Hypergraphs and Information Fusion for Term Representation Enrichment. Applications to Named Entity Recognition and Word Sense Disambiguation

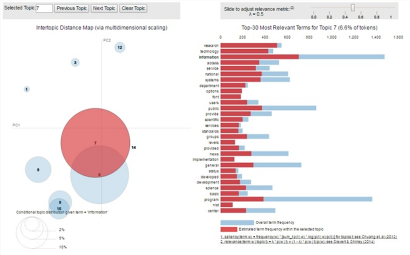
Ph.D. Thesis Defense

Pavel Soriano-Morales

Supervised by Sabine Loudcher and Julien Ah-Pine

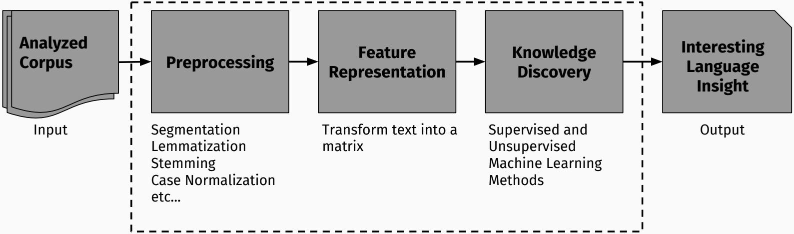
February 7th, 2018

Why is it useful to us to understand text?



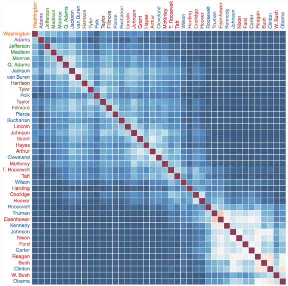
How do we extract meaning from text?

We use **Natural Language Processing** (NLP), a field of computer science interested in making computers extract useful information from text



Feature Representation and Knowledge Discovery

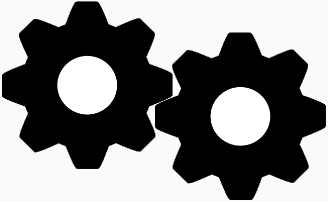
How do we represent text for the machine to understand?



Dealing with data sparsity

Leveraging heterogeneity

What techniques do we use to discover meaning from text?



Finding semantic communities

Representing Text

- **Common ways to represent text**
    - Lexical
    - Syntactic
        - Constituency Tree
        - Dependency Tree

Representing Text

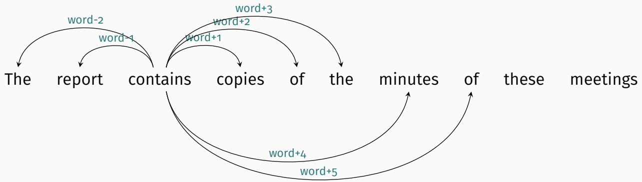
- **Common ways to represent text**
  - Lexical
  - Syntactic
    - Constituency Tree
    - Dependency Tree
- **Example Phrase**

  The report contains copies of the minutes of these meetings

Representing Text

- **Common ways to represent text**
  - Lexical
  - Syntactic
    - Constituency Tree
    - Dependency Tree
- **Example Phrase**

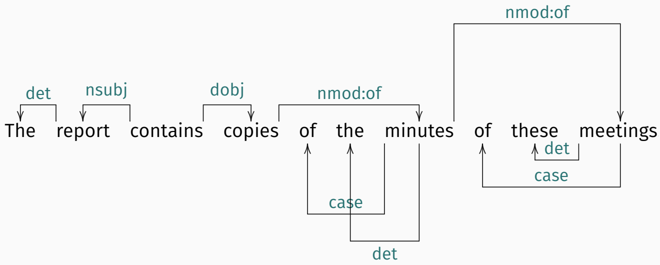
  The report contains copies of the minutes of these meetings



Representing Text

- **Common ways to represent text**
  - Lexical
  - Syntactic
    - Constituency Tree
    - Dependency Tree
- **Example Phrase**

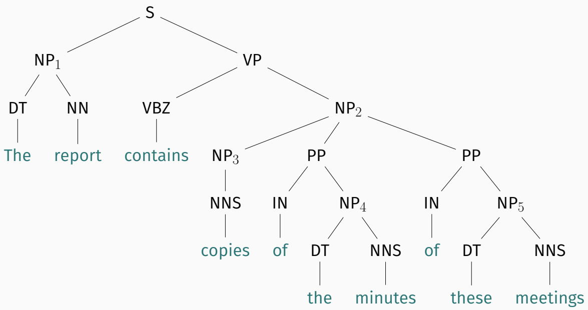
  The report contains copies of the minutes of these meetings



Representing Text

- **Common ways to represent text**
  - Lexical
  - Syntactic
    - Constituency Tree
    - Dependency Tree
- **Example Phrase**

  The report contains copies of the minutes of these meetings

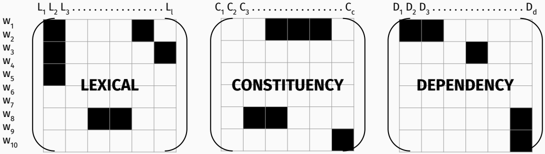


Represention Models

- **Two classic models**
  - Graph-based
  - Matrix-based
- **Leveraging the network structure**
  - We can find communities of similiar words according to their meaning

Represention Models

- **Two classic models**
  - Graph-based
  - Matrix-based
- **Leveraging the network structure**
  - We can find communities of similiar words according to their meaning



Main Challenges and Contributions

1. What type of model can we employ to represent a corpus using heterogeneous features?
    - *Hypergraph model to hold different types of linguistic information*

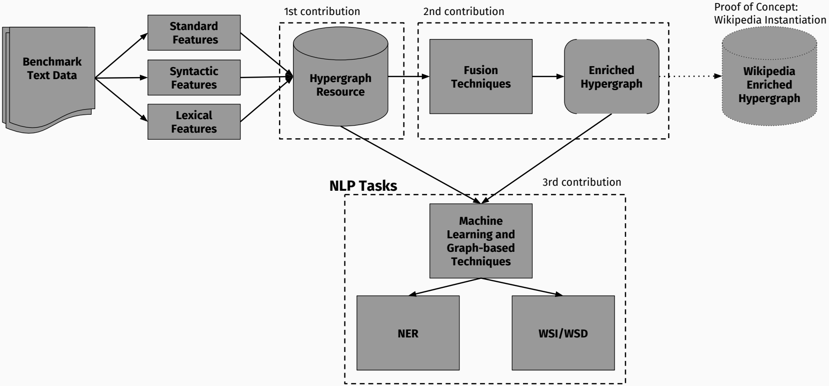
Main Challenges and Contributions

1. What type of model can we employ to represent a corpus using heterogeneous features?
    - *Hypergraph model to hold different types of linguistic information*
2. How can we combine these features while dealing with feature sparsity?
    - *Multimedia fusion techniques to combine and densify representation spaces*

Main Challenges and Contributions

1. What type of model can we employ to represent a corpus using heterogeneous features?
   - *Hypergraph model to hold different types of linguistic information*
2. How can we combine these features while dealing with feature sparsity?
   - *Multimedia fusion techniques to combine and densify representation spaces*
3. How can we find communities existing within the language networks?
   - *An alternative network-based algorithm to discover semantically related words within a text*

Work Overview



Contributions in Detail

Hypergraph Linguistic Model

Introduction

We extract useful information from a text based on the distributional hypothesis (a word is defined by its surroundings)

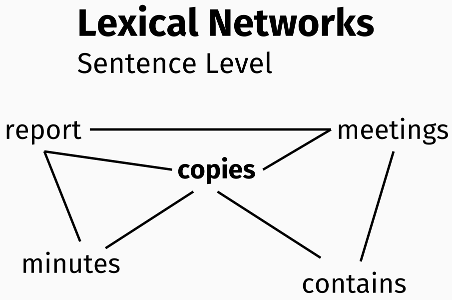
- **We choose network models**
    - Used in a large quantity of NLP tasks
    - Graphs structures can give us a clearer view into the relations of words within a text
    - Ultimately graphs are transformed to a vectorial representation through the adjacency/incidence matrices

Example phrase

The report contains copies of the minutes of these meetings

Example phrase

The report contains copies of the minutes of these meetings

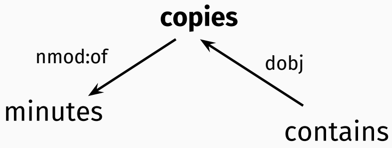


Example phrase

The report contains copies of the minutes of these meetings

Syntactic Networks

Dependency Tree

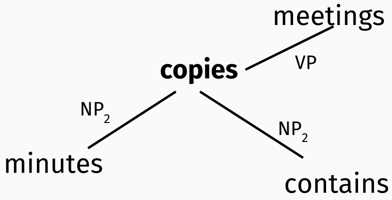


Example phrase

The report contains copies of the minutes of these meetings

Syntactic Networks

Constituency Tree



Limitations and Proposition

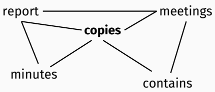
- **Limitations of existing representations**
  - Language networks generally employ a single type of textual information
  - The edges of the network may relate maximum two words at each time

Limitations and Proposition

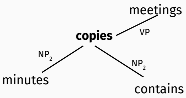
- **Limitations of existing representations**
  - Language networks generally employ a single type of textual information
  - The edges of the network may relate maximum two words at each time
- **Proposition**
  - Use a hypergraph model
  - Link together the different types of networks
  - Get a semantic overview at three different levels: short range, medium range, and long range

Proposed Model

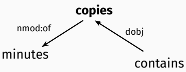
Lexical Networks
Sentence Level



Syntactic Networks
Constituency Tree

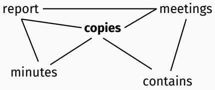


Syntactic Networks
Dependency Tree

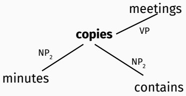


Proposed Model

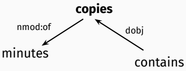
Lexical Networks
Sentence Level



Syntactic Networks
Constituency Tree



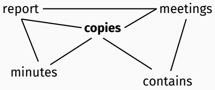
Syntactic Networks
Dependency Tree



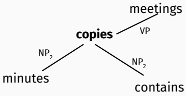
Hypergraph Model

Proposed Model

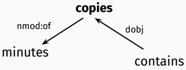
Lexical Networks
Sentence Level



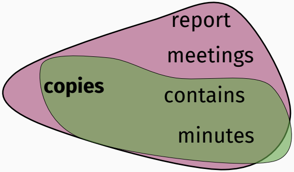
Syntactic Networks
Constituency Tree



Syntactic Networks
Dependency Tree



Hypergraph Model

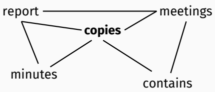


Lexical

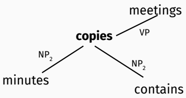
Constituency (NP_2)

Proposed Model

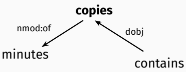
Lexical Networks
Sentence Level



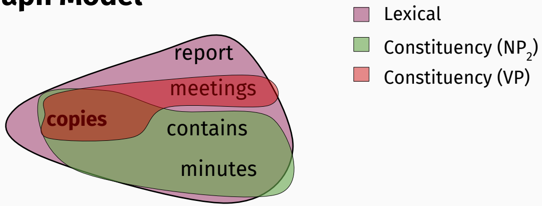
Syntactic Networks
Constituency Tree



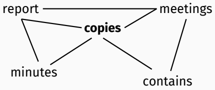
Syntactic Networks
Dependency Tree



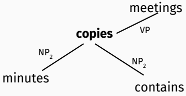
Hypergraph Model



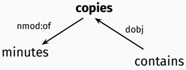
Lexical Networks
Sentence Level



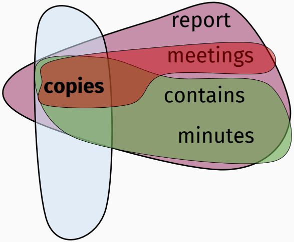
Syntactic Networks
Constituency Tree



Syntactic Networks
Dependency Tree



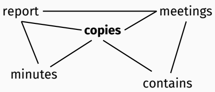
Hypergraph Model



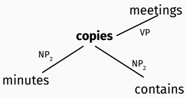
- Lexical
- Constituency (NP_2)
- Constituency (VP)
- Dependency (dobj:contains)

Proposed Model

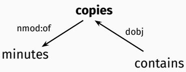
Lexical Networks
Sentence Level



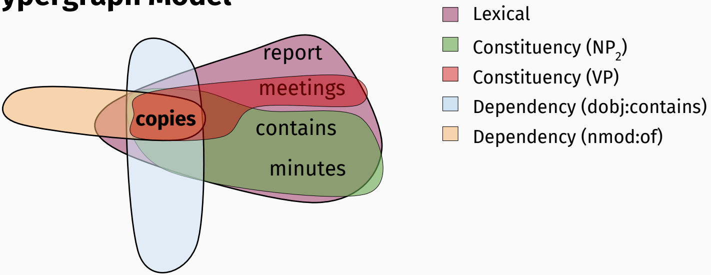
Syntactic Networks
Constituency Tree



Syntactic Networks
Dependency Tree



Hypergraph Model



Contributions in Detail

Combining Features and Dealing with Sparsity

Multimedia Fusion Techniques

- **Definition**
  - Set of techniques used in multimedia analysis tasks to integrate multiple media
  - The goal is to obtain rich insights about the data being treated
  - We adapt these techniques to our use case: textual information

Multimedia Fusion Techniques

- **Definition**
  - Set of techniques used in multimedia analysis tasks to integrate multiple media
  - The goal is to obtain rich insights about the data being treated
  - We adapt these techniques to our use case: textual information
- **Main fusion operators:**
  - Early Fusion $E_\alpha(\cdot)$,
  - Late Fusion $L_\beta(\cdot)$,
  - Cross Fusion $X_\gamma(\cdot)$

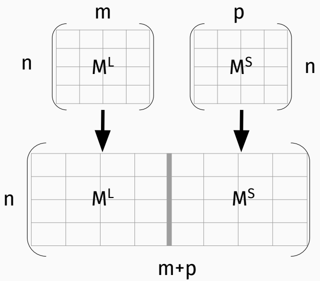
Early and Late Fusion

DEFINITIONS

M^L	Lexical features	M^S	Syntactic features
S^L	Lexical similarities	S^S	Syntactic similarities

EARLY FUSION

Matrices M^L and M^S have the same number of rows



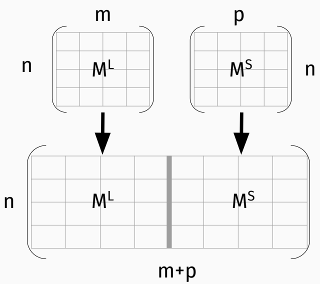
Early and Late Fusion

DEFINITIONS

M^L	Lexical features	M^S	Syntactic features
S^L	Lexical similarities	S^S	Syntactic similarities

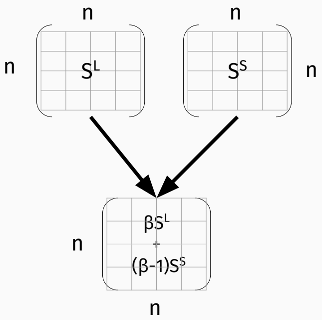
EARLY FUSION

Matrices M^L and M^S have the same number of rows

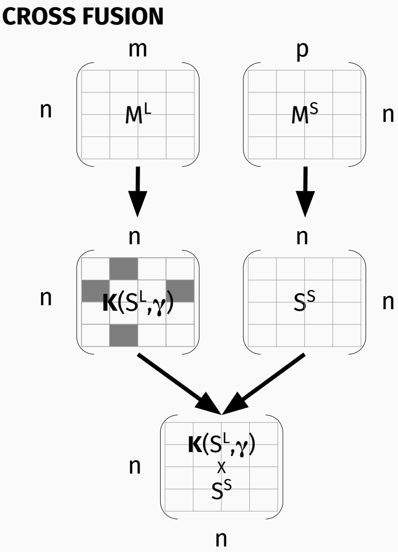


LATE FUSION: SIMILARITY FUSION

Matrices S^L and S^S have the same size



Cross Fusion



Hybrid Fusion

- **Combining fusion operators**
  - Applying one function to the result of another to produce a new fusion function

Hybrid Fusion

- **Combining fusion operators**
  - Applying one function to the result of another to produce a new fusion function

- **First Degree**
  - $E(M^L, M^S)$, $L(S^S, M^L)$
  - **Cross Feature Fusion**: $X_F(S^S, M^L)$
  - **Cross Similarity Fusion**: $X_S(S^S, S^L)$

$X_F(S^L, M^S)$ Cross Feature Fusion

$$\begin{array}{cc} & w_1\,w_2\,w_3 \\ \begin{matrix}w_1\\w_2\\w_3\end{matrix} & \left(\begin{matrix} S^L \end{matrix}\right)\end{array} \text{X} \begin{array}{cc} & f_{S1}\,f_{S2}\,f_{S3} \\ \begin{matrix}w_1\\w_2\\w_3\end{matrix} & \left(\begin{matrix} M^S \end{matrix}\right)\end{array} = \begin{array}{cc} & f_{S1}\,f_{S2}\,f_{S3} \\ \begin{matrix}w_1\\w_2\\w_3\end{matrix} & \left(\begin{matrix} X_F(S^L, M^S) \end{matrix}\right)\end{array}$$

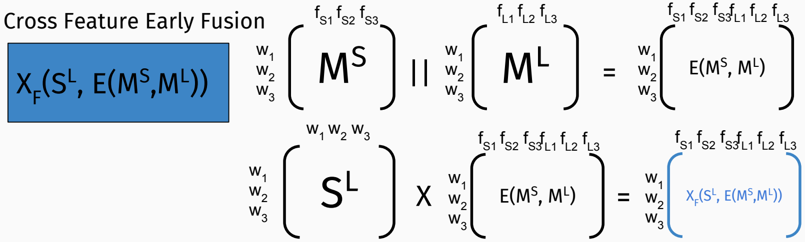
$X_S(S^L, S^S)$ Cross Similarity Fusion

$$\begin{array}{cc} & w_1\,w_2\,w_3 \\ \begin{matrix}w_1\\w_2\\w_3\end{matrix} & \left(\begin{matrix} S^L \end{matrix}\right)\end{array} \text{X} \begin{array}{cc} & w_1\,w_2\,w_3 \\ \begin{matrix}w_1\\w_2\\w_3\end{matrix} & \left(\begin{matrix} S^S \end{matrix}\right)\end{array} = \begin{array}{cc} & w_1\,w_2\,w_3 \\ \begin{matrix}w_1\\w_2\\w_3\end{matrix} & \left(\begin{matrix} X_S(S^L, S^S) \end{matrix}\right)\end{array}$$

Hybrid Fusion

- **Combining fusion operators**
  - Applying one function to the result of another to produce a new fusion function

- **Second Degree**
  - **Cross Feature Early Fusion**: $X_F(S^T, E(M^L, M^S))$
  - **Late Cross Feature Fusion**: $L(M^T, X_F(S^T, M^T))$



Hybrid Fusion

- **Combining fusion operators**
    - Applying one function to the result of another to produce a new fusion function
- **Higher Degree**
    - Triple Early Double Late Cross Feature Fusion: $E(M_L, E(E(M_T, L(M^T, X_F(S^T, M^T))), L(M^L, X_F(S^S, M^L))))$

Higher Degree Operator

$$E(M_L, E(E(M^T, L(M^T, X_F(S^T, M^T))), L(M^L, X_F(S^S, M^L))))$$

Higher Degree Operator

$$E(M_L, E(E(M^T, L(M^T, X_F(S^T, M^T))), L(M^L, X_F(S^S, M^L))))$$

Higher Degree Operator

$$E(M_L, E(E(M^T, L(M^T, X_F(S^T, M^T))), L(M^L, X_F(S^S, M^L))))$$

$$L(M^L, X_F(S^S, M^L))$$

$$\begin{array}{cc} & w_1\ w_2\ w_3 \\ \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} & \begin{pmatrix} S^S \end{pmatrix} \end{array} \times \begin{array}{cc} & f_{L1}\ f_{L2}\ f_{L3} \\ \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} & \begin{pmatrix} M^L \end{pmatrix} \end{array} = \begin{array}{cc} & f_{L1}\ f_{L2}\ f_{L3} \\ \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} & \begin{pmatrix} X_F(S^S, M^L) \end{pmatrix} \end{array}$$

$$\begin{array}{cc} & f_{L1}\ f_{L2}\ f_{L3} \\ \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} & \begin{pmatrix} M^L \end{pmatrix} \end{array} + \begin{array}{cc} & f_{L1}\ f_{L2}\ f_{L3} \\ \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} & \begin{pmatrix} X_F(S^S, M^L) \end{pmatrix} \end{array} = \begin{array}{cc} & f_{L1}\ f_{L2}\ f_{L3} \\ \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} & \begin{pmatrix} L(M^L, X_F(S^S, M^L)) \end{pmatrix} \end{array}$$

Higher Degree Operator

$$E(M_L, E(E(M^T, L(M^T, X_F(S^T, M^T))), L(M^L, X_F(S^S, M^L))))$$

$$L(M^T, X_F(S^T, M^T)))$$

$$\begin{array}{c c} & w_1\ w_2\ w_3 \\ \begin{array}{c} w_1 \\ w_2 \\ w_3 \end{array} & \begin{pmatrix} S^T \end{pmatrix} \end{array} \ X \ \begin{array}{c c} & f_{T1}\ f_{T2}\ f_{T3} \\ \begin{array}{c} w_1 \\ w_2 \\ w_3 \end{array} & \begin{pmatrix} M^T \end{pmatrix} \end{array} = \begin{array}{c c} & f_{T1}\ f_{T2}\ f_{T3} \\ \begin{array}{c} w_1 \\ w_2 \\ w_3 \end{array} & \begin{pmatrix} X_F(S^T, M^T) \end{pmatrix} \end{array}$$

$$\begin{array}{c c} & f_{T1}\ f_{T2}\ f_{T3} \\ \begin{array}{c} w_1 \\ w_2 \\ w_3 \end{array} & \begin{pmatrix} M^T \end{pmatrix} \end{array} + \begin{array}{c c} & f_{T1}\ f_{T2}\ f_{T3} \\ \begin{array}{c} w_1 \\ w_2 \\ w_3 \end{array} & \begin{pmatrix} X_F(S^T, M^T) \end{pmatrix} \end{array} = \begin{array}{c c} & f_{T1}\ f_{T2}\ f_{T3} \\ \begin{array}{c} w_1 \\ w_2 \\ w_3 \end{array} & \begin{pmatrix} L(M^T, X_F(S^T, M^T)) \end{pmatrix} \end{array}$$

Higher Degree Operator

$$E(M_L, E(E(M^T, L(M^T, X_F(S^T, M^T))), L(M^L, X_F(S^S, M^L))))$$

$$E(M^T, L(M^T, X_F(S^T, M^T)))$$

$$\begin{array}{cc} & f_{T1}\ f_{T2}\ f_{T3} \\ \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} & \left[\; M^T \; \right] \end{array} \; || \; \begin{array}{cc} & f_{T1}\ f_{T2}\ f_{T3} \\ \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} & \left[\; L(M^T, X_F(S^T, M^T)) \; \right] \end{array} = \begin{array}{cc} & f_{T1}\ f_{T2}\ f_{T3} \quad f_{T1}\ f_{T2}\ f_{T3} \\ \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} & \left[\; E(M^T, L(M^T, X_F(S^T, M^T))) \; \right] \end{array}$$

Higher Degree Operator

$$E(M_L, E(\boxed{E(M^T, \boxed{L(M^T, X_F(S^T, M^T))})}, \boxed{L(M^L, X_F(S^S, M^L))}))$$

$$E(E(M^T, L(M^T, X_F(S^T, M^T))), L(M^L, X_F(S^S, M^L)))$$

$$\begin{array}{cc} & f_{T1}\ f_{T2}\ f_{T3} \\ \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} & \left[E(M^T, L(M^T, X_F(S^T, M^T))) \right] \end{array} \; || \; \begin{array}{cc} & f_{L1}\ f_{L2}\ f_{L3} \\ \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} & \left[L(M^L, X_F(S^S, M^L)) \right] \end{array} =$$

$$\begin{array}{cc} & f_{L1}\ f_{L2}\ f_{L3}\quad f_{L1}\ f_{L2}\ f_{L3} \\ \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} & \left[E(E(M^T, L(M^T, X_F(S^T, M^T))), L(M^L, X_F(S^S, M^L))) \right] \end{array}$$

Higher Degree Operator

$$E(M_L, E(E(M^T, L(M^T, X_F(S^T, M^T))), L(M^L, X_F(S^S, M^L))))$$

$$E(M_L, E(E(M^T, L(M^T, X_F(S^T, M^T))), L(M^L, X_F(S^S, M^L))))$$

$$\begin{array}{cc} & f_{L1}\ f_{L2}\ f_{L3} \\ \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} & \begin{pmatrix} M^T \end{pmatrix} \end{array} \; || \; \begin{array}{cc} & f_{L1}\ f_{L2}\ f_{L3} \\ \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} & \begin{pmatrix} E(E(M^T, L(M^T, X_F(S^T, M^T))), L(M^L, X_F(S^S, M^L)))) \end{pmatrix} \end{array} =$$

$$\begin{array}{cc} & f_{L1}\ f_{L2}\ f_{L3}\ f_{L1}\ f_{L2}\ f_{L3} \\ \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} & \begin{pmatrix} E(M_L, E(E(M^T, L(M^T, X_F(S^T, M^T))), L(M^L, X_F(S^S, \\ M^L)))) \end{pmatrix} \end{array}$$

Contributions in Detail

Finding Communities in the Network

Introduction

- **Language networks tend to be scale-free**
    - There are certain nodes (hubs) that are very well connected forming communities within the network

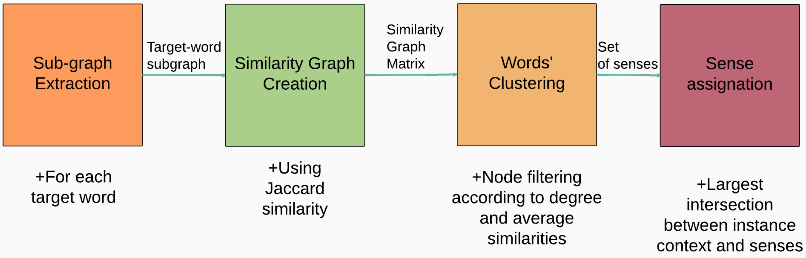
Introduction

- **Language networks tend to be scale-free**
  - There are certain nodes (hubs) that are very well connected forming communities within the network
- **Limitations of existing approaches**
  - Single typed networks
  - Large number of parameters

Introduction

- **Language networks tend to be scale-free**
    - There are certain nodes (hubs) that are very well connected forming communities within the network
- **Limitations of existing approaches**
    - Single typed networks
    - Large number of parameters
- **Proposition**
    - Be able to exploit different types of linguistic information (lexical or syntactic co-occurrence)
    - Keep the number of parameters low and allow for their automatic adjusting according to the network's nature

Proposed Method



Applications to NLP

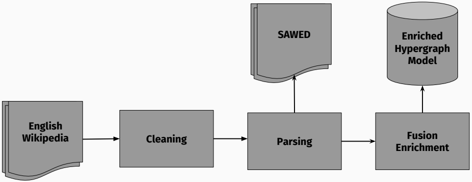
Hypergraph Model Instantiation

Hypergraph Model Instantiation

- **Apply our proposed linguistic model to a real world corpus**
  - Use the English Wikipedia as input and generate a textual structure following the proposed network model

Hypergraph Model Instantiation

- **Apply our proposed linguistic model to a real world corpus**
  - Use the English Wikipedia as input and generate a textual structure following the proposed network model
- **Steps performed**



SAEWD: Parsed sample

FILENAME wiki_00.parsed

token	**lemma**	**POS**	**constituency**	**head**	**dependency**
%%#PAGE Anarchism					
⋮	⋮	⋮	⋮	⋮	⋮
%%#SEN 25 9					
A	a	DT	NP_22,S_97	3	det
great	great	JJ	NP_22,S_97	3	amod
brigand	brigand	NN	NP_22,S_97	4	nsubj
becomes	become	VBZ	VP_44,S_97	0	root
a	a	DT	NP_18,NP_20,VP_44,S_97	6	det
ruler	ruler	NN	NP_18,NP_20,VP_44,S_97	4	xcomp
of	of	IN	PP_57,NP_20,VP_44,S_97	9	case
a	a	DT	NP_18,PP_57,NP_20,VP_44,S_97	9	det
Nation	nation	NN	NP_18,PP_57,NP_20,VP_44,S_97	6	nmod

Hypergraph Incidence Matrix

		CONSTITUENT			DEPENDENCY		SENTENCE
		NP_1 DT:NN	NP_2 NP:PP:PP	NP_3 NNS	nsubj contains	dobj contains	S_1
NN	report	1			1		1
	copies		1	1		1	1
	minutes		1				1
	meetings		1				1
VB	contains						1

Wikipedia Feature Enriched Space

- **Characteristics of the enriched space**
  - Sparsity is reduced
  - Semantic relatedness differs according to the representation space

	Lexical Features (5.49%) M^L	**Syntactic Features (4.97%)** M^S	**Early Fusion (5.23%)** $E(M^L, M^S)$	X_F **Fusion (16.75%)** $X_F(S^S, M^L)$	X_F **Fusion (13.45%)** $X_F(S^L, M^S)$
	priests	monk	sailor	vassal	sailor
	nun	regent	regent	regent	fluent
priest	canton	aedile	nuclei	nun	dean
	sailor	seer	nun	sailor	nuclei
	burial	meek	relic	monk	chorus

Applications to NLP

Solving Named Entity Recognition

- **NER Objective**
    - The goal is to automatically discover mentions that belong to a well-defined semantic category.

- **NER Objective**
  - The goal is to automatically discover mentions that belong to a well-defined semantic category.
- **Classic entities types**
  - Location (LOC)
  - Organization (ORG)
  - Person (PER)
  - Miscellaneous (MISC)
  - None (O)

- **NER Objective**
  - The goal is to automatically discover mentions that belong to a well-defined semantic category.
- **Classic entities types**
  - Location (LOC)
  - Organization (ORG)
  - Person (PER)
  - Miscellaneous (MISC)
  - None (O)
- **Our goal**
  - We assess the effectiveness of the classic fusion methods and propose new hybrid combinations

Example Phrase

Australian scientist discovers star with telescope

Three different types of features

Word	Features	Feature Type
Australian	word:Australian, word+1:scientist, ...	**Lexical (L)**
scientist	Australian/JJ/amod, discovers/VBZ/nsubj_inv	**Syntactic (S)**
discover	discover, no-capital-letter, prf:dis, suf:ver, VBZ	**Standard (T)**

Experimental Protocol

- **Preprocessing**
  - Normalize numbers

Experimental Protocol

- **Preprocessing**
  - Normalize numbers
- **Test Corpora**
  - CoNLL-2003 (CONLL): Train: 219,554 lines. Test: 50,350 lines
  - Wikiner (WNER): 3.5 million words.
  - Wikigold (WGLD): 41,011 words.

Experimental Protocol

- **Preprocessing**
    - Normalize numbers
- **Test Corpora**
    - CoNLL-2003 (CONLL): Train: 219,554 lines. Test: 50,350 lines
    - Wikiner (WNER): 3.5 million words.
    - Wikigold (WGLD): 41,011 words.
- **Learning Algorithm**
    - Structured Perceptron

Experimental Protocol

- **Preprocessing**
  - Normalize numbers
- **Test Corpora**
  - CoNLL-2003 (CONLL): Train: 219,554 lines. Test: 50,350 lines
  - Wikiner (WNER): 3.5 million words.
  - Wikigold (WGLD): 41,011 words.
- **Learning Algorithm**
  - Structured Perceptron
- **Evaluation Metric**
  - F-measure
  - Evaluated with a 5-fold CV (WNER and WGLD)

Evaluation

A	B	Early Fusion (EF)		
		CONLL	WNER	WGLD
M^L	M^S	72.01	70.59	59.38
M^L	M^T	78.13	79.78	61.96
M^S	M^T	77.70	78.10	60.93
M^L	$E(M^S, M^T)$	**78.90**	**80.04**	**63.20**
		Late Fusion (LF)		
		CONLL	WNER	WGLD
S^L	S^S	**61.65**	58.79	44.29
S^L	S^T	55.64	**67.70**	48.00
S^S	S^T	50.21	58.41	**49.81**

Evaluation

A	B	CONLL	WNER	WGLD
Early Fusion (EF)				
M^L	M^S	72.01	70.59	59.38
M^L	M^T	78.13	79.78	61.96
M^S	M^T	77.70	78.10	60.93
M^L	$E(M^S, M^T)$	**78.90**	**80.04**	**63.20**
Late Fusion (LF)				
S^L	S^S	**61.65**	58.79	44.29
S^L	S^T	55.64	**67.70**	48.00
S^S	S^T	50.21	58.41	**49.81**

		CONLL	WNER	WGLD
Cross Feature Fusion ($X_F F$)				
S^L	M^T	49.90	**70.27**	**62.69**
S^S	M^T	47.27	51.38	48.53
S^T	$b^*_{X_F F}$	**52.89**	62.21	50.15
Cross Similarity Fusion ($X_S F$)				
S^L	S^T	27.75	**59.12**	38.35
S^S	$b^*_{X_S F}$	36.87	40.92	39.62
S^T	$b^*_{X_S F}$	**41.89**	52.03	**39.92**

$b^*_{X_F F} \in \{M^L, M^T\}$

$b^*_{X_S F} \in \{S^L, S^S\}$

Evaluation

$$\boxed{E(M_L, \boxed{E(\boxed{E(M^T, \boxed{L(M^T, X_F(S^T, M^T))})}, \boxed{L(M^L, X_F(S^S, M^L))})})}$$

Evaluation

$$E(M_L, E(E(M^T, L(M^T, X_F(S^T, M^T))), L(M^L, X_F(S^S, M^L))))$$

		Triple Early Double Late Cross Feature Fusion ($EEELX_FLX_F$)		
		CONLL	WNER	WGLD
M^L	$\hat{b}_{EEELX_FLX_F}$	65.01	78.02	62.34
$M^L_{\alpha=0.95}$	$\hat{b}_{EEELX_FLX_F}$	**79.67**	**81.79**	**67.05**
EF Baseline		78.90	80.04	63.20

Analyzing the Best Fusion Operator

- **Understand how the evolution towards and enriched space helps the model take the correct decision**

Analyzing the Best Fusion Operator

- **Understand how the evolution towards and enriched space helps the model take the correct decision**
  - Decompose the large fusion operator into 4 separate representations

Analyzing the Best Fusion Operator

- **Understand how the evolution towards and enriched space helps the model take the correct decision**
    - Decompose the large fusion operator into 4 separate representations
    - Train a model with each individual operator (4 models: M_1, M_2, M_3, M_4)

Analyzing the Best Fusion Operator

- **Understand how the evolution towards and enriched space helps the model take the correct decision**
    - Decompose the large fusion operator into 4 separate representations
    - Train a model with each individual operator (4 models: M_1, M_2, M_3, M_4)
    - Investigate how the features added at each step help the model predict the correct class

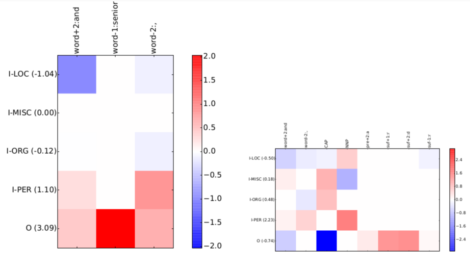
Analyzing the Best Fusion Operator

- **Understand how the evolution towards and enriched space helps the model take the correct decision**
  - Decompose the large fusion operator into 4 separate representations
  - Train a model with each individual operator (4 models: M_1, M_2, M_3, M_4)
  - Investigate how the features added at each step help the model predict the correct class

$$\overbrace{\underbrace{\overbrace{E_{\alpha=0.95}(\underbrace{M^L}_{1}, M^T}^{2}, L(M^T, X_F(S^S, M^T))}_{3}, L(M^L, X_F(S^S, M^L)))}^{4}$$

Analyzing the Best Fusion Operator

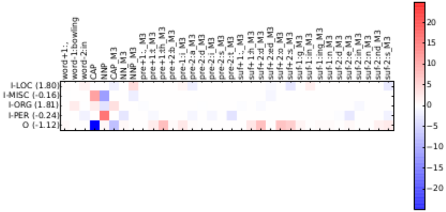
We focus on the word *Kory*, and its performance from model M_1 to M_2



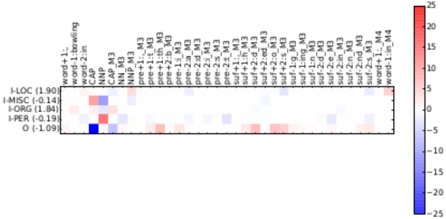
(a) The word *Kory* is predicted to be class O by model M_1 given the coefficients of the features indicated.

(b) The same word is correctly predicted to be class PER by model M_2 given the new features in the model.

We focus on the word *Green*, and its performance from model M_3 to M_4



(a) The word *Green* is predicted to be class ORG by model M_3 given the coefficients of the features indicated.



Applications to NLP

Solving Word Sense Induction and Disambiguation

- **WSI/WSD Objective**
  - The goal is to determine a set of possible senses to a given word according to its possible contexts (WSI). Then, assigning a correct sense to a particular instance of said word

- **WSI/WSD Objective**
  - The goal is to determine a set of possible senses to a given word according to its possible contexts (WSI). Then, assigning a correct sense to a particular instance of said word
- **Our goals**
  - Assess the effectiveness of the fusion enriched spaces
  - Evaluate the pertinence of our community discovering algorithm

- **Preprocessing**
  - Remove very frequent and very infrequent words

Experimental Protocol

- **Preprocessing**
  - Remove very frequent and very infrequent words
- **Test Corpora**
  - Semeval 2007 [SM03]: Train: 219,554 lines. Test: 50,350

Experimental Protocol

- **Preprocessing**
  - Remove very frequent and very infrequent words
- **Test Corpora**
  - Semeval 2007 [SM03]: Train: 219,554 lines. Test: 50,350
- **Clustering Algorithm**
  - Spectral Clustering [SM00]
  - Proposed Community Algorithm

Experimental Protocol

- **Preprocessing**
  - Remove very frequent and very infrequent words
- **Test Corpora**
  - Semeval 2007 [SM03]: Train: 219,554 lines. Test: 50,350
- **Clustering Algorithm**
  - Spectral Clustering [SM00]
  - Proposed Community Algorithm
- **Evaluation Metrics**
  - Supervised Recall
  - Unsupervised F-measure

Experimental Protocol

- **Preprocessing**
  - Remove very frequent and very infrequent words
- **Test Corpora**
  - Semeval 2007 [SM03]: Train: 219,554 lines. Test: 50,350
- **Clustering Algorithm**
  - Spectral Clustering [SM00]
  - Proposed Community Algorithm
- **Evaluation Metrics**
  - Supervised Recall
  - Unsupervised F-measure
  - Proposed: H-measure

$$\text{H-measure} = \frac{1}{2}\left(2 * \frac{SR * UF}{SR + UF} + \frac{\delta}{\delta + |\#cl - \delta|}\right)$$

δ is the average true number of senses of the words in a test corpus

Spectral Clustering Evaluation

Cross Feature Cross Similarity Fusion ($X_F X_S F$)					
$X_F(X_S(S^L, S^S), M^L)$	78.40	80.40	76.10	3.11	
$X_F(X_S(S^L, S^S), M^S)$	78.90	81.80	75.60	3.16	
Early Cross Feature Fusion ($EX_F F$)					
$E(M^L, X_F(S^L, M^L))$	79.20	82.40	75.70	3.57	**2F**
$E(M^S, X_F(S^L, M^L))$	78.30	80.50	75.80	1.95	
Late Cross Feature Fusion ($LX_F F$)					
$L(M^S, X_F(S^L, M^S))$	78.60	81.10	75.80	4.22	
$L(M^L, X_F(S^L, M^L))$	**79.50**	**82.80**	75.70	3.96	
Early Late Cross Feature Fusion ($ELX_F F$)					
$E(M^L, L(M^S, X_F(S^L, M^S)))$	78.50	81.40	75.40	4.26	**HF**
$E(M^L, L(M^L, X_F(S^L, M^L)))$	**79.50**	82.70	75.90	3.99	
Baseline MFS	78.70	80.90	76.20	1.00	

Figure 1: Supervised Recall

Spectral Clustering Evaluation

Cross Feature Cross Similarity Fusion ($X_F X_S F$)					
$X_F(X_S(S^L, S^S), M^L)$	78.40	80.40	76.10	3.11	
$X_F(X_S(S^L, S^S), M^S)$	78.90	81.80	75.60	3.16	
Early Cross Feature Fusion ($EX_F F$)					
$E(M^L, X_F(S^L, M^L))$	79.20	82.40	75.70	3.57	**2F**
$E(M^S, X_F(S^L, M^L))$	78.30	80.50	75.80	1.95	
Late Cross Feature Fusion ($LX_F F$)					
$L(M^S, X_F(S^L, M^S))$	78.60	81.10	75.80	4.22	
$L(M^L, X_F(S^L, M^L))$	**79.50**	**82.80**	75.70	3.96	
Early Late Cross Feature Fusion ($ELX_F F$)					
$E(M^L, L(M^S, X_F(S^L, M^S)))$	78.50	81.40	75.40	4.26	**HF**
$E(M^L, L(M^L, X_F(S^L, M^L)))$	**79.50**	82.70	75.90	3.99	
Baseline MFS	78.70	80.90	76.20	1.00	

Figure 1: Supervised Recall

Early Fusion (EF)					
$E(M^L, M^S)$	74.00	76.66	71.11	4.46	
Cross Feature Fusion ($X_F F$)					
$X_F(S^L, M^L)$	76.20	79.60	72.50	3.63	
$X_F(S^L, M^S)$	74.60	75.10	73.90	3.08	
$X_F(S^S, M^L)$	**78.90**	80.70	**76.90**	1.08	**1F**
$X_F(S^S, M^S)$	73.70	77.70	70.00	2.72	
Cross Similarity Fusion ($X_S F$)					
$X_S(S^S, S^L)$	**78.90**	**80.80**	76.80	1.01	
$X_S(S^L, S^S)$	78.70	80.50	76.80	1.33	

Figure 2: Unsupervised F-measure

Spectral Clustering Evaluation

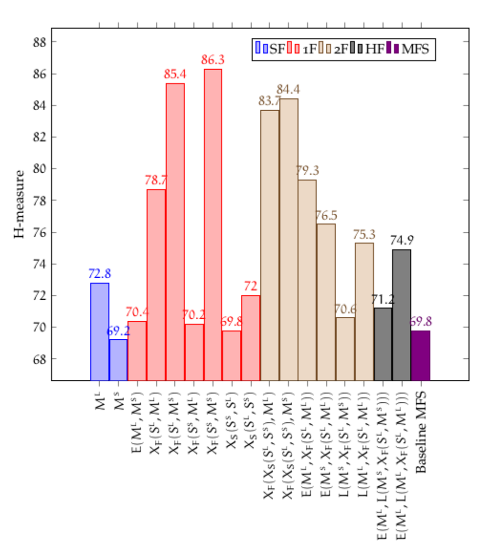


Figure 3: Proposed H-measure

Proposed Algorithm Evaluation

	Early Fusion (EF)				
$E(M^L, M^S)$	78.80	81.00	**76.40**	2.43	
	Cross Feature Fusion ($X_F F$)				
$X_F(S^L, M^L)$	78.70	80.90	76.20	3.11	
$X_F(S^L, M^S)$	78.50	81.10	75.60	1.92	
$X_F(S^S, M^L)$	**79.10**	**81.60**	**76.40**	1.73	**1F**
$X_F(S^S, M^S)$	78.60	80.90	76.00	1.81	
	Cross Similarity Fusion ($X_S F$)				
$X_S(S^S, S^L)$	78.60	80.80	76.20	1.44	
$X_S(S^L, S^S)$	78.70	80.90	76.20	1.10	

Figure 4: Supervised Recall

Proposed Algorithm Evaluation

				Early Fusion (EF)	
$E(M^L, M^S)$	78.80	81.00	**76.40**	2.43	
				Cross Feature Fusion (X_FF)	
$X_F(S^L, M^L)$	78.70	80.90	76.20	3.11	
$X_F(S^L, M^S)$	78.50	81.10	75.60	1.92	
$X_F(S^S, M^L)$	**79.10**	**81.60**	**76.40**	1.73	**1F**
$X_F(S^S, M^S)$	78.60	80.90	76.00	1.81	
				Cross Similarity Fusion (X_SF)	
$X_S(S^S, S^L)$	78.60	80.80	76.20	1.44	
$X_S(S^L, S^S)$	78.70	80.90	76.20	1.10	

Figure 4: Supervised Recall

				Early Fusion (EF)	
$E(M^L, M^S)$	76.90	**80.20**	73.10	2.43	
				Cross Feature Fusion (X_FF)	
$X_F(S^L, M^L)$	71.00	68.10	74.20	3.11	
$X_F(S^L, M^S)$	77.70	79.60	75.50	1.92	
$X_F(S^S, M^L)$	75.20	75.50	74.90	1.73	**1F**
$X_F(S^S, M^S)$	77.60	80.50	74.30	1.81	
				Cross Similarity Fusion (X_SF)	
$X_S(S^S, S^L)$	74.10	72.10	76.50	1.44	
$X_S(S^L, S^S)$	**78.30**	79.70	**76.80**	1.10	

Figure 5: Unsupervised F-measure

Proposed Algorithm Evaluation

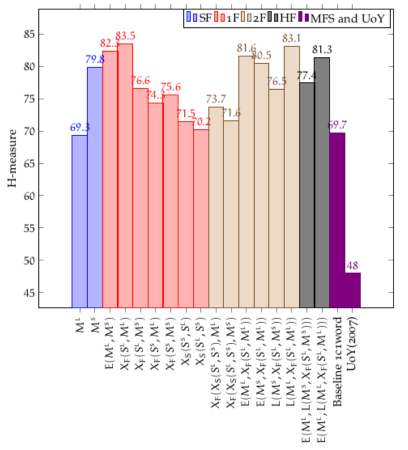


Figure 6: Proposed H-measure

Conclusions

Insights From our Contributions

- **Hypergraph linguistic model to hold heterogeneous information**
  - Hypergraphs allow a multi-layered representation of text within a single resource.
  - The Wikipedia-based instantiation serves as a NLP system starting point

Insights From our Contributions

- **Hypergraph linguistic model to hold heterogeneous information**
  - Hypergraphs allow a multi-layered representation of text within a single resource.
  - The Wikipedia-based instantiation serves as a NLP system starting point
- **Multimedia fusion techniques to combine and densify representations**
  - High-degree combinations of linguistic representations reduce sparsity
  - These fusion spaces achieve improvements on NER and WSI/WSD compared to single features and trivial fusion

Insights From our Contributions

- **Hypergraph linguistic model to hold heterogeneous information**
  - Hypergraphs allow a multi-layered representation of text within a single resource.
  - The Wikipedia-based instantiation serves as a NLP system starting point
- **Multimedia fusion techniques to combine and densify representations**
  - High-degree combinations of linguistic representations reduce sparsity
  - These fusion spaces achieve improvements on NER and WSI/WSD compared to single features and trivial fusion
- **Finding semantically-related communities on linguistic networks**
  - The proposed community finding method improves over similar algorithms while being simpler and allowing for heterogeneous features

Future Work

- **Hypergraph Linguistic Model**
  - A dataframe-like structure specialized on linguistic information based on the proposed model
  - Defining inter-features similarities measures within the network

Future Work

- **Hypergraph Linguistic Model**
    - A dataframe-like structure specialized on linguistic information based on the proposed model
    - Defining inter-features similarities measures within the network
- **Combining Features and Dealing with Sparsity**
    - Finding a more principled way to determine what type of context with what type of fusion operation according to the task at hand
    - Exploring fusion with other types of features (other modalities)

Future Work

- **Hypergraph Linguistic Model**
  - A dataframe-like structure specialized on linguistic information based on the proposed model
  - Defining inter-features similarities measures within the network
- **Combining Features and Dealing with Sparsity**
  - Finding a more principled way to determine what type of context with what type of fusion operation according to the task at hand
  - Exploring fusion with other types of features (other modalities)
- **Applications to NLP**
  - Comparison with other distributional representations (word embeddings)
  - Using the large Wikipedia-based network as a background corpus to further enrich domain-specific corpora
  - Test more feature weighting schemes, validate findings on more datasets

Publications Produced by our Research

- Edmundo-Pavel Soriano-Morales, Julien Ah-Pine, Sabine Loudcher: **Fusion Techniques for Named Entity Recognition and Word Sense Induction and Disambiguation**. DS 2017
- Edmundo-Pavel Soriano-Morales, Julien Ah-Pine, Sabine Loudcher: **Using a Heterogeneous Linguistic Network for Word Sense Induction and Disambiguation.** CICLING 2016
- Edmundo-Pavel Soriano-Morales, Julien Ah-Pine, Sabine Loudcher: **Hypergraph Modelization of a Syntactically Annotated English Wikipedia Dump**. LREC 2016
- Adrien Guille, Edmundo-Pavel Soriano-Morales, Ciprian-Octavian Truica: **Topic modeling and hypergraph mining to analyze the EGC conference history**. EGC 2016
- Adrien Guille, Edmundo-Pavel Soriano-Morales: **TOM: A library for topic modeling and browsing.** EGC 2016
- Julien Ah-Pine, Edmundo-Pavel Soriano-Morales: **A Study of Synthetic Oversampling for Twitter Imbalanced Sentiment Analysis**. DMNLP@PKDD/ECML 2016
- Sabine Loudcher, Wararat Jakawat, Edmundo-Pavel Soriano-Morales, Cécile Favre: **Combining OLAP and information networks for bibliographic data analysis: a survey**. Scientometrics 103(2) 2015

Thank you for your attention

- Edmundo-Pavel Soriano-Morales, Julien Ah-Pine, Sabine Loudcher: **Fusion Techniques for Named Entity Recognition and Word Sense Induction and Disambiguation**. DS 2017
- Edmundo-Pavel Soriano-Morales, Julien Ah-Pine, Sabine Loudcher: **Using a Heterogeneous Linguistic Network for Word Sense Induction and Disambiguation.** CICLING 2016
- Edmundo-Pavel Soriano-Morales, Julien Ah-Pine, Sabine Loudcher: **Hypergraph Modelization of a Syntactically Annotated English Wikipedia Dump**. LREC 2016
- Adrien Guille, Edmundo-Pavel Soriano-Morales, Ciprian-Octavian Truica: **Topic modeling and hypergraph mining to analyze the EGC conference history**. EGC 2016
- Adrien Guille, Edmundo-Pavel Soriano-Morales: **TOM: A library for topic modeling and browsing.** EGC 2016
- Julien Ah-Pine, Edmundo-Pavel Soriano-Morales: **A Study of Synthetic Oversampling for Twitter Imbalanced Sentiment Analysis**. DMNLP@PKDD/ECML 2016
- Sabine Loudcher, Wararat Jakawat, Edmundo-Pavel Soriano-Morales, Cécile Favre: **Combining OLAP and information networks for bibliographic data analysis: a survey**. Scientometrics 103(2) 2015