MLEnt: The Machine Learning Entailment System of the University of Alicante

Zornitsa Kozareva, Andrés Montoyo

{zkozareva,montoyo}@dlsi.ua.es

Departamento de Lenguajes y Sistemas Informáticos
University of Alicante

April 10th, 2006

Outline

1. Motivation
2. Overview of MLEnt System
3. Feature extraction
   - lexical feature extraction
   - semantic feature extraction
4. Classifier combination
5. Results of RTE2 data
6. Conclusions
   - Work in progress
   - Future work

Motivation

- avoid hand-made/empirically derived thresholds
- incorporate different knowledge information at the same time (e.g. lexical, semantic)
- apply techniques for classifier combination

Motivation

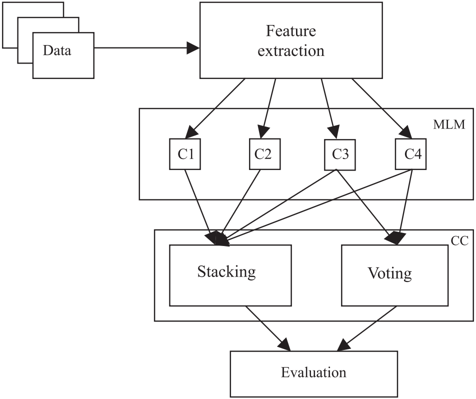
- avoid hand-made/empirically derived thresholds
- incorporate different knowledge information at the same time (e.g. lexical, semantic)
- apply techniques for classifier combination

Motivation

- avoid hand-made/empirically derived thresholds
- incorporate different knowledge information at the same time (e.g. lexical, semantic)
- apply techniques for classifier combination

Outline

Overview of MLEnt System



Outline

1. Motivation
2. Overview of MLEnt System
3. Feature extraction
    - lexical feature extraction
    - semantic feature extraction
4. Classifier combination
5. Results of RTE2 data
6. Conclusions
    - Work in progress
    - Future work

Outline

Lexical feature extraction

- f_1: n-gram

Example

<pair id="318" entailment="YES" task="QA">
<t>Mount Olympus towers up from the center of the earth.</t>
<h>Mount Olympus is in the center of the earth. </h>

- f_2: longest common subsequence
- f_3, f_4: skip-gram
- f_5, f_6: negation in T, H

$$A = \{f_1, \ldots, f_7\}, \text{ best features } f_2,\ f_4$$

Lexical feature extraction

- f_1: n-gram
- f_2: longest common subsequence

Example

<pair id="413" entailment="NO" task="QA">
<t>A male rabbit is called a buck and a female rabbit is called a doe, just like deer.</t>
<h>A female rabbit is called a buck. </h>

- f_3, f_4: skip-gram
- f_5, f_6: negation in T, H

$A = \{f_1, \ldots, f_7\}$, best features f_2, f_4

Lexical feature extraction

- f_1: n-gram
- f_2: longest common subsequence
- f_3, f_4: skip-gram

Example

<pair id="419" entailment="YES" task="QA">
<t>Elizabeth Dowdeswell is the Under Secretary General at the United Nations Offices at Nairobi and Executive Director of the United Nations Environment Programme.</t>
<h>Elizabeth Dowdeswell is Executive Director of the United Nations Environment Programme.</h>

- f_5, f_6: negation in T, H

Lexical feature extraction

- f_1: n-gram
- f_2: longest common subsequence
- f_3, f_4: skip-gram
- f_5, f_6: negation in T, H

$$A = \{f_1, \ldots, f_7\}, \text{ best features } f_2,\ f_4$$

Outline

Semantic feature extraction

- f_7: number

Example

four-thousand and 4000, 4-years-old and four-years-old
less than 5, more than 5

- f_8: proper name
- f_9: adverb
- f_{10}: adjective
- f_{11}, f_{12},f_{13}, f_{14}: noun_lin , noun_path, verb_lin, verb_path
- f_{15}, f_{16}: noun_verb_lin, noun_verb_path
- f_{17}: sim of all units in sentence

$$B = \{f_7, \ldots, f_{14}\}, \text{ best features } f_8 \text{ , } f_{14}$$

Semantic feature extraction

- f_7: number
- $f_8, \ldots, f_{14}$: proper name, adverb, adjective, noun_lin, noun_path, verb_lin, verb_path

Example

generate all t_i h_j adjective/noun/verb pairs
match lexically/semantically
normalize over the total number of generated t_i h_j word pairs

- f_{15}, f_{16}: noun_verb_lin, noun_verb_path
- f_{17}: sim of all units in sentence

$$B = \{f_7, \ldots, f_{14}\}, \text{ best features } f_8 , f_{14}$$

Semantic feature extraction

- f_7: number
- f_8: proper name
- f_9: adverb
- f_{10}: adjective
- f_{11}, f_{12},f_{13}, f_{14}: noun_lin , noun_path, verb_lin, verb_path
- f_{15}, f_{16}: noun_verb_lin, noun_verb_path
- f_{17}: sim of all units in sentence

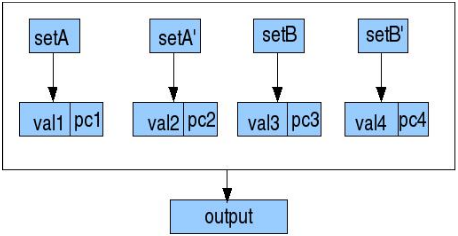
$$B = \{f_7, \ldots, f_{14}\}, \text{ best features } f_8 \text{ , } f_{14}$$

Outline

1. Motivation
2. Overview of MLEnt System
3. Feature extraction
   - lexical feature extraction
   - semantic feature extraction
4. Classifier combination
5. Results of RTE2 data
6. Conclusions
   - Work in progress
   - Future work

Classifier combination

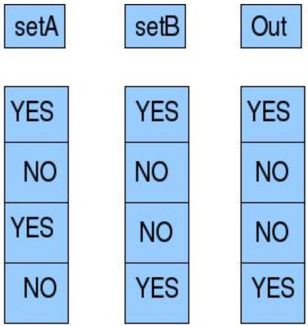
- stacking



- voting

Classifier combination

- stacking
- voting



Outline

1. Motivation
2. Overview of MLEnt System
3. Feature extraction
   - lexical feature extraction
   - semantic feature extraction
4. Classifier combination
5. **Results of RTE2 data**
6. Conclusions
   - Work in progress
   - Future work

Results of the RTE2 data

	Set	Acc.	IE	IR	QA	SUM
development	A	56.87	49.50	55.50	51.00	71.50
	A'	57.75	49.50	57.00	51.50	73.00
	B	60.12	54.00	61.00	59.00	66.50
	B'	57.13	54.50	61.00	49.50	63.50
	S	68.63	58.50	68.00	70.50	77.50
	V	62.12	54.50	61.50	62.00	70.50
test	A	51.75	52.00	53.50	55.50	46.00
	A'	56.75	46.00	55.50	56.50	69.00
	B	54.25	50.00	55.50	47.50	64.00
	B'	52.38	49.00	51.50	52.00	57.00
	S	54.87	54.00	56.00	50.00	59.50
	V	55.00	49.50	54.00	54.50	62.00

Outline

Conclusions

- high data sparsity problem
- need more examples
- levels of information e.g. lexical, semantic, syntactic
- performance depends on the data

Outline

1. Motivation
2. Overview of MLEnt System
3. Feature extraction
   - lexical feature extraction
   - semantic feature extraction
4. Classifier combination
5. Results of RTE2 data
6. Conclusions
   - Work in progress
   - Future work

Work in progress

- consider antonym relation
- measure similarity of inter-syntactic relations (e.g. noun-verb, adverb-noun, adjective-verb etc)
- LSA and cosine measure with corpus-based and Relevant Domains
- create multiple classifiers with this information

Work in progress

- consider antonym relation
- measure similarity of inter-syntactic relations (e.g. noun-verb, adverb-noun, adjective-verb etc)
- LSA and cosine measure with corpus-based and Relevant Domains
- create multiple classifiers with this information

Work in progress

- consider antonym relation
- measure similarity of inter-syntactic relations (e.g. noun-verb, adverb-noun, adjective-verb etc)
- LSA and cosine measure with corpus-based and Relevant Domains
- create multiple classifiers with this information

Work in progress

- consider antonym relation
- measure similarity of inter-syntactic relations (e.g. noun-verb, adverb-noun, adjective-verb etc)
- LSA and cosine measure with corpus-based and Relevant Domains
- create multiple classifiers with this information

Outline

1. Motivation
2. Overview of MLEnt System
3. Feature extraction
   - lexical feature extraction
   - semantic feature extraction
4. Classifier combination
5. Results of RTE2 data
6. Conclusions
   - Work in progress
   - Future work

Future work

- improve classifier's confidence score
- include syntactic information
- apply MLEnt to IR
- participate in the Answer Validation competition of CLEF

MLEnt: The Machine Learning Entailment System of the University of Alicante

Zornitsa Kozareva, Andrés Montoyo

{zkozareva,montoyo}@dlsi.ua.es

Departamento de Lenguajes y Sistemas Informáticos
University of Alicante

April 10th, 2006