Exploring New Frontiers of Web Search

Hang Li
Microsoft Research Asia
Our Projects on Web Search

• Whole Page Relevance
• Concept-based Search
• Context-aware Search
Whole Page Relevance
Whole Page Relevance

• Relevance, as well as diversity, document similarity, hyperlink relationship

• Beyond ten blue link (combination of categorization and ranking)
Global Ranking Using Continuous Conditional Random Fields

Qin et al. NIPS 2009
Global Ranking

- The rank of document is determined not only by itself, but also by the other documents

A set of docs
\[ x^{(q)} = \{x_1^{(q)}, x_2^{(q)}, \ldots, x_{n(q)}^{(q)} \} \]

A set of scores
\[ y^{(q)} = \{y_1^{(q)}, y_2^{(q)}, \ldots, y_{n(q)}^{(q)} \} \]

Conventional (local) ranking
\[ y_i^{(q)} = f(x_i^{(q)}) \]

Taking single document as input

Global ranking
\[ y^{(q)} = F(x^{(q)}) \]

Taking all documents as input
Global Ranking Tasks

• Example 1: Subtopic Retrieval
  – Query: programming language

    More Desirable
    1. Page about C++
    2. Page about C#
    3. Page about Java

    Less Desirable
    1. Page about C++
    2. Page about C++
    3. Page about C++
Global Ranking Tasks

• Example 2: Topic Distillation
  – A page is more preferred than its child page, if both of them are relevant to the query
Global Ranking Tasks

• Example 3: Pseudo Relevance Feedback
  – Similar documents should have similar relevance scores

  document $d_1$ is relevant to query $q$

  document $d_2$ is similar to document $d_1$

  $\downarrow$

  document $d_2$ is likely to be relevant to query $q$
Global Ranking Using Continuous CRF

Documents

Ranking scores

Relations among ranking scores
Continuous CRF (C-CRF)

\[
\Pr(y^{(q)}|x^{(q)}) = \frac{1}{Z(x^{(q)})} \exp \left\{ \sum_i \sum_{k=1}^{K_1} \alpha_k h_k(y^{(q)}_i, x^{(q)}) + \sum_{i,j} \sum_{k=1}^{K_2} \beta_k g_k(y^{(q)}_i, y^{(q)}_j, x^{(q)}) \right\}
\]

**Vertex feature function**    **Edge feature function**

**Learning**

\[
L(\alpha, \beta) = \sum_{q=1}^N \log \Pr(y^{(q)}|x^{(q)}; \alpha, \beta).
\]

**Prediction**

\[
F(x^{(q)}) = \arg \max_{y^{(q)}} \Pr(y^{(q)}|x^{(q)})
\]
C-CRF for Pseudo Relevance Feedback

\[
\Pr(y^{(q)}|x^{(q)}) = \frac{1}{Z(x^{(q)})} \exp \left\{ \sum_{i} \sum_{k=1}^{K_1} \alpha_k h_k(y_i^{(q)}, x^{(q)}) + \sum_{i,j} \sum_{k=1}^{K_2} \beta_k g_k(y_i^{(q)}, y_j^{(q)}, x^{(q)}) \right\}
\]

\[
\Pr(y|x) = \frac{1}{Z(x)} \exp \left\{ \sum_{i} \sum_{k=1}^{K_1} -\alpha_k (y_i - x_{i,k})^2 + \sum_{i,j} -\frac{\beta}{2} S_{i,j} (y_i - y_j)^2 \right\}
\]

S is the document similarity matrix, which is available in both training and testing.
C-CRF for Topic Distillation

\[
\Pr(y^{(q)}|x^{(q)}) = \frac{1}{Z(x^{(q)})} \exp \left\{ \sum_{i} \sum_{k=1}^{K_1} \alpha_k h_k(y^{(q)}_i, x^{(q)}) + \sum_{i,j} \sum_{k=1}^{K_2} \beta_k g_k(y^{(q)}_i, y^{(q)}_j, x^{(q)}) \right\}
\]

\[
\Pr(y|x) = \frac{1}{Z(x)} \exp \left\{ \sum_{i} \sum_{k=1}^{K_1} -\alpha_k (y_i - x_{i,k})^2 + \sum_{i,j} \beta R_{i,j} (y_i - y_j) \right\}
\]

*R is sitemap matrix, which is available in both training and testing.*
C-CRF for Global Ranking Using Multiple Types of Relations

\[
Pr(y|x) = \frac{1}{Z(x)} \exp \left\{ \sum_{i} \sum_{k=1}^{K_1} -\alpha_k(y_i - x_{i,k})^2 + \sum_{i,j} \left( \beta_1 R_{i,j} (y_i - y_j) - \beta_2 \frac{S_{i,j}}{2} (y_i - y_j)^2 \right) \right\}
\]

Topic distillation + Pseudo relevance feedback
Data Sets

• TD2004
  – For topic distillation
  – 75 queries, ~1000 documents per query
  – 5 fold cross validation

• OHSUMED
  – For pseudo relevance feedback
  – 106 queries, ~150 documents per query
  – 5 fold cross validation
Experimental Results on Pseudo Relevance Feedback

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>ndcg1</th>
<th>ndcg2</th>
<th>ndcg5</th>
</tr>
</thead>
<tbody>
<tr>
<td>BM25</td>
<td>0.3994</td>
<td>0.3931</td>
<td>0.3972</td>
</tr>
<tr>
<td>BM25-PRF</td>
<td>0.3962</td>
<td>0.4277</td>
<td>0.3981</td>
</tr>
<tr>
<td>RankSVM</td>
<td>0.4952</td>
<td>0.4755</td>
<td>0.4579</td>
</tr>
<tr>
<td>ListNet</td>
<td>0.5231</td>
<td>0.497</td>
<td>0.4662</td>
</tr>
<tr>
<td>C-CRF</td>
<td>0.5443</td>
<td>0.4986</td>
<td>0.4808</td>
</tr>
</tbody>
</table>
## Experimental Results on Topic Distillation

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>ndcg1</th>
<th>ndcg2</th>
<th>ndcg5</th>
</tr>
</thead>
<tbody>
<tr>
<td>BM25</td>
<td>0.3067</td>
<td>0.2933</td>
<td>0.2293</td>
</tr>
<tr>
<td>ST</td>
<td>0.3200</td>
<td>0.3133</td>
<td>0.3232</td>
</tr>
<tr>
<td>SS</td>
<td>0.3200</td>
<td>0.3200</td>
<td>0.3227</td>
</tr>
<tr>
<td>RankSVM</td>
<td>0.4400</td>
<td>0.4333</td>
<td>0.3935</td>
</tr>
<tr>
<td>ListNet</td>
<td>0.4400</td>
<td>0.4267</td>
<td>0.4209</td>
</tr>
<tr>
<td>C-CRF</td>
<td>0.5200</td>
<td>0.4733</td>
<td>0.4428</td>
</tr>
</tbody>
</table>
Future Work

• How to obtain scores of documents for global ranking training?
  – Using implicit feedback data?
• Whether possible to incorporate diversity relation?
Concept based Search
Concept based Search

- Better understand queries and documents
Query Understanding in Concept-based Search

• Can interpret acronym

• Can interpret synonym

• Can perform stemming

• Can process long query well
  – Query: “new york times square church” → Document “church near Times Square in New York”
Named Entity Recognition in Query Using WS-LDA

Guo et al. SIGIR 2009
Xu et al. KDD 2009
Named Entity Recognition in Query

- harry potter
  - harry potter – Movie (0.5)
  - harry potter – Book (0.4)
  - harry potter – Game (0.1)

- harry potter film
  - harry potter film
  - harry potter – Movie (0.95)

- harry potter author
  - harry potter author
  - harry potter – Book (0.95)
Challenges

- No rich context available like in documents
- No sufficient features available (e.g., first letter capitalized)
- Many new named entities
Our Approach

• Using Query Log Data (or Click-through Data)
• Using Topic Model
• Weakly Supervised Latent Dirichlet Allocation
• vs Pasca’s work (named entity mining from log data, deterministic approach)
Seed and Query Log

final fantasy
Movie Game

gone with the wind
Movie Book

harry potter
Movie Book Game

<table>
<thead>
<tr>
<th>Search Query</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>final fantasy</td>
<td>300</td>
</tr>
<tr>
<td>final fantasy movie</td>
<td>120</td>
</tr>
<tr>
<td>final fantasy wallpaper</td>
<td>50</td>
</tr>
<tr>
<td>gone with the wind movie</td>
<td>120</td>
</tr>
<tr>
<td>gone with the wind review</td>
<td>10</td>
</tr>
<tr>
<td>gone with the wind photos</td>
<td>10</td>
</tr>
<tr>
<td>harry potter</td>
<td>1000</td>
</tr>
<tr>
<td>harry potter book</td>
<td>650</td>
</tr>
<tr>
<td>gone with the wind book</td>
<td>80</td>
</tr>
<tr>
<td>gone with the wind summary</td>
<td>20</td>
</tr>
<tr>
<td>harry potter cheats</td>
<td>300</td>
</tr>
<tr>
<td>harry potter pics</td>
<td>200</td>
</tr>
<tr>
<td>harry potter summary</td>
<td>100</td>
</tr>
<tr>
<td>final fantasy xbox</td>
<td>10</td>
</tr>
<tr>
<td>final fantasy soundtrack</td>
<td>10</td>
</tr>
<tr>
<td>gone with the wind</td>
<td>250</td>
</tr>
<tr>
<td>harry potter movie</td>
<td>800</td>
</tr>
</tbody>
</table>

......
<table>
<thead>
<tr>
<th>Entity</th>
<th>Context</th>
<th>Class</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>final fantasy</td>
<td>#</td>
<td>Movie, Game</td>
<td>300</td>
</tr>
<tr>
<td>final fantasy</td>
<td># movie</td>
<td>Movie, Game</td>
<td>120</td>
</tr>
<tr>
<td>final fantasy</td>
<td># wallpaper</td>
<td>Movie, Game</td>
<td>50</td>
</tr>
<tr>
<td>final fantasy</td>
<td># xbox</td>
<td>Movie, Game</td>
<td>10</td>
</tr>
<tr>
<td>final fantasy</td>
<td># soundtrack</td>
<td>Movie, Game</td>
<td>10</td>
</tr>
<tr>
<td>gone with the wind</td>
<td>#</td>
<td>Movie, Book</td>
<td>250</td>
</tr>
<tr>
<td>gone with the wind</td>
<td># movie</td>
<td>Movie, Book</td>
<td>120</td>
</tr>
<tr>
<td>gone with the wind</td>
<td># book</td>
<td>Movie, Book</td>
<td>80</td>
</tr>
<tr>
<td>gone with the wind</td>
<td># summary</td>
<td>Movie, Book</td>
<td>20</td>
</tr>
<tr>
<td>gone with the wind</td>
<td># review</td>
<td>Movie, Book</td>
<td>10</td>
</tr>
<tr>
<td>gone with the wind</td>
<td># photos</td>
<td>Movie, Book</td>
<td>10</td>
</tr>
<tr>
<td>harry potter</td>
<td>#</td>
<td>Movie, Book, Game</td>
<td>1000</td>
</tr>
<tr>
<td>harry potter</td>
<td># movie</td>
<td>Movie, Book, Game</td>
<td>800</td>
</tr>
<tr>
<td>harry potter</td>
<td># book</td>
<td>Movie, Book, Game</td>
<td>650</td>
</tr>
<tr>
<td>harry potter</td>
<td># cheats</td>
<td>Movie, Book, Game</td>
<td>300</td>
</tr>
<tr>
<td>harry potter</td>
<td># pics</td>
<td>Movie, Book, Game</td>
<td>200</td>
</tr>
<tr>
<td>harry potter</td>
<td># summary</td>
<td>Movie, Book, Game</td>
<td>100</td>
</tr>
</tbody>
</table>
## Pseudo Documents of Named Entities

| Name               | #      | movie | wallpaper | xbox | soundtrack | summary | review | photos |
|--------------------|--------|-------|-----------|------|------------|---------|-------|--------|        |
| final fantasy      | 300    | 120   | 50        | 10   | 10         | 0       | 0     | 0      |         |
| gone with the wind | 250    | 120   | 80        | 20   | 10         | 10      | 10    | 10     |         |
| harry potter       | 1000   | 800   | 650       | 300  | 200        | 100     | 0     | 0      |         |

*Movie, Game*

*Movie, Book*

*Movie, Book, Game*
Latent Dirichlet Allocation Model

\[ z: \text{Movie, Book, Game} \]
\[ w: \# , \# \text{ movie, } \# \text{ book, } \ldots \]
\[ \theta: \text{distribution of classes for named entity} \]
\[ \beta: \text{distribution of contexts for class} \]
Weakly Supervised Latent Dirichlet Allocation

\[ p(D|\Theta) = \prod_{d=1}^{M} \int p(\theta_d|\alpha)(\prod_{n=1}^{N_d} p(z_{dn}|\theta_d)p(w_{dn}|z_{dn}, \beta))d\theta_d \]

\[ \log p(D|\Theta) + \lambda C(\Theta, y) \]

\[ = \sum_{d=1}^{M} \log \int p(\theta_d|\alpha)(\prod_{n=1}^{N_d} p(z_{dn}|\theta_d)p(w_{dn}|z_{dn}, \beta))d\theta_d \]

\[ + \sum_{d=1}^{M} \lambda \sum_{i=1}^{K} y_{di} \bar{z}_{di} \]

\[ \bar{z}_i = \frac{1}{N} \sum_{n=1}^{N} z_{in} \]

constraints
Learned Probabilities

$P(w \mid z, \beta)$

<table>
<thead>
<tr>
<th>Tag</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>0.5</td>
</tr>
<tr>
<td># movie</td>
<td>0.2</td>
</tr>
<tr>
<td># review</td>
<td>0.1</td>
</tr>
<tr>
<td># wallpaper</td>
<td>0.1</td>
</tr>
<tr>
<td># photos</td>
<td>0.1</td>
</tr>
</tbody>
</table>

$P(z \mid \theta)$

**Movie**
- final fantasy
  - Movie: 0.5
  - Game: 0.5

**Book**
- gone with the wind
  - Movie: 0.6
  - Book: 0.4

**Game**
- harry potter
  - Movie: 0.6
  - Book: 0.3
  - Game: 0.1
Inference

<table>
<thead>
<tr>
<th>#</th>
<th>250</th>
</tr>
</thead>
<tbody>
<tr>
<td># movie</td>
<td>100</td>
</tr>
<tr>
<td># wallpaper</td>
<td>20</td>
</tr>
<tr>
<td># walkthrough</td>
<td>10</td>
</tr>
<tr>
<td># review</td>
<td>10</td>
</tr>
</tbody>
</table>

**kung fu panda**

<table>
<thead>
<tr>
<th>#</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td># movie</td>
<td>150</td>
</tr>
<tr>
<td># summary</td>
<td>60</td>
</tr>
<tr>
<td># review</td>
<td>40</td>
</tr>
<tr>
<td># book</td>
<td>80</td>
</tr>
</tbody>
</table>

**beautiful mind**

<table>
<thead>
<tr>
<th>#</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td># movie</td>
<td>150</td>
</tr>
<tr>
<td># summary</td>
<td>60</td>
</tr>
<tr>
<td># review</td>
<td>40</td>
</tr>
<tr>
<td># book</td>
<td>80</td>
</tr>
</tbody>
</table>

**Movie, Game?**

<table>
<thead>
<tr>
<th>kung fu panda</th>
<th>Movie 0.9</th>
<th>Game 0.1</th>
</tr>
</thead>
</table>

**Movie, Book?**

<table>
<thead>
<tr>
<th>beautiful mind</th>
<th>Movie 0.7</th>
<th>Book 0.3</th>
</tr>
</thead>
</table>
Data Set

- Search log data from Bing: 12 million unique queries
- Four classes: book, movie, game, music
- 180 labeled training named-entities
- 60 labeled test named-entities
Accuracy of NERQ
Extension to Click-through Data

```
\# imdb.com 700
\# amazon.com 400
\# wikipedia.org 200
\# 100
\# movie imdb.com 300
\# movie 100
\# book amazon.com 650
\# book wikipedia.org 50
\# book 20
```

6/19/2009
Context Aware Search
Context-Aware Search

• Helping users accurately and quickly find information based on the context of search
Towards Context-Aware Search by Learning A Very Large Variable Length Hidden Markov Model from Search Logs

Cao et al. WWW 2009
Context of Search

• User usually conducts multiple related searches in a session

User --> query --> click --> query

Context

Current Search
## Context Information is Useful

- **Example of search sessions**

<table>
<thead>
<tr>
<th>SID</th>
<th>Search Sessions</th>
</tr>
</thead>
</table>
| S1  | Ford → Toyota → GMC → Allstate  
www.autohome.com |
| S2  | Ford cars → Toyota cars → GMC cars → Allstate  
www.autohome.com |
| S3  | Ford cars → Toyota cars → Allstate  
www.allstate.com |
| S4  | GMC → GMC dealers  
www.gmc.com |
Context Information is Useful

• 50% of users clicked car review site www.autohome.com after searching several car names.
Modeling Context by vHMM (variable length Hidden Markov Model)
Technical Details

• vlHMM Model
• Estimation Method
• Training Very Large vlHMM: Challenges and Solutions
• Application Tasks
  – Re-ranking
  – URL Suggestion
  – Query Suggestion
vHMM Model

• Given:
  – Set of hidden states \( \{s_1 \ldots s_{N_s}\} \);
  – Set of queries \( \{q_1 \ldots q_{N_q}\} \);
  – Set of URLs \( \{u_1 \ldots u_{N_u}\} \);
  – Maximal length \( T_{\text{max}} \) of sequence

• vHMM is probability model with
  – Transition probability distribution \( \Delta = \{P(s_i | s_j)\} \);
  – Probability distribution of initial states \( \Psi = \{P(s_i)\} \);
  – Probability distribution of emission from state \( \Lambda = \{P(q, U | S_j)\} \);
Estimation Method

- Let $X = \{O_1...O_N\}$ be set of training sessions
  - $O_n$ is a sequence of pairs $(q_{n,1}, U_{n,1}) \ldots (q_{n,T_n}, U_{n,T_n})$
  - $q_{n,t}$ and $U_{n,t}$ are t-th query and t-th set of clicked URLs respectively
  - $u_{n,t,k}$ is k-th URL in $U_{n,t}$.
- Goal is to find $\Theta^*$ such that
  \[
  \Theta^* = \arg \max_{\Theta} \ln P(X|\Theta) = \arg \max_{\Theta} \sum_n \ln P(O_n|\Theta)
  \]
- Learning with EM Algorithm
  \[
  Q(\Theta, \Theta^{(i-1)}) = \mathbb{E} \left[ \ln P(X, Y|\Theta)|X, \Theta^{(i-1)} \right]
  = \sum_{n,m} P(S_m|O_n, \Theta^{(i-1)}) \ln P(O_n, S_m|\Theta).
  \]
Training Very Large vHMM

• **Challenge 1:**
  - EM algorithm needs determined number of hidden states.
  - However, in our problem, hidden states correspond to search intents, for which the number is unknown.

• **Our Solution:**
  - Conduct clustering on click-bipartite graph and view clusters as hidden states.
Training Very Large vlHMM

• Challenge 2:
  – Search log data contains hundreds of millions of sessions.
  – It is impractical to train vlHMM from such huge training data on single machine.

• Our Solution:
  – Deploy learning task on distributed system under map-reduce model
Training Very Large vlHMM

• **Challenge 3:**
  – Each machine needs to hold the values of all parameters.
  – Since search log data contains millions of unique queries and URLs, the space of parameters is extremely large.

• **Our Solution:**
  – Employ special initialization strategy based on the clusters mined from click-through bipartite
Application Tasks

• Given observation $O$ consists of $q_1 \ldots q_t$ and $U_1 \ldots U_t$

• Document re-ranking:
  – Rank by $P(u \mid O) = \sum P(u \mid s_t) P(s_t \mid O)$

• Query suggestion & URL suggestion:
  – Suggest top k queries with $P(q \mid O) = \sum P(q \mid s_{t+1}) P(s_{t+1} \mid O)$
  – Recommend top k URLs with $P(u \mid O) = \sum P(u \mid s_{t+1}) P(s_{t+1} \mid O)$

• Advantages of our model: unified approach and accurate prediction.
Experiments

• Large-scale search log data from Bing
• Training data
  – 840,356,624 sessions
  – 1,812,563,301 search queries
  – 2,554,683,191 clicks
  – 151,869,102 unique queries
  – 114,882,486 unique URLs.
• Test data
  – 100,000 sessions held out from training data
Coverage

• For each test session \(<(q_1, U_1)\ldots(q_T, U_T)>\), vlHMM deals with each \(q_i\). When \(i > 1\), \(<(q_1, U_1)\ldots(q_{i-1}, U_{i-1})>\) is used as context.

• Total coverage is 58.3%.
• Denote test cases without context as Test0 and the others as Test1.
• For the covered cases in Test1, 25.5% contexts are matched.
Re-ranking

• Baseline:
  – Boost URLs with high click numbers given the query.

• Evaluation:
  – Sample 500 re-ranked URL pairs from Test0 and from the cases whose contexts are matched in Test1, respectively.
  
    – Each re-ranked URL pair is judged as Improved, Degraded or Unsure by three experts.
Experimental Results on Re-ranking

Better re-ranking by vlHMM than Baseline on (a) Test0 and (b) Test1.
## Examples of Re-ranking

<table>
<thead>
<tr>
<th></th>
<th>Context</th>
<th>Test query</th>
<th>Re-ranked document pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>online games</td>
<td>Disney channel</td>
<td>Games Disney Channel&lt;br&gt;<a href="http://tv.disney.go.com/disneychannel/games/index.html">http://tv.disney.go.com/disneychannel/games/index.html</a>&lt;br&gt;Disney Channel&lt;br&gt;<a href="http://www.disney.go.com/disneychannel">http://www.disney.go.com/disneychannel</a></td>
</tr>
<tr>
<td></td>
<td>▼</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><a href="http://games.yahoo.com">http://games.yahoo.com</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><a href="http://www.miniclip.com">http://www.miniclip.com</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>▼</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><a href="http://www.ask.com">http://www.ask.com</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Visited the homepage of Ask Jeeves**
- **Boosted the URL about the history of Ask Jeeves**
- **Searched for games**
- **Boosted the URL about game**
URL Suggestion

• Baseline:
  – Recommend the URLs with high click numbers following the current query.

• Evaluation:
  – “Leave-one-out" method: given \( (q_1, U_1) \ldots (q_T, U_T) \), we use \( q_{T-1} \) as the test query and consider \( U_T \) as the ground truth.

  – Suppose set of recommended URLs is \( R \), precision is \( \frac{|R \cap U_T|}{|R|} \) and recall is \( \frac{|R \cap U_T|}{|U_T|} \).
Experimental Results on URL Suggestion

Precision and recall of vlHMM and Baseline.
Example of URL Suggestion

Searched for online store about electronics

<table>
<thead>
<tr>
<th>Context:</th>
<th>circuit city → <a href="http://www.circuitcity.com">http://www.circuitcity.com</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Test query:</td>
<td>Walmart</td>
</tr>
<tr>
<td>URL recommendation</td>
<td></td>
</tr>
<tr>
<td>vlHMM</td>
<td>Baseline2</td>
</tr>
</tbody>
</table>

Online store about electronics
Online store about equipments
Query Suggestion

• Baseline:
  – CACB, a context-aware concept based approach of query suggestion.

• Evaluation:
  – The results of two approaches are comparable since they both consider contexts.
  – However, ratio of matched contexts is increased by 55% by vlHMM.
Summary
Summary

• Whole Page Relevance
  – Global Ranking Using Continuous CRF

• Concept based Search
  – Named Entity Recognition in Query

• Context-aware Search
  – Context Modeling Using Variable Length HMM