



Context Chain and Fractional Conditional Compressibility of Models

RESEARCH & DEVELOPMENT, THOMSON REUTERS

FRANK SCHILDER, RAVI KONADADI, HARSHA VEERAMACHANENI

2009-11-16



THOMSON REUTERS

Overview

- Update summarization results
- AESOP results:
 - Context Chain
 - FraCC
- Evaluating an automatic evaluation metric
 - Correlation precision and recall
- Conclusions

UPDATE SUMMARIZATION

TAC 2009 update summarization

- We submitted two runs:
 - FastSum (TAC 2008 system, re-trained)
 - SVR trained on a minimal set of easy to compute features (document frequency, content word frequency, sentence length)
 - First line baseline
 - Taking the first sentences of most recent articles, cosine similarity-based redundancy removal.
- Results:
 - Responsiveness
 - Pyramid
 - ROUGE
 - BE

UPDATE SUMMARIZATION RESULTS

System	Resp	Pyramid	Ling. Quality	ROUGE-2	BE
Best System	5.159	6.5	5.932	0.12184	0.06379
Baseline 1	3.636	3.182	6.705	0.06315	0.02916
Baseline 2	6.364	11.977	5.477	0.33133	0.2483
Baseline 3	6.341	6.341	7.477	0.10633	0.05333
FastSum	4.455	5.295	5.545	0.09366	0.04382
Rank	13	9	7	21	27
1st line Baseline	4.205	4.182	4.795	0.09307	0.04781
Rank	20	27	23	22	21

Table 1. TAC 2009 update summarization results: task A

System	Resp	Pyramid	Ling. Quality	ROUGE-2	BE
Best System	5.023	0.307	5.886	0.10417	0.06364
Baseline 1	4.318	0.16	6.455	0.05115	0.02417
Baseline 2	6.182	0.69	5.886	0.31932	0.25042
Baseline 3	6.182	0.69	7.25	0.09799	0.05669
FastSum	4.273	0.21	5.864	0.07586	0.04125
Rank	8	20	2	24	21
1st line Baseline	4.136	0.238	4.988	0.08819	0.05168
Rank	12	13	17	8	7

Table 2. TAC 2009 update summarization results: task B

AESOP

Two automatic metrics

- ContextChain (ICSC 2009)
- Fractional Conditional Compressibility of Models (FraCC)

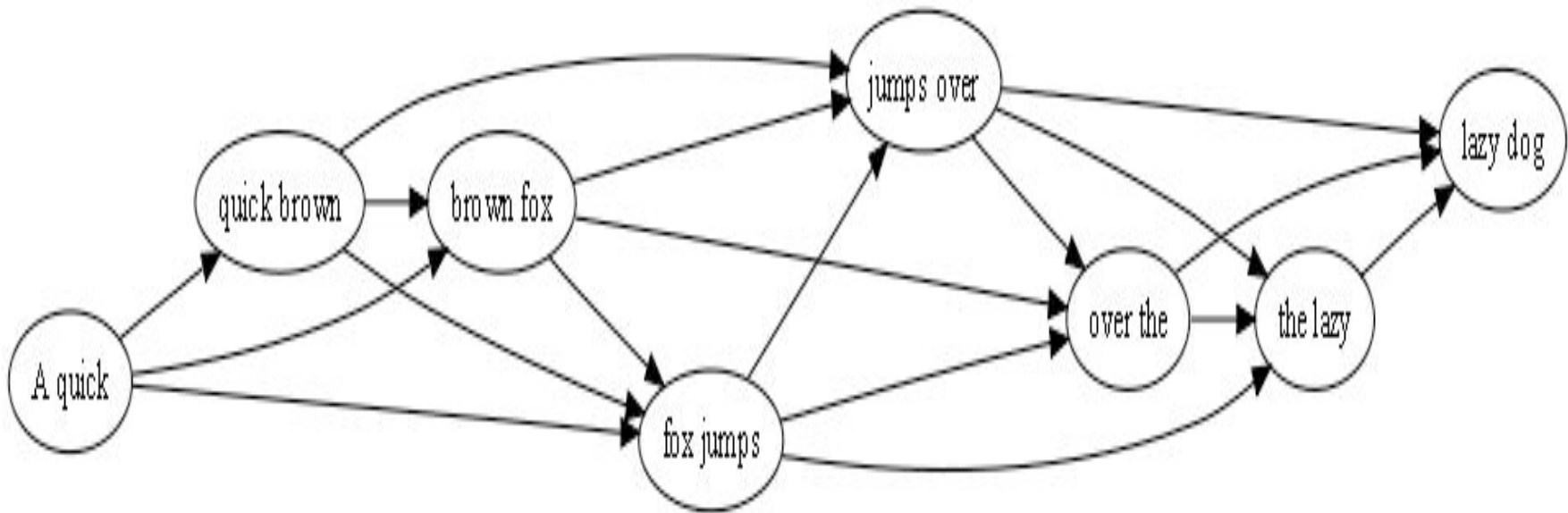
CONTEXTCHAIN

AutoSummENG

- ContextChain is based on AutoSummENG:
 - N-gram based (Character/Word)
 - N-gram graphs for model summary and system summary are compared
 - Vertices: n-grams
 - Edges: Adjacent n-grams in the neighborhood
 - Edge weights: Co-occurrence values

N-Gram Graph Example

- A quick brown fox jumps over the lazy dog
- Length = 2, window size = 5



Graph Similarity

- Size similarity
 - Containment similarity
 - Value similarity
- A combined score of these three similarity scores is the overall metric score

Context Chains

- Context Chains model the context around named entities.
- N-gram graphs are generated by the context window before and after an entity.
- Context Chains are generated on a sentence-by-sentence basis.
- Word n-grams only.

The Justice Department is conducting an anti-trust trial against *Microsoft Corp* with evidence that the company is increasingly attempting to crush competitors. *Microsoft* is accused of trying to forcefully buy into markets...

bi-grams context chains generated:

anti-trust_trial → accused_trying

evidence_company → accused_trying

increasingly_attempting → accused_trying

FRACTIONAL CONDITIONAL COMPRESSIBILITY OF MODELS

Fractional Conditional Compressibility (FraCC)

- The metric is based on the compressibility of the model summary, if a system summary is given:
 - Given system summary S and model summary M
 - We define $r(M|S)$ as the compressibility of M if S were to be observed.
 - $r(M|S)$ is defined as the difference in compressibility of M when S is seen to when S is not seen.
 - We use $r(M|S)$ as score for the similarity metric between the system and model summaries.

Fractional Conditional Compressibility (FraCC)

- Compressibility of string M is defined as

$$C(M) = H(M) / |M|$$

- $H(S)$: Entropy of a string
- $|S|$: Length of a string

- Conditional compressibility of M given S is

$$C(M|S) = (H(S+M) - H(S)) / |M|$$

- $r(M|S) = (C(M) - C(M|S)) / C(M)$

Estimating Entropy $H(\cdot)$

- We use the Burrows-Wheeler Transform (**BWT**) of the string followed by the computation of the move-to-front (**MTF**) entropy
- We compute the BWT on the string by encoding each word as a symbol (after some normalization)

BWT transformation

a	good	model	summary	a	good	system	summary
good	model	summary	a	good	system	summary	a
model	summary	a	good	system	summary	a	good
summary	a	good	system	summary	a	good	model
a	good	system	summary	a	good	model	summary
good	system	summary	a	good	model	summary	a
system	summary	a	good	model	summary	a	good
summary	a	good	model	summary	a	good	system

Sorting of rows

a	good	model	summary	a	good	system	summary
a	good	system	summary	a	good	model	summary
good	model	summary	a	good	system	summary	a
good	system	summary	a	good	model	summary	a
model	summary	a	good	system	summary	a	good
summary	a	good	model	summary	a	good	system
summary	a	good	system	summary	a	good	model
system	summary	a	good	model	summary	a	good

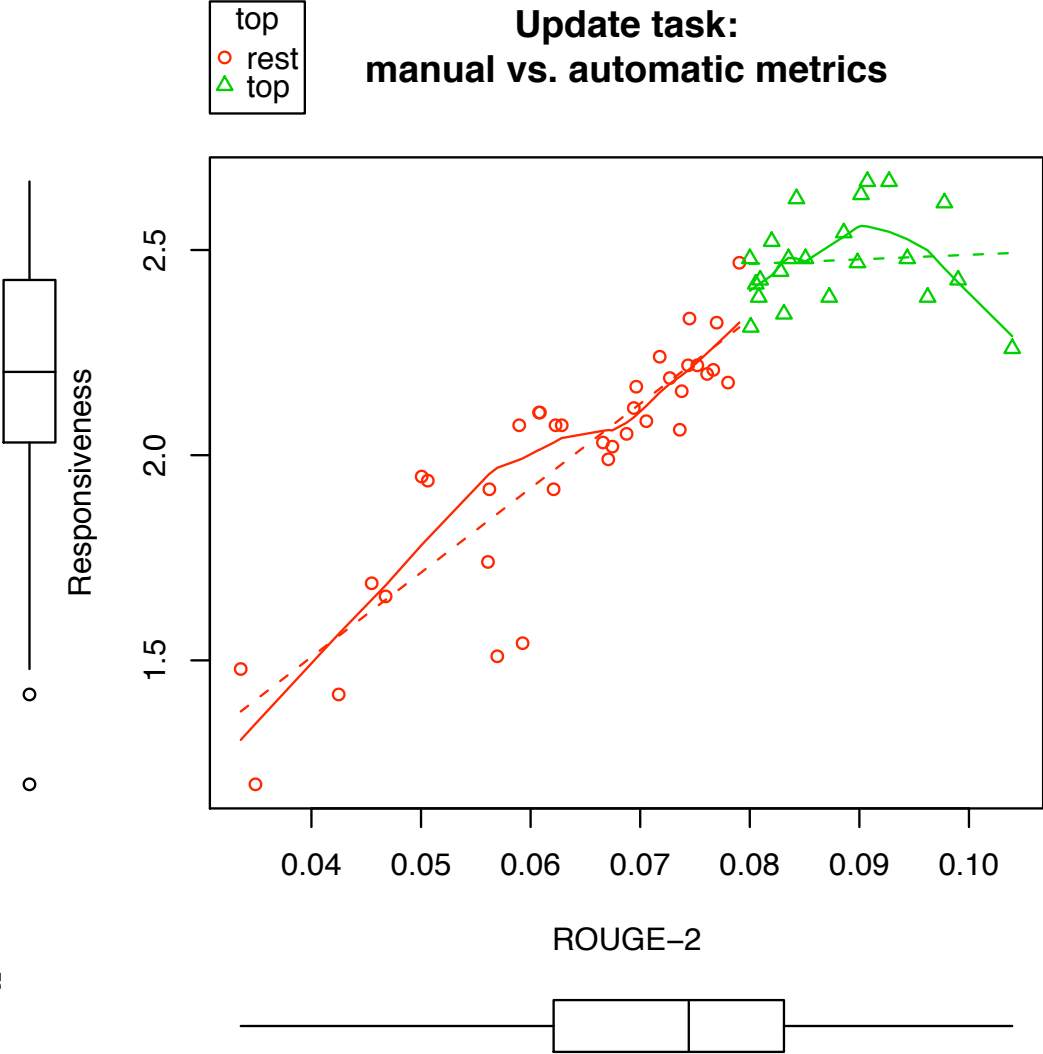
Local entropy based on last column

a	good	model	summary	a	good	system	summary
a	good	system	summary	a	good	model	summary
good	model	summary	a	good	system	summary	a
good	system	summary	a	good	model	summary	a
model	summary	a	good	system	summary	a	good
summary	a	good	model	summary	a	good	system
summary	a	good	system	summary	a	good	model
system	summary	a	good	model	summary	a	good

- Local Entropy (Move-To-Front, MTF) is derived from an integer value for each cell in the last column of the BWT transformation:
 - number of distinct symbols observed before its last occurrence.
- MTF entropy is calculated as the sum of the logs of 1+MTF encoding:
 - E.g. MTF entropy = $7 * \log(1) + 1 * \log(2)$

EVALUATING AN EVALUATION METRIC

TAC 2008: Responsiveness and ROUGE



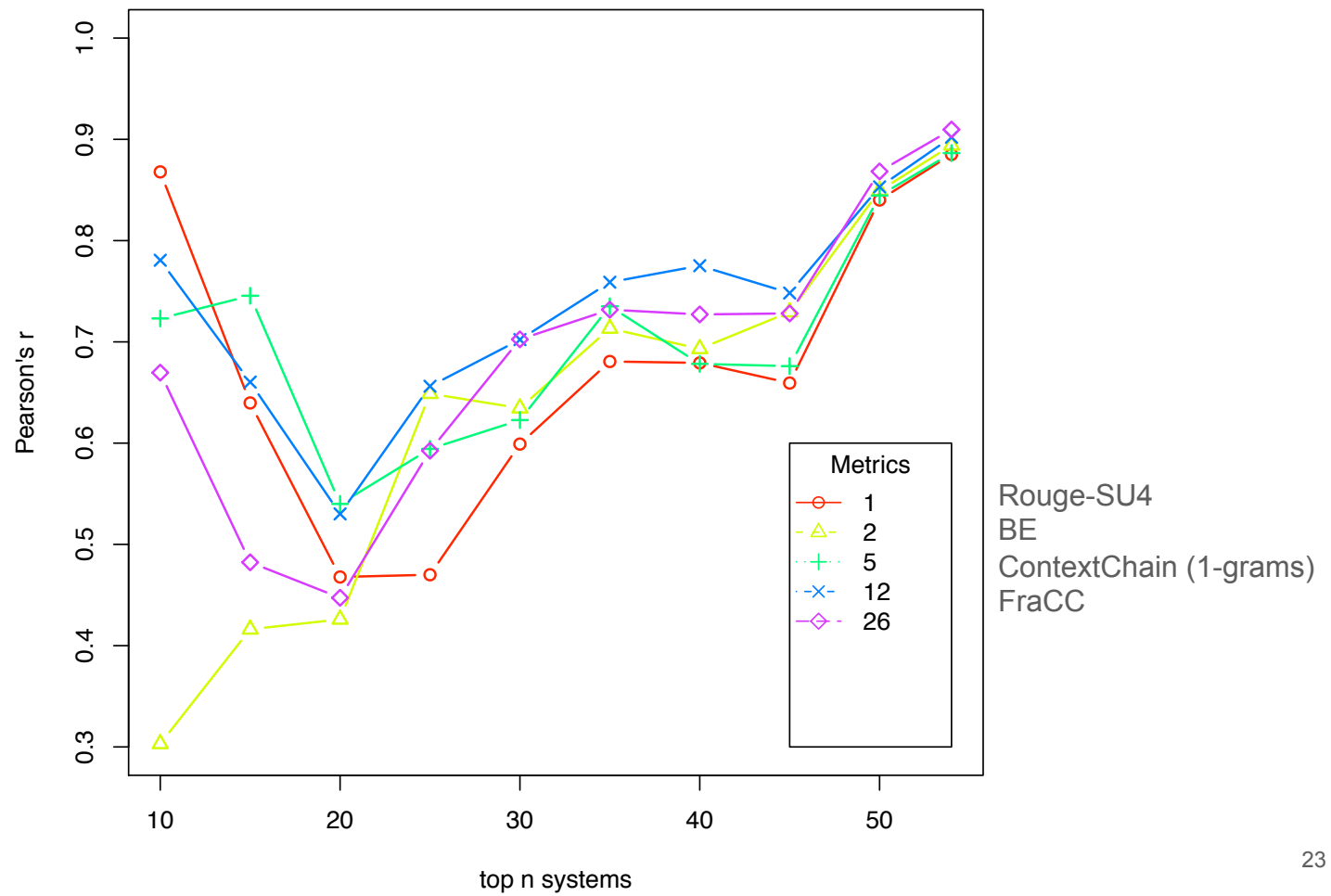
Meta Metrics: Correlation Precision & Recall

- Correlation Recall @ n
 - Correlation when the two metric vectors are sorted according to the manual metric
- Correlation Precision @ n
 - Correlation when the two metric vectors are sorted according to the automatic metric

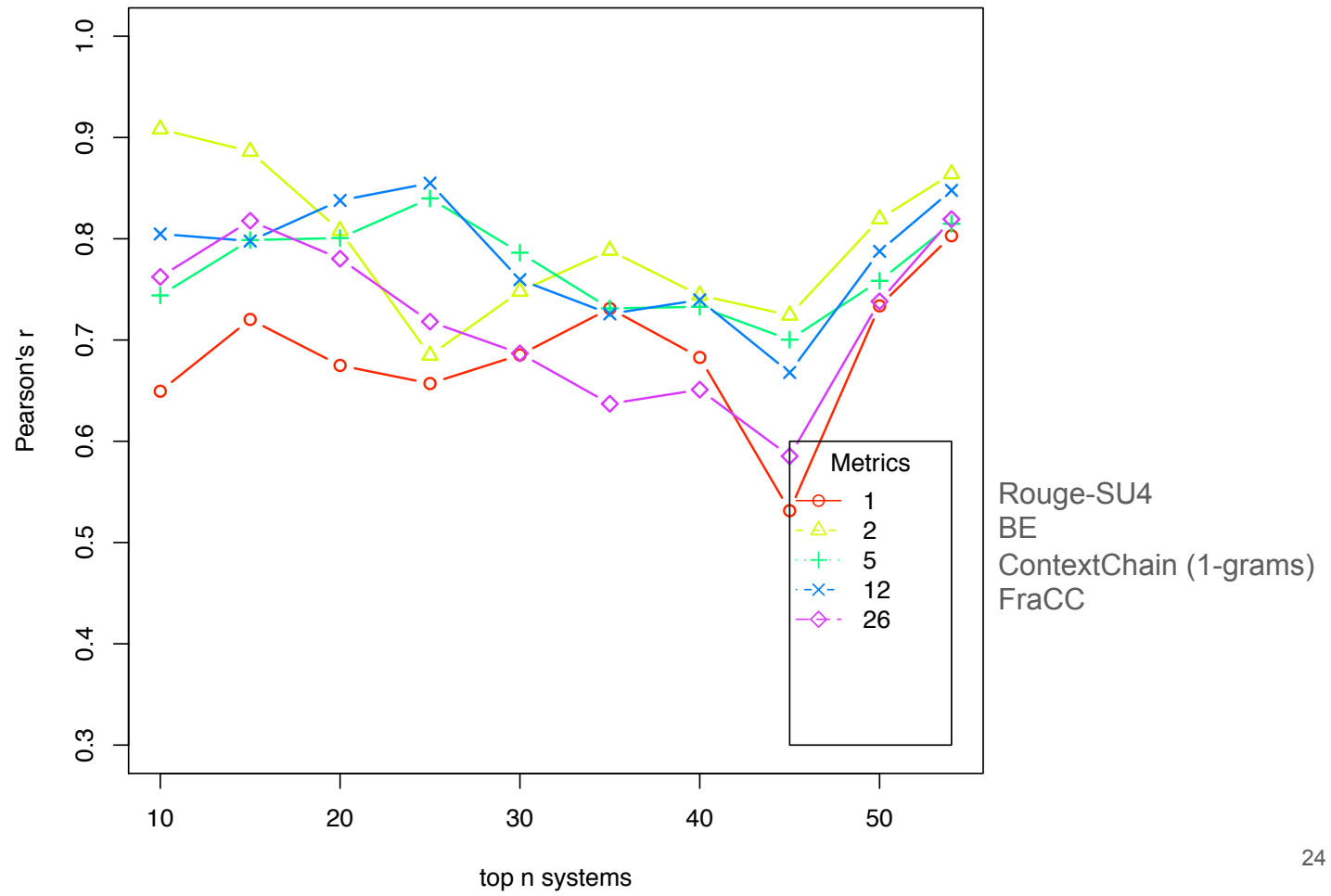
System ID	Responsiveness	Automatic metric	System ID	Responsiveness	Automatic metric
1	2.5	0.9	2	2.4	1.0
2	2.4	1.0	4	2.0	0.95
3	2.3	0.8	1	2.5	0.9
4	2.0	0.95	3	2.3	0.8
5	1.5	0.7	5	1.5	0.7



Correlation Precision: Responsiveness/Task A



Correlation Precision: Responsiveness/Task B



Conclusions

- Two new metrics:
 - ContectChain
 - FraCC
- Evaluating the metrics:
 - Correlation Precision and Recall

Questions



Backup slides

Burrows Wheeler Transform (BWT)

- A block-sorting transform of a string is a reversible permutation of the string.
- Sorts all the cyclical shifts of the string and uses the last column of the sorted array as the transformed string.
- Makes strings more amenable to compression by run-length coding.
- Used in BZIP2 compression algorithm.

BWT Example

Original string

THAT is no country for old men. The young In one another's arms, birds in the trees – Those dying generations – at their song, The salmon-falls, the mackerel-crowded seas, Fish, flesh, or fowl, commend all summer long Whatever is begotten, born, and dies. Caught in that sensual music all neglect Monuments of unageing intellect.

BWTransformed string

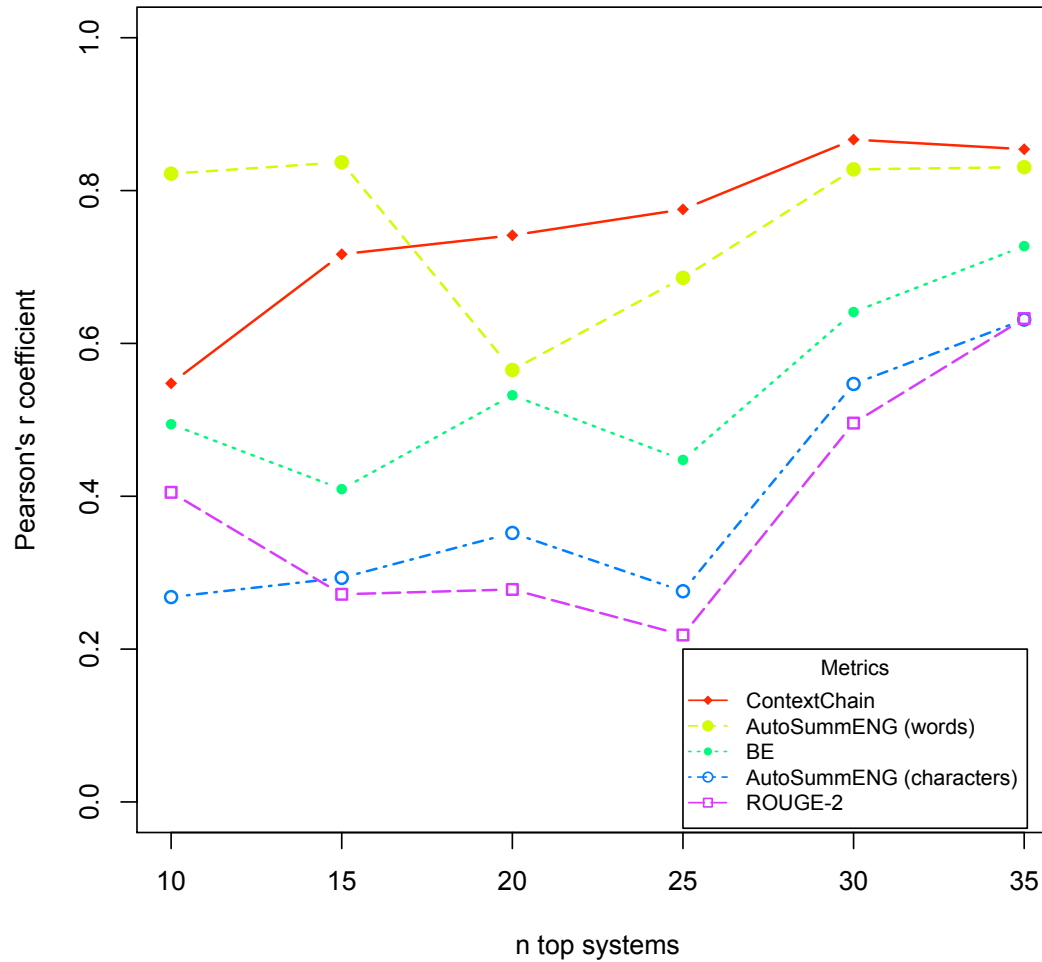
rsgnsnlhhs__IntsnH__T__.A____ss.,gt,.-gcd,es
s,,,ode,yrgtsgrTredllssrn,edtrln,ntefemnu__fs__eh_hrC__ia__-
eennlew_r_nshhhhslldrnbgrrttmmgsmhvmnkielto-
____nnnna_ueesstWtTtTtTgsd__ye_teb__Fcweallolgfaaea_l__mumoulr_re
oeliueao_eouoii_aoeiueon__cm_sliM_fbhngycrfeoeoieiteaoctamleen'idit_o
__ieu_n_cchaanta____oa_nnosans_oomeoord_

MTF Entropy (also called local entropy)

- Given a string S , the MTF encoding is calculated by giving each symbol in the string an integer value = number of distinct symbols observed before its last occurrence.
 - E.g. `abbcbabcbba` => `00001212102`
- MTF entropy is calculated as the sum of the logs of $1 + \text{MTF encoding}$.
 - E.g. $\text{MTF entropy} = 5 * \log(1) + 3 * \log(2) + 3 * \log(3)$
- Observe that if there are long runs of symbols the MTF entropy is small.

Correlation Recall for TAC 2008

Correlation Recall



Correlation Precision for TAC 2008

Correlation Precision

