Multi-dimensional feature merger for Question Answering

Apoorv Agarwal
Columbia University
Dec 13th 2012

Joint work with: William Murdock, Jennifer Chu-carroll, Adam Lally, Aditya Kalyanpur (IBM Research)

Question Answering

Question Answering

(Question) This large land animal also has large ears.

Question Answering

(Question) This large land animal also has large ears.

(Candidate Answer) African Elephant

Question Answering

(Question) This large land animal also has large ears.

(Candidate Answer) African Elephant

(Supporting passages)

Question Answering

(Question) This large land animal also has large ears.

(Candidate Answer) African Elephant

(Supporting passages)
1) African Elephant is a large land animal

Question Answering

(Question) This large land animal also has large ears.

(Candidate Answer) African Elephant

(Supporting passages)
1) African Elephant is a large land animal
2) African Elephants have large ears

Question Answering

(Question) This large land animal also has large ears.

(Candidate Answer) African Elephant

(Supporting passages)
1) African Elephant is a large land animal
2) African Elephants have large ears

-- Watson tries to find the best passage that supports the answer

Question Answering

(Question) This large land animal also has large ears.

(Candidate Answer) African Elephant

(Supporting passages)
1) African Elephant is a large land animal
2) African Elephants have large ears

-- Watson tries to find the best passage that supports the answer

-- But all the information may not be present in one passage

Question Answering

(Question) This large land animal also has large ears.

(Candidate Answer) African Elephant

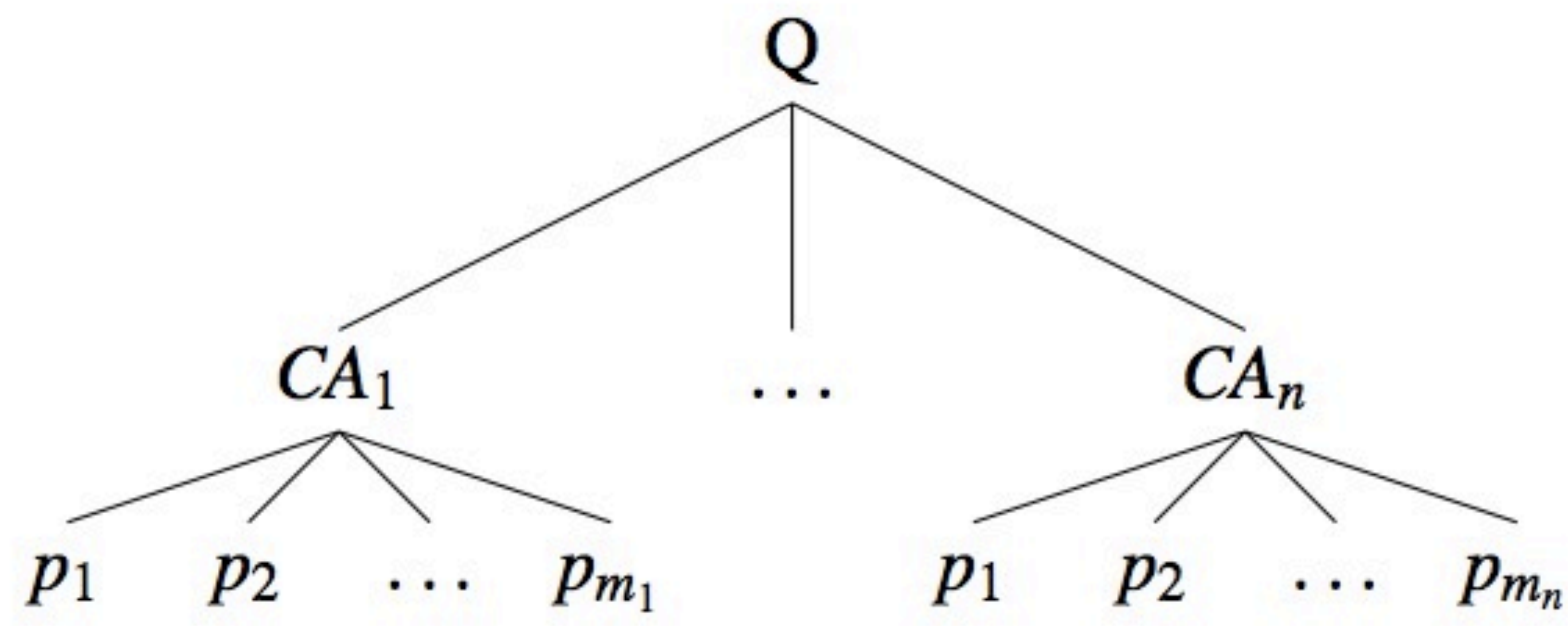
(Supporting passages)
1) African Elephant is a large land animal
2) African Elephants have large ears

-- Watson tries to find the best passage that supports the answer

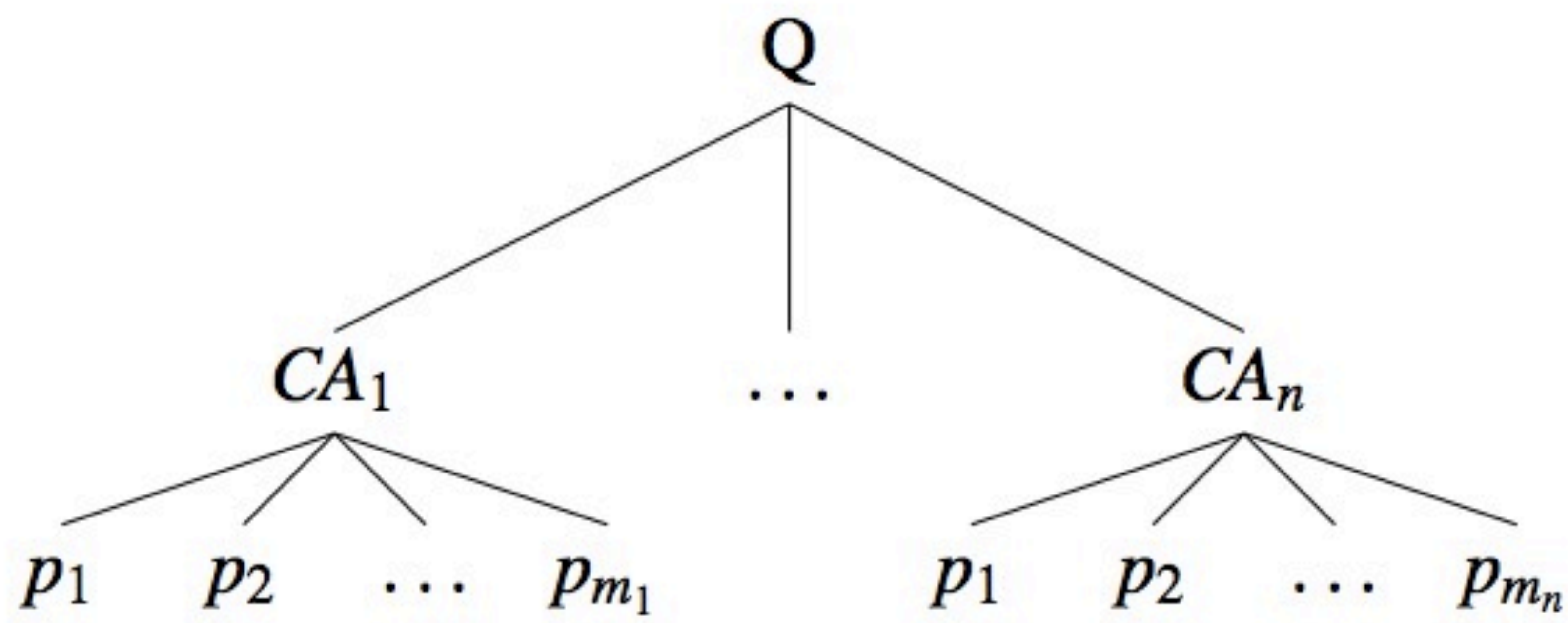
-- But all the information may not be present in one passage

-- In this work we present a framework to overcome this limitation

Question Answering Set-up



Question Answering Set-up



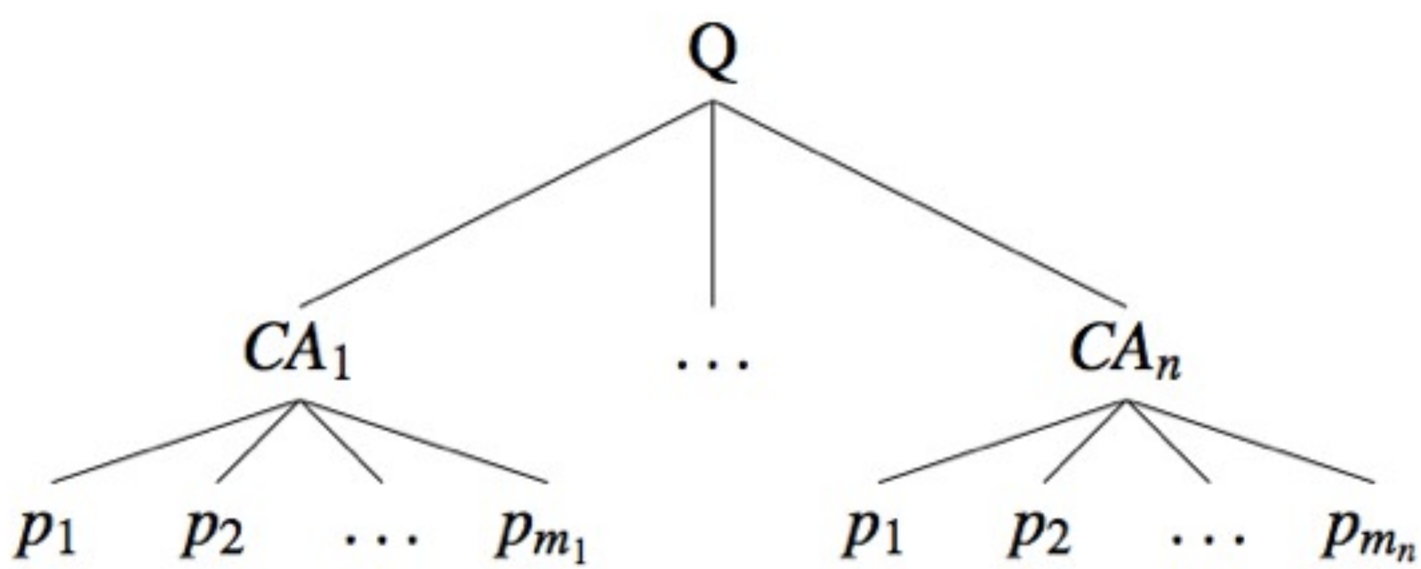
$$< Q_1, CA_1, -1 >, < Q_1, CA_2, -1 >, \ldots, < Q_1, CA_i, 1 >, \ldots, < Q_1, CA_{n_1}, -1 >$$

$$< Q_2, CA_1, -1 >, < Q_2, CA_2, -1 >, \ldots, < Q_2, CA_j, 1 >, \ldots, < Q_2, CA_{n_2}, -1 >$$

$$\ldots$$

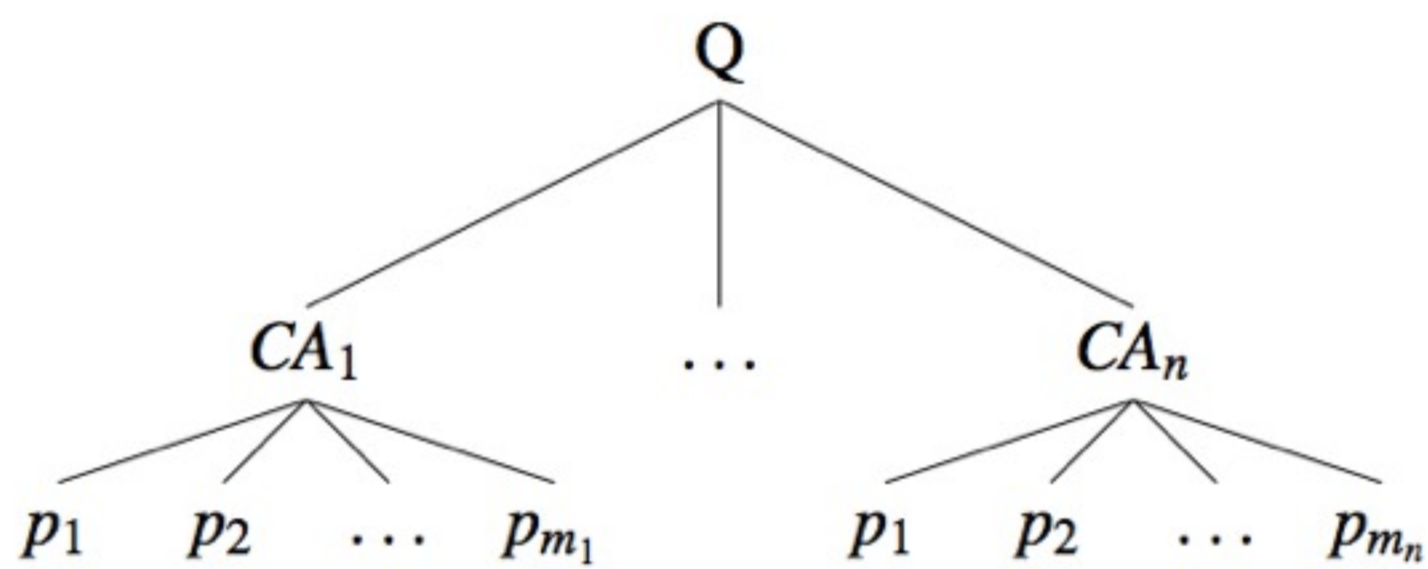
$$< Q_m, CA_1, -1 >, < Q_m, CA_2, -1 >, \ldots, < Q_m, CA_k, 1 >, \ldots, < Q_m, CA_{n_m}, -1 >$$

Outline



Outline

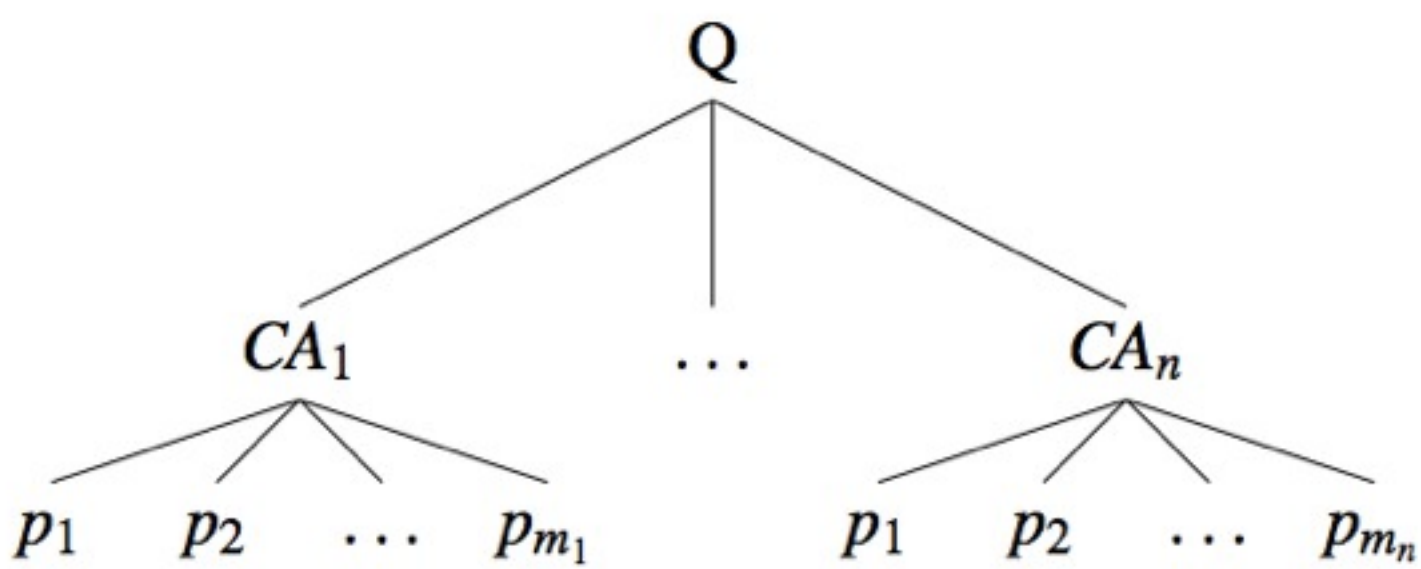
1. Brief overview of passage scorers in the system



Outline

1. Brief overview of passage scorers in the system

2. How features produced by passage scorers per passage are combined to get one feature for a candidate -- feature merger

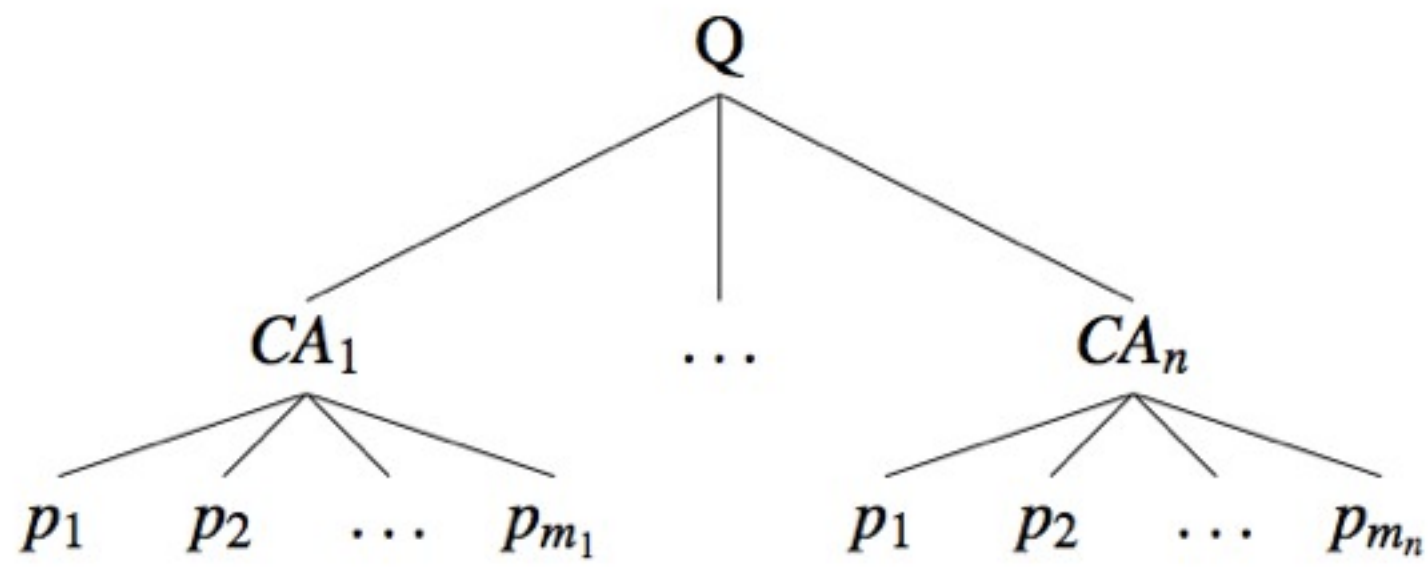


Outline

1. Brief overview of passage scorers in the system

2. How features produced by passage scorers per passage are combined to get one feature for a candidate -- feature merger

3. A new feature merger framework



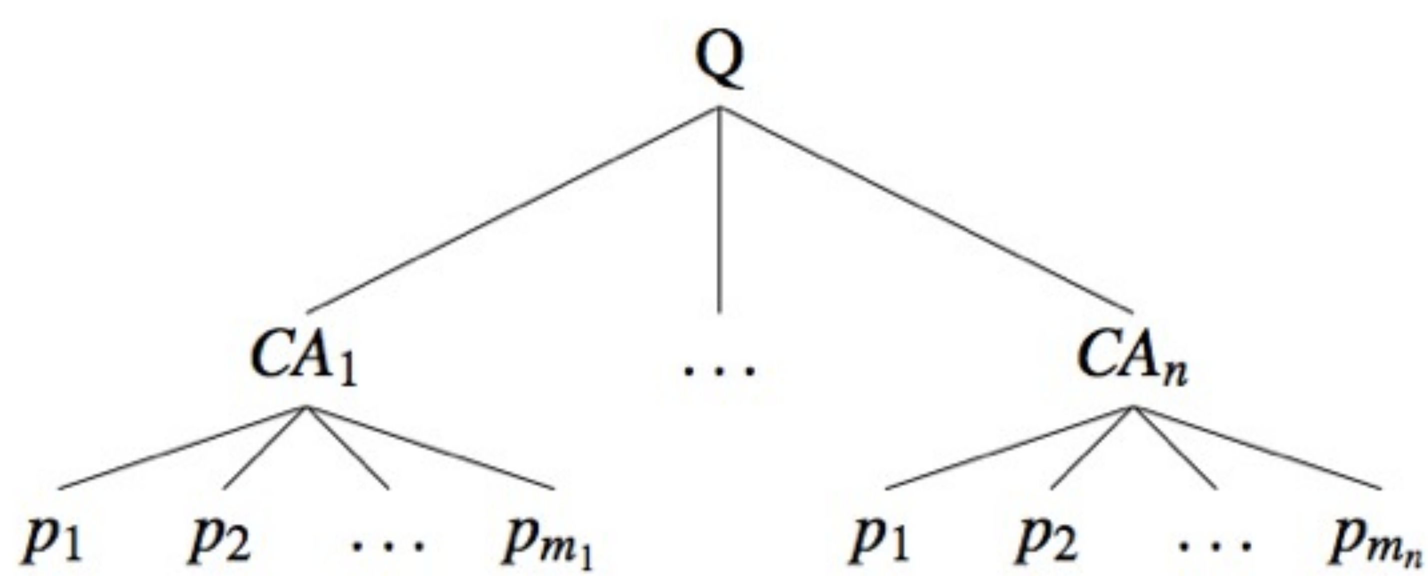
Outline

1. Brief overview of passage scorers in the system

2. How features produced by passage scorers per passage are combined to get one feature for a candidate -- feature merger

3. A new feature merger framework

4. Experiments and results



Passage Scoring Features

Passage Scoring Features

1. Capture how well Question matches supporting passages

Passage Scoring Features

1. Capture how well Question matches supporting passages

2. Four types:

Passage Scoring Features

1. Capture how well Question matches supporting passages

2. Four types:
   a. **Passage Term Match**: Assigns a score based on which question terms are included in the passage, regardless of word order or grammatical relationship.

Passage Scoring Features

1. Capture how well Question matches supporting passages

2. Four types:
   a. **Passage Term Match**: Assigns a score based on which question terms are included in the passage, regardless of word order or grammatical relationship.

   b. **Skip Bigram**: pairs of terms that are connected or nearly connected in the syntactic-semantic structure

Passage Scoring Features

1. Capture how well Question matches supporting passages

2. Four types:
   a. **Passage Term Match**: Assigns a score based on which question terms are included in the passage, regardless of word order or grammatical relationship.

   b. **Skip Bigram**: pairs of terms that are connected or nearly connected in the syntactic-semantic structure

   c. **Textual Alignment**: how well the word order of the passage aligns with that of the question

Passage Scoring Features

1. Capture how well Question matches supporting passages

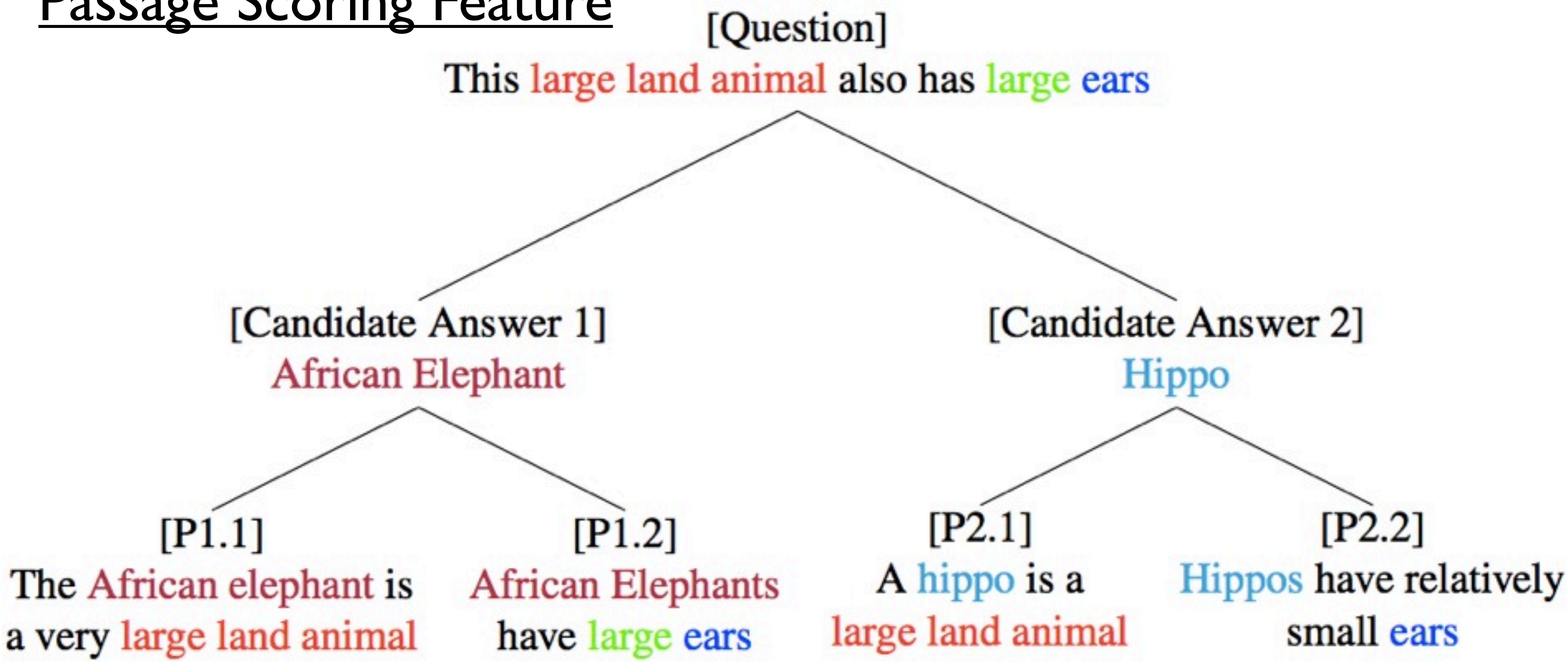
2. Four types:
   a. **Passage Term Match**: Assigns a score based on which question terms are included in the passage, regardless of word order or grammatical relationship.

   b. **Skip Bigram**: pairs of terms that are connected or nearly connected in the syntactic-semantic structure

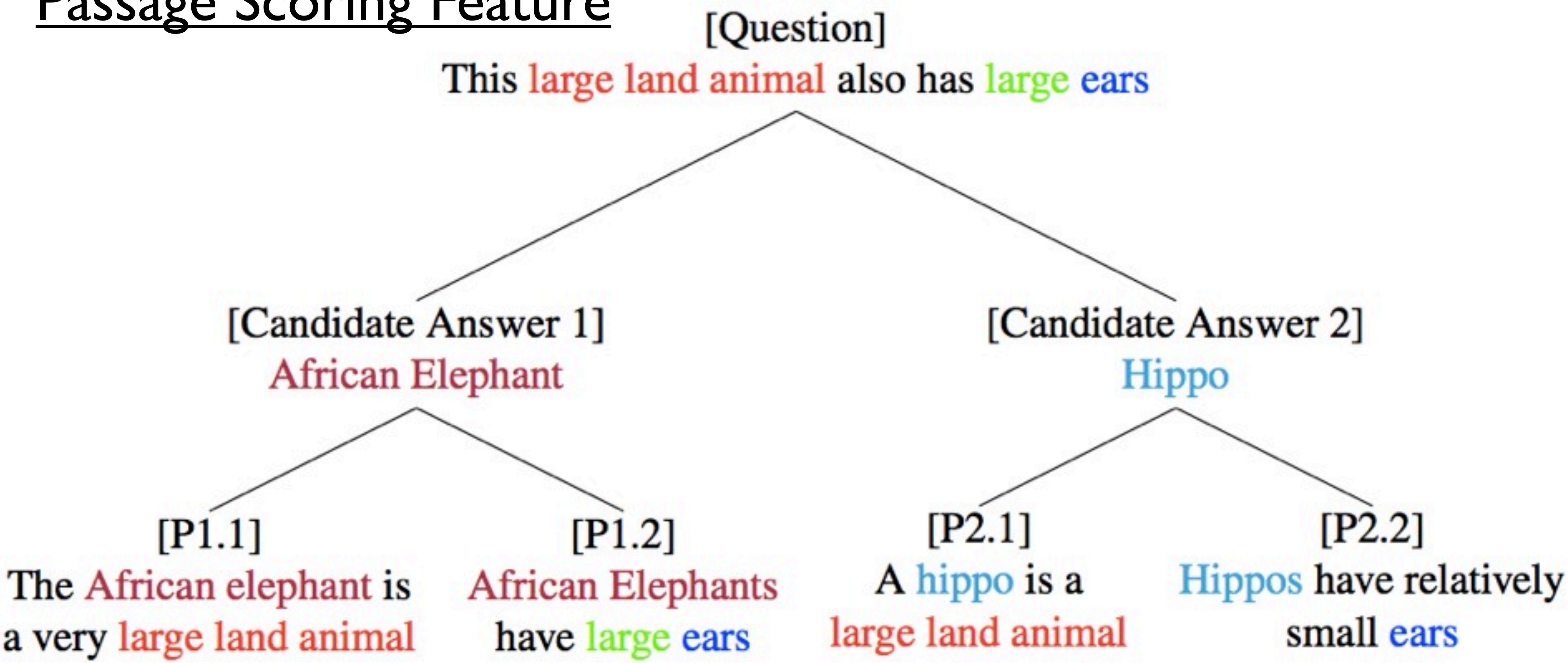
   c. **Textual Alignment**: how well the word order of the passage aligns with that of the question

   d. **Logical Form Answer Candidate Scorer (LFACS)**: Targets high-precision matching between the syntactic structures of passages and questions.

Passage Scoring Feature

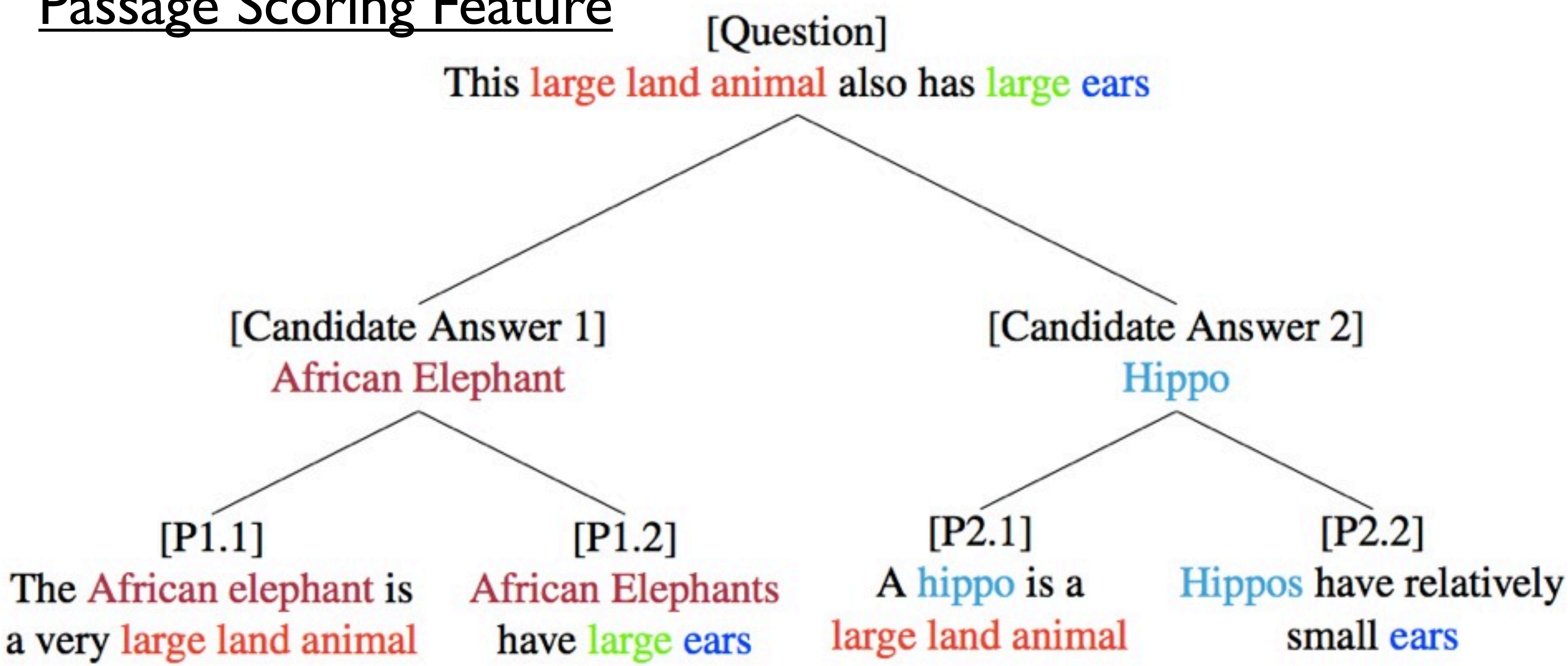


Passage Scoring Feature



Elephant	large	land	animal	large	ears
P1.1	1	1	1	0	0
P1.2	0.0	0.0	0.0	1	1

Passage Scoring Feature



Elephant	large	land	animal	large	ears
P1.1	1	1	1	0	0
P1.2	0.0	0.0	0.0	1	1

$$f = \max_i \left(\sum_j^N a_{i,j} \right)$$

Other Merger Strategies

Elephant	large	land	animal	large	ears
P1.1	1	1	1	0	0
P1.2	0	0	0	1	1

Other Merger Strategies

Elephant	large	land	animal	large	ears
P1.1	1	1	1	0	0
P1.2	0	0	0	1	1

1. MAX $\max_i(\sum_j^N a_{i,j})$

Other Merger Strategies

Elephant	large	land	animal	large	ears
P1.1	1	1	1	0	0
P1.2	0	0	0	1	1

1. MAX $\max_i(\sum_j^N a_{i,j})$ P1.1, score = 3

Other Merger Strategies

Elephant	large	land	animal	large	ears
P1.1	1	1	1	0	0
P1.2	0	0	0	1	1

1. MAX $\max_i (\sum_j^N a_{i,j})$ P1.1, score = 3
2. SUM $\sum_i \sum_j a_{i,j}$

Other Merger Strategies

Elephant	large	land	animal	large	ears
P1.1	1	1	1	0	0
P1.2	0	0	0	1	1

1. MAX $\max_i(\sum_j^N a_{i,j})$ P1.1, score = 3

2. SUM $\sum_i \sum_j a_{i,j}$ score = 5

Other Merger Strategies

Elephant	large	land	animal	large	ears
P1.1	1	1	1	0	0
P1.2	0	0	0	1	1

1. MAX $\max_i(\sum_j^N a_{i,j})$ P1.1, score = 3

2. SUM $\sum_i \sum_j a_{i,j}$ score = 5

3. Decaying Sum $\sum_i \alpha^i \sum_j a_{i,j} \quad 0 < \alpha < 1$

Other Merger Strategies

Elephant	large	land	animal	large	ears
P1.1	1	1	1	0	0
P1.2	0	0	0	1	1

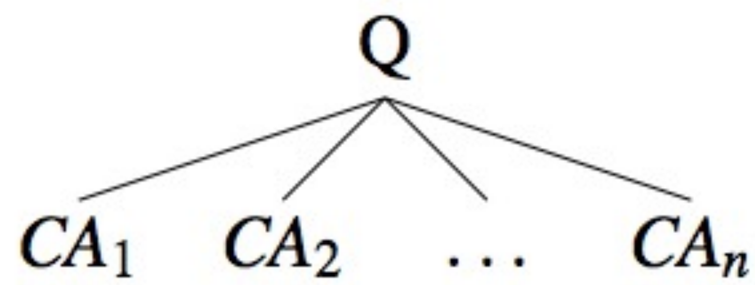
1. MAX $\max_i(\sum_j^N a_{i,j})$ P1.1, score = 3

2. SUM $\sum_i \sum_j a_{i,j}$ score = 5

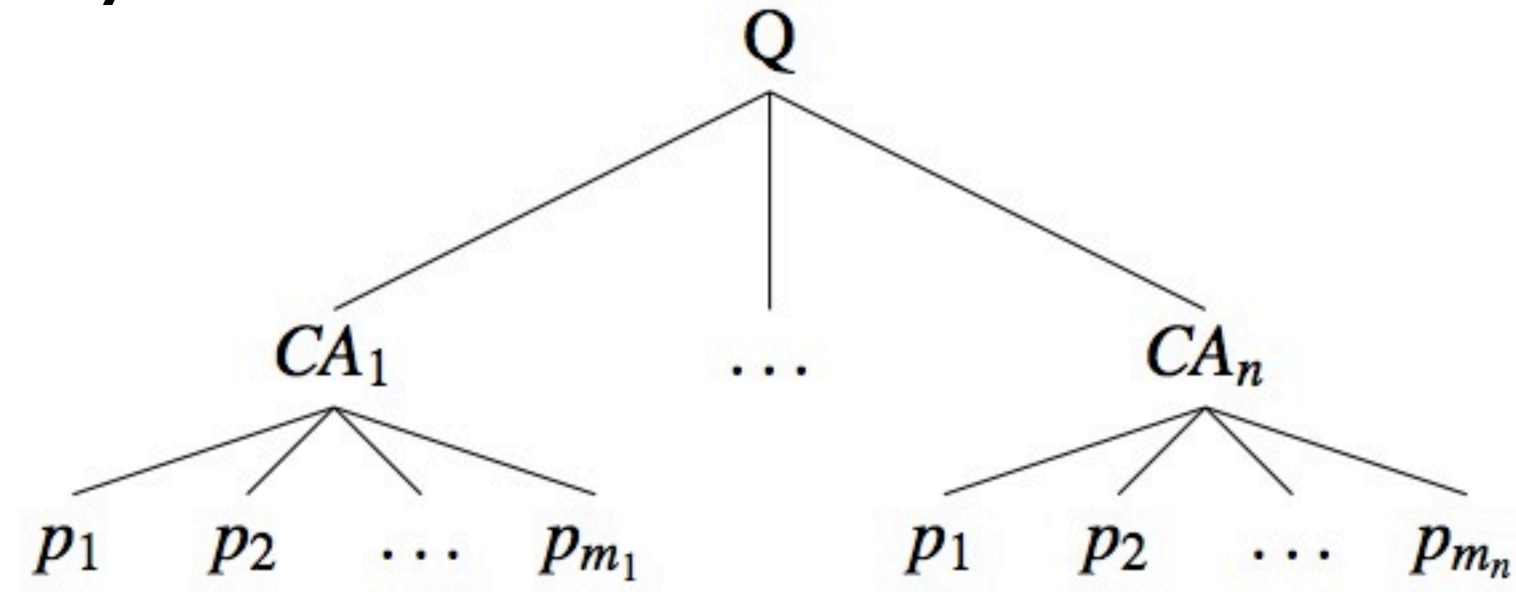
3. Decaying Sum $\sum_i \alpha^i \sum_j a_{i,j} \quad 0 < \alpha < 1$

Our contribution: Introduce a framework to capture the distribution of this matrix.

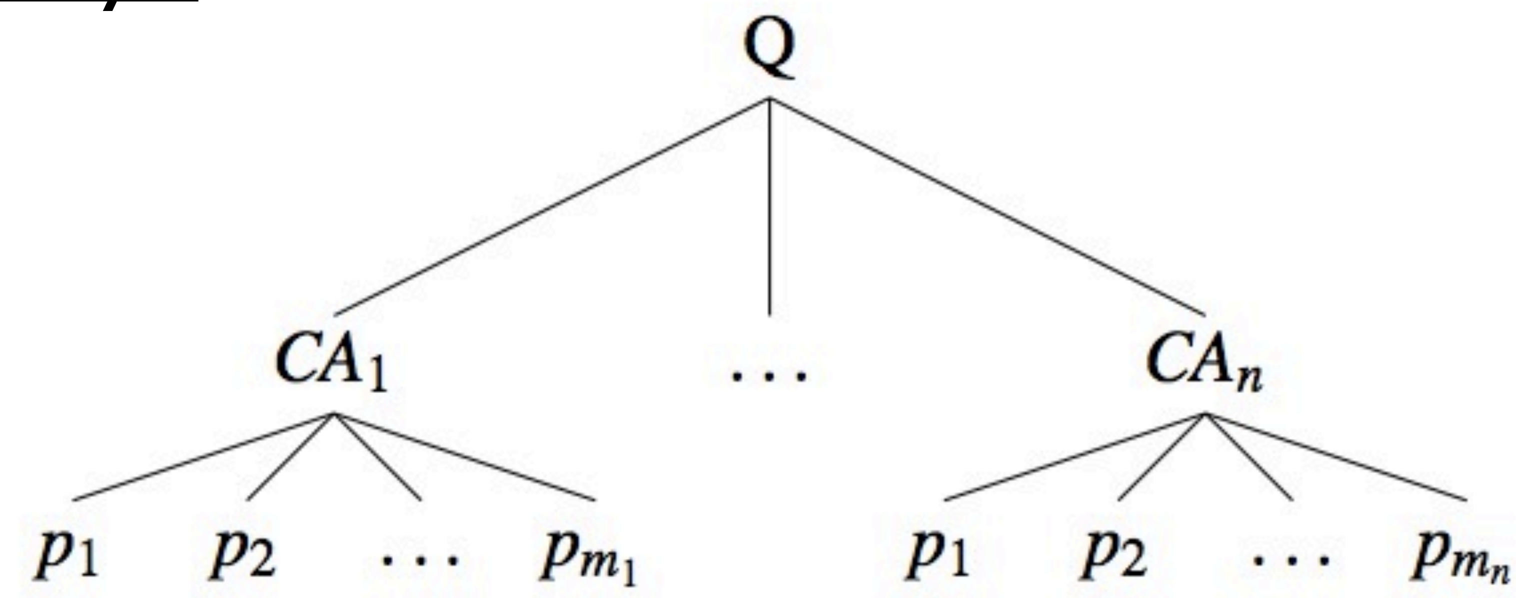
More Formally...



More Formally...

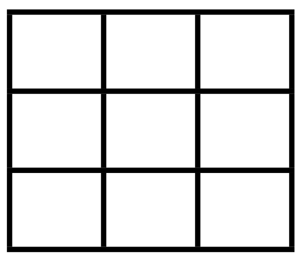


More Formally...

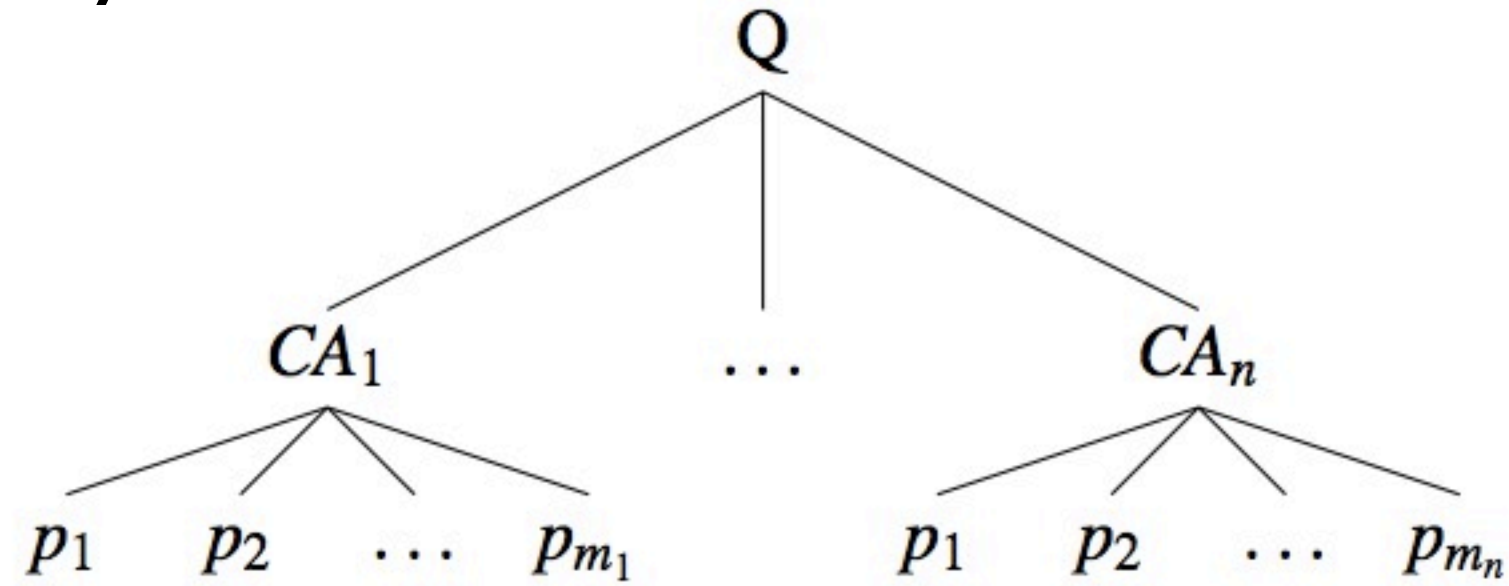


Question terms

Supporting passages

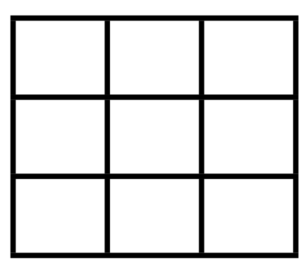


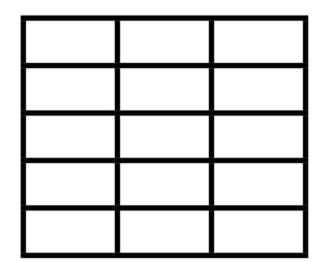
More Formally...



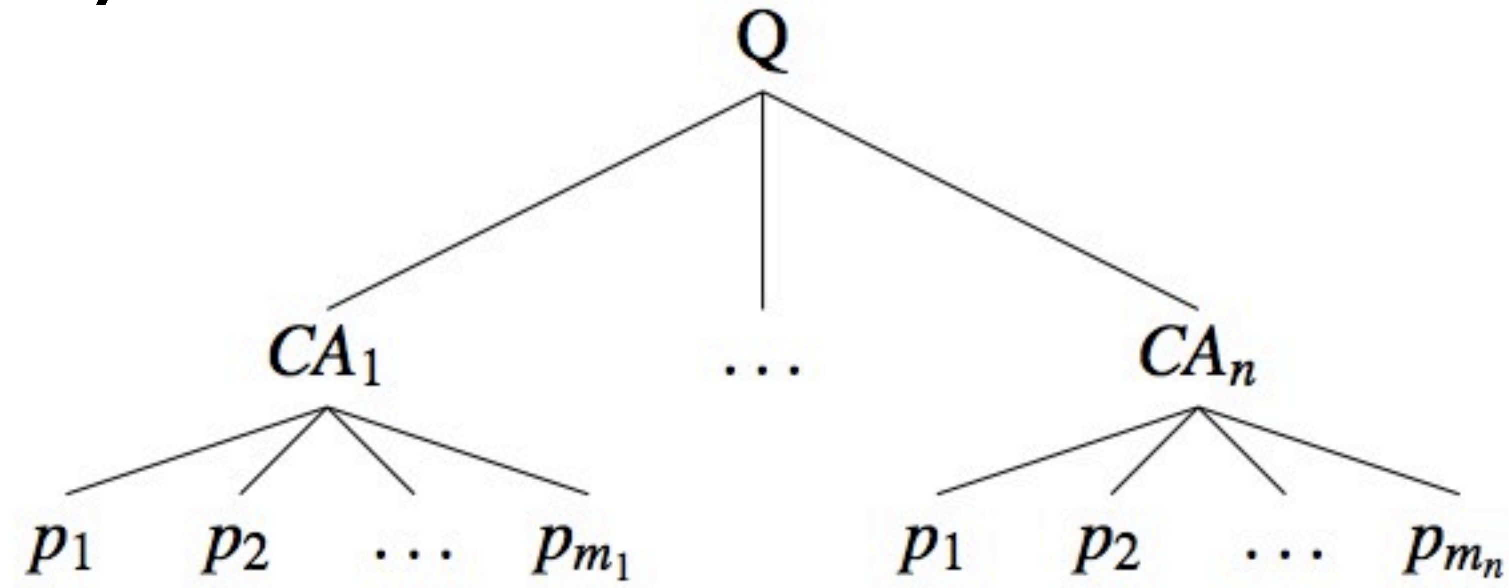
Question terms

Supporting passages





More Formally...



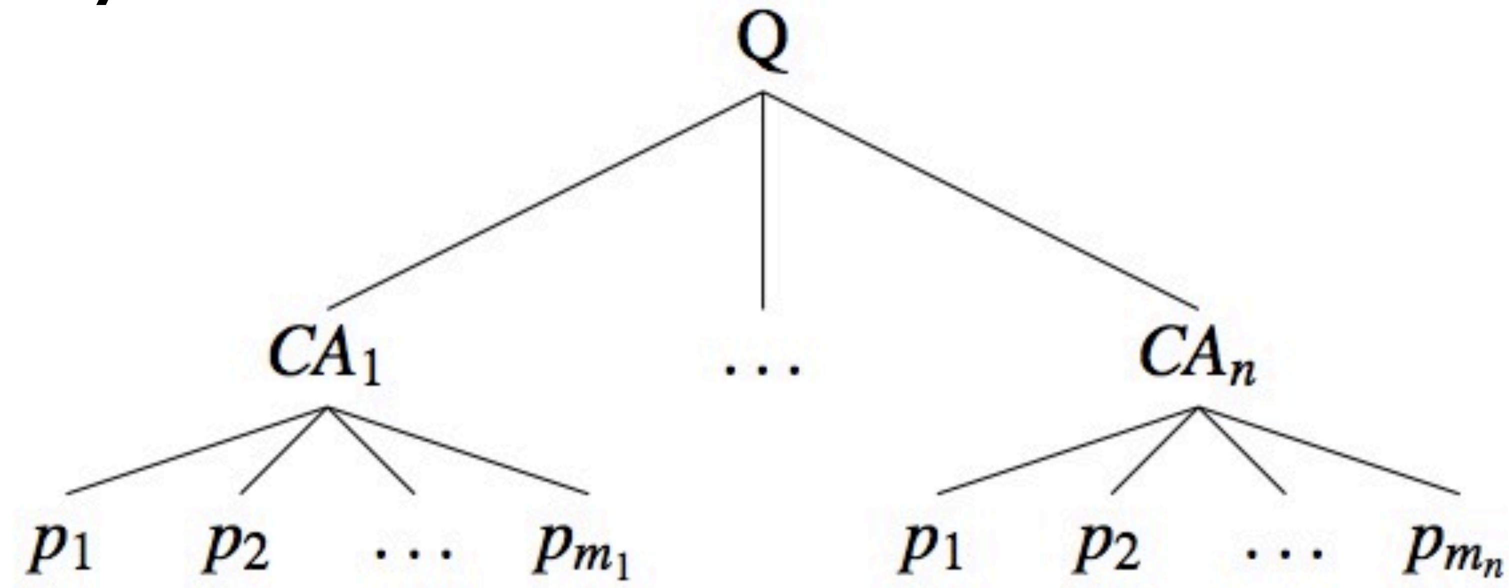
Question terms

Supporting passages

Question 1

Question 2

More Formally...



Question terms

Supporting passages

Question 1

Question 2

Cannot simply linearize this matrix

Multi-dimensional Feature Merger (MDM Features)

M =

	q1	q2	q3	q4	q5
p1					
p2					
p3					
p4					

Multi-dimensional Feature Merger (MDM Features)

Introduce features that capture the distribution of this matrix

M =

	q1	q2	q3	q4	q5
p1					
p2					
p3					
p4					

Multi-dimensional Feature Merger (MDM Features)

Introduce features that capture the distribution of this matrix

M =

	q1	q2	q3	q4	q5
p1					
p2					
p3					
p4					

Some terminology:

Multi-dimensional Feature Merger (MDM Features)

Introduce features that capture the distribution of this matrix

M =

	q1	q2	q3	q4	q5
p1					
p2					
p3					
p4					

Some terminology:

1. S = sum(M) = vector of column sums

Multi-dimensional Feature Merger (MDM Features)

Introduce features that capture the distribution of this matrix

M =

	q1	q2	q3	q4	q5
p1					
p2					
p3					
p4					

Some terminology:

1. S = sum(M) = vector of column sums

2. M' = transpose of matrix M

Multi-dimensional Feature Merger (MDM Features)

Some terminology:

1. S = sum(M) = vector of column sums

2. M' = transpose of matrix M

M =

	q1	q2	q3	q4	q5
p1					
p2					
p3					
p4					

Multi-dimensional Feature Merger (MDM Features)

Some terminology:

1. S = sum(M) = vector of column sums

2. M' = transpose of matrix M

M =

	q1	q2	q3	q4	q5
p1					
p2					
p3					
p4					

$$f : M \to \mathbf{R}^N$$

Multi-dimensional Feature Merger (MDM Features)

Some terminology:

1. S = sum(M) = vector of column sums
2. M' = transpose of matrix M

M =

	q1	q2	q3	q4	q5
p1					
p2					
p3					
p4					

$$f : M \to \mathbf{R}^N$$

$$f(M) =< g(M), g(M') >$$

Multi-dimensional Feature Merger (MDM Features)

Some terminology:

1. S = sum(M) = vector of column sums

2. M' = transpose of matrix M

M =

	q1	q2	q3	q4	q5
p1					
p2					
p3					
p4					

$$f : M \to \mathbf{R}^N$$

$$f(M) =< g(M), g(M') >$$

$$g : M \to \mathbf{R}^{\frac{N}{2}}$$

Multi-dimensional Feature Merger (MDM Features)

Some terminology:

1. S = sum(M) = vector of column sums

2. M' = transpose of matrix M

M =

	q1	q2	q3	q4	q5
p1					
p2					
p3					
p4					

$$f : M \rightarrow \mathbf{R}^N$$

$$f(M) =< g(M), g(M') >$$

$$g : M \rightarrow \mathbf{R}^{\frac{N}{2}}$$

$$g(M) =< sum(S), avg(S), std(S), max(S), min(S), dim(S), non{-}zero(S) >$$

Performance of new features

Performance of new features

Metrics:

Performance of new features

Metrics:

1. Correlation with the gold class in the training data

Performance of new features

Metrics:

1. Correlation with the gold class in the training data

2. Component level analysis

Performance of new features

Metrics:

1. Correlation with the gold class in the training data

2. Component level analysis
   -- Baseline = some basic features + passage scoring feature

Performance of new features

Metrics:

1. Correlation with the gold class in the training data

2. Component level analysis
    -- Baseline = some basic features + passage scoring feature
    -- New performance = some basic features + passage scoring feature + MDM features for that passage scoring feature

Performance of new features

Metrics:

1. Correlation with the gold class in the training data

2. Component level analysis
   -- Baseline = some basic features + passage scoring feature
   -- New performance = some basic features + passage scoring feature + MDM features for that passage scoring feature

3. End-to-end analysis

Performance of new features

Metrics:

1. Correlation with the gold class in the training data

2. Component level analysis
   -- Baseline = some basic features + passage scoring feature
   -- New performance = some basic features + passage scoring feature + MDM features for that passage scoring feature

3. End-to-end analysis

Data:

Performance of new features

Metrics:

1. Correlation with the gold class in the training data

2. Component level analysis
   -- Baseline = some basic features + passage scoring feature
   -- New performance = some basic features + passage scoring feature + MDM features for that passage scoring feature

3. End-to-end analysis

Data:

1. Jeopardy!

Performance of new features

Metrics:

1. Correlation with the gold class in the training data

2. Component level analysis
   -- Baseline = some basic features + passage scoring feature
   -- New performance = some basic features + passage scoring feature + MDM features for that passage scoring feature

3. End-to-end analysis

Data:

1. Jeopardy!

2. Doctor's Dilemma (Medical domain)

Performance of new features

Metrics:

1. Correlation with the gold class in the training data

2. Component level analysis
   -- Baseline = some basic features + passage scoring feature
   -- New performance = some basic features + passage scoring feature + MDM features for that passage scoring feature

3. End-to-end analysis

Data:

1. Jeopardy!

2. Doctor's Dilemma (Medical domain)

Eg: CARDIOLOGY: Murmur associated with this condition is harsh, systolic, diamond-shaped, and increases in intensity with Valsalva

Performance of new features

Metrics:

1. Correlation with the gold class in the training data

2. Component level analysis
   -- Baseline = some basic features + passage scoring feature
   -- New performance = some basic features + passage scoring feature + MDM features for that passage scoring feature

3. End-to-end analysis

Data:

1. Jeopardy!

2. Doctor's Dilemma (Medical domain)

Eg: CARDIOLOGY: Murmur associated with this condition is harsh, systolic, diamond-shaped, and increases in intensity with Valsalva

(A: Hypertrophic cardiomyopathy)

Distribution of data-sets

Training	#Q	#Pos	#Neg	#Avg cand per Q
Jeopardy!	11,520	12,173	2,555,396	222.87
DD	1,322	2,338	543,963	413.23

Distribution of data-sets

Training	#Q	#Pos	#Neg	#Avg cand per Q
Jeopardy!	11,520	12,173	2,555,396	222.87
DD	1,322	2,338	543,963	413.23

Test Data:

Distribution of data-sets

Training	#Q	#Pos	#Neg	#Avg cand per Q
Jeopardy!	11,520	12,173	2,555,396	222.87
DD	1,322	2,338	543,963	413.23

Test Data:

1. Jeopardy! 3,505

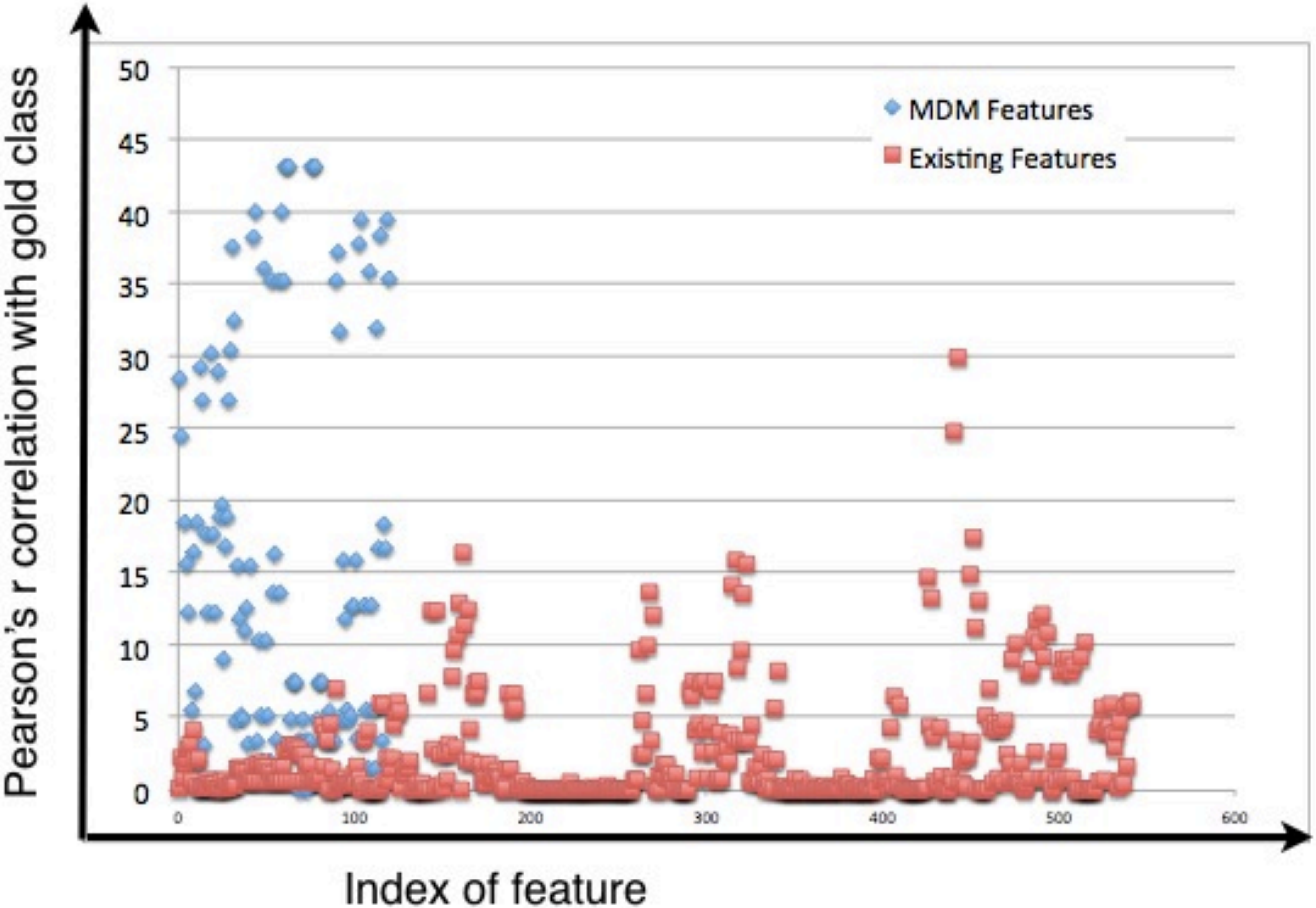
Distribution of data-sets

Training	#Q	#Pos	#Neg	#Avg cand per Q
Jeopardy!	11,520	12,173	2,555,396	222.87
DD	1,322	2,338	543,963	413.23

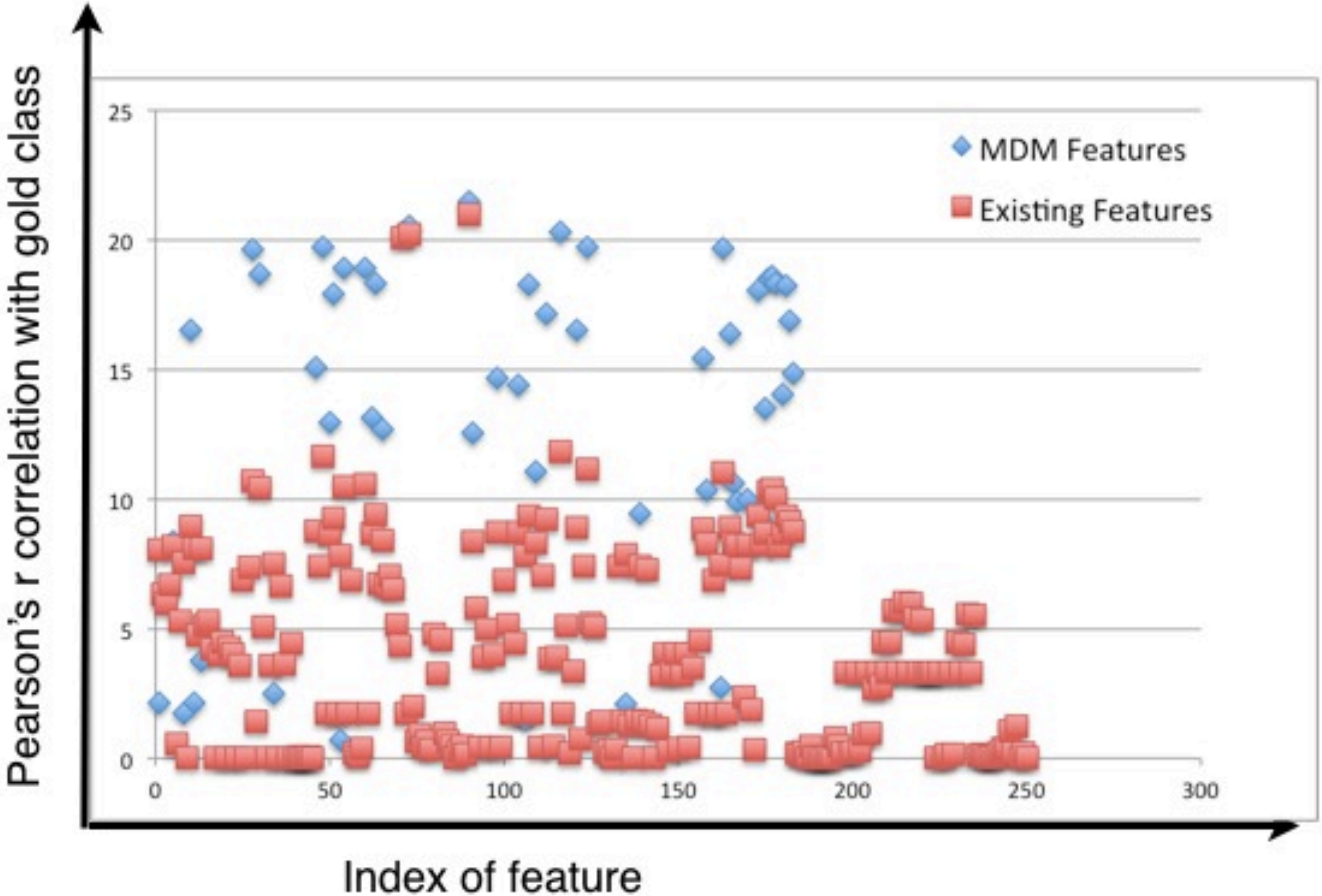
Test Data:

1. Jeopardy! 3,505
2. DD 905

Correlation with the gold class (Jeopardy!)

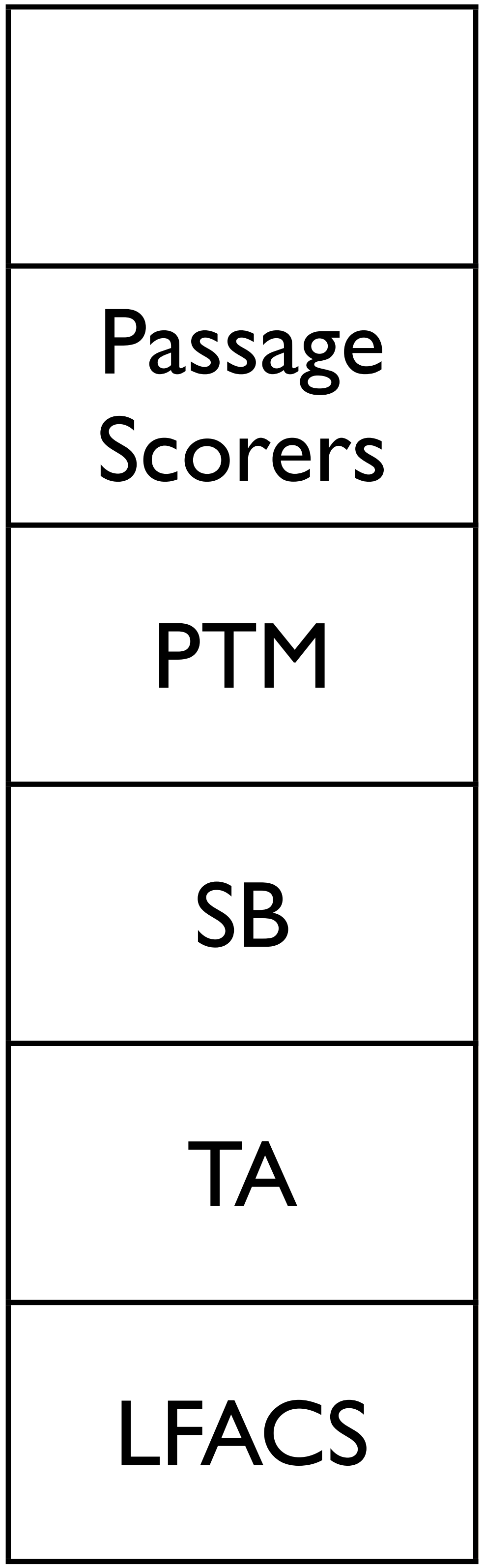


Correlation with the gold class (DD)



Component Level Analysis (DD)

Component Level Analysis (DD)



Component Level Analysis (DD)

	Component Level Baseline
Passage Scorers	Prec@70%
PTM	24.9
SB	26.8
TA	22.9
LFACS	25.7

Component Level Analysis (DD)

	Component Level Baseline	
Passage Scorers	Prec@70%	%Acc
PTM	24.9	20.2
SB	26.8	21.5
TA	22.9	18.8
LFACS	25.7	20.3

Component Level Analysis (DD)

	Component Level Baseline		With MDM
Passage Scorers	Prec@70%	%Acc	Prec@70%
PTM	24.9	20.2	29.2
SB	26.8	21.5	28.7
TA	22.9	18.8	25.7
LFACS	25.7	20.3	28.5

Component Level Analysis (DD)

	Component Level Baseline		With MDM	
Passage Scorers	Prec@70%	%Acc	Prec@70%	%Acc
PTM	24.9	20.2	29.2	23.4
SB	26.8	21.5	28.7	23.3
TA	22.9	18.8	25.7	21.1
LFACS	25.7	20.3	28.5	22.4

End-to-end Analysis (DD)

End-to-end Analysis (DD)

Baseline
Prec@70%
37.2

End-to-end Analysis (DD)

Baseline	
Prec@70%	%Acc
37.2	29.2

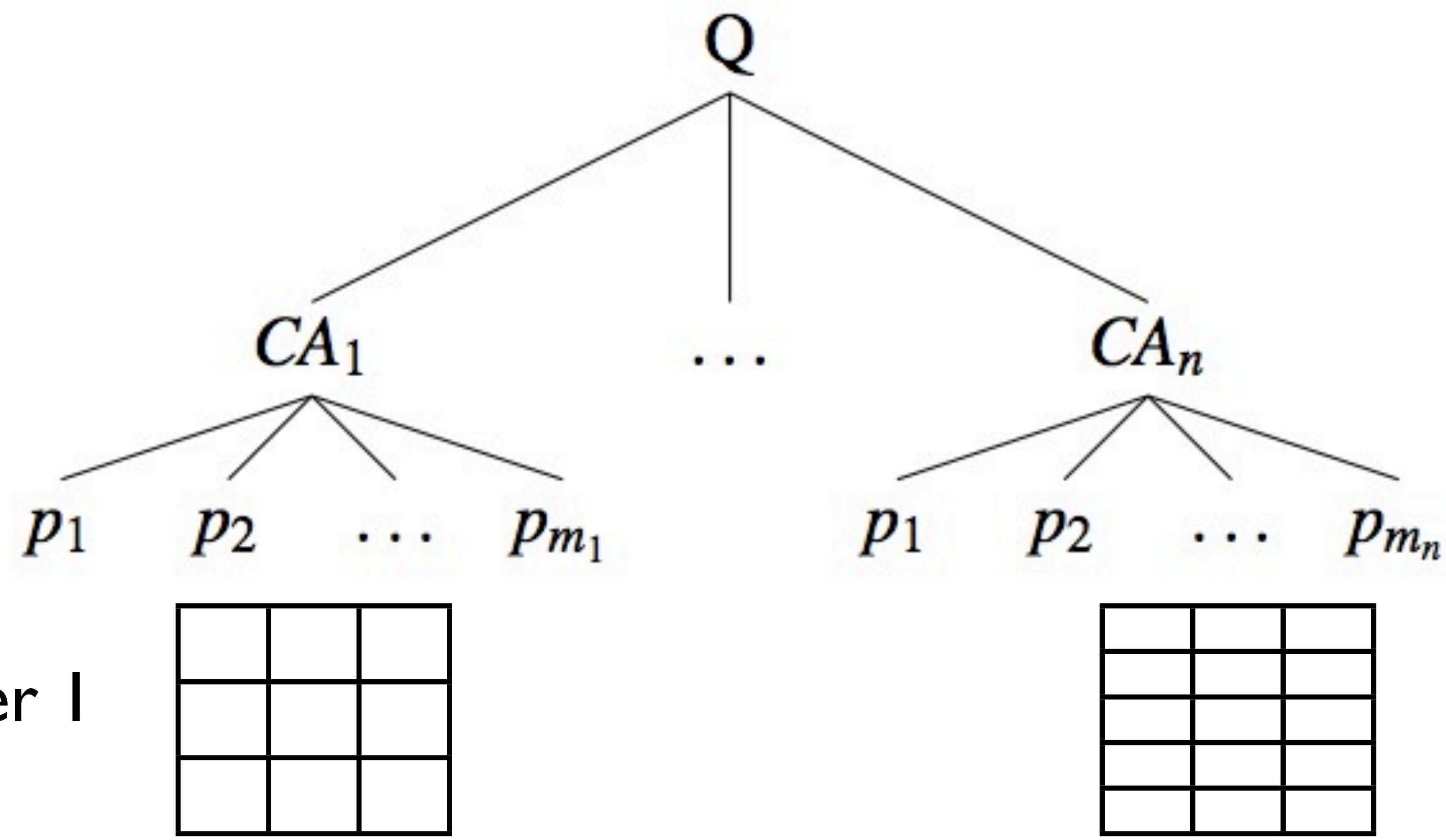
End-to-end Analysis (DD)

Baseline		With MDM
Prec@70%	%Acc	Prec@70%
37.2	29.2	40.2

End-to-end Analysis (DD)

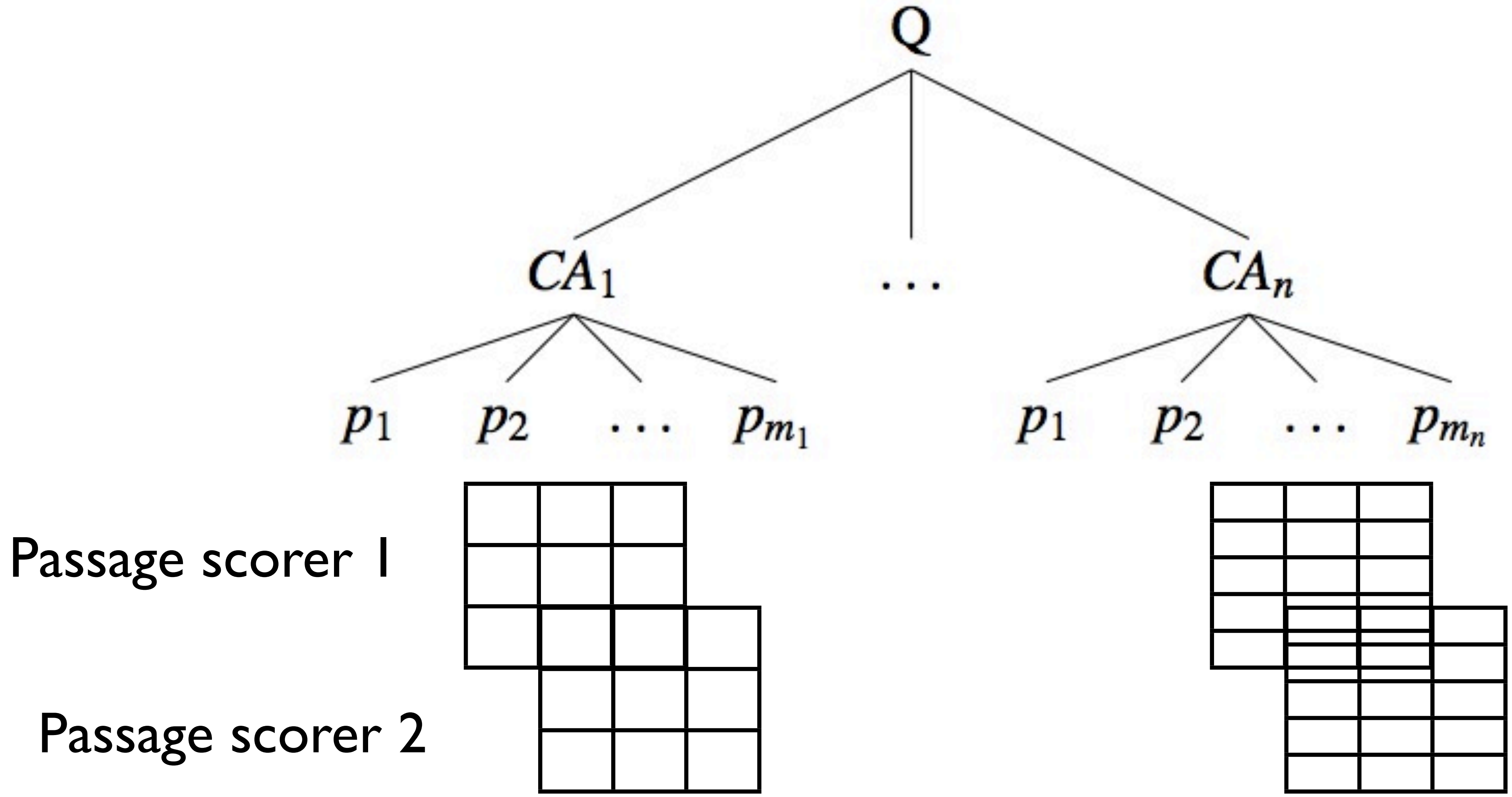
Baseline		With MDM	
Prec@70%	%Acc	Prec@70%	%Acc
37.2	29.2	40.2	31.3

Why the heck is it called Multi-dimensional?

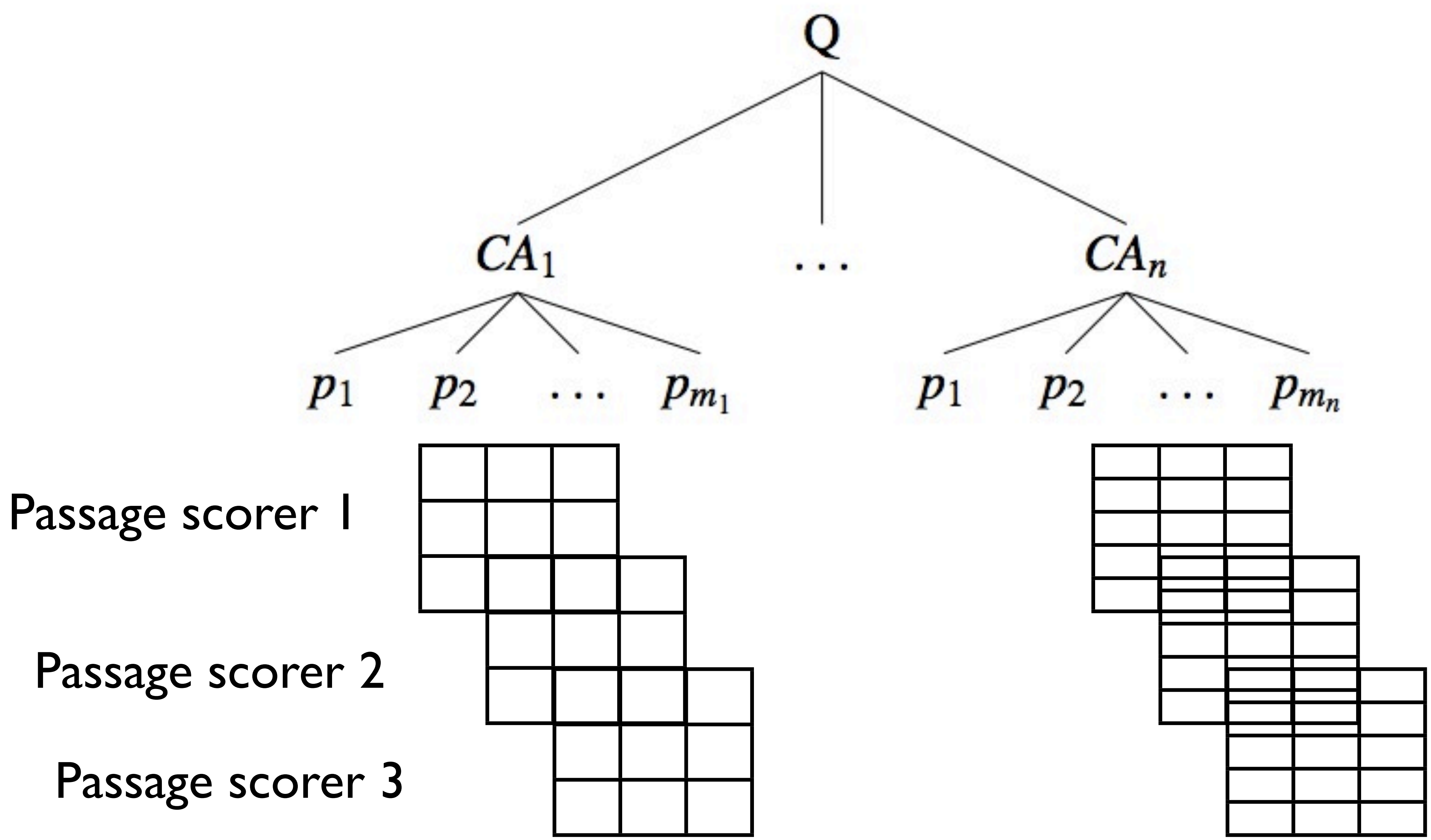


Passage scorer 1

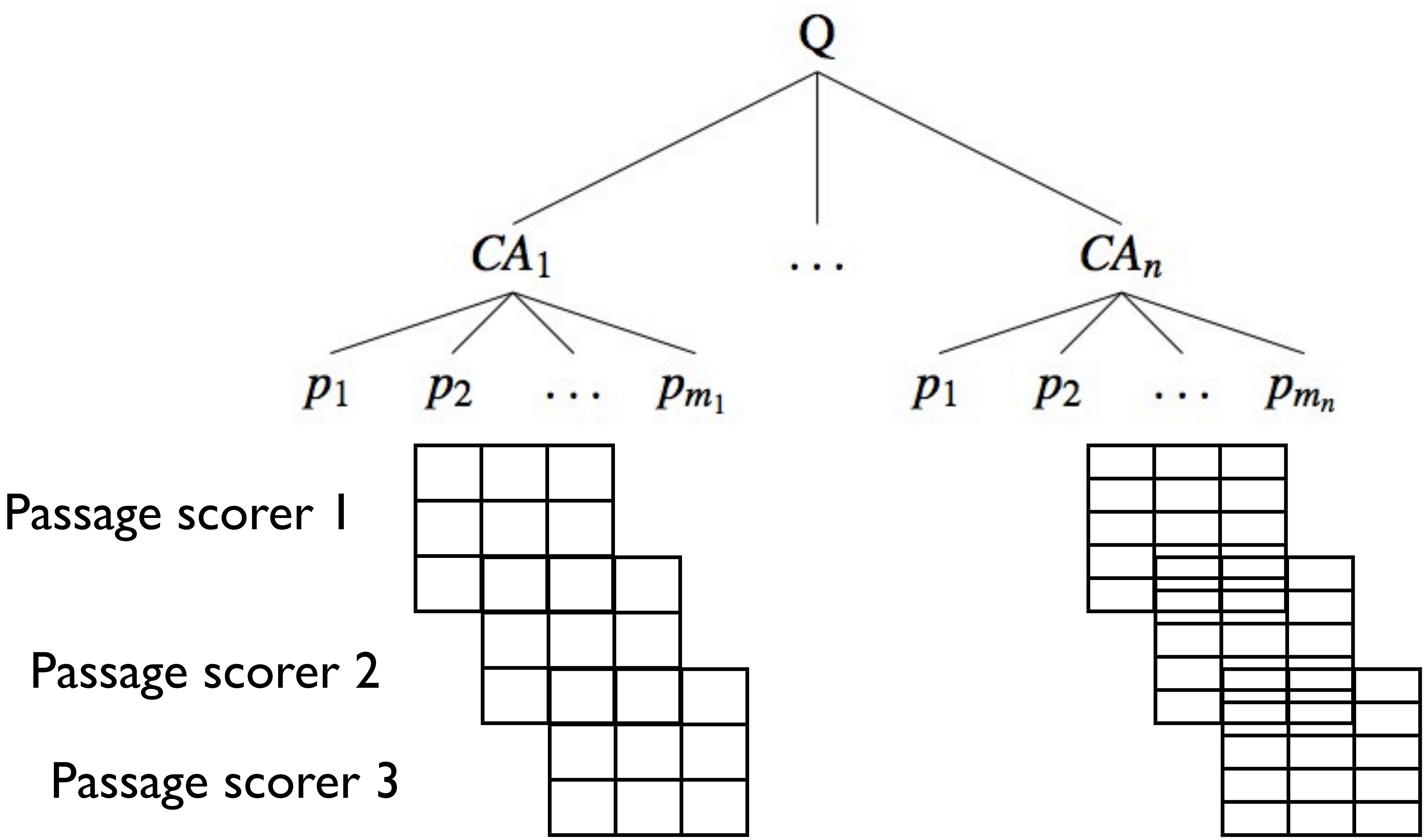
Why the heck is it called Multi-dimensional?



Why the heck is it called Multi-dimensional?



Why the heck is it called Multi-dimensional?



Future Work!

Conclusion

Conclusion

1. Introduced new features for QA (not specific to Watson)

Conclusion

1. Introduced new features for QA (not specific to Watson)

2. Show these features improve the performance of an existing state-of-the-art QA system statistically significantly

Conclusion

1. Introduced new features for QA (not specific to Watson)

2. Show these features improve the performance of an existing state-of-the-art QA system statistically significantly

Thanks and Questions?