ate his breakfast at the office', John is the subject NP and there are also two other object Nps breakfast and office. Ideally the subject NP is qualified to take the role of the Agent. The feature which exploits this grammatical function is the Governing category, which takes two values S and VP. S corresponds to subject and VP corresponds to object of verbs. This feature is restricted only to NP. This feature can be accessed by traversing up the parse tree from child to parent link from the constituent until an S or VP is encountered, and the value of the feature is assigned to what ever is encountered. Usually NP nodes corresponding to S are grammatical subjects and objects otherwise.

Parse Tree Paths: It is important to capture the syntactic relation of the constituent with rest of the sentence, apart from getting the catogory. This is achieved by the Parse tree path feature. The parse tree path feature gives the relationship between the predicate and any constituent in question. Essentially it is a path from the target word(predicate) to the constituent in question through the tree. (To write in more detail.)

Position : There are concerns about the accuracy of the parse trees, in order to overcome this Gildea and Jurafsky(2002) use the position of the constituent relative to the predicate as a feature called as position. The position feature gives tells if the constituents to be labeled occur before the predicate or after the predicate in the sentence. Previously it was mentioned about the grammatical function, and this feature has some correlation with it, since the grammatical subjects usually occur before the predicate and the grammatical objects occur after the predicate.

Voice : The voice feature identifies the disticinction between active and passive verbs. The interesting relation is that the objects of active words correspond to subjects of passive verbs in a semantic role. (Not clear, needs more explanation.)

Head Word :

The head word feature is important since the constituents with certain head words are more likely to be certain types of arguments. Head words of noun phrases can be used to restrict the semantic types of roles. For example for a communication frame, the noun phrases in which the head words are he, John, father are likely to be associated with the role of speaker, but for head words like story, question or proposal the role is more likely to be the Topic.

Subcategorization : This feature deals with the syntactic arguments of a verb, foe example consider the two sentences, 'The door closed' and 'He closed the door'. The subcatogorization of the first verb is {subject}, and the second one it is {subject,object}.

Argument Set: This is the set of all possible roles appearing for a verb in a given sentence, its also called Frame element group.

MACHINE LEARNING METHODS :

SRL is a challenging task and in order to attain high performance a combination of manual classifiers and machine learned classification is to be used. The learning systems need to be supplemented with sufficient linguistic information. Gildea and Jurafsky in the early stages used a backoff lattice, where the probabilities of a label is calculated given the set of features. But this way it was not possible to scale the number of features, later works which are more based on machine learning algorithms could make this possible. Fleischman et al use a log-linear classifier(Maximum entropy) which shows an improvement of 3.2% from Gildea & Jurafsky when using the same features. However with the Maximum entropy classifier it was possible to take advantage of more number of features, which further improved the system by 3%.

Some other works are of Surdeanu et el(2003) which used decision tree for SRL, Support Vector Machine based methods of Pradhan et al(2005). The functioning of the SRL mostly confirms with the general findings of machine learning techniques. The discriminative approaches can accommodate higher number of features than the direct methods like lattice backoff and decision trees. This is usually because in the direct methods suffer from sparsity in data as the data is partitioned on combinations of features.

Feature Combination:

The distinctive nature of SRL tasks using machine learning is how the features are used in ideal combinations either implicitly or explicitly. SVMs achieve slightly better performance than maximum entropy since it inherently considers feature combinations, at the cost of higher training time. But works by Xue and Palmer show that similar performance can be achieved by maximum entropy models by providing linguisitc information through feature combinations, and this takes lesser training time.

Some of the useful feature combinations are,

Predicate-phrase type combination: This is the combination of the predicate lemma and the type of the phrase of the constituent. This is useful since given predicate, the phrase type can be predictive of the semantic role of the constituent.

Predicate-Head word combination: This is the combination of the predicate lemma and the head word of the constituent of a feature. The head word is less informative when considered independently of the predicate.

Voice-Position combination: Here the voice of the predicate is combined with the position of the constituent. For example, in the feature combination "before-passive" which tries to captures the information as subject of a passive sentence, but it would give a different semantic role than the subject if it an active sentence.

THE ALGORITHMS:

This is a brief introduction to the most important works on Semantic Role Labeling. The task of automatic semantic role labeling is divided into two stages, first is role identification and classification of constituents, and second stage which is a combination of identification and classification in order to find the best overall role labeling of the constituents in the sentence. Gildea and Jurafsky(2002) divide the overall process as three different tasks,

1). Argument Identification – In this task all the constituents which are eligible to be represented as semantic arguments for a given predicate of a sentence are identified. The input is the parse tree, the nodes of which (with respect to a given predicate) are

classified as NON-NULL node(one that represents a semantic argument) and a NULL node(which does not represent any semantic argument).

2). Argument Classification – The appropriate arguments are assigned to the corresponding constituents.

3). Argument Identification and Classification – A combination of the first two tasks which is done on an overall level. This is usually referred to a joint inference.

The literature presented till here provide majority of the basic knowledge needed to build automatic semantic role labelers. Hence forth we present the overview for some of the most prominent algorithms which are Automatic Labeling of Semantic Roles (Daneil Gildea and Daniel Jurafsky,2002), A Global Joint Model for Semantic Role Labeling(Kristina Toutanova et el), Shallow Semantic Parsing using SVM(Sameer Pradhan et al) and Semantic Role Labeling via Integer Linear Programming(Vasin Punyakanok et al).

Automatic Labeling of Semantic Roles :

This system basically is for identifying the semantic relationships of the constituents in a sentence withing a semantic frame. The classification if based on statistical techniques and the FrameNet data is used fro training. All the sentences in the training data were parsed into a syntactic tree, features were extracted from the parse tree structure both lexical and syntactic. These features are than combined with the knowledge of the predicate verb and also with the information like prior probabilities of semantic role combinations. Here mainly two levels of classification is done, first it is the lexical clustering to generalize across possible role fillers. In the test phase the test sentences are parsed and annotated with the features, than the final classification is done.

Def: *A frame is a schematic representation of situations involving various participants, props, and other conceptual roles(Fillmore,1976).* FrameNet roles are defined for each semantic frame. Consider an example frame CONVERSATION is invoked by verbs like argue, chat, talk etc and also by nouns like dispute, discussion etc. The possible

roles for this frame would be PROTAGONIST (for the participants in the conversation), MEDIUM and TOPIC. These roles that are assigned to the frame are also referred to as frame elements.

Probability estimation of Roles:

A number of features at both the sentence level and the constituent level are used to calculate the probabilities for predicting frame element labels. The system consists of a statistical classifier trained by initially by an automatic syntactic parser which includes 36,995 sentences, then matching the frame elements(roles) to parse constituents. Next in the testing phase the sentences are parsed and the same features as in the training phase are extracted. Using these features the probabilities for the roles are calculated.

In the experiments 10 % of the annotated sentences were used as test set and another 10% was used as tuning set. In their used corpus the average number of sentences per target word was 34, and number of sentences per frame was 732.

To do the automatic semantic role labeling a probability distribution is estimated for the likelihood of the constituent to fill each role, as shown below.

$$P(r \vee h, pt, gov, position, voice, t) = \frac{Number\ of\ (r, h, gov, position, voice, t)}{Number\ of\ (h, pt, gov, position, voice, t)}$$

The inputs were the features and the predicate. Where the numerator is the the number of times every roles occur for a particular combination of features and the denominator is the total number of times this feature combination occurs. Note that these readings are from the test data.

Identification of the frame boundaries:

This work of Gildea & Jurafsky produced the possible sequences of role labels by combining the most likely labels for each constituent. They considered two kinds of probabilities, firstly the probability seen above I.e for individual constituent. The second is the probability for a set of roles for a given sentence predicate.

This approach is nothing but the reranking but is limited by the frequency based probability distributions.

Toutanova et al(2005) have given another reranking system trained from PropBank data. Here the strong dependencies between the arguments is taken into consideration and this is the reason it is the Global Joint Model for Semantic Role Labeling. *Even though previous work has modeled some correlations between the labels of parse tree nodes (see Section 2), many important phenomena have not been modeled.*

The key properties needed to model this joint structure are: (1) no finite Markov horizon assumption for dependencies among node labels, (2) features looking at the labels of multiple argument nodes and internal features of these nodes, and (3) a statistical model capable of incorporating these long-distance dependencies and generalizing well.(2)

In this work they show how to build a joint model using discriminative log-linear model incorporated with the features. They claim the error reduction to be 24.1% and 36.8% on ALL and CORE arguments respectively.

They propose a graphical model over a set of m variables(one each for the nodes in the parse tree), these variables denote the labels of the nodes and its dependencies between each other. The statistical tendency of some semantic roles to occur at most once they say that there must be some kind of link between any two variables. The probability that a node gets a role ex AGENT, is estimated by knowing is there are any other nodes which are labeled with this same role.

In the next aproach which is by Pradhan et al they propose a machine learning algorithm for shallow semantic parsing, . Their algorithm is based on Support Vector Machines which replaced the statistical classification algorithms used previously. The experiments in this work are performed on the PropBank data. They divide the task of shallow semantic parsing into three different tasks -

Argument Identification – To identify the parsed constituents in a sentence which represents the semantic arguments.

Argument classification – Assign appropriate argument labels to the constituents which are known to represent a predicate.

And a third task which is the combination of the two above tasks.

The first thing given a parse tree each nodes can be classified as one that represents a semantic argument and one which does not. They are called NON-NULL nodes and NULL nodes. Then the NON-NULL nodes are classified onto a set of argument labels.

They use some features from the works of Gildea & Jurafsky namely Predicate, Path, Phrase type, Position, Voice, Head word, and subcategorization.

Apart from these features they also use some new features like Named Entities in Constituents, Head word POS, Verb Clustering, Partial Path, Verb Sense Information, Head Word of Prepositional Phrases, First and Last Word/POS in Constituent.

Integer Linear Programming – In this work Punyakanok et al. Use the framework of an Integer Linear Programming (ILP) problem.They model the sentence level constraints on role labels the framework which can then be solved with a general-purpose ILP algorithm. An ILP problem is to maximize an objective function bound by certain constraints, the objective function and the constraint must be linear functions. It can be written in matrix form as

$\max_x c^T x$ subject to constraints $Ax < 0 \wedge Bx = 0$

Since it is ILP the x values are restricted to integers only. These problems are NP-complete in general. In order to apply ILP to the problem of SRL the have a binary variable $x_{ir} \in 0,1$ where I is the constituent and r is the possible role. The vector c which is the objective function consists of scores for every label from a local classifier. They introduce the following constraints to ensure that the constituents have only one label in the final result: $\forall i \quad \sum x_{ir} = 1$, the summation is over r.

They also specify a condition such that each role appears at max one time in a sentence by $\forall r \quad \sum x_{ir} \leqslant 1$, summation over I.

Discussion :

The reranking approaches discussed have drawbacks due to the large fixed number of parameters that are passed from the initial stage to the second stage which is the reranking stage. The SRL systems of Gildea and Jurafsky achieves 82% of accuracy in identifying the semantic roles for pre-segmented constituents. They draw comparisons between automatically derived and hand built semantic sources. However their observation is that use of probabilistic models to integrate the syntactic parsing and semantic interpretation the gains were not too high. Their system does not consider frame disambiguation ,i.e they assume knowledge of the correct frame to the target word, it is not a generic way.

The work of Toutanova et al claim to have achieved gains of 24.1% error reduction on all arguments and 36.8% for core arguments. The main high point of this work is the joint modeling of the argument frames of verbs. The two types of features used were, firstly the features of complete sequence of the arguments labels, and secondly the features modeling dependencies between argument labels and syntactic features of other arguments.

Punyakanok et al claim to achieve competetive results using the approach of shallow parsing rather than complete syntactic tree parses as used by others but under similar conditions.

BIBLIOGRAPHY:

1. 1. Automatic Labeling of Semantic Roles, Daniel Gildea Daniel Jurafsky

   Proceedings of the 38th Annual Conference of the Association for Computational Linguistics , 2000.

2. A Global Joint Model for Semantic Role Labeling

   Kristina Toutanova Aria Haghighi Christopher D. Manning

   Computational Linguistics Volume 34 Issue 2 2008

3. The Importance of Syntactic Parsing and Inference in Semantic Role Labeling

   Vasin Punyakanok Dan Roth Wen-tau Yih

   Computational Linguistics Volume 34 Issue 2 2008

4. Towards Robust Semantic Role Labeling

   Sameer S. Pradhan Wayne Ward James H. Martin

   Computational Linguistics Volume 34 Issue 2, June 2008

5. Semantic Role Labeling via Integer Linear Programming Inference

   Vasin Punyakanok Dan Roth Wen-tau Yih Dav Zimak

   COLING '04 Proceedings of the 20th international conference on Computational Linguistics

6. Shallow Semantic Parsing using Support Vector Machines

   Sameer Pradhan et al.