

|  |  |
| --- | --- |
| **ASSIGNMENT** | |
| **Module Code** | MIS503 |
| **Module Name** | Probabilistic Graphical Models |
| **Course** | M.Tech. |
| **Department** | Computer Science |
| **Faculty** | FET |

#### 

|  |  |
| --- | --- |
| **Name of the Student :** | Vishnu Prasad P |
| **Reg. No :** | 17ETCS075004 |
| **Batch :** | Full-Time 2017 |
| **Module Leader :** | Dr. Hariharan Ramasangu |

**Ramaiah University of Applied Sciences University House, Gnanagangothri Campus, New BEL Road, M S R Nagar, Bangalore, Karnataka, INDIA - 560 054**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Declaration Sheet** | | | | | | | | |
| Student Name | Vishnu Prasad P | | | | | | | |
| Reg. No | 17ETCS075004 | | | | | | | |
| Course | M-Tech in Machine Learning and Intelligent Systems | | | | | Batch | Full-Time 2017 | |
| Module Code | MIS503 | | | | | | | |
| Module Title | Probabilistic Graphical Models | | | | | | | |
| Module Date | 04/06/2018 | | to | | 07/07/2018 | | | |
| Module Leader | Dr. Hariharan Ramasangu | | | | | | | |
| **Declaration**  The assignment submitted herewith is a result of my own investigations and that I have conformed to the guidelines against plagiarism as laid out in the Student Handbook. All sections of the text and results, which have been obtained from other sources, are fully referenced. I understand that cheating and plagiarism constitute a breach of University regulations and will be dealt with accordingly. | | | | | | | | |
| Signature of the student | |  | | | | | Date | 07/07/2018 |
| Submission date stamp  (by Examination & Assessment Section) | |  | | | | | | |
| Signature of the Module Leader and date | | | | Signature of Reviewer and date | | | | |
|  | | | |  | | | | |

A**bstract**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The different structural graphical model are looked in perspective of generative and discriminative model. These study is done with a application of this models in a practical problem solving application of natural language processing.Both the study of solution using both the models are analysed in detail and the stance of generative models is taken as it perofrms very well that discriminative in the larger data set as the problem of NLP also lies within the Large Data to be handeled.

The story is choosed and firstly the story is tokenised according to word or sentence for the further processing Then syntactic analysis is done with the help of POS tagger and the important words are taken out followed by the semantic Role labelling and all the nodes are identified and used for modelling the system.

The modelling is done from the features of extracted from the prepossessing and the inference is done using the probalistic graphical models python library.

# **Contents**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

[**Declaration Sheet** ii](#_Toc518720180)

[**Contents** iv](#_Toc518720181)

[**List of Tables** vi](#_Toc518720182)

[**List of Figures** 7](#_Toc518720183)

[**PART-A CHAPTER 1** 8](#_Toc518720184)

[**Generative model is better than discriminative model in structured classification problems** 8](#_Toc518720185)

[1.1 Identification of a specific application regarding a structured classification problem 8](#_Toc518720186)

[1.2 Need for models in structured classification in context of the Natural Language processing 8](#_Toc518720187)

[1.3 Development of a block diagram for steps in structural classification with functional description of input and output at each stage 8](#_Toc518720188)

[1.4 Advantages and limitations of Generative models 8](#_Toc518720189)

[1.5 Advantages and limitations of Discriminative models 9](#_Toc518720190)

[1.6 Your stance and justification 9](#_Toc518720191)

[**PART-B CHAPTER 2** 11](#_Toc518720192)

[**Develop an algorithm to extract syntactic cues and semantic labels from a narrative text such that they form an input to knowledge representation module.** 11](#_Toc518720193)

[2.1 Tokenization of the narrative text 11](#_Toc518720194)

[2.2 Develop an algorithm to extract the syntactic cues from the text. 12](#_Toc518720195)

[2.3 Formation of a semantic label table 13](#_Toc518720196)

[2.4 Development of an algorithm to extract the semantic labels 13](#_Toc518720197)

[2.5 Development of an algorithm for coreference resolution for the extracted syntactic and semantic information 14](#_Toc518720198)

[2.6 Feature identification by integrating syntactic, semantic, and coreference information 15](#_Toc518720199)

[**PART-C CHAPTER 3** 16](#_Toc518720200)

[**Develop an algorithm for action description of a narrative text using Bayesian network** 16](#_Toc518720201)

[3.1 Development of a knowledge representation algorithm using the features identified in Part-B. 16](#_Toc518720202)

[3.2 Identification of threshold patterns for joint probability distribution 16](#_Toc518720203)

[3.3 Development of a Bayesian network model 16](#_Toc518720204)

[3.4 Development of a reasoning scheme for inferring action description 17](#_Toc518720205)

[3.5 Analysis of results and comment on the influence of representation, inference, and learning algorithm 18](#_Toc518720206)

[**References** 22](#_Toc518720207)

# **List of Tables**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Table No. Title of the table Page. No.**

Table 2.1‑1:Output from the syntactic extractor 12

Table 2.4‑1:Output of the semantic Extractor 14

# **List of Figures**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Figure No. Title of the figure Page. No.**

Figure 0‑1:Block Diagram of structural classification of NLP 8

Figure 2.5‑1:Output of the coreference Coefficient of the first two sentences. 14

Figure 3.3‑1:Bayesian Model of the before conversation of the story 17

Figure 3.3‑2:Bayesian model of the while conversation of the story 18

# **PART-A** **CHAPTER 1**

# **Generative model is better than discriminative model in structured classification problems**

## Identification of a specific application regarding a structured classification problem

Structured Classification Problem in **Natural Language Processing.**

**“Natural language processing** (**NLP**) is an area of [computer science](https://en.wikipedia.org/wiki/Computer_science) and [artificial intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence) concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyse large amounts of [natural language](https://en.wikipedia.org/wiki/Natural_language) data.**”**

## Need for models in structured classification in context of the Natural Language processing

Models Enables us to extract the unseen information from the input and generalization of the procedure and also,

* Probability estimates can be smoothed to accommodate unseen events
* Redundancy in language supports effective statistical inference procedures \ the stimulus is richer than it might seem.
* Statistical learning theory: generalization ability of a model class can be measured independently of model representation.

## Development of a block diagram for steps in structural classification with functional description of input and output at each stage

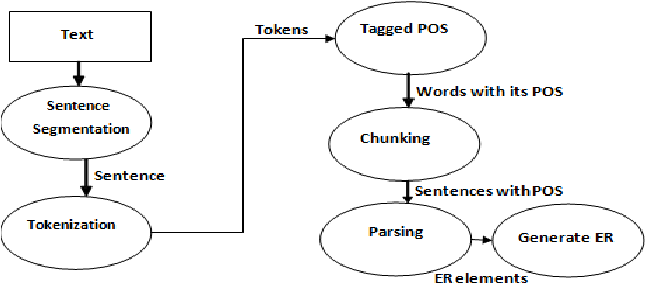


Figure ‑:Block Diagram of structural classification of NLP

## Advantages and limitations of Generative models

In the generative algorithms have discriminative properties, since you can get P(Y|X) once you have P(X|Y) and P(Y) (by Bayes’ Theorem), though discriminative algorithms don’t really have generative properties.

Generative algorithms make some kind of structure assumptions on your model (For example, Naive Bayes assumes conditional independence of your features).Generative models are typically specified as probabilistic graphical models, which offer rich representations of the independence relations in the dataset.

* Relax assuming conditional independence of observed data given the labels
* Can contain arbitrary feature functions

Generative models often outperform discriminative models on smaller datasets because their generative assumptions place some structure on your model that prevent overfitting. For example, let’s consider Naive Bayes vs. Logistic Regression. The Naive Bayes assumption is of course rarely satisfied, so logistic regression will tend to outperform Naive Bayes as your dataset grows (since it can capture dependencies that Naive Bayes can’t).

## Advantages and limitations of Discriminative models

SVMs and decision trees are discriminative models because they learn explicit boundaties between classes. SVM is a maximal margin classifier, meaning that it learns a decision boundary that maximizes the distance between samples of the two classes, given a kernel. The distance between a sample and the learned decision boundary can be used to make the SVM a “soft” classifier. DTs learn the decision boundary by recursively partitioning the space in a manner that maximizes the information gain (or another criterion).

Discriminative models do not generally function for outlier detection, though generative models generally do. What’s best for a specific application should, of course, be evaluated based on the application.

Discriminative models do not offer clear representations of relations between features and classes in the dataset. Instead of using resources to fully model each class, they focus on richly modeling the boundary between classes. Given the same amount of capacity (say, bits in a computer program executing the model), a discriminative model thus may yield more complex representations of this boundary than a generative model.

## Your stance and justification

Discriminative models are more powerful than the generative models and hence work better for larger datasets than smaller datasets. The only problem is of Overfitting in smaller Datasets.

# **PART-B CHAPTER 2**

\_\_\_\_­­­­­­­­\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# **Develop an algorithm to extract syntactic cues and semantic labels from a narrative text such that they form an input to knowledge representation module.**

**“””Story chosen**: One day Buddha was walking through a village. A very angry and rude young man came up and began insulting him. “You have no right teaching others,” he shouted. “You are as stupid as everyone else. You are nothing but a fake.”

Buddha was not upset by these insults. Instead he asked the young man “Tell me, if you buy a gift for someone, and that person does not take it, to whom does the gift belong?”

The man was surprised to be asked such a strange question and answered, “It would belong to me, because I bought the gift.”

The Buddha smiled and said, “That is correct. And it is exactly the same with your anger.

If you become angry with me and I do not get insulted, then the anger falls back on you.**”””**

## Tokenization of the narrative text

*Tokenization* is the process of demarcating and possibly classifying sections of a string of input characters. The resulting tokens are then passed on to some other form of processing. The process can be considered a sub-task of [parsing](https://en.wikipedia.org/wiki/Parsing) input.

There are different types of tokenization on which sentence tokenization and word tokenization are some:

In Sentence tokenisation the whole story is fragmented into each sentences.

Sentence tokenization output:

['One day Buddha was walking through a village.', 'A very angry and rude young man came up and began insulting him.', '“You have no right teaching others,” he shouted.', '“You are as stupid as everyone else.', 'You are nothing but a fake.”Buddha was not upset by these insults.', 'Instead he asked the young man “Tell me, if you buy a gift for someone, and that person does not take it, to whom does the gift belong?”The man was surprised to be asked such a strange question and answered, “It would belong to me, because I bought the gift.”The Buddha smiled and said, “That is correct.', 'And it is exactly the same with your anger.If you become angry with me and I do not get insulted, then the anger falls back on you.']

In Word tokenisation the whole story is fragmented into each words.

Word tokenization output:

['One', 'day', 'Buddha', 'was', 'walking', 'through', 'a', 'village', '.', 'A', 'very', 'angry', 'and', 'rude', 'young', 'man', 'came', 'up', 'and', 'began', 'insulting', 'him', '.', '“', 'You', 'have', 'no', 'right', 'teaching', 'others', ',', '”', 'he', 'shouted', '.', '“', 'You', 'are', 'as', 'stupid', 'as', 'everyone', 'else', '.', 'You', 'are', 'nothing', 'but', 'a', 'fake.', '”', 'Buddha', 'was', 'not', 'upset', 'by', 'these', 'insults', '.', 'Instead', 'he', 'asked', 'the', 'young', 'man', '“', 'Tell', 'me', ',', 'if', 'you', 'buy', 'a', 'gift', 'for', 'someone', ',', 'and', 'that', 'person', 'does', 'not', 'take', 'it', ',', 'to', 'whom', 'does', 'the', 'gift', 'belong', '?', '”', 'The', 'man', 'was', 'surprised', 'to', 'be', 'asked', 'such', 'a', 'strange', 'question', 'and', 'answered', ',', '“', 'It', 'would', 'belong', 'to', 'me', ',', 'because', 'I', 'bought', 'the', 'gift.', '”', 'The', 'Buddha', 'smiled', 'and', 'said', ',', '“', 'That', 'is', 'correct', '.', 'And', 'it', 'is', 'exactly', 'the', 'same', 'with', 'your', 'anger’,’.’,If', 'you', 'become', 'angry', 'with', 'me', 'and', 'I', 'do', 'not', 'get', 'insulted', ',', 'then', 'the', 'anger', 'falls', 'back', 'on', 'you', '.']

* 1. Develop an algorithm to extract the syntactic cues from the text.

A paragraph is given and there should some information in the syntactic structure of the given data. As humans we find that task easy until when we think of that being thought to the machine.

In this section with the help of the syntactic rules/information in the given text is first analysed and the grammar is set to extract only the required words that will get the overview of the story.

**Algorithm :**

* Read text.
* Identify the Special Characters and flag them(Not included for POS tagging)
* Creation of the dictionary the only picks the required words
  + Words that starts with the Noun=>Adjective=>Noun
  + Words ending (Leaf Ending) which has a conjunction as parent.
* Perform tokenisation eliminating the special characters.
* Run through parts of speech tagger.
* With the help of the dictionary only select the words that fit into and print.

**Output:**

Table ‑:Output from the syntactic extractor

|  |
| --- |
| **Sentence 1** |
| day |
| buddha |
| village |
| **Sentence 2** |
| rude |
| young |
| man |
| **Sentence 3** |
| right |
| teach |
| other |
| **Sentence 4** |
| everyone |
| **Sentence 5** |
| nothing |
| fake |
| buddha |
| insult |
| **Sentence 6** |
| young |
| man |
| tell |
| gift |
| someone |
| person |
| gift |
| belong |
| man |
| strange |
| question |
| **Sentence 7** |
| gift |
| buddha |
| anger |
| anger |

Tabel 2.1-1 Represents the extraction of the key words that determines the story. In the above table it is observed that the word obtained are of the name of the people, things and places are recognised and extracted.

* 1. Formation of a semantic label table

Semantics is the meaning of words in sentences. it refers to meaning that does not depend on the context where it appears. Semantic table consists of [Context1, Intent1, Dependencies, Dep].

Context has the words in the sentence ,Intent represents to the Action that is performed in the sentence, Dependencies represents the dependence of other words in the same sent to the word in the context, Dep is the semantic role label.

Semantic label table of the first Sentence.

0 One day [] nummod

1 day walking [One] npadvmod

2 Buddha walking [] nsubj

3 was walking [] aux

4 walking walking [day, Buddha, was, through, .] ROOT

5 through walking [village] prep

6 a village [] det

7 village through [a] pobj

8 . walking [] punct

## Development of an algorithm to extract the semantic labels

Semantic labelling tells the meaning of the text given as the input. This is done with the following steps. In this section the sematic label is considered as the Intent Recognition. Hence the Intent tells the whole action that is happening in the sentence.

**Algorithm:**

* Read text.
* Tokenize the paragraph into sentences.
* Identify the ‘and’ and ‘also’ words in the input and combine if wrongly tokenised in tokenisation.
* Links between the words are identified.
* The word with maximum links is the word that contains the action(Meaning) of the sentences.
* The root word is extracted from the whole sentence.

This Root word gives the picture of what happened in that line.

Tabel 2.4-1 consists of the words that represent the actions that is performed in each of the sentences.

**Output:**

Table ‑:Output of the semantic Extractor

|  |  |
| --- | --- |
| Sentence 1 | Walking |
| Sentence 2 | Angry, Came, Began |
| Sentence 3 | Shouted |
| Sentence 4 | Are Stupid |
| Sentence 5 | Are Fake |
| Sentence 6 | Take, gift, Buy, Belong |
| Sentence 7 | Falls, Anger |

## Development of an algorithm for coreference resolution for the extracted syntactic and semantic information

Coreference resolution is the task of determining linguistic expressions that refer to the same real-world entity in natural language.

The Coreference is done with the help of nueralCoref which gives the graphical coreference of the story.

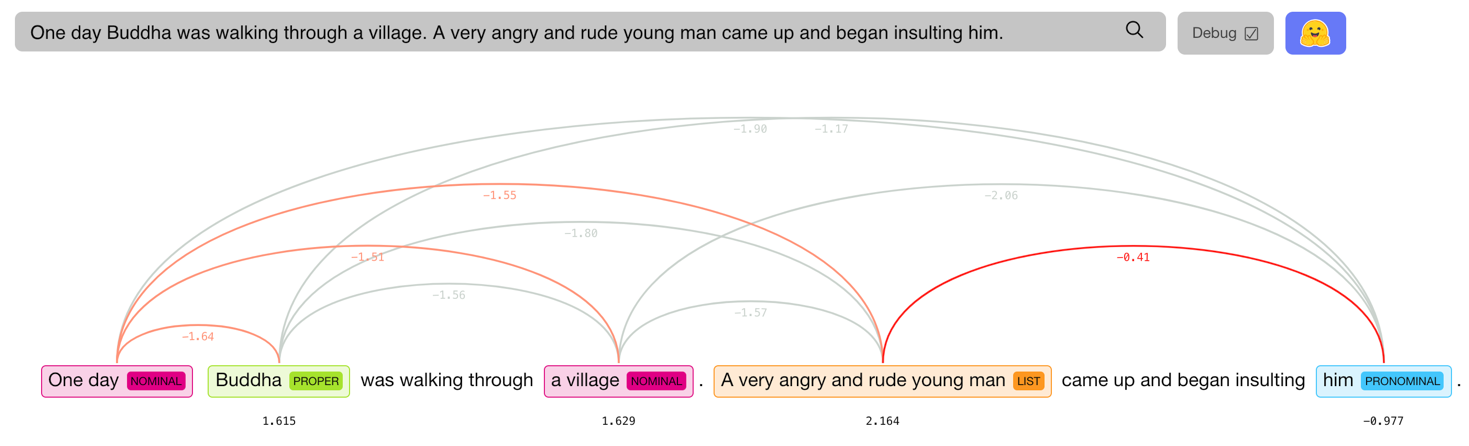


Figure ‑:Output of the coreference Coefficient of the first two sentences.

Figure 2.5-1 shows the coreference coefficient of the first two sentences and from which the relationship of the words is evaluated as coefficient values.

With the output of Syntactic and semantic information obtained fed into the Coreference Coefficient calculator it gives the relationship between the important words obtained from the syntactic extractor and the words from the semantic extractor.

## Feature identification by integrating syntactic, semantic, and coreference information

Each every step that is achieved until now has given some important information for further processing to be precise with the information contributed by the individuals are as follows.

* **Syntactic Extractor: Base words that form the story.**
* **Semantic Extractor: Main Theme of the sentence.**
* **Coreference resolution: Relationship between linguistic expression and real world word.**

By combining the output of the three previous steps sentence wise it gives an model that is good for the analysis purpose.

Syntactic Extractor: gives out all the person ,things and sometimes feelings from the story.

Semantic Extractor: gives the happening I.e. the root/cause.

Coreference coefficient: helps in framing he, him, that, etc.

The Syntactic and semantic words considered as the nodes of the model. The Conditional probability distribution is assigned to the Semantic extractor nodes. These nodes are constructed sentence wise.

# **PART-C CHAPTER 3**

\_\_\_\_­­­­­­­­\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# **Develop an algorithm for action description of a narrative text using Bayesian network**

## Development of a knowledge representation algorithm using the features identified in Part-B.

* The actual story the main context of the story lies in the conversation of Buddha and the young Man.
* The part before the conversation tells the thinking of the young man about Buddha.
* The Conversation refers to the Intent call gift but ultimately it is referred to the anger and the hurt.
* The complete reference to gift has to be taken to the Anger and Hurt.

## Identification of threshold patterns for joint probability distribution

As the story is seen in perspective of two parts I.e.

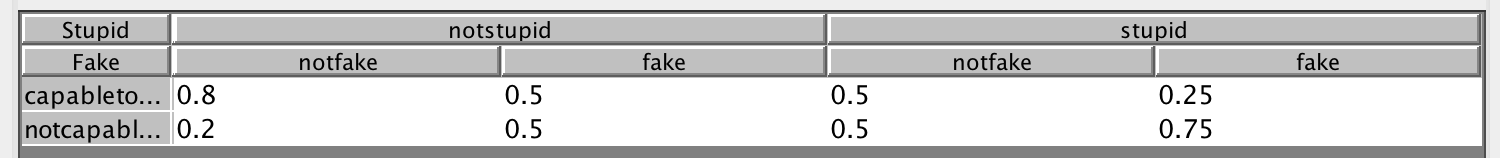
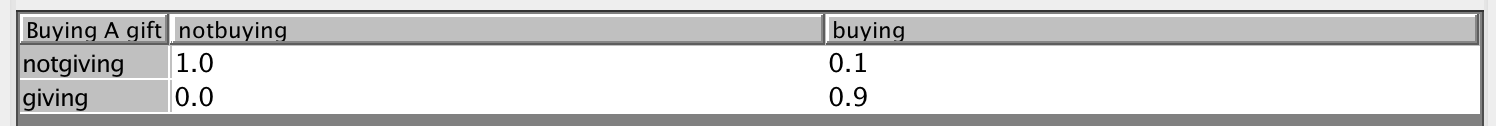
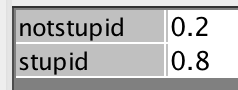
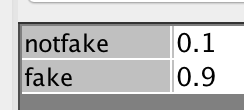
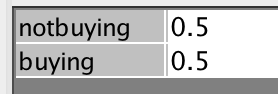
* Before conversation
* While conversation

Before Conversation the nature of Buddha is explained

While Conversation about Purchase and acceptance of Gift is referred

This demands us to model the problem with two different graphical model for before and while conversation.

The conditional probability distribution of the each node are as follows



## Development of a Bayesian network model

The Bayesian Network model is designed as per the identification of the threshold patterns Hence there are two Bayesian network graphs.

* Bayesian Graph for before conversation That represents the Nature of Buddha.

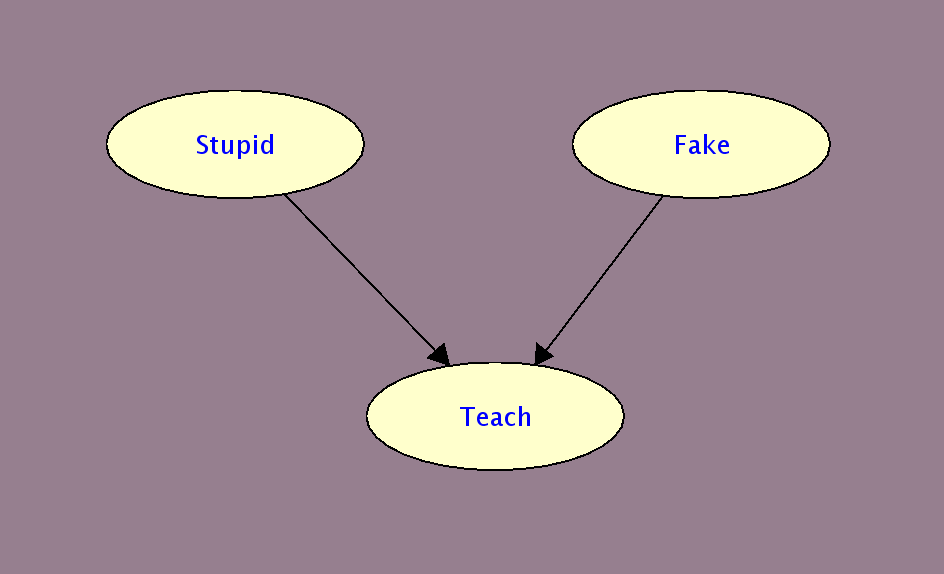


Figure ‑:Bayesian Model of the before conversation of the story

In this part the Ability of Buddha is inferenced weather he is a valid teacher or not with the quality that was mentioned by the young man.

* Bayesian Network for while conversation the dictates about gift that ultimately means Anger and Hurt.

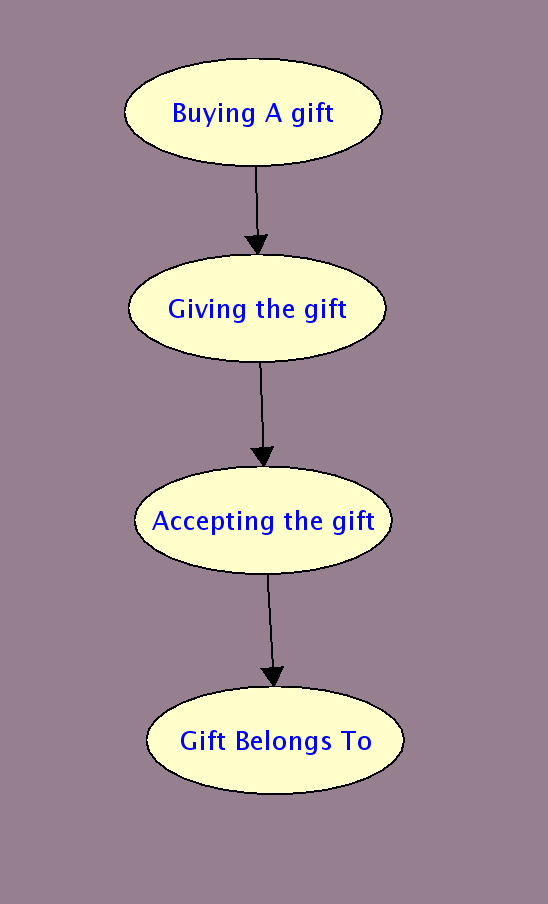


Figure ‑:Bayesian model of the while conversation of the story

In this part the acceptance of Anger and Hurt is expressed with the intent Gift

## Development of a reasoning scheme for inferring action description

Reasoning scheme for inferring the action is done using the python library for probabilistic graphical model **pgmpy**.

**from** **pgmpy.models** **import** BayesianModel

storybeforeconv\_model = BayesianModel([('Stupid', 'Teach'),

('Fake', ' Teach')])

*# Now defining the parameters.*

**from** **pgmpy.factors.discrete** **import** TabularCPD

cpd\_stu = TabularCPD(variable='Stupid', variable\_card=2,

values=[[0.2], [0.8])

cpd\_fak = TabularCPD(variable='Fake', variable\_card=2,

values=[[0.1], [0.9]])

cpd\_teach = TabularCPD(variable='Teach', variable\_card=2,

values=[[0.8, 0.5, 0.5, 0.25],

[0.2, 0.5, 0.5, 0.75]],

evidence=['Stupid', 'Fake'],

evidence\_card=[2, 2])

*# Computing the probability of teaching given stupid.*

q = storybeforeconv\_model.query(variables=['Teach'], evidence={'Stupid': 0})

**print**(q['Teach'])

+---------+--------------+

| Teach | phi(Stupid) |

|---------+--------------|

| Teach\_0 | 0.7000 |

| Teach\_1 | 0.3000 |

+---------+--------------+

Inferencing of the While conversation part of the story is also built on same way above.

## Analysis of results and comment on the influence of representation, inference, and learning algorithm

*Computing the probability of teaching given stupid.*

+---------+--------------+

| Teach | phi(Stupid) |

|---------+--------------|

| Teach\_0 | 0.7000 |

| Teach\_1 | 0.3000 |

+---------+--------------+

Looking at the above result it clearly made an inference of if the person is stupid there is no point of him teaching.

In this model the inference of the anger and is directly obtained it is from the Knowledge representation only.

Similarly as given all the inference can be made from the above graphical models.

# 

# **References**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Xavier Carreras and Lu´ıs Marquez. 2005. Introduction to ` the CoNLL-2005 shared task: Semantic role labeling. In Proceedings of CoNLL.

Eugene Charniak and Mark Johnson. 2005. Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In Proceedings of ACL.

Eugene Charniak. 2000. A maximum-entropy-inspired parser. In Proceedings of NAACL, pages 132–139.

Michael Collins and James Brooks. 1995. Prepositional attachment through a backed-off model. In Proceedings of the Third Workshop on Very Large Corpora, pages 27–38. Michael Collins. 1997. Three generative, lexicalised models for statistical parsing. In Proceedings of ACL, pages 16 – 23. Michael Collins. 2000. Discriminative reranking for natural language parsing. In Proceedings of ICML, pages 175– 182.

Stephen Della Pietra, Vincent J. Della Pietra, and John D. Lafferty. 1997. Inducing features of random fields. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(4):380–393. Nir Friedman and Moises Goldszmidt. 1996. Learning Bayesian Networks with local structure. In Proceeding of UAI, pages 252–262.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. Computational Linguistics, 28(3):245– 288. Joshua T. Goodman. 2001. A bit of progress in language modeling. In MSR Technical Report MSR-TR-2001-72.

Daniel Grossman and Pedro Domingos. 2004. Learning Bayesian Network classifiers by maximizing conditional likelihood. In Proceedings of ICML, pages 361—368.