Unsupervised Relation Extraction by Massive Clustering

Edgar Gonzàlez Jordi Turmo

TALP Research Center

7 December 2009

Relation Detection

Thousands of people were in the streets and in the basilica to pay tribute. Former president Jimmy Carter represented the United States.

Relation Detection

Thousands of people *were in* **the streets** *and in* **the basilica** *to pay tribute. Former president* **Jimmy Carter** *represented* **the United States***.*

Entities

Relation Detection

Thousands of people *were in* **the streets** *and in* **the basilica** *to pay tribute. Former president* **Jimmy Carter** *represented* **the United States**.

people $\longleftrightarrow$ *streets*

Relation Detection

Thousands of people *were in* **the streets** *and in* **the basilica** *to pay tribute. Former president* **Jimmy Carter** *represented* **the United States**.

$people \longleftrightarrow basilica$

Relation Detection

Thousands of people *were in* **the streets** *and in* **the basilica** *to pay tribute. Former president* **Jimmy Carter** *represented* **the United States**.

Jimmy Carter $\longleftrightarrow$ *United States*

Machine Learning for Information Extraction

- Relation Detection $\subset$ Information Extraction
  - Uses specific linguistic knowledge
  - Adaptation requires costly human effort

Machine Learning for Information Extraction

- Relation Detection $\subset$ Information Extraction
  - Uses specific linguistic knowledge
  - Adaptation requires costly human effort
- Machine Learning $\rightarrow$ Adaptative IE
  - Supervised approaches
  - Weakly supervised approaches

Machine Learning for Information Extraction

- Relation Detection $\subset$ Information Extraction
  - Uses specific linguistic knowledge
  - Adaptation requires costly human effort
- Machine Learning $\rightarrow$ Adaptative IE
  - Supervised approaches
  - Weakly supervised approaches
    - Supervision $\rightarrow$ Bias

Machine Learning for Information Extraction

- Relation Detection $\subset$ Information Extraction
  - Uses specific linguistic knowledge
  - Adaptation requires costly human effort
- Machine Learning $\rightarrow$ Adaptative IE
  - Supervised approaches
  - Weakly supervised approaches
    - Supervision $\rightarrow$ Bias
  - Unsupervised approaches
    - Avoid biases

Machine Learning for Information Extraction

- Relation Detection $\subset$ Information Extraction
  - Uses specific linguistic knowledge
  - Adaptation requires costly human effort
- Machine Learning $\rightarrow$ Adaptative IE
  - Supervised approaches
  - Weakly supervised approaches
    - Supervision $\rightarrow$ Bias
  - Unsupervised approaches
    - Avoid biases
    - Use clustering techniques

Our Proposal

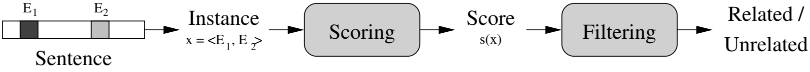
- New unsupervised approach to learning for relation extraction
  - Using probabilistic clustering models
- Evaluation in ACE Relation Mention Detection task
  - Popular evaluation framework

Approach

Overview


E1
E2
Sentence
Instance
x = <E1, E2>
Classifier
Related /
Unrelated

Overview


E1
E2
Sentence
Instance
x = <E1, E2>
Scoring
Score
s(x)
Filtering
Related /
Unrelated

Assumptions

- Scoring:
  - Clustering → point of view
  - Cluster → shared sets of features → relatedness
  - Cluster → reliability → score

Assumptions

- Scoring:
  - Clustering $\rightarrow$ point of view
  - Cluster $\rightarrow$ shared sets of features $\rightarrow$ relatedness
  - Cluster $\rightarrow$ reliability $\rightarrow$ score
  - $\sum$ clusterings with scored clusters $\Rightarrow$ Scorer

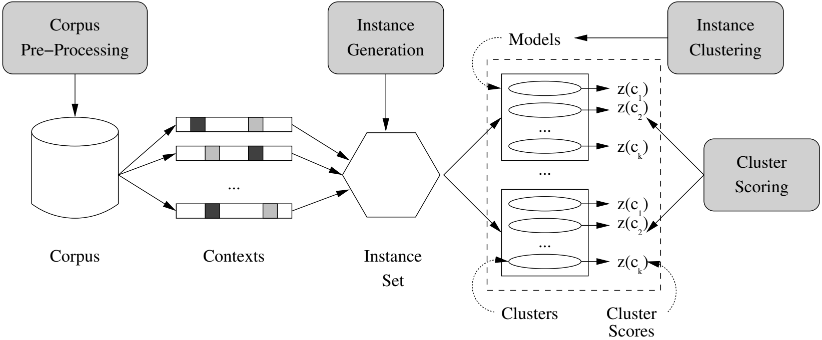
Assumptions

- Scoring:
  - Clustering $\rightarrow$ point of view
  - Cluster $\rightarrow$ shared sets of features $\rightarrow$ relatedness
  - Cluster $\rightarrow$ reliability $\rightarrow$ score
  - $\sum$ clusterings with scored clusters $\Rightarrow$ Scorer
- Filtering:
  - Unsupervised learning $\rightarrow$ $\exists$ non-related instances
  - Highly scored instances $\rightarrow$ related pairs

Assumptions

- Scoring:
  - Clustering $\rightarrow$ point of view
  - Cluster $\rightarrow$ shared sets of features $\rightarrow$ relatedness
  - Cluster $\rightarrow$ reliability $\rightarrow$ score
  - $\sum$ clusterings with scored clusters $\Rightarrow$ Scorer
- Filtering:
  - Unsupervised learning $\rightarrow$ $\exists$ non-related instances
  - Highly scored instances $\rightarrow$ related pairs
  - Threshold value $\Rightarrow$ Filterer

Scorer Learning



Scorer Learning

- Corpus Pre-Processing
  - Tokenization, POS-Tagging, NERC

Scorer Learning

- Corpus Pre-Processing
  - Tokenization, POS-Tagging, NERC
- Instance Generation
  - $\mathcal{X} = \{x_i\}$
  - Pairs of entities co-occurring within a sentence
    - Distance threshold
  - Generation of binary features
    - Context window
    - Pattern-based $\rightarrow$ `dist_%d`, `left_%d_%t`...
    - Frequency threshold

Scorer Learning

- Instance Clustering
  - $p(c_{pq}|x_i;\Theta_p)$
  - Mixture of Bernoulli distributions
  - Expectation-Maximization algorithm
  - Massive repeated randomization $\longrightarrow$ Robustness

Scorer Learning

- Instance Clustering
  - $p(c_{pq}|x_i;\Theta_p)$
  - Mixture of Bernoulli distributions
  - Expectation-Maximization algorithm
  - Massive repeated randomization $\rightarrow$ Robustness
- Cluster Scoring
  - $z(c_{pq})$
  - Cluster Measures
    - Size
    - Homogeneousness $\rightarrow$ Radius

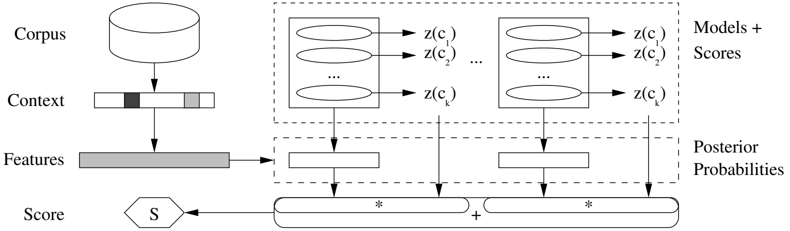
Scorer Learning

- Instance Clustering
  - $p(c_{pq}|x_i;\Theta_p)$
  - Mixture of Bernoulli distributions
  - Expectation-Maximization algorithm
  - Massive repeated randomization $\rightarrow$ Robustness
- Cluster Scoring
  - $z(c_{pq})$
  - Cluster Measures
    - Size (Normalized by number of non-empty clusters in clustering)
    - Homogeneousness $\rightarrow$ Radius

Scorer Learning

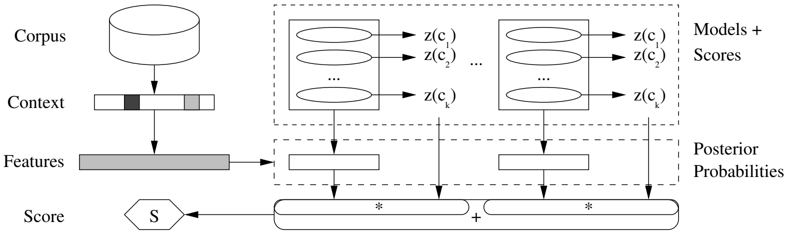
- Instance Clustering
  - $p(c_{pq}|x_i;\Theta_p)$
  - Mixture of Bernoulli distributions
  - Expectation-Maximization algorithm
  - Massive repeated randomization $\rightarrow$ Robustness
- Cluster Scoring
  - $z(c_{pq})$
  - Cluster Measures
    - Size (Normalized by number of non-empty clusters in clustering)
    - Homogeneousness $\rightarrow$ Radius
  - Formulae
    - NSiz, Rad, NDns

Scoring



Corpus
Context
Features
Score
S
z(c 1)
z(c 2)
z(c k)
...
Models +
Scores
Posterior
Probabilities
*
+

Scoring



$$s(x_i) = \sum_{\hat{\Theta}_p} \sum_{q=1}^{k_p} p(c_{pq} \mid x_i) \cdot z(c_{pq})$$

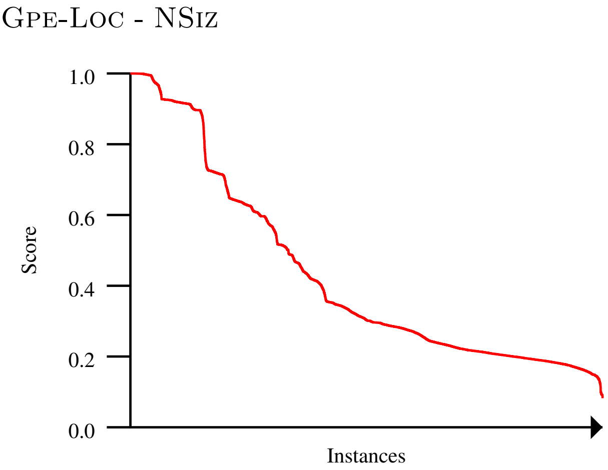
Filterer Learning

- Determination of Threshold Score
  - s_{th} such that $x_i \in R^+ \leftrightarrow s(x_i) \geq s_{th}$

Filterer Learning

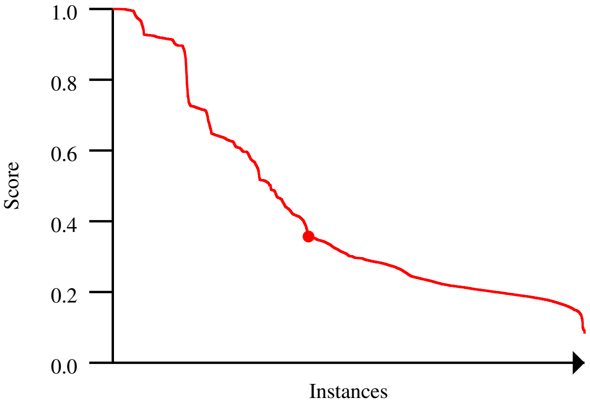
- Determination of Threshold Score
    - s_{th} such that $x_i \in R^+ \leftrightarrow s(x_i) \geq s_{th}$
    - Heuristic-based
        1. Obtain scores of instances in training corpus
        2. Sort instances by score, obtaining a decreasing convex function
        3. Find a cut-off point

Filterer Learning



Filterer Learning

GPE-LOC - NSIZ



Evaluation

Evaluation Framework

- Corpora
  - AQUAINT (APW 2000) $\rightarrow$ 29Mw
  - ACE 2003–2008 $\rightarrow$ 500kw, 98k entities, 18k relations
- Task
  - Relation Mention Detection
  - Recall, Precision, F1
- Approaches
  - Grams-Ub
  - Single
  - Mass

Average Results

		Rec	Prc	F1
GRAMS-UB	-	43.5	65.6	**51.0**
SINGLE	NSIZ	52.8	54.3	**52.3**
SINGLE	RAD	52.1	54.2	**50.3**
SINGLE	NDNS	53.4	54.1	**52.5**
MASS	NSIZ	59.5	53.7	**55.8**
MASS	RAD	62.8	51.7	**56.0**
MASS	NDNS	59.1	54.2	**55.9**

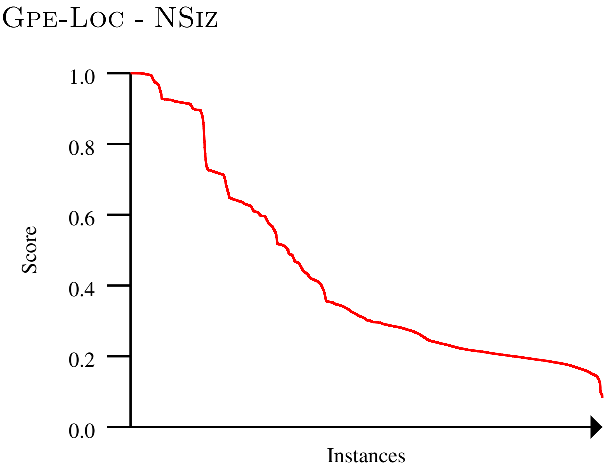
Average Results

		Rec	Prc	F1
GRAMS-UB	-	43.5	65.6	**51.0**
SINGLE	NSIZ	52.8	54.3	**52.3**
SINGLE	RAD	52.1	54.2	**50.3**
SINGLE	NDNS	53.4	54.1	**52.5**
MASS	NSIZ	59.5	53.7	**55.8**
MASS	RAD	62.8	51.7	**56.0**
MASS	NDNS	59.1	54.2	**55.9**

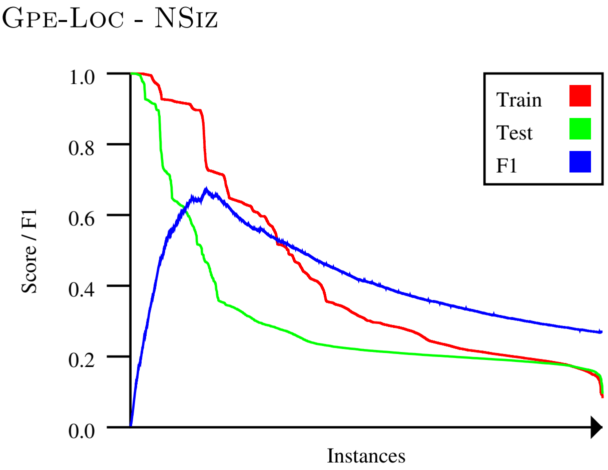
Average Results

		Rec	Prc	F1
GRAMS-UB	-	43.5	65.6	**51.0**
SINGLE	NSIZ	52.8	54.3	**52.3**
SINGLE	RAD	52.1	54.2	**50.3**
SINGLE	NDNS	53.4	54.1	**52.5**
MASS	NSIZ	59.5	53.7	**55.8**
MASS	RAD	62.8	51.7	**56.0**
MASS	NDNS	59.1	54.2	**55.9**

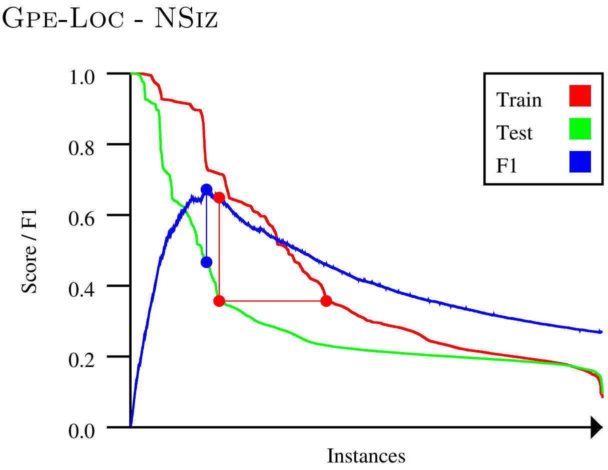
Filtering



Filtering



Filtering



Conclusions

Conclusions

- New unsupervised approach to learning for relation extraction
  - Using probabilistic clustering models
- Evaluation in ACE Relation Mention Detection task
  - Popular evaluation framework

Conclusions

- New unsupervised approach to learning for relation extraction
    - Using probabilistic clustering models
- Evaluation in ACE Relation Mention Detection task
    - Popular evaluation framework
    - 4-point F1 increase above state-of-the-art upper bound
    - Inclusion of richer features $\longrightarrow$ Greater flexibility
    - Benefits of massive combination
    - Robustness to cluster score function

Thank you!