Fuzzy Coreference Resolution for Summarization

René Witte and Sabine Bergler

Concordia University
Montréal, Canada

Context

Fuzzy Coreference Resolution for Summarization:

- 10-word summaries of newspaper articles for DUC *(Document Understanding Conference)* 2003

  ```
  People: construction project, Schulz's
  work, voices, a repository, his "Peanuts"
  strip
  ```

- Summarization algorithm based (solely) on noun phrase coreference chains
- Fuzzy resolution algorithm for explicit uncertainty representation
- Implementation notes and results

Outline

1. Imperfect Information and Fuzzy Theory
2. Fuzzy Coreference Resolution:
   - Fuzzy Coreference Chains
   - Fuzzy Heuristics
   - Fuzzy Chain Merging
   - Evaluation
3. Coreference Chain-based Summarization
4. Evaluation & Future Work

Imperfect Information

Information encountered in NLP are (almost) always *imperfect:* vague, uncertain, imprecise, inconsistent

- Coreference Resolution: mainly *Uncertainty*
- Imperfections cannot be captured with "standard" (crisp) data models
- this leads to *impedence mismatch:* reality is semantically richer than data model used for representation
- consequence: information loss through premature (and unnecessary) interpretation of uncertain data in precise data model

Imperfect Knowledge Representation

How can we deal with imperfect information?

- Statistical models (only uncertainty)
- Fuzzy Set Theory (uncertainty and vagueness)
- "Home-made" models (ad-hoc representations: weights, biases, magic numbers, . . .)
- None (ignore imperfections)

Proposal: choose an appropriate representation formalism as basis for NLP algorithms

Fuzzy Set Theory

Fuzzy logic (fuzzy set theory) is one kind of multi-valued logic that can represent imperfections (both vague and uncertain) explicitly

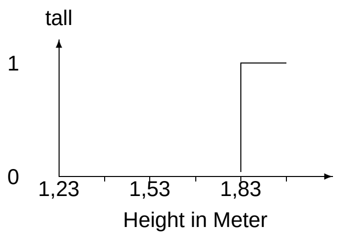
- multi-valued logics first introduced by Jan Lukasiewicz in the 1930s
- promoted as "fuzzy logic" by Lotfi Zadeh since 1965
- now has widespread industrial applications in the area of process control *(fuzzy control)*
- other kinds of applications still in research & development stage

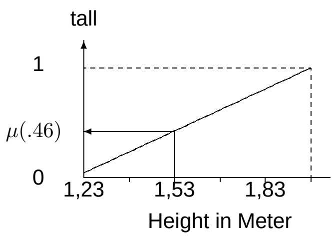
Fuzzy Theory Primer

Fuzzy set μ of Ω:

$$\mu : \Omega \to [0,1]$$

Example: crisp vs. fuzzy set *"tall person"* (vague data)





Fuzzy Set Semantics

Possibilistic Interpretation of fuzzy sets:

- a fuzzy set $\mu_C(\Omega)$ gives an *elastic restriction* of a concept C (predicate) over a set Ω
- and a single value $\mu_C(\omega)$ gives a degree of compatibility of ω with the concept C:

$\mu_C(\omega) = 0$	ω is impossible in C
$\mu_C(\omega) = 1$	ω is completely compatible with C
$\mu_C(\omega) \in (0,1)$	compatibiliy degree of ω with C

Fuzzy Set Operators

Translation of set operators *intersection, union,* and *complement* to fuzzy sets:

- $(\mu_1 \cap \mu_2)(\omega) := \min\{\mu_1(\omega), \mu_2(\omega)\}, \omega \in \Omega$
- $(\mu_1 \cup \mu_2)(\omega) := \max\{\mu_1(\omega), \mu_2(\omega)\}, \omega \in \Omega$
- $\overline{\mu}(\omega) := 1 - \mu(\omega), \omega \in \Omega$

Note: $\mu \cup \overline{\mu} = \Omega$ and $\mu \cap \overline{\mu} = \emptyset$ no longer hold for fuzzy sets!

However, $\mu \cap \mu = \mu$ and $\mu \cup \mu = \mu$ still holds.

Modeling Coreference

Core idea: coreference between noun phrases is almost never "100% certain"

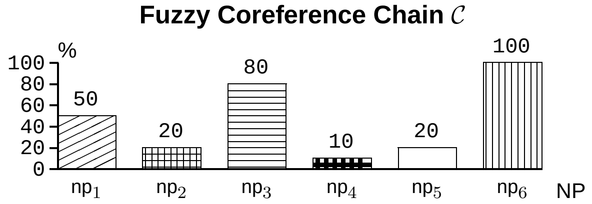
- fuzzy model: represent certainty of coreference *explicitly* with a membership degree
- formally: represent fuzzy chain $\mathcal{C}$ with a fuzzy set $\mu_{\mathcal{C}}$, mapping the domain of all NPs in a text to the $[0,1]$-interval
- then, each noun phrase np_i has a corresponding membership degree $\mu_{\mathcal{C}}(\text{np}_i)$, indicating how certain this NP is a member of chain $\mathcal{C}$

Fuzzy Coreference Chain

Fuzzy chain $\mathcal{C}$: each noun phrase np_i in the text is a member of chain $\mathcal{C}$ with degree $\mu_{\mathcal{C}}(\text{np}_i) \in [0,1]$

Example:

Fuzzy Coreference Chain $\mathcal{C}$



Fuzzy Coreference Chains

Properties of fuzzy chains:

- each chain holds *all* noun phrases in a text
- i.e., each NP is a member of every chain (but with very different certainties)
- we don't have to reject inconsistencies right away — they can be reconciled later through suitable fuzzy operators
- also, there is no arbitrary boundary for discriminating between *"corefering"* and *"not corefering"*
- thus, in this step we don't lose information we might need later

Fuzzy Heuristics

How can we *build* fuzzy chains?

- Use knowledge-poor heuristics to check for coreference between NP pairs
- Examples: Substring, Synonym/Hypernym, Pronoun, CommonHead, Acronym...
- Fuzzy heuristic: return a *degree* of coreference $\in [0, 1]$
- Formally: fuzzy heuristic $\mathcal{H}_i$ is a mapping

$$\mathcal{H}_i(\mathrm{np}_j, \mathrm{np}_k) \mapsto \mu^{\mathcal{H}_i}_{(\mathrm{np}_j, \mathrm{np}_k)}$$

Designing Fuzzy Heuristics

How can we compute a coreference degree $\mu^{\mathcal{H}_i}_{(\mathrm{np}_j, \mathrm{np}_k)}$?

Fuzzy Substring Heuristic: (character n-gram match) return coreference degree of 1.0 if two NP string are identical, 0.0 if they share no substring. Otherwise, select longest matching substring and set coreference degree to its percentage of first NP.

Fuzzy Synonym/Hypernym Heuristic: Synonyms (determined through *WordNet*) receive a coreference degree of 1.0. If two NPs are hypernyms, set the coreference degree depending on distance in the hierarchy (i.e., longer paths result in lower certainty degrees).

Building Fuzzy Chains

So far, we only have results for single NP pairs, i.e., an m-dimensional fuzzy coreference matrix.
Next step: construct *fuzzy chains*

- for a text with n NPs $\langle \mathrm{np}_1, \ldots, \mathrm{np}_n \rangle$, initialize n fuzzy chains $\mathcal{C}_1, \ldots, \mathcal{C}_n$, each represented by a fuzzy set $\mu_{\mathcal{C}_j}$, by computing the fuzzy-*or* for all fuzzy heuristics $\mathcal{H}_1, \ldots, \mathcal{H}_m$:

$$\begin{aligned}\mu_{\mathcal{C}_j} \quad := \quad & \mu^{\mathcal{H}_1}_{(\mathrm{np}_j,\mathrm{np}_1)} \cup \mu^{\mathcal{H}_1}_{(\mathrm{np}_j,\mathrm{np}_2)} \cup \ldots \cup \mu^{\mathcal{H}_1}_{(\mathrm{np}_j,\mathrm{np}_n)} \cup \\ & \mu^{\mathcal{H}_2}_{(\mathrm{np}_j,\mathrm{np}_1)} \cup \mu^{\mathcal{H}_2}_{(\mathrm{np}_j,\mathrm{np}_2)} \cup \ldots \cup \mu^{\mathcal{H}_2}_{(\mathrm{np}_j,\mathrm{np}_n)} \cup \\ & \ldots \cup \\ & \mu^{\mathcal{H}_m}_{(\mathrm{np}_j,\mathrm{np}_1)} \cup \mu^{\mathcal{H}_m}_{(\mathrm{np}_j,\mathrm{np}_2)} \cup \ldots \cup \mu^{\mathcal{H}_m}_{(\mathrm{np}_j,\mathrm{np}_n)}\end{aligned}$$

Merging Fuzzy Chains

Results of the chain building algorithm is a list of fuzzy chains, one for each NP in the text

- but the coreference relation is symmetric and transitive
- hence, if $\mathcal{C}_1$ shows a coreference of np_1 and np_3 and $\mathcal{C}_2$ for np_3 and np_5, we also want the coreference of np_1 and np_5 in the result
- this is achieved by a *chain merging algorithm*
- merging again depends on coreference certainty

Merging Algorithm Outline

Two fuzzy chains $\mathcal{C}_i, \mathcal{C}_j$ are *merged* if their fuzzy-*and* combination reaches a prescribed consistency degree γ:

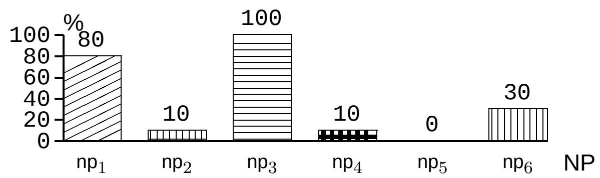
$$\text{if } max(\mu_{\mathcal{C}_i} \cap \mu_{\mathcal{C}_j}) \geq \gamma, \text{ then } \mu_{\mathcal{C}_{(i,j)}} := \mu_{\mathcal{C}_i} \cup \mu_{\mathcal{C}_j}$$

Note:

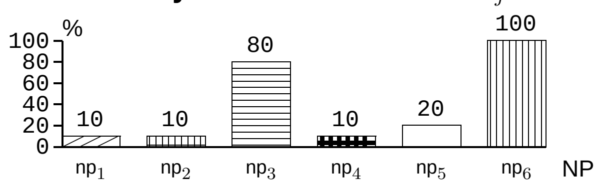
- if $\gamma = 0.0$, all chains are merged, resulting in a single (but useless) coreference chain
- if $\gamma = 1.0$, only chains with 100% certain coreferences *(note: closed-world assumption)* are created
- values in between result in chains that have at least a coreference certainty of γ

Merging Example (I/II)

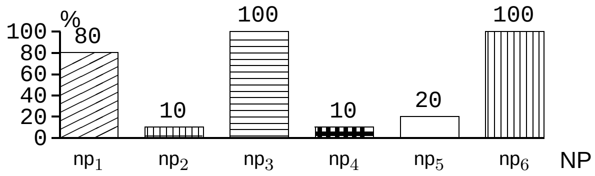
Fuzzy Coreference Chain $\mathcal{C}_i$



Fuzzy Coreference Chain $\mathcal{C}_j$



Merging Example (II/II)



Defuzzification of Fuzzy Chains

Result of the merging phase is a set of (complete) fuzzy chains:

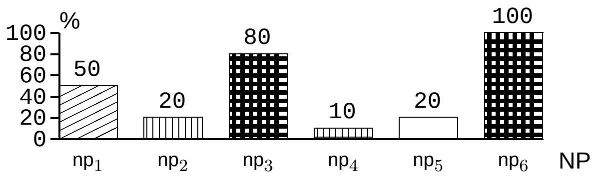
- but not all processing components can deal with fuzzy information
- re-writing components takes time. . .
- thus, we need a gentle transition for components that still expect conventional (crisp) coreference chains

Solution: defuzzify *the fuzzy chains, i.e., convert them to crisp chains*

Defuzzification Example

To (γ-)defuzzify a fuzzy chain, take all elements np_i from the chain with a membership degree $\mu_{\mathcal{C}}(\text{np}_i) \geq \gamma$

Example:



With a certainty degree of $\gamma = 0.8$ we get the crisp result set $c = \{\text{np}_3, \text{np}_6\}$.

Evaluation of the Fuzzy Algorithm

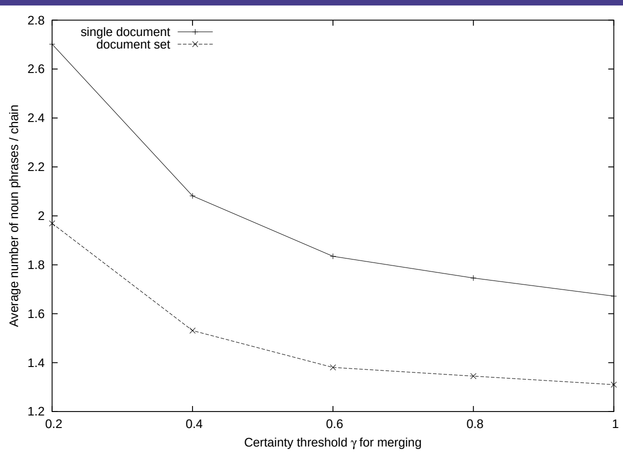
There are two things to evaluate:

1. the *fuzzy algorithm* itself, i.e., the properties of the chain building and chain merging functions; and
2. the *fuzzy heuristics*, i.e., the quality of their rules and resources

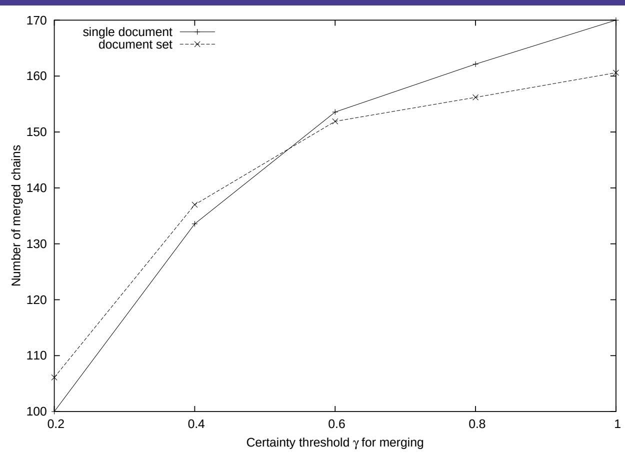
Here, we only show results for 1.
(2. is work in progress)

- Evaluation based on DUC 2002 (development) and 2003 (testing) corpus

Evaluation: Number of NPs/Chain



Evaluation: Number of Chains



ERSS Summarizer

System built with GATE (University of Sheffield); main processing components are:

Preprocessing Tokenizer, Sentence Splitter, POS Tagger, . . .

NPE a noun phrase chunker (Earley-type chart parser with grammar for partial parsing) that performs above 85%

Fuzzy-ERS the fuzzy coreference resolution algorithm

Classifier a naive Bayes classifier for multi-dimensional text categorization

ERSS the summarization system

Summarization

Create 10-word summaries (DUC 2003 Task 1):

1. find most important entities in the text
2. extract textual designator for each entity

Algorithm:

1. build coreference chain; chain length corresponds with importance
2. extract longest NP from each chain until 10-word limit has been reached

Summarization Results

In most cases, summaries give a suprisingly good indication of the text:

Business & Economics: [eBay, 1.8 million auctions, an auction site, frequent service outages]

Space News: [the shuttle Discovery's Hubble repair mission, the observatory's central computer]

- NIST assessors evaluated ERSS as (slightly above) average wrt. coverage and usefulness
- Fuzzy coreference algorithm allows to trade certainty with completeness, overcoming chain fragmentation

Conclusions

- explicit representation of uncertainty allows for more flexible and more robust algorithms
- exchange of information across processing components becomes simpler because of uniform representation model
- choosing appropriate representation formalism allows access to existing research for (here) uncertainty management
- coreference resolution can give a baseline for summarization!

Further Work

Work in progress:

- enhancements to the fuzzy algorithm *(modeling exceptions, incorporating fuzzy belief revision operators)*
- detailed evaluation
- multi-lingual coreference resolution *(English, French, German)*
- multi-document coreference resolution
- enabling more components for the fuzzy model *(parser, summarizer, . . .)*

More Information

- the paper. . .
- http://www.cs.concordia.ca/CLAC/
- http://rene-witte.net/
- http://www.cs.concordia.ca/~faculty/bergler/