Predicting Prerequisite Structure in Wikipedia

Partha Pratim Talukdar (CMU)
William W. Cohen (CMU)

7th Workshop on Innovative Use of NLP for Building Educational Applications (BEA7)
NAACL-HLT 2012, June 7, 2012
Need to Comprehend: Conditional Random Field
Need to Comprehend: Conditional Random Field

Scholarly articles for conditional random field
Conditional random fields: Probabilistic models for ... - Lafferty - Cited by 4507
A conditional random field word segmenter for sighan ... - Tseng - Cited by 122
... segmentation system with conditional random field - Zhao - Cited by 75

Conditional random field - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Conditional_random_field
Conditional random fields (CRFs) are a class of statistical modelling method often
applied in pattern recognition and machine learning, where they are used for ...
   • Description - Software - See also - References

[PDF] Conditional Random Fields: An Introduction
www.inference.phy.cam.ac.uk/hmw26/papers/crf_intro.pdf
File Format: PDF/Adobe Acrobat - Quick View
by HM Wallach - 2004 - Cited by 169 - Related articles
1 Labeling Sequential Data. The task of assigning label sequences to a ...

Conditional Random Fields
www.inference.phy.cam.ac.uk/hmw26/crf/
This page contains material on, or relating to, conditional random fields. I shall
continue to update this page as research on conditional random fields advances, ...

[PDF] Conditional Random Fields: Probabilistic Models for Segmenti...
www.cis.upenn.edu/~pereira/papers/crf.pdf
File Format: PDF/Adobe Acrobat - Quick View
by J Lafferty - Cited by 4495 - Related articles
This paper introduces conditional random fields (CRFs), a sequence modeling ... each
other. We can also think of a CRF as a finite state model with un- ...
Need to Comprehend: Conditional Random Field

What to read first and how to go about?

Scholarly articles for conditional random field
Conditional random fields: Probabilistic models for segmentation - Lafferty - Cited by 4507
A conditional random field word segmenter for sighan - Tseng - Cited by 122
... segmentation system with conditional random field - Zhao - Cited by 75

Conditional random fields - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Conditional_random_field
Conditional random fields (CRFs) are a class of statistical modelling method often applied in pattern recognition and machine learning, where they are used for ...
→ Description - Software - See also - References

This page contains material on, or relating to, conditional random fields. I shall continue to update this page as research on conditional random fields advances, ...

File Format: PDF/Adobe Acrobat - Quick View
by J Lafferty - Cited by 4495 - Related articles
This paper introduces conditional random fields (CRFs), a sequence modeling ... each other. We can also think of a CRF as a finite state model with un- ...
Conditional random fields (CRFs) are a class of statistical modelling method often applied in pattern recognition and machine learning, where they are used for structured prediction. Whereas an ordinary classifier predicts a label for a single sample without regard to "neighboring" samples, a CRF can take context into account; e.g., the linear chain CRF popular in natural language processing predicts sequences of labels for sequences of input samples.

CRFs are a type of discriminative undirected probabilistic graphical model. It is used to encode known relationships between observations and construct consistent interpretations. It is often used for labeling or parsing of sequential data, such as natural language text or biological sequences\(^1\) and in computer vision.\(^2\) Specifically, CRFs find applications in shallow parsing,\(^3\) named entity recognition\(^4\) and gene finding, among other tasks, being an alternative to the related hidden Markov models. In computer vision, CRFs are often used for object recognition and image segmentation.

### Contents

1. Description
   1.1 Inference
   1.2 Parameter Learning
   1.3 Examples
   1.4 Higher-order CRFs and semi-Markov CRFs
2. Software
3. See also
4. References

---

This article provides insufficient context for those unfamiliar with the subject. Please help improve the article with a good introductory style. (June 2011)
Conditional random fields (CRFs) are a class of statistical modelling method often applied in pattern recognition and machine learning, where they are used for structured data prediction. Whereas an ordinary classifier predicts a label for a single sample without regard to "neighboring" samples, a CRF can take context into account; e.g., the linear chain CRF popular in natural language processing predicts sequences of labels for sequences of input samples.

CRFs are a type of discriminative undirected probabilistic graphical model. It is used to encode known relationships between observations and construct consistent interpretations. It is often used for labeling or parsing of sequential data, such as natural language text or biological sequences[1] and in computer vision.[2] Specifically, CRFs find applications in shallow parsing,[3] named entity recognition[4] and gene finding, among other tasks, being an alternative to the related hidden Markov models. In computer vision, CRFs are often used for object recognition and image segmentation.
Random variable

In probability and statistics, a random variable or stochastic variable is a variable whose value is subject to variations due to chance (i.e. randomness, in a mathematical sense). As opposed to other mathematical variables, a random variable conceptually does not have a single, fixed value (even if unknown); rather, it can take on a set of possible different values, each with an associated probability. The interpretation of a random variable depends on the interpretation of probability:

- The objectivist viewpoint: as the outcome of an experiment or event where randomness is involved (e.g. the result of rolling a die, which is a number between 1 and 6, with all equal probability; or the sum of the results of rolling two dice, which is a number between 2 and 12, with some numbers more likely than others).
- The subjectivist viewpoint: the formal encoding of one's beliefs about the various potential values of a quantity that is not known with certainty (e.g. a particular person's belief about the net worth of someone like Bill Gates after Internet research on the subject, which might have possible values ranging between about $50 billion and $100 billion, with values near the center more likely).

Conditional random fields (CRFs) are a class of statistical modelling method often applied in pattern recognition and machine learning, where they are used for structured prediction. Whereas an ordinary classifier predicts a label for a single sample without regard to "neighboring" samples, a CRF can take context into account, e.g., the linear chain CRF popular in natural language processing predicts sequences of labels for sequences of input samples. CRFs are a type of discriminative undirected probabilistic graphical model. It is used to encode known relationships between observations and construct consistent interpretations. It is often used for labeling or parsing of sequential data, such as natural language text or biological sequences and in computer vision. Specifically, CRFs find applications in shallow parsing, named entity recognition and gene finding, among other tasks, being an alternative to the related hidden Markov models. In computer vision, CRFs are often used for object recognition and image segmentation.
Random variable

From Wikipedia, the free encyclopedia

In probability and statistics, a random variable or stochastic variable is a variable whose value is subject to variations due to chance (i.e. randomness, in a mathematical sense). As opposed to other mathematical variables, a random variable conceptually does not have a single, fixed value (even if unknown); rather, it can take on a set of possible different values, each with an associated probability. The interpretation of a random variable depends on the interpretation of probability:

- The objectivist viewpoint: the outcome of an experiment or event where randomness is involved (e.g. the result of rolling a die, which is a number between 1 and 6, all with equal probability; or the sum of the results of rolling two dice, which is a number between 2 and 12, with some numbers more likely than others).
- The subjectivist viewpoint: the formal encoding of one's beliefs about the various potential values of a quantity that is not known with certainty (e.g. a particular person's belief about the net worth of someone like Bill Gates after Internet research on the subject, which might have possible values ranging between about $50 billion and $100 billion, with values near the center more likely).

Conditional random fields (CRFs) are a class of statistical machine learning models often used in pattern recognition and natural language processing for prediction. Where as an ordinary statistician models a label for a single sample without regard to "neighboring" samples, a CRF can take context into account, e.g., the linear chain CRF popular in natural language processing precision.

CRFs are a type of discriminative undirected probabilistic graphical model. It is used to encode known relationships between observations and construct consistent interpretations. It is often used for labeling or parsing of sequential data, such as natural language text or biological sequences[1] and in computer vision.[2] Specifically, CRFs find applications in the following areas:[3] named entity recognition[4] and solving finding problems such as word segmentation[5] in natural language processing. An example of a conditional random field is the hidden Markov model.
Random variable

In probability and statistics, a random variable or stochastic variable is a variable whose value is subject to variations due to chance (i.e., randomness, in a mathematical sense). As opposed to other mathematical variables, a random variable conceptually does not have a single, fixed value (even if unknown); rather, it can take on a set of possible different values, each with an associated probability. The interpretation of a random variable depends on the interpretation of probability:

- The objective viewpoint: as the outcome of an experiment or event where randomness is involved (e.g., the result of rolling a die, which is a number between 1 and 6, all with equal probability; or the sum of the results of rolling two dice, which is a number between 2 and 12, with some numbers more likely than others).
- The subjective viewpoint: the formal encoding of one’s beliefs about the various potential values of a quantity that is not known with certainty (e.g., a particular person’s belief about the net worth of someone like Bill Gates after Internet research on the subject, which might have possible values ranging between about $50 billion and $100 billion, with values near the center more likely).

Conditional random fields (CRFs) are a class of statistical models and machine learning, where they are used for structured prediction. Where as an ordinary classifier predicts a label for a single sample without regard to “neighboring” samples, a CRF can take context into account, e.g., the linear chain CRF popular in natural language processing prediction. CRFs are a type of discriminative undirected probabilistic graphical model. It is used to encode known relationships between observations and construct consistent interpretations. It is often used for labeling or parsing of sequential data, such as natural language text or biological sequences[1] and in computer vision.[2] Specifically, CRFs find applications in computer vision, [3] named entity recognition [4] and many other natural language processing tasks. Conditional random fields have been applied to tasks such as optical character recognition, part-of-speech tagging, statistical machine translation, and many others, where they have shown substantial advantages over traditional approaches.

Markov random field

In the domain of physics and probability, a Markov random field (often abbreviated as MRF), Markov network or undirected graphical model is a set of random variables having the Markov property described by an undirected graph. A Markov random field is similar to a Bayesian network in its representation of dependencies; the differences being that Bayesian networks are directed and acyclic, whereas Markov networks are undirected and may be cyclic. Thus, a Markov network can represent certain dependencies that a Bayesian network cannot (such as cyclic dependencies); on the other hand, it cannot represent certain dependencies that a Bayesian network can (such as induced dependencies).
Attempting to comprehend a topic top-down can easily become difficult to manage: need better tools!
Comprehension Plan for Conditional Random Field (CRF)
Comprehension Plan for Conditional Random Field (CRF)
Comprehension Plan for Conditional Random Field (CRF)

Topics the user is already familiar with
Comprehension Plan for Conditional Random Field (CRF)

Topics the user is already familiar with

- Conditional independence
- Statistical model
- Probability distribution
- Random variable
- Variable mathematics

Topics of interest

- Graphical model
- Discriminative model
- Markov random field
- Hidden Markov Model
- Maximum likelihood
- Viterbi algorithm
- Markov chain
- Baum-Welch algorithm
- Inference
- Dynamic programming
- Expectation maximization algorithm
- Gradient descent
Comprehension Plan for Conditional Random Field (CRF)

Prerequisite
- Discriminative model
- Markov_random_field
  - Graphical_model
    - Conditional_independence
    - Statistical_model
      - random_variable
        - probability_distribution
        - variable_mathematics
- Markov_random_field
- Hidden_Markov_Model
  - Maximum_likelihood
  - Viterbi_algorithm
    - Markov_chain
      - Baum_Welch_algorithm
- Parameter_learning
  - Inference
    - Gradient_descent
    - dynamic_programming
      - Expectional_maximization_algorithm

Topics the user is already familiar with

Topic of interest
Ingredients for a Comprehension Plan
Ingredients for a Comprehension Plan

- Estimation of user’s current knowledge
Ingredients for a Comprehension Plan

• Estimation of user’s current knowledge
• Determining prerequisite structure among topics
Ingredients for a Comprehension Plan

• Estimation of user’s current knowledge
• Determining prerequisite structure among topics
• Plan over prerequisite structure to connect user’s current knowledge with target topic
Ingredients for a Comprehension Plan

• Estimation of user’s current knowledge

• Determining prerequisite structure among topics

• Plan over prerequisite structure to connect user’s current knowledge with target topic
Prerequisites $\subseteq$ Hyperlinks
Prerequisites ⊆ Hyperlinks

CRF

- Graphical Model
- HMM
- Computer Vision
- Shallow Parsing

Hyperlink from CRF’s Wikipedia page
Prerequisites $\subseteq$ Hyperlinks

- **CRF**
  - Prereq
  - Not Prereq

- **Graphical Model**
  - Prereq

- **HMM**
  - Not Prereq

- **Computer Vision**
  - Not Prereq

- **Shallow Parsing**
  - Hyperlink from CRF's Wikipedia page
All hyperlinks are not prerequisites. Given a hyperlink, classify whether it is a prerequisite.

Hyperlink from CRF’s Wikipedia page
Outline

• Motivation

• Prerequisite Classification
  • Data Preprocessing
  • Classifier
  • Prerequisite Judgements using MTurk
  • Features

• Experiments

• Conclusion
Data Preprocessing:
Target Concepts
Data Preprocessing: Target Concepts

<table>
<thead>
<tr>
<th>Target Concept</th>
<th>#Nodes</th>
<th>#Edges</th>
<th>#Edits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Warming</td>
<td>19,170</td>
<td>501,608</td>
<td>1,490,967</td>
</tr>
<tr>
<td>Meiosis</td>
<td>19,811</td>
<td>444,100</td>
<td>880,684</td>
</tr>
<tr>
<td>Newton’s Laws of Motion</td>
<td>15,714</td>
<td>436,035</td>
<td>795,988</td>
</tr>
<tr>
<td>Parallel Postulate</td>
<td>14,966</td>
<td>363,462</td>
<td>858,785</td>
</tr>
<tr>
<td>Public-key cryptography</td>
<td>16,695</td>
<td>371,104</td>
<td>1,003,181</td>
</tr>
</tbody>
</table>

Table 1: Target concepts used in the experiments.

Sample up to 400 edges from each subgraph
Gold Prerequisite Judgements using Mechanical Turk
Gold Prerequisite Judgements using Mechanical Turk

**Determine the Prerequisite Concept**

In many situations, we need to read and understand the main concept in one Wikipedia page before we can understand the main concept in another Wikipedia page. For example, in order to understand \textit{Acceleration}, one should first read and understand \textit{Velocity}, as \textit{Velocity} is a prerequisite for \textit{Acceleration}. However, the reverse is not true. Similarly, \textit{Atom} is a prerequisite for \textit{molecule}.

- Below, you will be asked to compare two concepts from Wikipedia, and determine whether one is a prerequisite for the other. Clicking on a concept will take you to its corresponding Wikipedia page.
- If you are unsure about the concepts, or whether one is a prerequisite of the other, please select "I don't know the answer" option.
Gold Prerequisite Judgements using Mechanical Turk

Determine the Prerequisite Concept

In many situations, we need to read and understand the main concept in one Wikipedia page before we can understand the main concept in another Wikipedia page. For example, in order to understand Acceleration, one should first read and understand Velocity, as Velocity is a prerequisite for Acceleration. However, the reverse is not true. Similarly, Atom is a prerequisite for molecule.

- Below, you will be asked to compare two concepts from Wikipedia, and determine whether one is a prerequisite for the other. Clicking on a concept will take you to its corresponding Wikipedia page.
- If you are unsure about the concepts, or whether one is a prerequisite of the other, please select "I don't know the answer" option.

Select only ONE of the following:

- Is Chromosome a prerequisite for Spirochaete?
- Is Spirochaete a prerequisite for Chromosome?
- Related but one is not a prerequisite of the other.
- The two concepts are unrelated.
- I do not know the answer.

Please provide comments or feedback below:

Submit
Gold Prerequisite Judgements using Mechanical Turk

Determine the Prerequisite Concept

In many situations, we need to read and understand the main concept in one Wikipedia page before we can understand the main concept in another Wikipedia page. For example, in order to understand Acceleration, one should first read and understand Velocity, as Velocity is a prerequisite for Acceleration. However, the reverse is not true. Similarly, Atom is a prerequisite for molecule.

- Below, you will be asked to compare two concepts from Wikipedia, and determine whether one is a prerequisite for the other. Clicking on a concept will take you to its corresponding Wikipedia page.
- If you are unsure about the concepts, or whether one is a prerequisite of the other, please select "I don't know the answer" option.

Select only ONE of the following:

- Is Chromosome a prerequisite for Spirochaete?
- Is Spirochaete a prerequisite for Chromosome?
- Related but one is not a prerequisite of the other.
- The two concepts are unrelated.
- I do not know the answer.

Please provide comments or feedback below:

Submit
Agreement among Turkers
Agreement among Turkers

<table>
<thead>
<tr>
<th>Domain</th>
<th>Time (s) / Evaluation</th>
<th>Worker / HIT</th>
<th># HITs</th>
<th>$\kappa$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meiosis</td>
<td>38</td>
<td>3</td>
<td>400</td>
<td>0.50</td>
</tr>
<tr>
<td>Public-key Cryp.</td>
<td>26</td>
<td>3</td>
<td>200</td>
<td>0.63</td>
</tr>
<tr>
<td>Parallel Postulate</td>
<td>41</td>
<td>3</td>
<td>200</td>
<td>0.55</td>
</tr>
<tr>
<td>Newton’s Laws</td>
<td>20</td>
<td>5</td>
<td>400</td>
<td>0.47</td>
</tr>
<tr>
<td>Global Warming</td>
<td>14</td>
<td>5</td>
<td>400</td>
<td>0.56</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>27.8</strong></td>
<td><strong>-</strong></td>
<td><strong>-</strong></td>
<td><strong>0.54</strong></td>
</tr>
</tbody>
</table>

Table 2: Statistics about the Gold-standard data prepared using Amazon Mechanical Turk. Also shown are the averaged $\kappa$ statistics-based inter-annotator agreement in each domain. The last row corresponds to the $\kappa$ value averaged across all five domains.
## Agreement among Turkers

<table>
<thead>
<tr>
<th>Domain</th>
<th>Time (s) / Evaluation</th>
<th>Worker / HIT</th>
<th># HITs</th>
<th>$\kappa$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meiosis</td>
<td>38</td>
<td>3</td>
<td>400</td>
<td>0.50</td>
</tr>
<tr>
<td>Public-key Cryp.</td>
<td>26</td>
<td>3</td>
<td>200</td>
<td>0.63</td>
</tr>
<tr>
<td>Parallel Postulate</td>
<td>41</td>
<td>3</td>
<td>200</td>
<td>0.55</td>
</tr>
<tr>
<td>Newton’s Laws</td>
<td>20</td>
<td>5</td>
<td>400</td>
<td>0.47</td>
</tr>
<tr>
<td>Global Warming</td>
<td>14</td>
<td>5</td>
<td>400</td>
<td>0.56</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>27.8</strong></td>
<td><strong>-</strong></td>
<td><strong>-</strong></td>
<td><strong>0.54</strong></td>
</tr>
</tbody>
</table>

Table 2: Statistics about the Gold-standard data prepared using Amazon Mechanical Turk. Also shown are the averaged $\kappa$ statistics-based inter-rater agreement for each domain. The last row contains the averaged across all five domains.

Total cost: $278, completed in a week.
Classifier: Maximum Entropy

\[ p(y|x) = \frac{\exp(w \cdot \phi(x, y))}{\sum_{y' \in Y} \exp(w \cdot \phi(x, y'))}, \ y \in Y = \{-1, +1\} \]
Classifier: Maximum Entropy

\[ p(y|x) = \frac{\exp(w \cdot \phi(x, y))}{\sum_{y' \in Y} \exp(w \cdot \phi(x, y'))}, \quad y \in Y = \{-1, +1\} \]
Classifier: Maximum Entropy

\[ p(y|x) = \frac{\exp(w \cdot \phi(x, y))}{\sum_{y' \in Y} \exp(w \cdot \phi(x, y'))}, \quad y \in Y = \{-1, +1\} \]
Classifier: Maximum Entropy

\[ p(y|x) = \frac{\exp(w \cdot \phi(x, y))}{\sum_{y' \in Y} \exp(w \cdot \phi(x, y'))}, \quad y \in \{ -1, +1 \} \]
Features

Wiki Editor

CRF (target)

HMM

Wiki Editor
Features

Features over PageContent:
- Wiki category of HMM page
- Name of first CRF section in which the hyperlink occurs
- Category overlap between pages
- #times HMM is referenced from CRF...

Wiki Editor

CRF (target)

HMM

PageContent
Features

Features over Hyperlink Graph:
- Random Walk with Restart (RWR) score of a topic
- Pagerank score of source
- Pagerank score of target

Features over PageContent:
- Wiki category of HMM page
- Name of first CRF section in which the hyperlink occurs
- Category overlap between pages
- #times HMM is referenced from CRF

Wiki Editor

CRF (target)

HMM
Features

Feature over WikiEdits Graph:
• RWR score of a topic from CRF

Feature over Hyperlink Graph:
• Random Walk with Restart (RWR) score of a topic
• Pagerank score of source
• Pagerank score of target

Features over PageContent:
• Wiki category of HMM page
• Name of first CRF section in which the hyperlink occurs
• Category overlap between pages
• #times HMM is referenced from CRF
...
Experiments
Experiments

- Can we train a classifier for prerequisite classification?
Experiments

• Can we train a classifier for prerequisite classification?

• How effective is out-of-domain training?
Experiments

• Can we train a classifier for prerequisite classification?
• How effective is out-of-domain training?
• What are the effects of different features?
Experiments

• Can we train a classifier for prerequisite classification?
• How effective is out-of-domain training?
• What are the effects of different features?

All evaluations in leave-one-target-out setting
Performance Comparison for Prerequisite Prediction

- **Random Baseline**
- **MaxEnt Classifier**

Accuracy

<table>
<thead>
<tr>
<th>Topic</th>
<th>Random Baseline</th>
<th>MaxEnt Classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meiosis</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Public Key</td>
<td>50</td>
<td>70</td>
</tr>
<tr>
<td>Para. Postulate</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>Newton’s Law</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Global Warming</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>Average</td>
<td>50</td>
<td>63.33</td>
</tr>
</tbody>
</table>
Trained classifier achieves 8.6% absolute improvement in accuracy
Effect of Out of Domain vs In Domain Training

Accuracy

- Meiosis
- Public Key
- Para. Postulate
- Newton’s Laws
- Global Warming
- Average

Out of Domain Training
In Domain Training
Effect of Out of Domain vs In Domain Training

Out of Domain Training
In Domain Training

Close to optimal (93%) performance possible in out-of-domain training.
Effect of Different Features
Effect of Different Features

Accuracy

<table>
<thead>
<tr>
<th>Feature</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>50</td>
</tr>
<tr>
<td>WikiEdits</td>
<td>53</td>
</tr>
<tr>
<td>Hyperlinks</td>
<td>56</td>
</tr>
<tr>
<td>PageContent</td>
<td>59</td>
</tr>
<tr>
<td>All</td>
<td>50</td>
</tr>
</tbody>
</table>
Effect of Different Features

Accuracy

59
56
53
50

Random WikiEdits Hyperlinks PageContent All
Effect of Different Features

Accuracy

Random  WikiEdits  Hyperlinks  PageContent  All

59  56  53  50
Effect of Different Features

Accuracy

59

56

53

50

Random  WikiEdits  Hyperlinks  PageContent  All
Effect of Different Features

- Random: 50
- WikiEdits: 53
- Hyperlinks: 56
- PageContent: 59
- All: 59

Accuracy
Effect of Different Features

Accuracy

Random | WikiEdits | Hyperlinks | PageContent | All
--- | --- | --- | --- | ---
50 | 53 | 56 | 59
Effect of Different Features

MaxEnt model with different features

Accuracy

Random     WikiEdits     Hyperlinks     PageContent     All

Accuracy: 50, 53, 56, 59
Incorrect Classifications
Incorrect Classifications

• Mass is not a prerequisite for Physics

• Quantum mechanics is not a prerequisite for Bohr-Einstein debates

• Global warming is a prerequisite for Nitrous Oxide

• Carbon Dioxide is not a prerequisite for Carbon Sequestration
Examples of Discriminative Features
Examples of Discriminative Features

• Is Target title a substring of Source title?
Examples of Discriminative Features

• Is Target title a substring of Source title?

• Is there category overlap between source and target pages?
Examples of Discriminative Features

• Is Target title a substring of Source title?
• Is there category overlap between source and target pages?
• Identity of category overlap between pages.
Examples of Discriminative Features

- Is Target title a substring of Source title?
- Is there category overlap between source and target pages?
- Identity of category overlap between pages.
- Identity of source section from where target is linked.
Conclusion
Conclusion

• Novel task: prerequisite structure prediction
  – Demonstrated that relatively reliable features exist
Conclusion

• Novel task: prerequisite structure prediction
  - Demonstrated that relatively reliable features exist

• Future work
  - feature engineering
  - controlled data collection
Conclusion

• Novel task: prerequisite structure prediction
  - Demonstrated that relatively reliable features exist

• Future work
  - feature engineering
  - controlled data collection

• Ongoing work
  - full comprehension plan generation using predicted prerequisite structure
Thank You!