Multi-Dimensional Matrix Feature Merger and Tree Kernels

Apoorv Agarwal

IBM Research

September 1, 2011
Part I: Multi-Dimensional Matrix Features

1. MDM features
2. Impact on Full DeepQA system for Jeopardy!
3. Joint work with Jennifer, Adam, Bill, Aditya, Siddharth
Part I: Multi-Dimensional Matrix Features

1. MDM features
2. Impact on Full DeepQA system for Jeopardy!
3. Joint work with Jennifer, Adam, Bill, Aditya, Siddharth

Part II: Tree Kernel Package

1. Tree decoration operations in the package
2. Preliminary results on RTE and Event detection
3. Joint work with Alfio, Siddharth, Alessandro, Jennifer, Aditya
DeepQA System

Q

CA₁  CA₂  ...  CAₙ
DeepQA System

Q

CA_1

p_1 \quad p_2 \quad \cdots \quad p_{m_1}

\cdots

CA_n

p_1 \quad p_2 \quad \cdots \quad p_{m_n}

Typical merger:

\text{MAX: } CA \rightarrow R;
\text{MAX}(p_1, p_2, \ldots, p_m) = \ldots

Training data:

<Q_1, CA_1, F>, <Q_1, CA_2, T>, \ldots, <Q_1, CA_{n_1}, F>

<Q_2, CA_1, T>, <Q_2, CA_2, F>, \ldots, <Q_2, CA_{n_2}, F>, \ldots
DeepQA System

Typical merger: $MAX : CA \rightarrow \mathbb{R}; MAX(p_1, p_2, \ldots, p_m) = 9.3$

Training data:

$< Q_1, CA_1, F >, < Q_1, CA_2, T >, \ldots, < Q_1, CA_{n_1}, F >$

$< Q_2, CA_1, T >, < Q_2, CA_2, F >, \ldots, < Q_2, CA_{n_2}, F > \ldots$
**Question:** NOW THAT'S BIG!: This largest land animal also has the largest ears at 5 feet across each

**Passage p1:** Apart from their massive size their most striking features are a long trunk, or proboscis, a flexible nose strong enough to lift objects, their huge ivory tusks, and their large flapping ears, used to keep them cool. Not only is the African elephant the largest animal you’ll find in Etosha it’s the largest living land animal in the world.
**Question:** NOW THAT’S BIG!: This largest land animal also has the largest ears at 5 feet across each

**Passage p1:** Apart from their massive size their most striking features are a long trunk, or proboscis, a flexible nose strong enough to lift objects, their huge ivory tusks, and their large flapping ears, used to keep them cool. Not only is the African elephant the largest animal you’ll find in Etosha it’s the largest living land animal in the world.
Merger for LFACS scorer

**Question:** NOW THAT’S BIG!: This largest land animal also has the largest ears at 5 feet across each

**Passage p3:** African elephants also have larger ears, which are used for cooling their massive bodies. The African elephant is the largest living land mammal

<table>
<thead>
<tr>
<th></th>
<th>large</th>
<th>land</th>
<th>animal</th>
<th>also</th>
<th>have</th>
<th>large</th>
<th>ear</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1$</td>
<td>1.95</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
</tr>
<tr>
<td>$p_2$</td>
<td>1.95</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>0.05</td>
<td>xx</td>
<td>3.98</td>
</tr>
<tr>
<td>$p_3$</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>0.05</td>
<td>0.05</td>
<td>1.95</td>
<td>3.98</td>
</tr>
</tbody>
</table>

**Table: Focus Matrix $M$**

- Ignores NonFocus match
- $LFACS = max(sum(M^T)) = 6.03$
1. Can LFACS NonFocus score be incorporated in a meaningful way?
Goal

1. Can LFACS NonFocus score be incorporated in a meaningful way?
2. Can we leverage from looking at all the passages for a candidate answer together?

Answer to these questions is Yes!

Main Result: Achieved a significant improvement ($p < 0.88\%$) on the full system for hit-list-normalization phase.
Goals

1. Can LFACS NonFocus score be incorporated in a meaningful way?
2. Can we leverage from looking at all the passages for a candidate answer together?

Answer to these questions is Yes!
Goals

1. Can LFACS NonFocus score be incorporated in a meaningful way?
2. Can we leverage from looking at all the passages for a candidate answer together?

Answer to these questions is Yes!

Main Result: Achieved a significant improvement ($p < 1.88\%$) on the full system for hit-list-normalization phase.
**Question:** NOW THAT’S BIG!: This largest land animal also has the largest ears at 5 feet across each

**Passage p2:** ELEPHANT FACTOIDS: Elephants are of two types, African and Asian. African elephants are larger and have much bigger ears. The largest elephants grow to be about 12 feet tall and weigh about 9 tons. Elephant tusks grow throughout an elephant’s life and can weigh more than 200 pounds.
New Merger: Transfer scores from NonFocus Matrix

<table>
<thead>
<tr>
<th></th>
<th>large</th>
<th>land</th>
<th>animal</th>
<th>also</th>
<th>have</th>
<th>large</th>
<th>ear</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1$</td>
<td>1.95</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
</tr>
<tr>
<td>$p_2$</td>
<td>1.95</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>0.05</td>
<td>xx</td>
<td>3.98</td>
</tr>
<tr>
<td>$p_3$</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>0.05</td>
<td>0.05</td>
<td>1.95</td>
<td>3.98</td>
</tr>
</tbody>
</table>

Table: Focus Matrix $M$

<table>
<thead>
<tr>
<th></th>
<th>large</th>
<th>land</th>
<th>animal</th>
<th>also</th>
<th>have</th>
<th>large</th>
<th>ear</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1$</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>1.95</td>
<td>3.98</td>
</tr>
<tr>
<td>$p_2$</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
</tr>
<tr>
<td>$p_3$</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
</tr>
</tbody>
</table>

Table: NonFocus Matrix $NM$
New Merger: Transfer scores from NonFocus Matrix

### Table: Focus Matrix $M$

<table>
<thead>
<tr>
<th></th>
<th>large</th>
<th>land</th>
<th>animal</th>
<th>also</th>
<th>have</th>
<th>large</th>
<th>ear</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1$</td>
<td>1.95</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
</tr>
<tr>
<td>$p_2$</td>
<td>1.95</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>0.05</td>
<td>xx</td>
<td>3.98</td>
</tr>
<tr>
<td>$p_3$</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>0.05</td>
<td>0.05</td>
<td>1.95</td>
<td>3.98</td>
</tr>
</tbody>
</table>

### Table: NonFocus Matrix $NM$

<table>
<thead>
<tr>
<th></th>
<th>large</th>
<th>land</th>
<th>animal</th>
<th>also</th>
<th>have</th>
<th>large</th>
<th>ear</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1$</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>1.95</td>
<td>3.98</td>
</tr>
<tr>
<td>$p_2$</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
</tr>
<tr>
<td>$p_3$</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
</tr>
</tbody>
</table>
New Merger: Transfer scores from NonFocus Matrix

Table: Focus Matrix $M$

<table>
<thead>
<tr>
<th></th>
<th>large</th>
<th>land</th>
<th>animal</th>
<th>also</th>
<th>have</th>
<th>large</th>
<th>ear</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1$</td>
<td>1.95</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>1.95</td>
<td>3.98</td>
</tr>
<tr>
<td>$p_2$</td>
<td>1.95</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>0.05</td>
<td>xx</td>
<td>3.98</td>
</tr>
<tr>
<td>$p_3$</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>0.05</td>
<td>0.05</td>
<td>1.95</td>
<td>3.98</td>
</tr>
</tbody>
</table>

Table: NonFocus Matrix $NM$

<table>
<thead>
<tr>
<th></th>
<th>large</th>
<th>land</th>
<th>animal</th>
<th>also</th>
<th>have</th>
<th>large</th>
<th>ear</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1$</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>1.95</td>
<td>3.98</td>
</tr>
<tr>
<td>$p_2$</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
</tr>
<tr>
<td>$p_3$</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
</tr>
</tbody>
</table>

New LFACS score = 7.88
New Features: Multi-Dimensional Matrix Features (MDM)

<table>
<thead>
<tr>
<th></th>
<th>large</th>
<th>land</th>
<th>animal</th>
<th>also</th>
<th>have</th>
<th>large</th>
<th>ear</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1$</td>
<td>1.95</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>1.95</td>
<td>3.98</td>
</tr>
<tr>
<td>$p_2$</td>
<td>1.95</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>0.05</td>
<td>xx</td>
<td>3.98</td>
</tr>
<tr>
<td>$p_3$</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>0.05</td>
<td>0.05</td>
<td>1.95</td>
<td>3.98</td>
</tr>
</tbody>
</table>

Table: Matrix $M$ in final merger, one for each scorer (LFACS, SkipBigram, TextualAlignment, PassageTermMatch)

- \( \min, \max, \sum, \text{mean}, \text{stddev}, \text{num-max of } \sum(M) \text{ and } \sum(M^T) \)
- \#ofColumns, \#ofNon-ZeroColumns
- All these features and their normalized versions
Result 1: Transferring score helps

Features derived from combined $M$ has better Inform Analysis
Result 2: Looking at all the passages together helps
Comparison with existing effort

![Inform Analysis Chart](chart.png)
Comparison with existing effort

![Inform Analysis](chart1)

- LFACS
- LFACS + Class0.5
- #MatchingPassages

![% Accuracy Gain and Precision @ 70 Gain](chart2)

- LFACS + Class0.5
- LFACS MDM features
Inform Analysis of MDM features and Existing features
Main Result: Full system improvement (only hit-list-normalization phase)

- Experimental set-up: 2011 Week 19, Test set T14, only hit-list-normalization phase
Main Result: Full system improvement (only hit-list-normalization phase)

- Experimental set-up: 2011 Week 19, Test set T14, only hit-list-normalization phase
- Questions gained = 28, significant with $p < 1.88\%$
Future Work

- **Running**: Full system all phases
- **Running**: Feature selection
Future Work

- **Running**: Full system all phases
- **Running**: Feature selection
- Look for more linguistically motivated transfer of scores
Future Work

- **Running**: Full system all phases
- **Running**: Feature selection
- Look for more linguistically motivated transfer of scores
- Current system is highly optimized for T14. Test on T20
Future Work

- **Running**: Full system all phases
- **Running**: Feature selection
- Look for more linguistically motivated transfer of scores
- Current system is highly optimized for T14. Test on T20
- Look at all candidate answer matrices together?
Conclusion

- Non-focus score can be incorporated in a meaningful way
- There is value in looking at all the passages together
Questions?

Thanks to Adam and Bill!
Part II: Tree Kernels

- Design of Tree and Word Kernels
- Application 1: Recognizing Textual Entailment (RTE3, with Alfio, Aditya, Jennifer)
- Application 2: Event Detection (with Siddharth)
Introduction to Kernel Methods

- What is a Kernel?
  \[ K : X \times X \rightarrow \mathbb{R} \]

- What is a Tree Kernel?
  \[ K : T \times T \rightarrow \mathbb{R} \]
  \[ K(T_1, T_2) = \sum_s h_s(T_1)h_s(T_2) \]
## A Simplified Example

A simplified example of trees and their corresponding matrix features is shown below. The table represents the trees and their associated columns for A, B, C, and D. Each tree is illustrated with a binary structure, and the columns represent different paths through the trees.

<table>
<thead>
<tr>
<th>Trees</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th></th>
<th>A</th>
<th>D</th>
<th>C</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B</td>
</tr>
</tbody>
</table>

Each row represents a tree, and the columns correspond to the paths through the tree. The tree structure is illustrated alongside the matrix representation.
## A Simplified Example

<table>
<thead>
<tr>
<th>Tree</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>C</th>
<th>A</th>
<th>A</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td></td>
<td></td>
<td>B</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
K(T_1, T_2) = 1 + 1 + 1 + 0 + 0 + 0 + 0 + 0 = 3
\]
Tree Kernel Package: Operations on a Single tree

was

Bikal sold

| to

| Rosneft
Tree Kernel Package: Operations on a Single tree

was
Bikal sold
  to
  Rosneft

was
ORG sold
  to
  Bikal to
  Rosneft
Tree Kernel Package: Operations on a Single tree

- **was**
  - **Bikal**
  - **sold**
  - **to**
  - **Rosneft**

- **was**
  - **ORG**
  - **sold**
  - **to**
  - **Bikal**
  - **Rosneft**

- **top**
  - **ORG**
  - **pred**
  - **subj**
  - **iobj**
  - **ORG**
  - **objprep**

- **Comm**
  - **goods-transfer**
  - **iobj**
  - **subj**
  - **Comm**
  - **goods-transfer**
  - **iobj**
  - **subj**
Tree Kernel Package: Operations on a Single tree

was

Bikal sold

to

Rosneft

was

ORG sold

to

Bikal to

ROS

Rosneft

top

ORG pred

subj

iobj

ORG

objprep

top

ORG Comm_goods-transfer

subj

iobj

ORG

objprep
Tree Kernel Package: Operations on a Single tree

was
Bikal sold
to Rosneft

was
ORG sold
to Bikal to
ORG Rosneft

top
ORG pred
subj iobj
ORG objprep

top
Comm_goods-transfer
subj iobj
ORG objprep

Comm_goods-transfer
subj iobj
Tree Kernel Package: Operations on a Single tree

Part I: Multi-Dimensional Matrix Features

Part II: Tree Kernels

was
Bikal sold
to Rosneft

was
ORG sold
to Bikal to
ORGBikal to
subj Rosneft

was
ORGPred
subj obj
subj Bikal
OBJ comm_goods-transfer
subj Bikal
OBJ objprep
ORGRosneft
OBJ comm_goods-transfer
OBJ objprep
ORGRosneft
OBJ comm_goods-transfer
OBJ objprep
ORGRosneft
Hypothesis: **Bikal** was sold to **Rosneft**
Text: **Bikal** was bought by **Rosneft**

<table>
<thead>
<tr>
<th>Representation</th>
<th>F1-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>was</code> <code>Bikal</code> <code>sold</code> <code>to</code> <code>Rosneft</code> <code>was</code> <code>Bikal</code> <code>bought</code> <code>by</code> <code>Rosneft</code></td>
<td>54.5</td>
</tr>
<tr>
<td><code>was</code> <code>PLHD</code> <code>sold</code> <code>to</code> <code>Bikal</code> <code>by</code> <code>Rosneft</code> <code>was</code> <code>PLHD</code> <code>bought</code> <code>by</code> <code>Rosneft</code></td>
<td>65.6</td>
</tr>
</tbody>
</table>
### Abstraction helps – Empirical Evidence (RTE)

<table>
<thead>
<tr>
<th>Representation</th>
<th>F1-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>was PLHD sold Bikal to Bikal by PLHD Rosneft</td>
<td>65.6</td>
</tr>
<tr>
<td>top PLHD pred subj iobj subj objprep</td>
<td>66.25</td>
</tr>
</tbody>
</table>
Path Enclosed Tree (Moschitti 2004): For relation extraction, F-measure goes up from 40% to 68%
Tree Kernel Package: Pruning 2

- Animal
  - Large
  - Land
- At
  - 5-foot
  - Across
  - Each

- Ear
  - Large

- Animal
  - Large
- African
  - Type
  - Asian
  - Too
  - Much
- Have
  - Ear
  - Big
Motivation/Hypothesis: Relevant part of the XSG lies around and above the “matching” terms

For event detection, F-measure goes up from 56% to 66.58%
Motivation is to capture how text and hypothesis (or question and answer) overlap with each other
**Tree Kernel Package: Operations on a Pair of trees (Merge)**

- Motivation is to capture how text and hypothesis (or question and answer) overlap with each other.
- Basic Idea: merging two trees results in a graph. No usable graph kernel. Take projections of the graph on certain nodes (e.g. focus/matching nodes).
Tree Kernel Package: Operations on a Pair of trees (Merge)

- Motivation is to capture how text and hypothesis (or question and answer) overlap with each other.
- Basic Idea: merging two trees results in a graph. No usable graph kernel. Take projections of the graph on certain nodes (e.g. focus/matching nodes).
- Fairly complicated and no results for it right now. Can discuss if people are interested and if there is time.
Conclusion

- Implemented a comprehensive package for decorating trees
Conclusion

- Implemented a comprehensive package for decorating trees
- For RTE3, we gain over 5% over the LSA baseline
• Implemented a comprehensive package for decorating trees
• For RTE3, we gain over 5% over the LSA baseline
• Preliminary results on Event detection:
Future Work

- Create representations and apply to various tasks
Future Work

- Create representations and apply to various tasks
- Develop a new tree kernel that skips over nodes/sub-structures
Future Work

- Create representations and apply to various tasks
- Develop a new tree kernel that skips over nodes/sub-structures
- Analyzed in detail the RTE task. Looked at dependency paths between matching nodes in text and hypothesis. It seems that dependency paths encode all the information we need to decide entailment. Come up with a meaningful similarity function between all pairs of dependency paths. For a detailed analysis and examples look at twiki page:
  /twiki/bin/view/BlueJay/SequenceKernelsForRTE
Figure: Result of running a block modeling algorithm with number of clusters = 6 on the network.