

Evaluation Metrics for Machine Reading Comprehension: Prerequisite Skills and Readability

Saku Sugawara♠, Yusuke Kido♠, Hikaru Yokono♣, Akiko
Aizawa◇♠

♠University of Tokyo

♣Fujitsu Laboratories Ltd.

◇National Institute of Informatics

August 1, ACL 2017

Reading Comprehension (RC) Task

ID: MCTest MC160.dev.29 (1) multiple:

Context: The princess climbed out the window of the high tower and climbed down the south wall when her mother was sleeping. She wandered out a good ways. Finally she went into the forest where there are no electric poles but where there are some caves.

Question: Where did the princess wander to after escaping?

Answer: A) Mountain *B) Forest C) Cave D) Castle

Reading Comprehension (RC) Task

ID: MCTest MC160.dev.29 (1) multiple:

C1: The **princess** **climbed** out the window of the high tower and **climbed down** the south wall when her mother was sleeping.

C2: **She** **wandered** out a good ways.

C3: **Finally** **she** went into the forest where there are no electric poles but where there are some caves.

Q: Where did the **princess** wander to **after escaping**?

A: A) Mountain *B) Forest C) Cave D) Castle

Coreference resolution (*she = princess*)

Commonsense reasoning (*escaping = climbed down*)

Temporal relation (*climbed → wandered*)

RC and Related Datasets

- ❖ DeepRead (1999)
- ❖ QA4MRE (2013)
- ❖ MCTest (2013)
- ❖ CNN/Daily Mail (2015)
- ❖ SQuAD (2016)
- ❖ LAMBADA (2016)
- ❖ Who-did-What (2016)
- ❖ MS MARCO (2016)
- ❖ NewsQA (2016)
- ❖ WikiSuggest (ACL 2017)
- ❖ TriviaQA (ACL 2017)
- ❖ RACE (EMNLP 2017)

RC and Related Datasets

❖ DeepRead (1999)	300
❖ QA4MRE (2013)	240
❖ MCTest (2013)	2640
❖ CNN/Daily Mail (2015)	1.4M
❖ SQuAD (2016)	100K
❖ LAMBADA (2016)	10K
❖ Who-did-What (2016)	200K
❖ MS MARCO (2016)	100K
❖ NewsQA (2016)	120K
❖ WikiSuggest (ACL 2017)	3.5M
❖ TriviaQA (ACL 2017)	650K
❖ RACE (EMNLP 2017)	100K

RC and Related Datasets

❖ DeepRead (1999)	300
❖ QA4MRE (2013)	240
❖ MCTest (2013)	2640
❖ CNN/Daily Mail (2015)	1.4M
❖ SQuAD (2016)	100K
❖ LAMBADA (2016)	10K
❖ Who-did-What (2016)	200K
❖ MS MARCO (2016)	100K
❖ NewsQA (2016)	120K
❖ WikiSuggest (ACL 2017)	3.5M
❖ TriviaQA (ACL 2017)	650K
❖ RACE (EMNLP 2017)	100K

Over 6.0M
Questions!!

System Analysis by Accuracy

Dataset A	System X
Q1	x
Q2	o
Q3	x
⋮	⋮
Q100	o
Accuracy	75.0%

- Only with accuracy, we cannot tell what the systems understand and what they don't.
- ✦ Chen⁺ (2016) shows: CNN/Daily Mail datasets contain unanswerable or ambiguous questions

Our Research Question

- ✦ How can we evaluate and analyze our RC systems?
 - Propose evaluation metrics for RC
 - Focus on prerequisite skills and readability

Motivation: Two Types of Difficulties

ID: SQuAD (2016), United_Methodist_Church

Context: The United Methodist Church (UMC) practices infant and adult baptism. Baptized Members are those who have been baptized as an infant or child, but who have not subsequently professed their own faith.

Question: What are members who have been baptized as an infant or child but who have not subsequently professed their own faith?

Answer: Baptized Members

ID: MCTest (2013), mc160.dev.29

Context: The princess climbed out the window of the high tower and climbed down the south wall when her mother was sleeping. She wandered out a good ways. Finally she went into the forest where there are no electric poles but where there are some caves.

Question: Where did the princess wander to after escaping?

Answer: A) Mountain *B) Forest C) Cave D) Castle

Motivation: Two Types of Difficulties

ID: SQuAD (2016), United_Methodist_Church

Context: The United Methodist Church (UMC) practices infant and adult baptism. **Baptized Members** are those who have been baptized as an infant or child, but who have not subsequently professed their own faith.

Question: What are members who have been baptized as an infant or child but who have not subsequently professed their own faith?

Answer: Baptized Members

ID: MCTest (2013), mc160.dev.29

Context: The princess climbed out the window of the high tower and climbed down the south wall when her mother was sleeping. She wandered out a good ways. Finally she went into the forest where there are no electric poles but where there are some caves.

Question: Where did the princess wander to after escaping?

Answer: A) Mountain *B) Forest C) Cave D) Castle

Motivation: Two Types of Difficulties

ID: SQuAD (2016), United_Methodist_Church

Context: The United Methodist Church (UMC) practices infant and adult baptism. **Baptized Members** are those *who have been baptized as an infant or child, but who have not subsequently professed their own faith.*

Question: What are members *who have been baptized as an infant or child but who have r*

Answer: Baptized M → Answerble simply by noticing one sentence

ID: MCTest (2013), mc160.dev.29

Context: The princess climbed out the window of the high tower and climbed down the south wall when her mother was sleeping. She wandered out a good ways. Finally she went into the forest where there are no electric poles but where there are some caves.

Question: Where did the princess wander to after escaping?

Answer: A) Mountain *B) Forest C) Cave D) Castle

Motivation: Two Types of Difficulties

ID: SQuAD (2016), United_Methodist_Church

Context: The United Methodist Church (UMC) practices infant and adult baptism. **Baptized Members** are those who have been baptized as an infant or child, but who have not subsequently professed their own faith.

Question: What are members who have been baptized as an infant or child but who have r

Answer: Baptized M

→ Answerable simply by noticing one sentence

ID: MCTest (2013), mc160.dev.29

Context: The princess climbed out the window of the high tower and climbed down the south wall when her mother was sleeping. She wandered out a good ways. Finally she went into the forest where there are no electric poles but where there are some caves.

Question: Where did the princess wander to after escaping?

Answer: A) Mountain *B) Forest C) Cave D) Castle

Motivation: Two Types of Difficulties

ID: SQuAD (2016), United_Methodist_Church

Context: The United Methodist Church (UMC) practices infant and adult baptism. **Baptized Members** are those who have been baptized as an infant or child, but who have not subsequently professed their own faith.

Question: What are members who have been baptized as an infant or child but who have r

Answer: Baptized M

→ Answerable simply by noticing one sentence

ID: MCTest (2013), mc160.dev.29

Context: The princess climbed out the window of the high tower and climbed down the south wall when her mother was sleeping. She wandered out a good ways. Finally she went into the forest where there are no electric poles but where there are some caves.

Question: Where d

Answer: A) Mount:

→ Require reading multiple sentence with skills

Motivation: Two Types of Difficulties

ID: SQuAD (2016), United_Methodist_Church

Context: The United Methodist Church (UMC) practices infant and adult baptism. **Baptized Members** are those who have been baptized as an infant or child, but who have not subsequently professed their own faith.

Question: What are members who have been baptized as an infant or child but who have not subsequently professed their own faith?

Answer: Baptized Members

Difficult-to-read & Easy-to-answer

ID: MCTest (2013), mc160.dev.29

Context: The princess climbed out the window of the high tower and climbed down the south wall when her mother was sleeping. She wandered out a good ways. Finally she went into the forest where there are no electric poles but where there are some caves.

Question: Where did the princess go when she was alone?

Answer: A) Mountain *B) Forest

Easy-to-read & Difficult-to-answer

Our study: Evaluation Metrics for RC

1. Define prerequisite skills and readability metrics
 - 13 prerequisite skills and 10 readability measures

Our study: Evaluation Metrics for RC

1. Define prerequisite skills and readability metrics
 - ✦ 13 prerequisite skills and 10 readability measures
2. Annotate RC datasets with skills
 - ✦ 6 datasets: QA4MRE, MCTest, SQuAD, Who-did-What, MS MARCO, NewsQA

Our study: Evaluation Metrics for RC

1. Define prerequisite skills and readability metrics
 - ✦ 13 prerequisite skills and 10 readability measures
2. Annotate RC datasets with skills
 - ✦ 6 datasets: QA4MRE, MCTest, SQuAD, Who-did-What, MS MARCO, NewsQA
3. Calculate readability of the datasets
 - ✦ Readability of “context sentences necessary for answering” (selected in the annotation) (\neq whole context)

Our study: Evaluation Metrics for RC

1. Define prerequisite skills and readability metrics
 - 13 prerequisite skills and 10 readability measures
2. Annotate RC datasets with skills
 - 6 datasets: QA4MRE, MCTest, SQuAD, Who-did-What, MS MARCO, NewsQA
3. Calculate readability of the datasets
 - Readability of “context sentences necessary for answering” (selected in the annotation) (\neq whole context)
4. Analyze the datasets on two types of difficulties
 - See the relation between skills and readability

System Analysis by Accuracy

Dataset A	System X
Q1	x
Q2	o
Q3	x
⋮	⋮
Q100	o
Accuracy	75.0%

System Analysis by the Skills and Readability

Question	Dataset A								System X
	Prerequisite Skills				Readability Metrics				
	Skill 1	Skill 2	...	Skill 13	RM 1	RM 2	...	RM 10	
Q1	x	-	...	x	5.1	27.1	...	0.17	x
Q2	-	o	...	-	3.9	13.5	...	0.11	o
Q3	x	x	...	-	4.6	26.9	...	0.08	x
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Q100	o	o	...	o	4.3	16.9	...	0.12	o
Accuracy	40.0%	90.0%	...	70.0%	-	-	...	-	75.0%

This study shows statistics of datasets & an observation on the relation between skills (difficulty in answering) and readability (difficulty in reading)

Prerequisite Skills

-
- | | |
|----------------------------|--------------------------------|
| 1. Object tracking | 8. Ellipsis |
| 2. Mathematical reasoning | 9. Bridging |
| 3. Coreference resolution | 10. Elaboration |
| 4. Logical reasoning | 11. Meta-knowledge |
| 5. Analogy | 12. Schematics clause relation |
| 6. Causal relation | 13. Punctuation |
| 7. Spatiotemporal relation | |
-

- ✦ New knowledge reasoning skills in this study
 - “Commonsense reasoning” is updated to new 4 skills for more detailed analysis

Prerequisite Skills

- | | |
|----------------------------|--------------------------------|
| 1. Object tracking | 8. Ellipsis |
| 2. Mathematical reasoning | 9. Bridging |
| 3. Coreference resolution | 10. Elaboration |
| 4. Logical reasoning | 11. Meta-knowledge |
| 5. Analogy | 12. Schematics clause relation |
| 6. Causal relation | 13. Punctuation |
| 7. Spatiotemporal relation | |
-

- ✦ Previous study (Sugawara⁺, AACL 2017): skills are based on existing NLU tasks in NLP
- Analyzed MCTest dataset and three systems, and showed that *“the more skills are required, the more difficult to answer (lower accuracy).”*
- ✦ We regard this as the difficulty of answering

Annotated RC Datasets: 100 Qs for each

RC dataset	Genre	Query sourcing	Task formulation
QA4MRE (2013)	Technical documents	Handcrafted by experts	Multiple choice
MCTest (2013)	Narratives by crowd workers	Crowd sourced	Multiple choice
SQuAD (2016)	Wikipedia articles	Crowd sourced	Text span selection
Who-did-What (2016)	News articles (Gigaward v5)	Automated from other articles	Cloze
MS MARCO (2016)	Segmented web pages	Search engine queries	Description
NewsQA (2016)	News articles	Crowd sourced	Text span selection

Annotation with the Prerequisite Skills

- ✦ 13 skills: mult-label annotation
- ✦ 6 datasets: QA4MRE, MCTest, SQuAD, Who-did-What, MS MARCO, NewsQA
- ✦ 100 questions for each dataset
- ✦ 4 annotators: graduate NLP students
- ✦ For 62 randomly sampled questions, 90.1% agreement
- ✦ Annotation: choose skill labels and “necessary context sentences” for answering
- ✦ Sentences are used to calculate readability measures

Result: Frequencies (%) of Prerequisite Skills

Skills	QA4MRE	MCTest	SQuAD	WDW	MARCO	NewsQA
1. Tracking	11.0	6.0	3.0	8.0	6.0	2.0
2. Math.	4.0	4.0	0.0	3.0	0.0	1.0
3. Coref. resol.	32.0	49.0	13.0	19.0	15.0	24.0
4. Logical rsng.	15.0	2.0	0.0	8.0	1.0	2.0
5. Analogy	7.0	0.0	0.0	7.0	0.0	3.0
6. Causal rel.	1.0	6.0	0.0	2.0	0.0	4.0
7. Sptemp rel.	26.0	9.0	2.0	2.0	0.0	3.0
8. Ellipsis	13.0	4.0	3.0	16.0	2.0	15.0
9. Bridging	69.0	26.0	42.0	59.0	36.0	50.0
10. Elaboration	60.0	8.0	13.0	57.0	18.0	36.0
11. Meta	1.0	1.0	0.0	0.0	0.0	0.0
12. Clause rel.	52.0	40.0	28.0	42.0	27.0	34.0
13. Punctuation	34.0	1.0	24.0	20.0	14.0	25.0
Nonsense	10.0	1.0	3.0	27.0	14.0	1.0

Result: Frequencies (%) of Prerequisite Skills

Skills	QA4MRE	MCTest	SQuAD	WDW	MARCO	NewsQA
1. Tracking	11.0	6.0	3.0	8.0	6.0	2.0
2. Math.	4.0	4.0	0.0	3.0	0.0	1.0
3. Coref. resol.	32.0	49.0	13.0	19.0	15.0	24.0
4. Logical rsng.	15.0	2.0	0.0	8.0	1.0	2.0
5. Analogy	7.0	0.0	0.0	7.0	0.0	3.0
6. Causal rel.	1.0	6.0	0.0	2.0	0.0	4.0
7. Sptemp rel.	26.0	9.0	2.0	2.0	0.0	3.0
8. Ellipsis	13.0	4.0	3.0	16.0	2.0	15.0
9. Bridging	69.0	26.0	42.0	59.0	36.0	50.0
10. Elaboration	60.0	8.0	13.0	57.0	18.0	36.0
11. Meta	1.0	1.0	0.0	0.0	0.0	0.0
12. Clause rel.	52.0	40.0	28.0	42.0	27.0	34.0
13. Punctuation	34.0	1.0	24.0	20.0	14.0	25.0
Nonsense	10.0	1.0	3.0	27.0	14.0	1.0

Result: Frequencies (%) of Prerequisite Skills

Skills	QA4MRE	MCTest	SQuAD	WDW	MARCO	NewsQA
1. Tracking	11.0	6.0	3.0	8.0	6.0	2.0
2. Math.	4.0	4.0	0.0	3.0	0.0	1.0
3. Coref. resol.	32.0	49.0	13.0	19.0	15.0	24.0
4. Logical rsng.	15.0	2.0	0.0	8.0	1.0	2.0
5. Analogy	7.0	0.0	0.0	7.0	0.0	3.0
6. Causal rel.	1.0	6.0	0.0	2.0	0.0	4.0
7. Sptemp rel.	26.0	9.0	2.0	2.0	0.0	3.0
8. Ellipsis	13.0	4.0	3.0	16.0	2.0	15.0
9. Bridging	69.0	26.0	42.0	59.0	36.0	50.0
10. Elaboration	60.0	8.0	13.0	57.0	18.0	36.0
11. Meta	1.0	1.0	0.0	0.0	0.0	0.0
12. Clause rel.	52.0	40.0	28.0	42.0	27.0	34.0
13. Punctuation	34.0	1.0	24.0	20.0	14.0	25.0
Nonsense	10.0	1.0	3.0	27.0	14.0	1.0

Result: Frequencies (%) of Prerequisite Skills

Skills	QA4MRE	MCTest	SQuAD	WDW	MARCO	NewsQA
1. Tracking	11.0	6.0	3.0	8.0	6.0	2.0
2. Math.	4.0	4.0	0.0	3.0	0.0	1.0
3. Coref. resol.	32.0	49.0	13.0	19.0	15.0	24.0
4. Logical rsng.	15.0	2.0	0.0	8.0	1.0	2.0
5. Analogy	7.0	0.0	0.0	7.0	0.0	3.0
6. Causal rel.	1.0	6.0	0.0	2.0	0.0	4.0
7. Sptemp rel.	26.0	9.0	2.0	2.0	0.0	3.0
8. Ellipsis	13.0	4.0	3.0	16.0	2.0	15.0
9. Bridging	69.0	26.0	42.0	59.0	36.0	50.0
10. Elaboration	60.0	8.0	13.0	57.0	18.0	36.0
11. Meta	1.0	1.0	0.0	0.0	0.0	0.0
12. Clause rel.	52.0	40.0	28.0	42.0	27.0	34.0
13. Punctuation	34.0	1.0	24.0	20.0	14.0	25.0
Nonsense	10.0	1.0	3.0	27.0	14.0	1.0

Result: Numbers of Required Skills

#Skills	QA4MRE	MCTest	SQuAD	WDW	MARCO	NewsQA
0	2.0	18.0	27.0	2.0	15.0	13.0
1	13.0	36.0	33.0	5.0	35.0	26.0
2	13.0	24.0	24.0	14.0	29.0	23.0
3	20.0	15.0	6.0	22.0	6.0	25.0
4	14.0	4.0	6.0	16.0	2.0	9.0
5	13.0	1.0	1.0	6.0	0.0	2.0
6	10.0	1.0	0.0	6.0	0.0	1.0
7	1.0	0.0	0.0	2.0	0.0	0.0
8	1.0	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	0.0
10	3.0	0.0	0.0	0.0	0.0	0.0
Ave.	3.25	1.56	1.28	2.43	1.19	1.99

Result: Numbers of Required Skills

#Skills	QA4MRE	MCTest	SQuAD	WDW	MARCO	NewsQA
0	2.0	18.0	27.0	2.0	15.0	13.0
1	13.0	36.0	33.0	5.0	35.0	26.0
2	13.0	24.0	24.0	14.0	29.0	23.0
3	20.0	15.0	6.0	22.0	6.0	25.0
4	14.0	4.0	6.0	16.0	2.0	9.0
5	13.0	1.0	1.0	6.0	0.0	2.0
6	10.0	1.0	0.0	6.0	0.0	1.0
7	1.0	0.0	0.0	2.0	0.0	0.0
8	1.0	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	0.0
10	3.0	0.0	0.0	0.0	0.0	0.0
Ave.	3.25	1.56	1.28	2.43	1.19	1.99

Calculation of Readability

- ❖ Ave. Num. of characters per word (*NumChar*)
- ❖ Ave. Num. of syllables per word (*NumSyll*)
- ❖ Ave. sentence length in words (*MLS*)
- ❖ Proportion of words in AWL (*AWL*)
- ❖ Modifier variation (*ModVar*)
- ❖ Num. of coordinate phrases per sentence (*CoOrd*)
- ❖ Coleman-Liau index (*Coleman*)
- ❖ Dependent clause to clause ratio (*DC/C*)
- ❖ Complex nominals per clause (*CN/C*)
- ❖ Adverb variation (*AdvVar*)

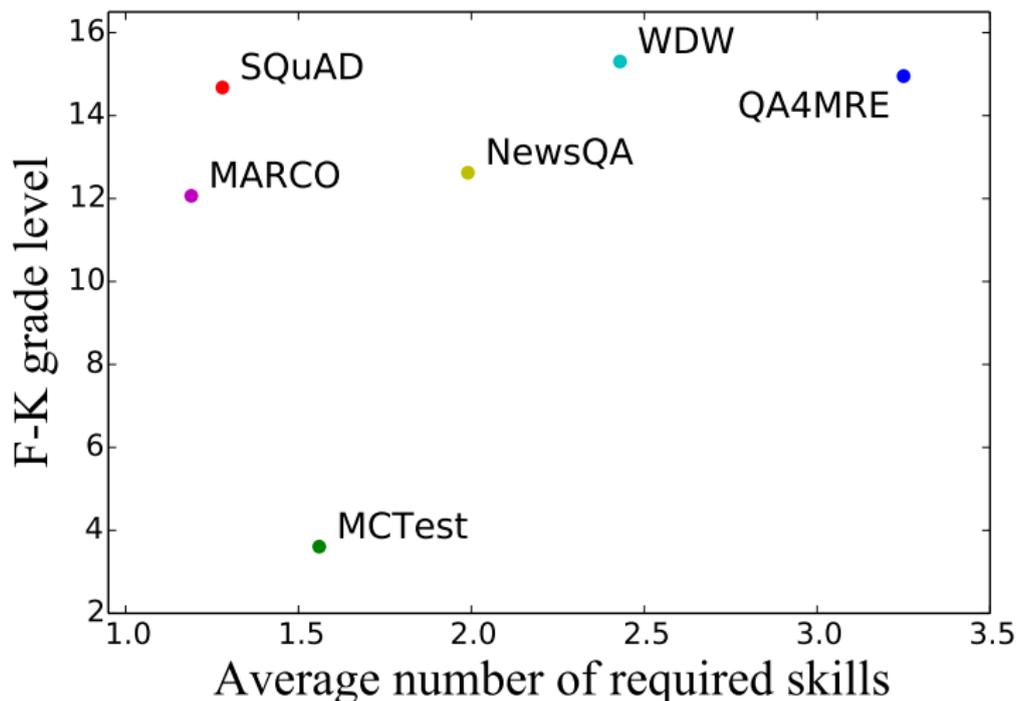
Figure: 10 readability measure from Vajjala and Meurers (2012).

Result: Readability Metrics

Mesures	QA4MRE	MCTest	SQuAD	WDW	MARCO	NewsQA
NumChar	5.026	<u>3.892</u>	5.378	4.988	5.016	5.017
NumSyll	1.663	<u>1.250</u>	1.791	1.657	1.698	1.635
MLS	28.488	<u>11.858</u>	23.479	29.146	19.634	22.933
AWL	0.067	<u>0.003</u>	0.071	0.033	0.047	0.038
ModVar	0.174	<u>0.114</u>	0.188	0.150	0.186	0.138
CoOrd	0.922	<u>0.309</u>	0.722	0.467	0.651	0.507
Coleman	12.553	<u>4.333</u>	14.095	12.398	11.836	12.138
DC/C	0.343	0.223	0.243	0.254	<u>0.220</u>	0.264
CN/C	1.948	<u>0.614</u>	1.887	2.310	1.935	1.702
AdvVar	0.038	0.035	0.032	<u>0.019</u>	0.022	<u>0.019</u>
F-K	14.953	<u>3.607</u>	14.678	15.304	12.065	12.624
Words	1545.7	174.1	130.4	253.7	<u>70.7</u>	638.4

*F-K = Flesch-Kincaid grade level
= education level required to understand the text.

Relation btwn Skills and Readability

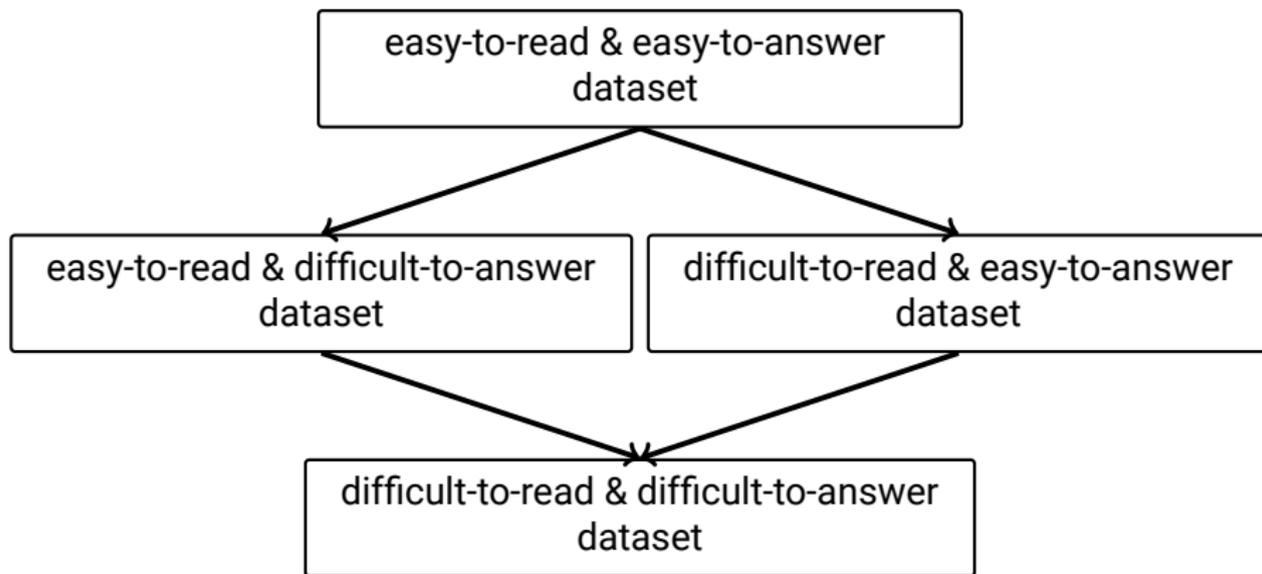


Observation

- ✦ There is only a weak correlation between readability metrics and numbers of required skills
- Difficult to read \nRightarrow difficult to answer (and vice versa)
- It is possible to create a dataset that consists of an easy-to-read context and difficult-to-answer questions.

How to Use This Observation?

For development of RC systems, select datasets in the following steps:



Discussion: Answerability and RC as TE

Answerability of questions

- ✦ We cannot evaluate the difference among “truly difficult,” “non sense,” and “no answer” when we encounter systems’ incorrect answers (and even in human performance!)
- ✦ It is not easy to maintain the quality of questions especially in crowd-based sourcing.

Discussion: Answerability and RC as TE

Answerability of questions

- ✦ We cannot evaluate the difference among “truly difficult,” “non sense,” and “no answer” when we encounter systems’ incorrect answers (and even in human performance!)
- ✦ It is not easy to maintain the quality of questions especially in crowd-based sourcing.

Reading comprehension seen as textual entailment

Multiple premises (from context) \Rightarrow hypothesis (from Q+A)

- ✦ Our methodology cannot deal with the competence of:
 1. Collecting evidences from context
 2. Generating hypothesis from question and answer candidates

Discussion: Genre and other NLP fields

Corpus genre

- ✦ Are Wikipedia or news articles enough?

Example: narratives are close to our everyday experience (characters' emotions, intentions, and actions)

Discussion: Genre and other NLP fields

Corpus genre

- ✦ Are Wikipedia or news articles enough?

Example: narratives are close to our everyday experience (characters' emotions, intentions, and actions)

Relation between RC and other NLP fields

- ✦ How can RC systems contribute to other fields in which natural language understanding is required?
- ✦ RC may be able to lay a foundation of “contextual” NLU.
 - ✦ Example: dialogue systems
dialogue history = context, utterance = question

Summary

Proposed evaluation metrics for RC

1. Defined two classes of metrics:
prerequisite skills and readability
 2. Annotated RC datasets with the skills
 3. Calculated readability of datasets
 4. Analyzed datasets
- Results can be used for evaluation of systems

Observation

- ✦ There is only a weak correlation between readability metrics and numbers of required skills

Appendix: Correlation btwn Readability and the Skills

Metrics	r	p
NumChar	0.067	0.161
NumSyll	0.057	0.235
MLS	0.411	0.000
AWL	0.160	0.001
ModVar	0.063	0.189
CoOrd	0.194	0.000
Coleman	0.147	0.002
DC/C	0.174	0.000
CN/C	0.167	0.000
AdvVar	0.007	0.882
F-K	0.348	0.000

Table: Pearson's correlation coefficients (r) with the p-values (p) in all RC datasets

Appendix: Textual Entailment and RC

- ✦ Textual entailment
 - ✦ Recognizing and testing:
premise → hypothesis
- ✦ Reading comprehension as textual entailment
 - ✦ Recognizing and testing:
Multiple premises → hypothesis

Issue:

Our metrics cannot evaluate the following processes in RC:

- ✦ Multiple premises ← gathered from **context sentences**
 - ✦ hypothesis ← generated from **answer candidates**
- ⇒ depending on context length and question formulation

Appendix: Knowledge reasoning

- ✦ Ellipsis

- ✦ Recognizing implicit/omitted information
- ✦ e.g. *She is a smart student* → *She is a student*

Appendix: Knowledge reasoning

✦ Ellipsis

- ✦ Recognizing implicit/omitted information
- ✦ e.g. *She is a smart student* → *She is a student*

✦ Bridging

- ✦ Inferences supported by grammatical and lexical information
- ✦ e.g. *She loves sushi.* → *She likes sushi.*

Appendix: Knowledge reasoning

❖ Ellipsis

- ❖ Recognizing implicit/omitted information
- ❖ e.g. *She is a smart student* → *She is a student*

❖ Bridging

- ❖ Inferences supported by grammatical and lexical information
- ❖ e.g. *She loves sushi.* → *She likes sushi.*

❖ Elaboration

- ❖ Inference using known facts and general knowledge
- ❖ e.g. *The writer of Hamlet was Shakespeare* → *Shakespeare wrote Hamlet*

Appendix: Knowledge reasoning

❖ Ellipsis

- ❖ Recognizing implicit/omitted information
- ❖ e.g. *She is a smart student* → *She is a student*

❖ Bridging

- ❖ Inferences supported by grammatical and lexical information
- ❖ e.g. *She loves sushi.* → *She likes sushi.*

❖ Elaboration

- ❖ Inference using known facts and general knowledge
- ❖ e.g. *The writer of Hamlet was Shakespeare* → *Shakespeare wrote Hamlet*

❖ Meta-knowledge

- ❖ Inference using external knowledge including a reader, writer, and text genre
- ❖ e.g. *Who is the main character in this story?*