ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 7, January 2016

Survey of WordSense Disambiguation Approaches To Identify Sense of Context

Dhivya.A
Department of Information Technology
Kongu Engineering College
Perundurai,Erode
mail-id:dhivyaarumugam02@gmail.com

RajaDevi.R
Department of Information Technology
Kongu Engineering College
Perundurai,Erode
mail-id: rajdevi14@gmail.com

***Abstract*- Word Sense Disambiguation is the difficult problem in Natural language processing. Word sense disambiguation finds the correct sense of the word in a sentence or the query has multiple meanings. WSD is used in question answering, machine translation, text summarization, text classification and information retrieval. Some approaches and algorithm are available for WSD .**

***Keywords*- Natural Language processing, Machine Translation, Word Sense Disambiguation, Information retrieval, Text classification**

I. INTRODUCTION

There are many words in Natural languages which have different meaning for different context those words are known as polysemous words. Words will have multiple meaning. Word sense disambiguation (WSD) is used to find the correct context of the given word. Context is the text or words which are ambiguous word. Using the context, human can sense the correct meaning of the word in that context. Computer need to follow some rules using which the system can estimate the absolute meaning out of multiple meanings of the word. The various methods have been proposed like dictionary-based methods that use knowledge in lexical resources, supervised machine learning method works on classifiers and unsupervised learning method supports clusters. Precision and recall are two measures of performance for WSD.

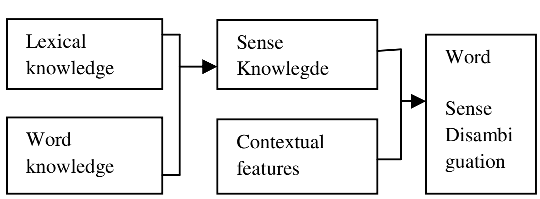

Fig1.Conceptual model for Word Sense Disambiguation

Precision is the amount of relevant information retrieved to the total amount of information retrieved, while recall is the amount of relevant information retrieved to the total amount of relevant information. Lexical Knowledge is expressed in words.Lexical Knowledge includes resources such as machine readable dictionaries and encyclopedia.WSD is key for lexical knowledge and word knowledge.Word sense disambiguation include words and word knowledge to identify dictionaries. Sense knowledge can be denoted by a vector(sense ID, features). The contextual features consider the unigrams and bigrams .contextual features includes parts-of-speech and lemmas .

II. APPROACHES

Some approaches are applied to the problem of Word Sense Disambiguation(WSD).

- Supervised approach
- Unsupervised approach
- Overlap approach
- Knowledge approach

III. SUPERVISED APPROACH

The supervised approaches applied to Word Sense Disambiguation systems use machine-learning technique from manually created sense-annotated data. [4]
Training set will be used for classification to learn and this training set consist examples related to target word. Basically this WSD algorithm gives better result than other approaches. Supervised approaches are as follows

- Decision Tree
- Neural Networks
- Navie Bayes
- Decision List

1. *Decision Tree*

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 7, January 2016

A decision tree divide the training data in a recursive method and represent the rules for classification in a tree structure. The internal nodes represents test on the features and each branch show how the decision is being made and the leaf node refers to the prediction. It is frequently regarded as a prediction tool. Algorithms for learning decision trees are ID3 and C4.5.[5]

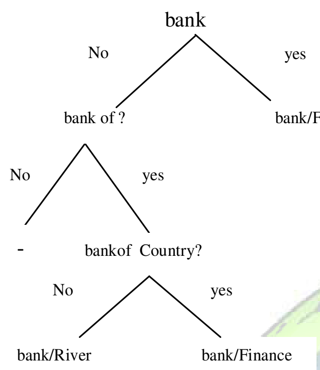

Fig 2. A example of a Decision tree

On comparing with other machine learning algorithms it was found that some supervised approaches perform better than the decision tree obtained C4.5 algorithm. An example of decision tree for WSD is describe in the Figure 2. Noun sense of the ambiguous word "bank" is classified in the sentence, "I will be at the bank of Narmada River in the afternoon". In the Figure 1, the tree is created and traversed and the selection of sense bank/RIVER is made[6]. Empty value of leaf node says that no selection is available for that feature value.

2. *Neural Networks*

Neural networks process the information based on computational model of connection approach[5].The training dataset is partition into non-overlapping sets based on desired responses. Learning in neural networks is eventually updating of weights. The network can have weights both positive and negative corresponding to correct or wrong sense choice.

When the network finds new input pairs the weights are adjusted so that the output unit giving the target output has the larger activation. The input can easily be propagate from input layer to output layer through the intermediate layer. It is difficult to compute a clear output from a network connections are extend in all directions and form the loops.

3. *Navie Bayes*

Naive Bayes based on Bayes theorem[5]. Navie Bayes algorithm is one of the supervised approach to find the sense using

$$s^{\wedge} = \operatorname{argmax}_{s\epsilon sense} \Pr(S|V_w)$$

V_w is a feature vector of

- Part of Speech of Word(W).
- Semantic and Syntactic features of Word(W).
- Collocation Vector(set of words around it) → typically consist of next word (+1),next to next word(+2),-2,-1 and their part –of –speech.
- Cooccurence Vector(number of times w occur in bag of words)

Apply bayes rule and navie independence as

$$S^{\wedge} = \operatorname{argmax}_{s\epsilon senses} \Pr(s)\pi_{i=1}^{n} \Pr(V_w^i|S)$$

Example: I went to the bank to withdraw some money

Withdraw(clue1),money(clue2) based on the clues bank is disambiguated.

$S^{\wedge}$is a target word.V is a feature vector is constructed based on W.

$\overline{V}_{bank}$= feature vector

<I, went, to, the, bank, to, withdraw, some, money>

a) *Part-of-speech(bank)-noun*

b) *Bank has both syntactic and semantic features*

- Syntactic (grammar)features→ take's for plural, ends with consonant, noun.
- Semantic features (bank)→ place, organization, has value, concrete noun.

c) *Collocation Vector*

Collocation vector will be important in Natural Language Processing.

Ex: I went to the bank to withdraw some money

- Ignore the function words such as preposition, conjunction etc.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 7, January 2016

- Pickup content words :+1,+2,-2,-1

- <I, went, withdraw. money>
- Here the collocation words are:
- I→2,went→1,bank,withdraw→+1,money→+2.

d) Cooccurence Vector

Cooccurence Vector contains Window.

- Ex: I went to the bank to withdraw some money.
- Window has two words before the "bank" → (to, the) and two words after the "bank"→(to, withdraw).
- Bank appears one time in window.

Bayes Rule and independence Assumption

$$s^{\wedge} = \operatorname{argmax}_{s\in sense} \Pr(S|V_w)$$

Apply bayes rule

$$\Pr(S|V_w) = \Pr(S).\Pr(V_w|s)/\Pr(V_w)$$

$P(\overline{V}_{bank}|bank)$=P(<Noun(N),organisation,went,withdraw,money>/bank);

$\Pr(S|V_w)$ can be approximated by independence assumption:

$\Pr(V_w \mid S) =$
$\Pr(V_w^1 \mid S),\Pr(V_w^2 \mid S, V_w^1)......\Pr(V_w^n \mid S, V_w^1, \dots\dots, V_w^{n-1})$

$\Pr(V_w \mid S) = \pi_{i=1}^{n} \Pr(V_w^i|S)$

Then,

$P(\overline{V}_{bank}|bank)$=P(N/bank),P(organisation/bank), P(money/bank),P(withdraw/bank),P(went/bank).

$P(\overline{V}_{bank}|sense(bank))$=P(N/ sense(bank)),P(organisation/ sense(bank)),P(money/ sense(bank)),

P(withdraw/ sense(bank)),P(went/ sense(bank)).

$$P\left(\overline{V}_{bank}\middle|sense(bank)\right) = P(N|Sense_{bank}^1), P(organisation|Sense_{bank}^1), P(money|Sense_{bank}^1), P(withdraw|Sense_{bank}^1), P(went \mid Sense_{bank}^1)$$

Sense are marked with respect to the WordNet

$$Sense_{bank}^1 = \text{Finance bank}$$

$$Sense_{bank}^2 = \text{River bank}$$

$$P(organisation|Sense_{bank}^1) \gg P(organisation|Sense_{bank}^2)$$

"Organisation → related to Finance"

Thus

$$S^{\wedge} = \operatorname{argmax}_{s\in senses} \Pr(s)\,\pi_{i=1}^{n} \Pr(V_w^i|S)$$

Pr(S)=count(S,W)/count(w)

$$\Pr(V_w^i|S) = \Pr(V_w^i|S)/\Pr(S)$$

$\Pr(V_w^i|S) = C(V_w^i|S)/C(S,W)$

W→ word, S →Finance sense, V_w^i →Organisation so count is high.
W→ word,S→Riversense, V_w^i →Organisation so count is Zero.

4. *Decision List*

DecisionList is based on the collocation property. Nearby words provide strong and consistent clue as to the sense of a target word. Once the features are get from the corpus, rules of the form (feature value, sense, score) are generated. These rules are embed into a table, one record for each sense. Table is sorted in descending order of scores. The resulting data structure, *i.e.*, the sorted table is decision list.

$$S = \operatorname{argmax}_{s\in senses_D(w)} Score(S_i)$$

$$score(S_i) = \max_f \log\left(\frac{P(S_i|f)}{P(S_j|f)}\right)$$

f →collocation(feature value), S_i → Sense A, S_j → SenseB . Collect a large set of collocation for ambiguous word. If Pr(sense B) is less than one , then the log number will be negative otherwise it will be positive. Higher the probability is equal to more Predictive evidence .Collocation are ordered in a decision list, with most predictive collocation ranked highest.

Advantage:

- Supervised approach find correct context using the classifiers.

- This supervised approach perform better when comparing with other approach.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 7, January 2016

IV. UNSUPERVISED APPROACH

Unsupervised approaches do not need manual training data because it uses online training data. So unsupervised algorithm is expensive. Unsupervised algorithm can disambiguate word sense accurately in unannotated corpus. Unsupervised approaches are

- Hyperlex
- Context clustering
- Word Clustering

1. Hyperlex

The words that co-occurred with the target word is represented as nodes in the graph. Hyperlex extracting the sense from the corpus instead using of using the dictionary defined senses.[7]

This hyperlex is based on a corpus. The words that co-occur with the target word. Then the edge will connect the nodes. Hubs represent the senses. The target word becomes the vertex.

$$W_{i,j} = 1 - \max\{P(W_i|W_j), P(W_j|W_i)\}$$

Weights are given to each nodes. Minimum Spanning Tree find the distance between the word in the context and for the particular component.[7]

$$P(W_i|W_j) = \frac{\text{frequency of cooccurence of words } W_i \text{ and} W_j}{\text{frequency of occurence of } W_j}$$

Each node in the minimum spanning tree is assigned a score vector with many dimension and there components.

Advantage

- Hyperlex require any tagged data.
- This hyperlex make use of small world structure of co-occurrence graph.

2. Context Clustering

Context Clustering is one of the clustering techniques in which context vectors are generated and grouped into clusters to find the meaning of the word.

Context cluster use word space as vector space and dimensions are words. Word in a corpus will be represented as vector and how many times the word occurred will be count within its context. Co-occurrence matrix is generated and similarity measures are applied.[5]

3. Word Clustering

Word Clustering methods cluster the words which are semantically related and thus express a specific meaning[5]. A known approach to word clustering is Lin approach consists of words $W=(w_{1,......}w_k)$ similar to a target word .

The similarity between the w_0 and w_i is firmed based on the information content of their single features. If their is more dependencies between two words the information content is high.

Disadvantage:

- The disadvantage of unsupervised algorithm is that it cannot exploit any dictionary.
- Unsupervised approach does not relay on the inventory of senses.

V. OVERLAP APPROACH

Overlap Based Approach requires Machine Readable Dictionary(MRD).Finding the overlap between the features of context bags and features of sense bag.

- Features such as Sense definition, hypernyms, hyponyms.
- Sense which have maximum overlap is selected as appropriate senses.
- Context Bag contains the word in the definition of each sense of the context word.

Sense Bag contains the word in the definition of each sense of the ambiguous word.

Overlap based approach

- Lesk algorithm
- Extended Lesk's Algorithm

1. Lesk Algorithm

Disambiguating the word by comparing the gloss of each of its senses to the glosses of every other word.[8]

Example "On burning coal we get ash"(ash) NOUN:

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 7, January 2016

- Sense1:

ash→ The residue that remains when something is burned.

- Sense2:

ash, ashtree→ any of various deciduous pinnate-leaved ornamental or timber trees of the genus Fraxinus.

- Sense3:

ash→ strong elastic wood of any of various ash trees, used for furniture tool handles and sporting goods such as baseball bats.

- burned→ overlap with→ burning.
- So Sense1 indicates the correct sense.
- No other sense are not overlap.
- Here the context clues is burning and coal.
- Context**:** on burning the ash we found that the root has deep into the ground.
- Tree is meronym to the root.
- And the context satisfy the Sense2 of ash(noun).

2. Extended Lesk's Algorithm

- Original algorithm is sensitive towards exact words in the definition

.

- It includes glosses of semantically related senses from WordNet (hypernyms, hyponyms, etc).[2]
- Scoring Function:

$$\mathrm{Score}(S) = \sum \left|\mathrm{context}(W) \cap \mathrm{gloss}\left(S^{|}\right)\right| S = S^{|}$$

- Gloss(S) :S is from the Lexical resources.
- Context(W):gloss of each sense of each context word.

Example:"On combustion of coal we get ash"

Ash has three senses from the WordNet.

- Sense1:

ash→ The residue that remains when something is burned.

- Sense2:

ash, ashtree→ any of various deciduous pinnate-leaved ornamental or timber trees of the genus Fraxinus.

- Sense3:

- ash→ strong elastic wood of any of various ash trees, used for furniture tool handles and sporting goods such as baseball bats.
- Here the combustion will not present so the count value will be zero.
- Fly ash→ fine solid particle of ash that are carried into the air when fuel is combusted.
- Bone ash→ ash left when bones burn;

High is calcium phosphate; used as fertilizer and in bone china.

- Combustion and combusted are overlapped. Combustion is used as clue.
- So sense2(ash) and Fly ash are overlapped based on hyponymy.
- Ash has hyponymy(fly ash, bone ash).

Advantage:

- In Overlap based approach the overlap will be based on the dictionary definition and WordNet.
- Sense with maximum overlap is known as correct sense.

.

VI. KNOWLEDGE BASED APPROACH

Knowledge Based is lexical resources and it is the component of word sense disambiguation (WSD).Knowledge based resources have Machine Readable Dictionary(MRD). Knowledge Based approach

- SSI algorithm

SSI Algorithm

Structural semantic Interconnection algorithm is to disambiguate polysemous words by structural specifications of senses of each word and selecting the correct sense using structural grammar.[3]

- T→ [t_1, t_2, t_3,...... t_n] t is list of co-occurring terms to be disambiguated and n is total number of noun types (word)
- $S_1^t, S_2^t \ldots \ldots \ldots \ldots . . S_k^t$ are structural specifications of the possible concepts for the given t.
- I→ $[S^{t_1}, \ldots., S^{t_n}]$ is a list of disambiguate senses.
- P→[t_i $|S^{t_i}$ = null],p is list of pending terms.
- G=(E,N,S_G,P_G) here G is Context Free Grammar(CFG),
- E is edge labels to point out semantic relation between senses.

The WordNet definition of the t is an monosemous word.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 7, January 2016

Disadvantage:

- Knowledge based approach has low performance when compared to the supervised approach.

VII. CONCLUSION

Supervised approach performs better than the unsupervised approach, Overlap based approach and knowledge based approach. Supervised word sense disambiguation use machine learning techniques. There are different classifiers of supervised approach to classify an appropriate sense to instance of a single word.

REFERENCES

[1] Alok Ranjan Pal and Diganta Sahaz, “Word Sense Disambiguation: A Survey”, International Journal of Control Theory and Computer Modeling (IJCTCM) Vol.5, No.3, pp 56-82, 2015.

[2] Jagdeep Kaur, “Word Sense Disambiguation(WSD)”,International Journal For Technological Research In Engineering,Vol.1, No. 5, pp. 2347 – 4718, 2014.

[3] Navigli R and Velardi, P, “Structural semantic interconnections: A knowledge based approach to word sense disambiguation”, IEEE Transactions on pattern analysis and machine intelligence ,Vol. 27,No 3 ,pp. 1075–1086,2005.

[4] Nirali Patel and Bhargesh Patel and Rajvi Parikh and Brijesh Bhatt, “Word Sense Disambiguation (WSD)”, International Journal of Advance Foundation and Research in Computer (IJAFRC),Vol 2, pp.2348 – 4853.2015.

[5] Pranjal Protim Borah and Gitimoni Talukdar and Arup Baruah, “ Approaches for Word Sense Disambiguation – A Survey”, International Journal of Recent Technology and Engineering (IJRTE), Vol-3, No.1, pp. 2277-3878,2014.

[6] Roberto Navigli, “Word Sense Disambiguation: A Survey”, ACM Computing Surveys, Vol. 41, No. 2, pp.1-68,2009

[7] Shallu and Vishal Gupta, “A Survey of Word-sense Disambiguation Effective Techniques and Methods for Indian Languages”, Journal Of Emerging Technologies in web intelligence, Vol. 5, No. 4,pp.354-360,2013.

[8] Sreedhar J, Viswanadha Raju S, Vinaya Babu A, Amjan Shaik, Pavan Kumar P, “Word Sense Disambiguation: An Empirical Survey”, International Journal of Soft Computing and Engineering (IJSCE) , Vol.2, No.2, pp.2231-2307,2012.