Thesaurus
- Expanding the range of possible terms in information retrieval
- Example-based machine translation system
- Most technical terminology is not included

Automatic thesaurus construction
- Determining whether a word pair has a hierarchical relation or a disrelation

Exponential weight method (Suzuki, 2003)

Hyponym
Word Y
Estimation using linear discrimination analysis (proposed method)

The CYCLONE CORPUS
- Collection procedure (Fujii, 2005)
  - Searching the Web for pages including a target term
  - Analyzing the layout of the pages and identifying paragraphs that potentially describe the target term
  - Classifying multiple paragraphs to predefined domains

Hierarchical relation estimation method
- Descriptions of words have "directionality"

Lion
A large wild animal of the cat family with yellowish brown fur ...

Animal
A living creature such as a dog or cat that is generally distinguished from plants ...

Persian cat
A type of cat with long hair, short thick legs and a round face ...

Cat
A small four-legged furry animal with a tail and claws ...

A collection of descriptions for a headword shares common hyponyms but may not share hyponyms

Target function
\[ H(X|Y) = C(Y|X) - C(Y) \]

\[ C(X|Y) : \text{Probability of } X \text{ given the descriptions of } Y \]

Semantic expansion technique for descriptions

Persian cat is a type of cat with long hair, short thick legs and a round face ...

Cat is a small four-legged furry animal with a tail and claws ...

Headword (Target term)
- Description 1
- Description 2

New terms and rarer terms
Many descriptions for each headword
Semantic relations between a headword and terms in its description

Extracting hierarchical relations effectively

Hierarchical relation data

Evaluation

Word pairs in the computer-related domain (2074 words)

<table>
<thead>
<tr>
<th>Correctness of description</th>
<th>Average number of descriptions per headword</th>
<th>Test sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6.6</td>
<td>136 pairs</td>
</tr>
<tr>
<td>A+B</td>
<td>10.4</td>
<td>301 pairs</td>
</tr>
<tr>
<td>ALL</td>
<td>80.7</td>
<td>168 pairs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>366 pairs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>206 pairs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>497 pairs</td>
</tr>
</tbody>
</table>

Example

ISO, MPEG
Bit error rate
MS-DOS

ISO, MPEG
Bit error rate
MS-DOS

Method 1
Estimation using linear discrimination analysis (proposed method)

Method 2
Exponential weight method (Suzuki, 2003)

\[ C = \lim_{b \to 0} b (a + b)^2 \]

\[ 0 < a \leq 1, \quad b = \frac{1-a}{a} \]

Hierarchical relation or disrelation?

Weights for the expanded matrices

(Weights are normalized by each first-order weight)

Using only the top correct description

Hierarchical discrimination

Effectiveness of using multiple and correct descriptions

Conclusion and future work

Conclusion
- Discrimination for the hierarchical relation of a word pair using an encyclopedic corpus called the CYCLONE corpus

In order not to miss an indirect relationship, a semantic expansion technique for descriptions is used

The proposed method is able to detect 66.1% of relations

Future work
- Discrimination between hierarchical and synonymous relation

MOTIVATION

PREVIOUS WORK

To extract hyponyms, synonyms, and hypernyms,
Sentences that have specific syntactic patterns
("a part of" "is a" "such as")
Descriptions in a dictionary (Suzuki, 2003)
Specific document structure (Shinzato, 2004) are used

It is difficult to cover such an enormous vocabulary
Less frequent words are not expected in the desired expressions

CONCLUSION AND FUTURE WORK

EXPERIMENTAL RESULTS

Effectiveness of using multiple and correct descriptions

The proposed method is able to detect 66.1% of relations

Future work
- Discrimination between hierarchical and synonymous relation

Expanding probability matrix

Identifying the relation of a word pair \((w, w)\) by calculating \(H(w | w)\)