

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



UNIVERSITÉ
LUMIÈRE
LYON 2



UNIVERSITÉ
DE LYON

INSTITUT
DES SCIENCES
DE L'HOMME



Introduction

Why is it useful to us to understand text?



Who invented Python?

All

Images

Shopping

Videos

News

More

Settings

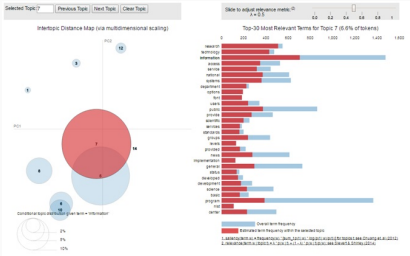
Tools

About 520,000 results (0.63 seconds)

Guido van Rossum

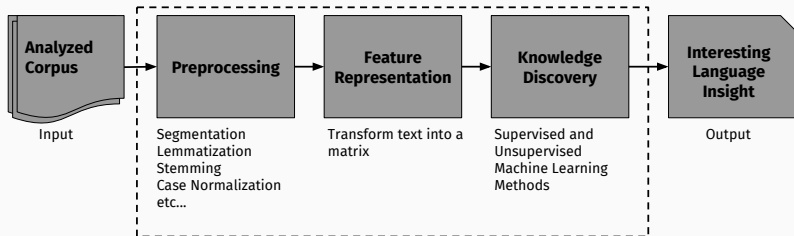
Python was conceived in the late 1980s, and its implementation began in December 1989 by **Guido van Rossum** at Centrum Wiskunde & Informatica (CWI) in the Netherlands as a successor to the ABC language (itself inspired by SETL) capable of exception handling and interfacing with the operating system Amoeba. **Van Rossum** is ...



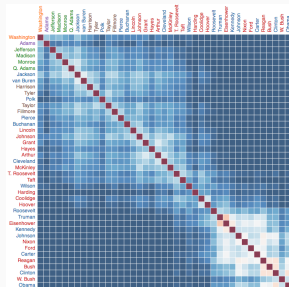


How do we extract meaning from text?

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

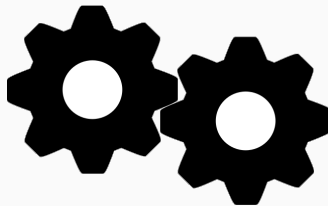


How do we represent text for the machine to understand?



Dealing with **data sparsity**
Leveraging **heterogeneity**

What techniques do we use to discover meaning from text?



Finding **semantic communities**

- **Common ways to represent text**

- Lexical
- Syntactic
 - Constituency Tree
 - Dependency Tree

- **Common ways to represent text**

- Lexical
- Syntactic
 - Constituency Tree
 - Dependency Tree

- **Example Phrase**

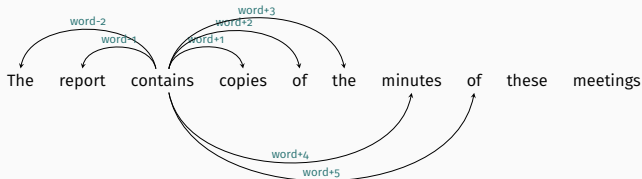
The report contains copies of the minutes of these meetings

- **Common ways to represent text**

- Lexical
- Syntactic
 - Constituency Tree
 - Dependency Tree

- **Example Phrase**

The report contains copies of the minutes of these meetings

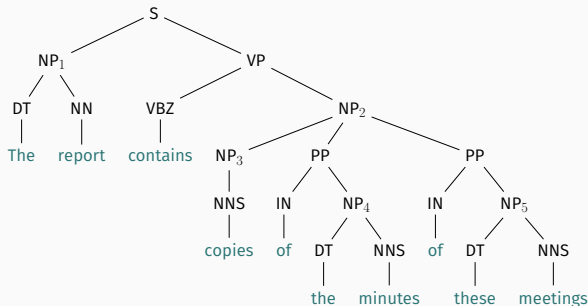


- **Common ways to represent text**

- Lexical
- Syntactic
 - Constituency Tree
 - Dependency Tree

- **Example Phrase**

The report contains copies of the minutes of these meetings

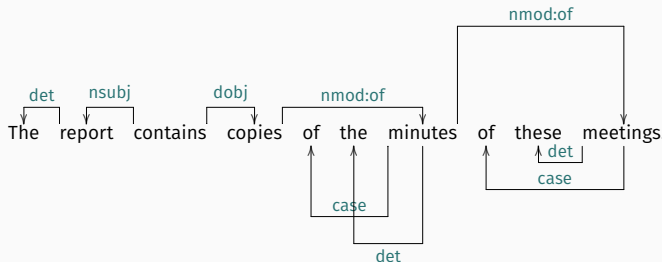


- **Common ways to represent text**

- Lexical
- Syntactic
 - Constituency Tree
 - Dependency Tree

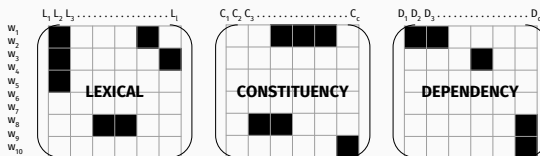
- **Example Phrase**

The report contains copies of the minutes of these meetings



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

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



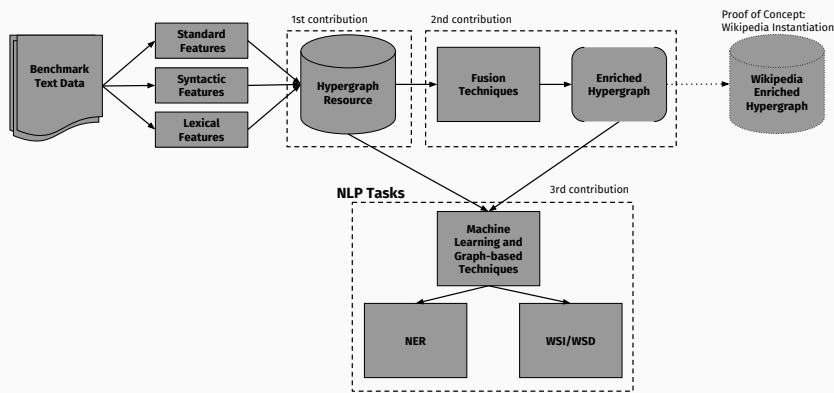
1. What type of model can we employ to represent a corpus **using heterogeneous features**?
 - *Hypergraph model to hold different types of linguistic information*

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*

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*

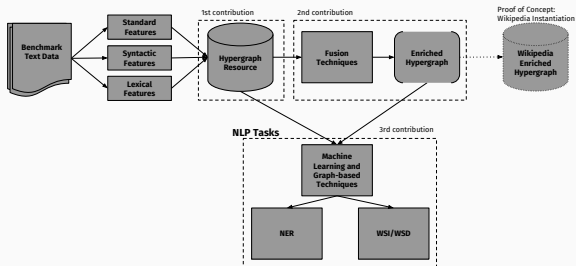
Introduction

Work Overview



Contributions in Detail

Hypergraph Linguistic Model



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

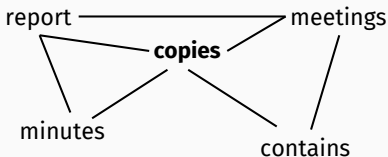
The report contains copies of the minutes of these meetings

Example phrase

The report contains copies of the minutes of these meetings

Lexical Networks

Sentence Level

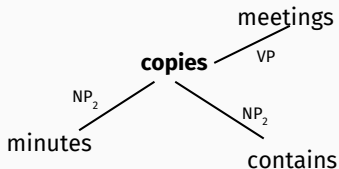


Example phrase

The report contains copies of the minutes of these meetings

Syntactic Networks

Constituency Tree

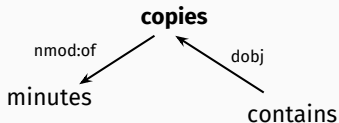


Example phrase

The report contains copies of the minutes of these meetings

Syntactic Networks

Dependency Tree



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

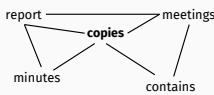
- **Limitations of existing representations**
 - Language networks generally employ a single type of textual information
 - The edges of the network relate maximum two words at each time
- **Proposition**
 - Use a hypergraph model to link together the different types of networks
 - This allows for a semantic overview at three different layers: short range, medium range, and long range at once
 - Relating more than two words at the same time

Hypergraph Linguistic Model

Proposed Model

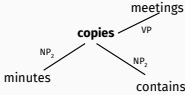
Lexical Networks

Sentence Level



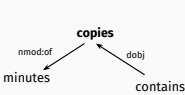
Syntactic Networks

Constituency Tree



Syntactic Networks

Dependency Tree

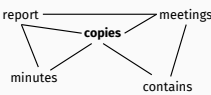


Hypergraph Linguistic Model

Proposed Model

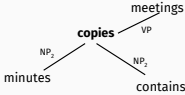
Lexical Networks

Sentence Level



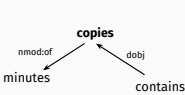
Syntactic Networks

Constituency Tree



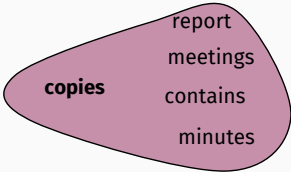
Syntactic Networks

Dependency Tree



Hypergraph Model

■ Lexical

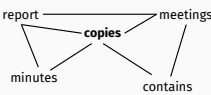


Hypergraph Linguistic Model

Proposed Model

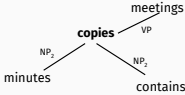
Lexical Networks

Sentence Level



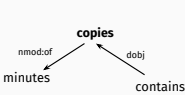
Syntactic Networks

Constituency Tree

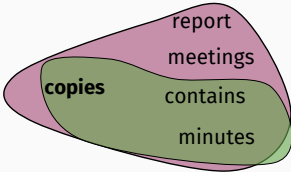


Syntactic Networks

Dependency Tree



Hypergraph Model



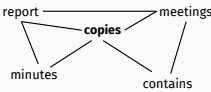
- Lexical
- Constituency (NP₂)

Hypergraph Linguistic Model

Proposed Model

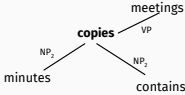
Lexical Networks

Sentence Level



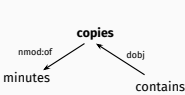
Syntactic Networks

Constituency Tree

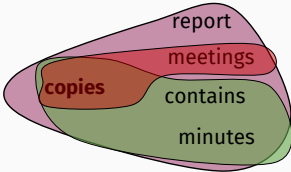


Syntactic Networks

Dependency Tree



Hypergraph Model



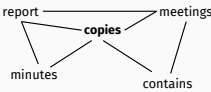
- Lexical
- Constituency (NP₂)
- Constituency (VP)

Hypergraph Linguistic Model

Proposed Model

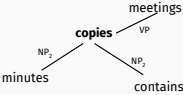
Lexical Networks

Sentence Level



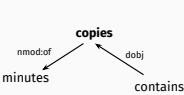
Syntactic Networks

Constituency Tree

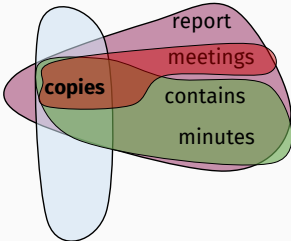


Syntactic Networks

Dependency Tree



Hypergraph Model



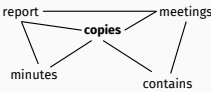
- Lexical
- Constituency (NP₂)
- Constituency (VP)
- Dependency (dobj:contains)

Hypergraph Linguistic Model

Proposed Model

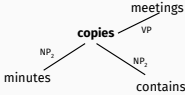
Lexical Networks

Sentence Level



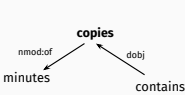
Syntactic Networks

Constituency Tree

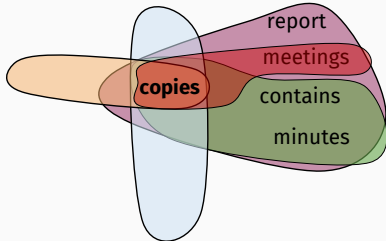


Syntactic Networks

Dependency Tree



Hypergraph Model



- Lexical
- Constituency (NP₂)
- Constituency (VP)
- Dependency (dobj:contains)
- Dependency (nmod:of)

Contributions in Detail

**Combining Features and Dealing with
Sparsity**

- **Definition**

- Used in multimedia analysis tasks to integrate multiple media
- We adapt them to combine textual information
- The goal is to obtain rich insights about the data being treated
- By creating a single representation from heterogeneous information

- **Definition**

- Used in multimedia analysis tasks to integrate multiple media
- We adapt them to combine textual information
- The goal is to obtain rich insights about the data being treated
- By creating a single representation from heterogeneous information

- **Main fusion operators:**

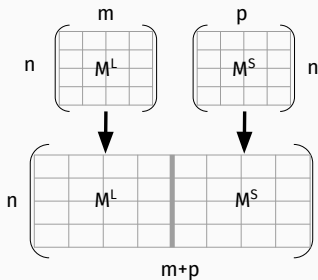
- Early Fusion $E_{\alpha}(\cdot)$,
- Late Fusion $L_{\beta}(\cdot)$,
- Cross Fusion $X_{\gamma}(\cdot)$

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

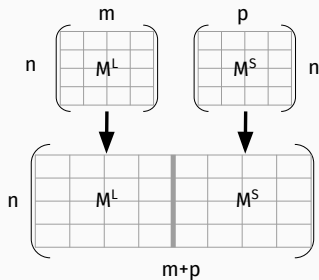


DEFINITIONS

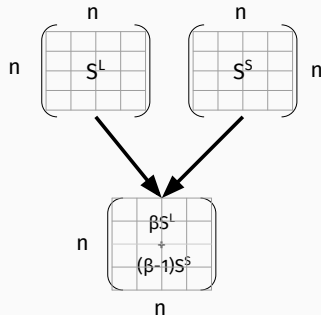
M^L	Lexical features	M^S	Syntactic features
S^L	Lexical similarities	S^S	Syntactic similarities

EARLY FUSION

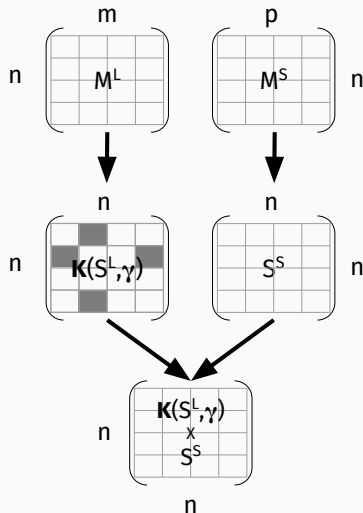
Matrices M^L and M^S have the same number of rows

**LATE FUSION: SIMILARITY FUSION**

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



CROSS FUSION



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

- **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

$$\begin{matrix} w_1 & w_2 & w_3 \\ \begin{pmatrix} S^L \end{pmatrix} \end{matrix} \times \begin{matrix} f_{S1} & f_{S2} & f_{S3} \\ \begin{pmatrix} M^S \end{pmatrix} \end{matrix} = \begin{matrix} f_{S1} & f_{S2} & f_{S3} \\ \begin{pmatrix} X_F(S^L, M^S) \end{pmatrix} \end{matrix}$$

$X_S(S^L, S^S)$

Cross Similarity Fusion

$$\begin{matrix} w_1 & w_2 & w_3 \\ \begin{pmatrix} S^L \end{pmatrix} \end{matrix} \times \begin{matrix} w_1 & w_2 & w_3 \\ \begin{pmatrix} S^S \end{pmatrix} \end{matrix} = \begin{matrix} w_1 & w_2 & w_3 \\ \begin{pmatrix} X_S(S^L, S^S) \end{pmatrix} \end{matrix}$$

- **Combining fusion operators**

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

- **Second Degree**

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

Cross Feature Early Fusion

$$\begin{array}{c}
 \boxed{X_F(S^L, E(M^S, M^L))} \\
 \begin{array}{c}
 \begin{matrix} f_{S1} & f_{S2} & f_{S3} \\ w_1 & w_2 & w_3 \end{matrix} \begin{pmatrix} M^S \end{pmatrix} \parallel \begin{matrix} f_{L1} & f_{L2} & f_{L3} \\ w_1 & w_2 & w_3 \end{matrix} \begin{pmatrix} M^L \end{pmatrix} = \begin{matrix} f_{S1} & f_{S2} & f_{S3} & f_{L1} & f_{L2} & f_{L3} \\ w_1 & w_2 & w_3 \end{matrix} \begin{pmatrix} E(M^S, M^L) \end{pmatrix} \\
 \begin{matrix} w_1 & w_2 & w_3 \end{matrix} \begin{pmatrix} S^L \end{pmatrix} \times \begin{matrix} f_{S1} & f_{S2} & f_{S3} & f_{L1} & f_{L2} & f_{L3} \\ w_1 & w_2 & w_3 \end{matrix} \begin{pmatrix} E(M^S, M^L) \end{pmatrix} = \begin{matrix} f_{S1} & f_{S2} & f_{S3} & f_{L1} & f_{L2} & f_{L3} \\ w_1 & w_2 & w_3 \end{matrix} \begin{pmatrix} X_F(S^L, E(M^S, M^L)) \end{pmatrix}
 \end{array}
 \end{array}$$

- **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



The diagram shows a nested expression for a Higher Degree Operator. It consists of several nested boxes: a large light blue box containing the expression $E(M_L, E(E(M^T, L(M^T, X_F(S^T, M^T))), L(M^L, X_F(S^S, M^L))))$. Inside this, there is a light purple box containing $E(E(M^T, L(M^T, X_F(S^T, M^T))), L(M^L, X_F(S^S, M^L)))$. This is further nested with a green box containing $E(M^T, L(M^T, X_F(S^T, M^T)))$, and finally a red box containing $L(M^T, X_F(S^T, M^T))$. The yellow box contains $L(M^L, X_F(S^S, M^L))$.

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

Higher Degree Operator

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

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

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

Higher Degree Operator

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

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

$$\begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} \begin{pmatrix} w_1 w_2 w_3 \\ S^T \end{pmatrix} \times \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} \begin{pmatrix} f_{T1} f_{T2} f_{T3} \\ M^T \end{pmatrix} = \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} \begin{pmatrix} f_{T1} f_{T2} f_{T3} \\ X_F(S^T, M^T) \end{pmatrix}$$

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

Higher Degree Operator

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

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

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

Higher Degree Operator

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

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

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

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

Higher Degree Operator

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

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

$$\begin{aligned} \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} \left(\begin{matrix} f_{L1} & f_{L2} & f_{L3} \\ M^T \end{matrix} \right) & \parallel \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} \left(\begin{matrix} f_{L1} & f_{L2} & f_{L3} \\ E(E(M^T, L(M^T, X_F(S^T, M^T))), L(M^L, X_F(S^S, M^L))) \end{matrix} \right) = \\ & \begin{matrix} w_1 \\ w_2 \\ w_3 \end{matrix} \left(\begin{matrix} f_{L1} & f_{L2} & f_{L3} & f_{L1} & f_{L2} & f_{L3} \\ E(M_L, E(E(M^T, L(M^T, X_F(S^T, M^T))), L(M^L, X_F(S^S, M^L)))) \end{matrix} \right) \end{aligned}$$

Contributions in Detail

Finding Communities in the Network

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

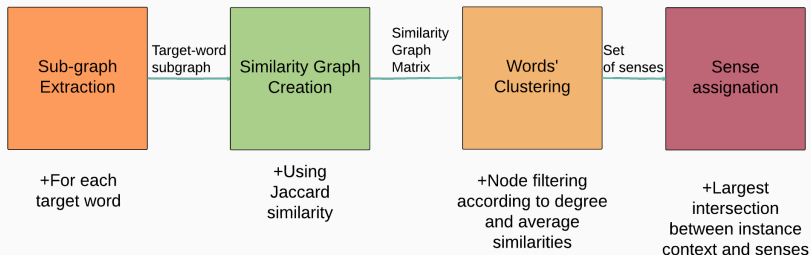
- **Language networks tend to be scale-free**
 - There are certain nodes (hubs) that are very well connected forming communities within the network
- **Seminal approaches**
 - Hyperlex [VÓ4]
 - University of York (UoY) [KMo7]

- **Language networks tend to be scale-free**
 - There are certain nodes (hubs) that are very well connected forming communities within the network
- **Seminal approaches**
 - Hyperlex [VÓ4]
 - University of York (UoY) [KM07]
- **Limitations of existing approaches**
 - Single typed networks
 - Large number of parameters

- **Language networks tend to be scale-free**
 - There are certain nodes (hubs) that are very well connected forming communities within the network
- **Seminal approaches**
 - Hyperlex [VÓ4]
 - University of York (UoY) [KMo7]
- **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

Finding Communities in the Network

Proposed Method

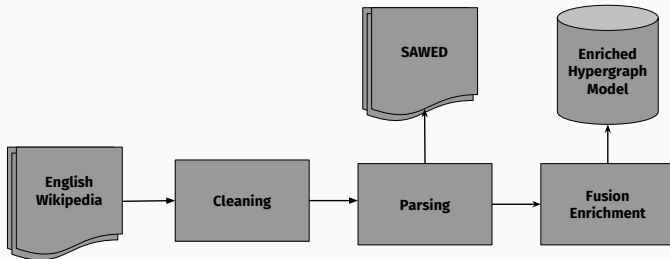


Applications to NLP

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

- **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**



Hypergraph Model Instantiation

Syntactically Annotated Wikipedia

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 Model Instantiation

Hypergraph Incidence Matrix

		CONSTITUENT			DEPENDENCY	SENTENCE
		NP ₁ DT:NN	NP ₂ NP:PP:PP	NP ₃ NNS	nsubj contains dobj contains	S ₁
N N	report	1			1	1
	copies		1	1	1	1
	minutes		1			1
	meetings		1			1
VB	contains					1

- **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%) $X_F(S^L, M^S)$
priest	priests	monk	sailor	vassal	sailor
	nun	regent	regent	regent	fluent
	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)

- **Preprocessing**
 - Normalize numbers

- **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.

- **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

- **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)

Solving Named Entity Recognition

Evaluation Baselines (F-measure)

A	Single Features		
	CONLL	WNER	WGLD
M^T	77.41	77.50	59.66
M^L	69.40	69.17	52.34
M^S	32.95	28.47	25.49

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

Solving Named Entity Recognition Evaluation (F-measure)

A	B	Baseline (EF)		
		CONLL	WNER	WGLD
M^L	$E(M^S, M^T)$	78.90	80.04	63.20

First Degree Fusion

		Cross Feature Fusion ($X_F F$)		
		CONLL	WNER	WGLD
S^L	M^T	49.90	70.27	62.69
S^S	M^T	47.27	51.38	48.53
S^T	M^L	52.89	62.21	50.15

A	B	Baseline (EF)		
		CONLL	WNER	WGLD
M^L	$E(M^S, M^T)$	78.90	80.04	63.20

Second Degree Fusion

A	B	Early Cross Feature Fusion (EX _F F)		
		CONLL	WNER	WGLD
M^T	$X_F(S^S, M^L)$	49.58	77.32	61.69

Solving Named Entity Recognition Evaluation (F-measure)

A	B	Baseline (EF)		
		CONLL	WNER	WGLD
M^L	$E(M^S, M^T)$	78.90	80.04	63.20

Second Degree Fusion

A	B	Late Cross Feature Fusion ($LX_F F$)		
		CONLL	WNER	WGLD
M^T	$X_F(S^S, M^T)$	56.53	62.27	52.39

Solving Named Entity Recognition Evaluation (F-measure)

A	B	Baseline (EF)		
		CONLL	WNER	WGLD
M^L	$E(M^S, M^T)$	78.90	80.04	63.20

High Degree Fusion

Triple Early Double Late Cross Feature Fusion (EEELX _F LX _F)				
		CONLL	WNER	WGLD
$M^L_{\alpha=0.95}$	$\hat{b}_{EEELX_F LX_F}$	79.67	81.79	67.05

$$\hat{b}_{EEELX_F LX_F} = E(E(M^T, L(M^T, X_F(S^S, M^T))), L(M^L, X_F(S^S, M^L)))$$

- Split the operator in four different models

$$\begin{array}{c}
 \overbrace{\hspace{15em}}^{M_4} \\
 \overbrace{\hspace{10em}}^{M_2} \\
 E_{\alpha=0.95}(\underbrace{M^L, M^T}_{M_1}, L(M^T, X_F(S^S, M^T)), L(M^L, X_F(S^S, M^L))) \\
 \underbrace{\hspace{15em}}_{M_3}
 \end{array}$$

- Split the operator in four different models

$$\begin{array}{c}
 \overbrace{\hspace{15em}}^{M_4} \\
 \overbrace{\hspace{10em}}^{M_2} \\
 E_{\alpha=0.95}(\underbrace{M^L, M^T}_{M_1}, L(M^T, X_F(S^S, M^T)), L(M^L, X_F(S^S, M^L))) \\
 \underbrace{\hspace{15em}}_{M_3}
 \end{array}$$

$$M_1 \quad M^L$$

$$M_2 \quad E_{\alpha}(M^L, M^T)$$

$$M_3 \quad E_{\alpha}(M^L, M^T, L(M^T, X_F(S^S, M^T)))$$

$$M_4 \quad E_{\alpha}(M^L, M^T, L(M^T, X_F(S^S, M^T)), L(M^L, X_F(S^S, M^L)))$$

- **Error Analysis Model**

- To facilitate the interpretation, we change the prediction model to a logistic regression with L_1 normalization, which also benefits from the enriched spaces

- **Error Analysis Model**

- To facilitate the interpretation, we change the prediction model to a logistic regression with L_1 normalization, which also benefits from the enriched spaces

- **Procedure**

- We find an error on a model and then see if this error was fixed in the next evolved model
- We study the weights assigned to each feature and see if those added by the fusion make the model change its decision

- **Error Analysis Model**

- To facilitate the interpretation, we change the prediction model to a logistic regression with L_1 normalization, which also benefits from the enriched spaces

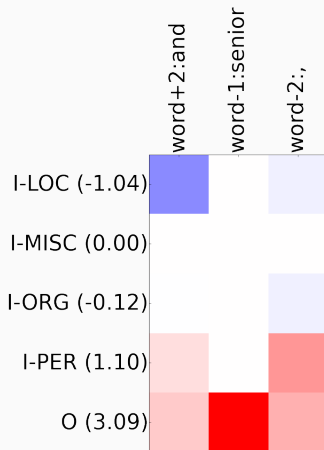
- **Procedure**

- We find an error on a model and then see if this error was fixed in the next evolved model
- We study the weights assigned to each feature and see if those added by the fusion make the model change its decision

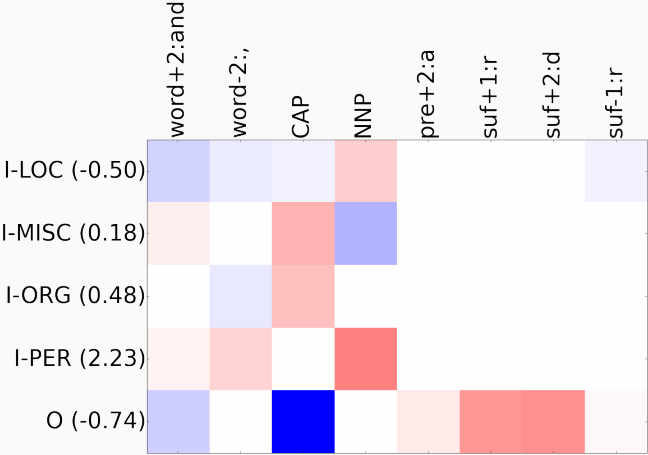
- **Experiment**

- We follow the proper name *Kory* from M_1 (incorrectly classified as O) to M_2 (correctly classified as PER)
- Similarly, we follow the proper name *Green* from M_3 (incorrectly classified as ORG) to M_4 (correctly classified as PER)

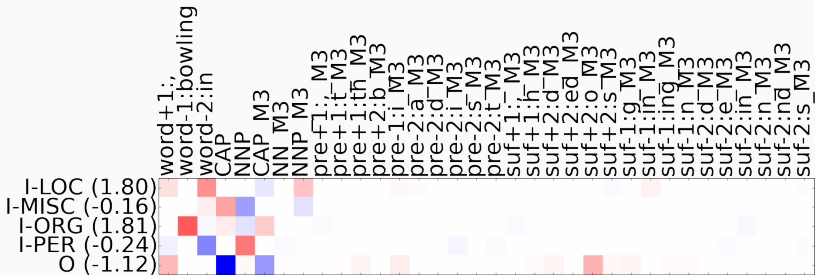
We focus on the word *Kory*, and its performance from model M_1 to M_2



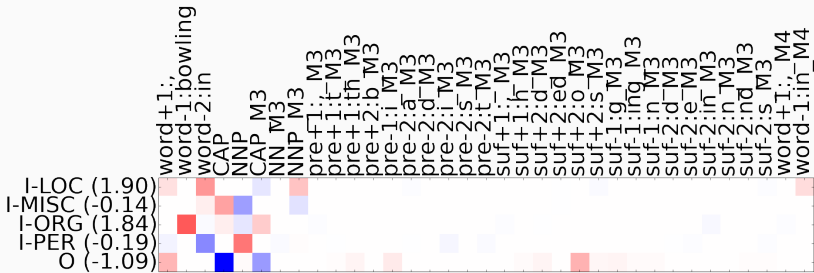
We focus on the word *Kory*, and its performance from model M_1 to M_2



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



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



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

Experimental Protocol

- **Feature Space**
 - Lexical (L) and Syntactic (S) Features
- **Preprocessing**
 - Remove very frequent and very infrequent words

Experimental Protocol

- **Feature Space**

- Lexical (L) and Syntactic (S) Features

- **Preprocessing**

- Remove very frequent and very infrequent words

- **Test Corpora**

- Semeval Competition 2007: Train 219,554 lines. Test 50,350 lines

- **Feature Space**
 - Lexical (L) and Syntactic (S) Features
- **Preprocessing**
 - Remove very frequent and very infrequent words
- **Test Corpora**
 - Semeval Competition 2007: Train 219,554 lines. Test 50,350 lines
- **Clustering Algorithm**
 - Spectral Clustering
 - Proposed Community Algorithm

- **Feature Space**
 - Lexical (L) and Syntactic (S) Features
- **Preprocessing**
 - Remove very frequent and very infrequent words
- **Test Corpora**
 - Semeval Competition 2007: Train 219,554 lines. Test 50,350 lines
- **Clustering Algorithm**
 - Spectral Clustering
 - Proposed Community Algorithm
- **Evaluation Metrics**
 - Supervised Recall (SR)
 - Unsupervised F-measure (UF)

- **Feature Space**
 - Lexical (L) and Syntactic (S) Features
- **Preprocessing**
 - Remove very frequent and very infrequent words
- **Test Corpora**
 - Semeval Competition 2007: Train 219,554 lines. Test 50,350 lines
- **Clustering Algorithm**
 - Spectral Clustering
 - Proposed Community Algorithm
- **Evaluation Metrics**
 - Supervised Recall (SR)
 - Unsupervised F-measure (UF)
 - **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

Supervised Recall

Fusion Operation / System	Recall (%)			#cl
	all	nouns	verbs	
Single Features				
M^L	79.20	82.10	75.80	4.13
M^S	79.10	81.60	76.20	4.47
Early Fusion (EF)				
$E(M^L, M^S)$	78.70	81.11	76.10	4.46
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

Unsupervised F-measure

Fusion Operation / System	F-measure (%)			#cl
	all	nouns	verbs	
Single Features				
M^L	72.70	76.90	67.90	4.13
M^S	69.30	69.40	69.20	4.47
Early Fusion (EF)				
$E(M^L, M^S)$	74.00	76.66	71.11	4.46
Cross Feature Fusion (X_F F)				
$X_F(S^S, M^L)$	78.90	80.70	76.90	1.08

Solving Word Sense Induction and Disambiguation

Spectral Clustering Evaluation

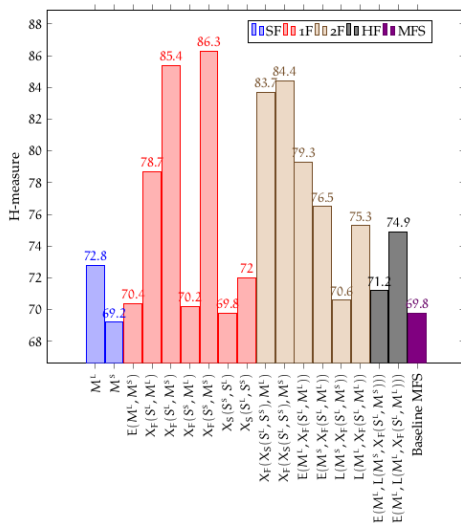


Figure 1: Proposed H-measure

Solving Word Sense Induction and Disambiguation

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
$X_F(S^S, M^S)$	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 2: Supervised Recall

Solving Word Sense Induction and Disambiguation

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
$X_F(S^S, M^S)$	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 2: Supervised Recall

Early Fusion (EF)				
$E(M^L, M^S)$	76.90	80.20	73.10	2.43
Cross Feature Fusion ($X_F F$)				
$X_F(S^L, M^L)$	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
$X_F(S^S, M^S)$	77.60	80.50	74.30	1.81
Cross Similarity Fusion ($X_S F$)				
$X_S(S^S, S^L)$	74.10	72.10	76.50	1.44
$X_S(S^L, S^S)$	78.30	79.70	76.80	1.10

Figure 3: Unsupervised F-measure

Solving Word Sense Induction and Disambiguation

Proposed Algorithm Evaluation

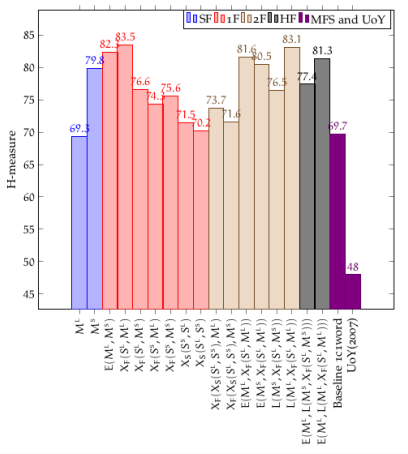


Figure 4: Proposed H-measure

Conclusions

- **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

- **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

- **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

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

- **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)

- **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