Reasoning about Quantities in Natural Language

Subhro Roy, Tim Vieira, Dan Roth

TACL 2015

読み手 : 稲田 和明 (東北大学)

2015/08/30

Overview

e.g. Textual Entailment

✓ Quantity reasoning is important to understand NL.

-> Little work in NLP has analyzed the use of quantities

Quantities Extraction

Key steps necessary to facilitate quantity reasoning

RTEの部分タスク

Quantity Entailment

Currency Range Search

Elementary Math Word Problem

Tasks of numerical reasoning investigated in this work

Overview

e.g. Textual Entailment

✓ Quantity reasoning is important to understand NL.

-> Little work in NLP [illegible] the use of quantities

Quantities Extraction

Contribution 1

Detecting and normalizing quantities in unrestricted English text

(問題点: 正規化の評価なし)

Key steps necessary to facilitate quantity reasoning

Quantity Entailment

Currency Range Search

Elementary Math Word Problem

Contribution 2

Proposing the tasks and the methods for these and creating new dataset of the evaluations

Tasks of numerical reasoning investigated in this work

Contribution 3

High performance on these tasks.

(問題点: 既存研究との比較なし)

Overview

e.g. Textual Entailment

✓ Quantity reasoning is important to understand NL.

-> Little work in NLP has analyzed the use of quantities

A Quantities Extraction

B Quantity Entailment

~~Currency Range Search~~

C Elementary Math Word Problem

Key steps necessary to facilitate quantity reasoning

時間の都合上省略

Tasks of numerical reasoning investigated in this work

A: Quantities Extraction

Input: Text

"… killed five Americans …"

1. Segmentation
   - Taking raw test and finds segments of contiguous text which describe quantities.
2. Standardization
   - Derive the quantity-value representation(QVR)
   - Converting value and extracting unit

Output: quantity-value representation(QVR)

- A quantity is represented as a triple (value, unit, change)

" (5.0, American, -)"

Representing Quantity (QVR)

A quantity is represented as a triple (value, unit, change)

- ✓***value***: a numeric value, range, or set of values which measure the aspect.
  - e.g. more than 500, one or two, March 18, 1986.
  - more than 500 → (500,+∞) (convert to a set or range)
- ✓***unit***: a noun phrase that describes what the value is associated with.
  - e.g. inches, minutes, bananas, American
- ✓*change*: specifies how the parameter is changing
  - e.g. increasing

A: Quantities Extraction

Algorithm 1 QuantityExtraction(T)

Input: Text T

Output: Set of Quantity-value triples extracted from T

```
1: Q ← ∅
2: S ← Segmentation( T )
3: for all segment s ∈ S do
4:     q ← Standardization( s )
5:     if unit of q not inferred then
6:         q ← InferUnitFromSemantics( q, s, T )
7:     end if
8:     Q ← Q ∪ {q}
9: end for
10: return Q
```

1

Find segments of contiguous text which describe quantities

2-1

Convert to Quantity-value

2-2

Extraction of Unit with inference

1. Segmentation

… [nine people], including [five Americans].

Find quantities representation

✓Model

- Semi-CRF (Sarawagi and Cohen 2004)
- a bank of classifiers model (Punyakanok and Roth 2001)

✓Feature

- Word class features
  - Appear in List of scientific units (e.g. meters), written numbers (two, fifteen), name of temporal word (today, tomorrow), etc
- Character-based
  - Contains digit, all digits, has suffix (st, nd, rd, th)
- Part of speech tags

上述の素性を前後3単語に適用

Experiment: Segmentation

Model	P%	R%	F%
Semi-CRF (SC)	75.6	77.7	76.6
C+I (PR)	80.3	79.3	79.8

↑ the bank of classifiers model (Punyakanok and Roth 2001)

参考 P:0.86, R:0.78
日本語の数量抽出（ルールベース）と同程度
（https://github.com/nullnull/normalizeNumexp）

- ✓Task:
  - boundary recognizer of quantities representation
- ✓Data:
  - 384 text-hypothesis pairs and 600 sentences (agreement 0.91)
- ✓Method:
  - 10 fold cross-validation

2-1. Standardization

[nine people] → (9.0, people, -) Convert to QVR

✓Model: Rule-based (全ルール・詳細は § 4.2 参照)

1. Convert written numbers to floating point:

   three thousand five hundred twenty → 3520.0

3. Replace known names for ranges

   teenage → [13, 19] years-old

4. Convert all scientific units to a standard base unit

   1 mile → 1609.344 meters

6. Rewrite known units to a standard unit

   USD, US$, dollars → US$

7. Standardize changing quantity

   additional 10 → + 10

unitを推定やchangeの決定

2-2. Extraction of Units with inference

✓Most units adjacent to numeric value

→ Some cases need to infer the unit

Example 4

A report from UNAIDS, the Joint United Nations Program on HIV/AIDS, released on Tuesday, shows the number of adults and children with HIV/AIDS reached 39.4 million in 2004.

unit

quanitity

Semantic Role Labeling

Example 5

unit

The number of member nations was 80 in 2000, and then it increased to 95.

quanitity

SRL +

Coreference Resolution

the segmented chunk does not have adequate information to infer the unit.

↓

The Illinois Coreference Resolver (Bengtson and Roth, 2008; Chang et al., 2013)

The Illinois SRL (Punyakanok et al., 2008)

Experiment: Standardization

✓本来あるべきと思うが、評価なし。

・著者の主張（原文ママ）

We do not directly evaluate our system's ability to map raw text segments into our representation, but instead evaluate this capability extrinsically, in the context of the aforementioned tasks, since good Standardization is necessary to perform quantitative inference.

- ✓ 必要性は感じているっぽい
- ✓「Extraction of Units with inference」の効果はQEのタスクで確認している

読み手の主観

アノテーション時のコストが大きな問題？（agreement, annotate design）
Quantity Reasoning が目的なので、中間情報への関心は乏しい？

B: Quantity Entailment (QE)

Input: Text T and a QVR h(v, u, c)

T: "killed [five Americans] and [four Israelis].
h: (9.0, people, -)

1. Quantity Extraction
   - Extracting quantities in T
2. Generate Implicit value
   - Producing implicit quantity by rule-based method
3. Compare Quantities
   - Compare h with T's quantities
   - Return to entail / contradict / no-relation

Output:

entails: there exists a quantity in T which entails h
contradict: no quantity in T entail h, but there is quantity in T which contradict h
no-relation: there exists no quantity in T, which is comparable with h

B: Quantity Entailment (QE)

Algorithm 3 QuantityEntailment(T, h)

Input: Text T and a quantity-value triples $h(v_h, u_h, c_h)$

Output: Returns whether T *entails*, *contradicts* or has *no relation* with h

```
1:  Q ← QuantityExtraction( T )
2:  Q' ← GenerateImplicitQuantities( Q )
3:  Q ← Q ∪ Q'
4:  contradict ← false
5:  for all quantity-value triple q ∈ Q do
6:      if QuantityComparison( q, h )= entails then
7:          return entails
8:      end if
9:      if QuantityComparison( q, h )= contradicts
        then
10:         contradict← true
11:     end if
12: end for
13: if contradict= true then
14:     return contradicts
15: else
16:     return no relation
17: end if
```

1 Extract Quantity representation — 説明済み

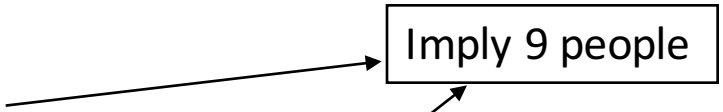
2 Infer implicit quantity

3 Compare Quantities

Main part of QE

Judge entail / contradict / no-relation by comparing h with T

2. Generate Implicit Value



T: *"…killed [five Americans] and [four Israelis]."*
H: *"…killed [nine people] …"* → Entail

1. Range may imply duration
   "lived in X from 1980 to 2000" → *lived* for a duration of 20 years.
2. Compatible terms may be combined and abstracted
   "3 bananas, 2 oranges" → *5* fruits
3. Ratios can imply percentages
   "9 out of the 10 people" → *90% people*
4. Composition
   "*six couples*" → 12 people

* 1, 3に関するルールは記述されているが、2，4に関しての記述はない

3. Quantity Comparison

Algorithm 2 QuantityComparison(t, h)

Input: Quantity-value triples $t(v_t, u_t, c_t)$ and $h(v_h, u_h, c_h)$

Output: Returns whether t *entails*, *contradicts* or has *no relation* with h

```
1:  if c_t ≠ c_h then
2:      return no relation
3:  end if
4:  if comparableUnits( u_t, u_h )= false then
5:      return no relation
6:  end if
7:  if checkMonotonicityOfUnits( u_t, u_h )= true
    then
8:      if checkMonotonicityOfValues( v_t, v_h )=
        true then
9:          return entails
10:     end if
11: end if
12: return contradicts
```

1
Comparison change

2
Check comparable of unit

3
Check monotonicity of unit
u_t is more specific → true

American ⊆ *People*

4
Check monotonicity of value
v_t is more specific → true

03/18/1986 ⊆ *March 1986*

Entailの条件: changeが同じ & unit が比較可能 & $u_t \subseteq u_h$ & $v_t \subseteq v_h$

Data: Quantity Entailment

✓Sub-corpus of the RTE Datasets (Dagan et al. 2006)

RTE2-4 datasets, which have quantity mentions in the h

- annotated entailment pairs with information about which quantities entail, in addition to the boundary information.

Annotation Example for RTE sub-corpus

T:*A bomb in a Hebrew University cafeteria killed [five Americans] and [four Israelis].*

H:*A bombing at Hebrew University in Jerusalem killed [nine people], including [five Americans].*

h_1 h_2

"nine people" : entails

"five Americans" : entails

Global entailment decision : entails

Entail: 309

Contradict: 71

No-relation: 56

Agreement: 0.95

For each quantity in H labeled entail, no-relation, or contradict

Experiment: Result of QE

Task	System	P%	R%	F%
Entailment	Baseline	100.0	43.3	60.5
	GOLDSEG	98.5	88.0	92.9
	+SEM	97.8	88.6	93.0
	PREDSEG	94.9	76.2	84.5
	+SEM	95.4	78.3	86.0
Contradiction	Baseline	16.6	48.5	24.8
	GOLDSEG	61.6	92.9	74.2
	+SEM	64.3	91.5	75.5
	PREDSEG	51.9	79.7	62.8
	+SEM	52.8	81.1	64.0
No Relation	Baseline	41.8	71.9	52.9
	GOLDSEG	81.1	76.7	78.8
	+SEM	80.0	78.5	79.3
	PREDSEG	54.0	75.4	62.9
	+SEM	56.3	72.7	63.5

1. GOLDSEG : Uses gold segmentation, and does not use SRL and Coreference Resolver.
2. GOLDSEG+SEM : Uses gold segmentation, and also uses SRL and Coreference Resolver to infer units.
3. PREDSEG : Performs segmentation, and does not use SRL and Coreference Resolver.
4. PREDSEG+SEM : Performs segmentation, and uses SRL and Coreference Resolver.

+SEM consistently improves performance

疑問: Generate Implicit Valueの効果は？

✓Baseline: an exact string match

"entails" if the quantity unit and value are present in the text, and answers "contradicts" if only the unit matches and the value does not. Otherwise, it returns "no relation".

C: Math Word Problem

Input: math word problem

"Roger reads 2 books in 1 day. How many books will he read in 3 days ?"

limitation

- The question mentions two or three quantities
- .The answer can be computed by choosing two quantities from the question and applying one of the four basic operations

0. Quantity Extraction
1. Quantity Pair Classifier
2. Operation Classifier
3. Order Classifier

a cascade of three classifiers by average perceptron

Output: Answer (2 * 3 = 6)

Features for Math Word Problem

1. Unigrams and bigrams from sentences containing quantities
2. POS tags from sentences with quantities
3. Relevant pair of quantities, and whether their units match and whether their units are present in the last sentence of the question

共通素性

4. Relevant operation for the problem (for Operation and Order classifiers)
5. Relevant order of quantities for the operation (for Order classifier)

分類器個別素性

6. Various conjunctions of the above feature

組み合わせ素性

Classifier of Math Word Problem

w: weight vector, Φ: feature function

1. Quantity Pair Classifier

$$(q_i, q_j) \leftarrow \arg\max_{p \in P} w_{qp}^T \phi_{qp}(Q, p)$$

where $P = \{(q_1, q_2), (q_2, q_3), (q_3, q_1)\}$, $\phi_{qp}(\cdot)$ is a feature function, and w_{qp} is a learned weight vector.

Input:
P: {(q1, q2) (q2, q3) (q3, q1)}, Q: problem text
Output:
p: (qi, qj)

2. Operation Classifier

$$op \leftarrow \arg\max_{op \in O} w_{opr}^T \phi_{opr}(Q, (q_i, q_j), op)$$

where $O = \{+, -, \times, /\}$.

Input:
Q, (qi, qj), basic-operator set (+, - , * , /)
Output:
/ (one of operator)

3. Order Classifier

$$(q_i', q_j') \leftarrow \arg\max_{p \in P} w_{or}^T \phi_{or}(Q, (q_i, q_j), op, p)$$

where $P = \{(q_i, q_j), (q_j, q_i)\}$

Input:
Q, (qi, qj) / (operator)
Output:
order (qi, qj) or (qj, qi)

Data: Math Word Problem

✓Collected from

- http://www.k5learning.com/
- http://www.dadsworksheets.com/

✓Limitation

- 2 or 3 quantities in a question.
- Remove problems requiring background knowledge
- Excluded rounding issues

#quantities	Relevant Operation			
	Add	Subtract	Multiply	Divide
2	228	214	257	260
3	107	132	75	131

Experiment: math word problem

Module	Accuracy
Quantity Pair	94.3
Operation	91.8
Order	95.9
Correct Answer	86.9

Individual evaluation

- ✓2-fold cross-validation
- ✓The accuracy of each classifier is based only on the relevant examples for that particular classifier.
- ✓the performance of individual classifiers without "Correct Answer".
- ✓Correct Answer denotes the end to end system

Conclusion & Future work

✓Conclusion

- Proposed for detecting and normalizing quantities in unrestricted English text.

- Support the reasoning required by Quantity Entailment and elementary school level math word problems.

✓Future work

- Focus on alleviating some of the limitations of the inference module to subtle and deeper language understanding

Scope of QE Inference

- ✓ we attribute these limitations to subtle and deeper language understanding.

T : Adam has exactly 100 dollars in the bank.
H_1 : Adam has 50 dollars in the bank.
H_2 : Adam's bank balance is 50 dollars.

- ✓ T imply H1, but not H2.

- ✓ Both H1 and H2, QE will infer that "50 dollars" is a contradiction to sentence T.

cannot make the subtle distinction required here

Related Work

- ✓Quantities have been recognized as an important part of a textual entailment system

  (de Marneffe et al., 2008; Maccartney and Manning, 2008; Garoufi, 2007; Sammons et al., 2010)

- ✓Perspective of Formal Semantics

  Reasoning about quantities often depends on reasoning about monotonicity (Barwise and Cooper, 1981)

- ✓In specific domains (temporal domain)

  TimeML project (Pustejovsky et al., 2003; Saur et al., 2005; Pratt- Hartmann, 2005; Do et al., 2012)

- ✓Automatically solving math word problems

  (Bobrow, 1964; Lev et al., 2004; Mukherjee and Garain, 2008; Kushman et al., 2014)