Short Introduction to Lexicalized Tree Kernels for NLP: SPTK and CSPTK

Roberto Basili,

a.a. 2015-16

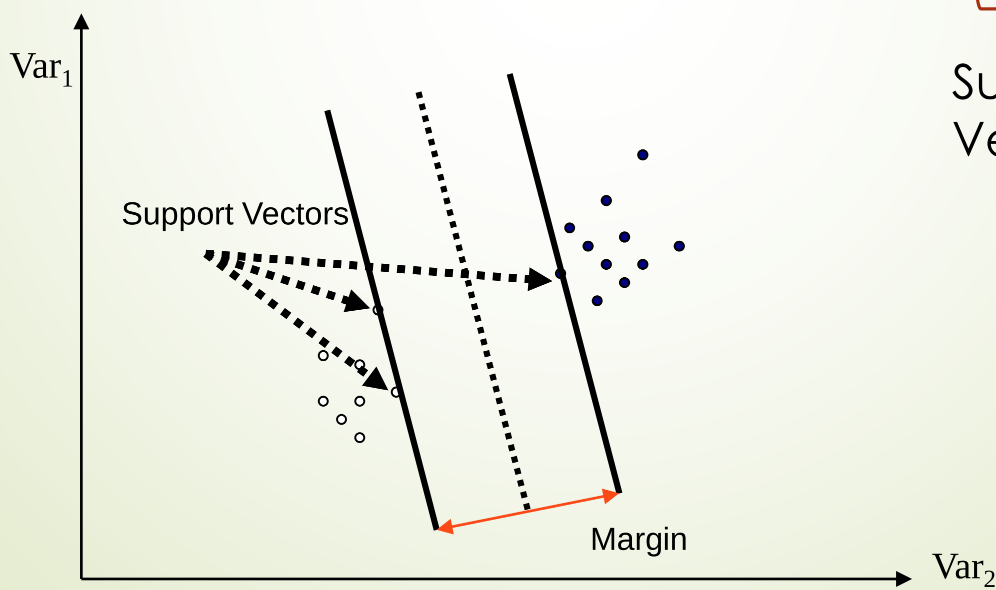
Outline

- Natural Language Learning, Compositional Semantics and Kernel based learning
  - Convolution Tree Kernels
  - Distributional Compositional Semantics
- Semantic Tree Kernels
  - The Compositionally Smoothed Partial
  - Experimental Evaluations
    - Question Classification
    - Paraphrase Identification
    - Metaphor Detection
  - Optimization of complex kernels: Nystrom method
- Industrial Applications of Kernel-based Learning
  - KELP: a Java-based framework for Kernel-based learning
- Conclusions

Supervised Learning from data: Support Vector Machines

- Support Vector Machines (SVMs) are machine learning algorithms based on statistical learning theory [Vapnik,1995]

$$h(x) = \mathrm{sgn}(\vec{w}\cdot\vec{x}+b) = \mathrm{sgn}(\underbrace{\sum_{j=1..\ell}\alpha_j y_j \vec{x}_j}_{\text{Support Vectors}} \cdot \vec{x}+b)$$



Representation and Kernel functions

$$h(x)=\mathrm{sgn}(\vec{w}\cdot\overbrace{\varphi(\vec{x})}^{\text{Projection Function}}+b)=\mathrm{sgn}(\sum_{j=1..\ell}\overbrace{\alpha_j y_j \varphi(\vec{x}_j)}^{\text{Support Vectors}})\cdot\varphi(\vec{x})+b)$$

Representation and Kernel functions

$$h(x) = \mathrm{sgn}(\vec{w}\cdot\overbrace{\varphi(\vec{x})}^{\text{Projection Function}} + b) = \mathrm{sgn}(\overbrace{\sum_{j=1..\ell}\alpha_j y_j}^{\text{Support Vectors}}\boxed{\varphi(\vec{x}_j)\cdot\varphi(\vec{x})} + b) =$$

$$= \mathrm{sgn}(\sum_{i=1..\ell}\alpha_j y_j \boxed{k(\vec{x}_j,\vec{x})} + b)$$

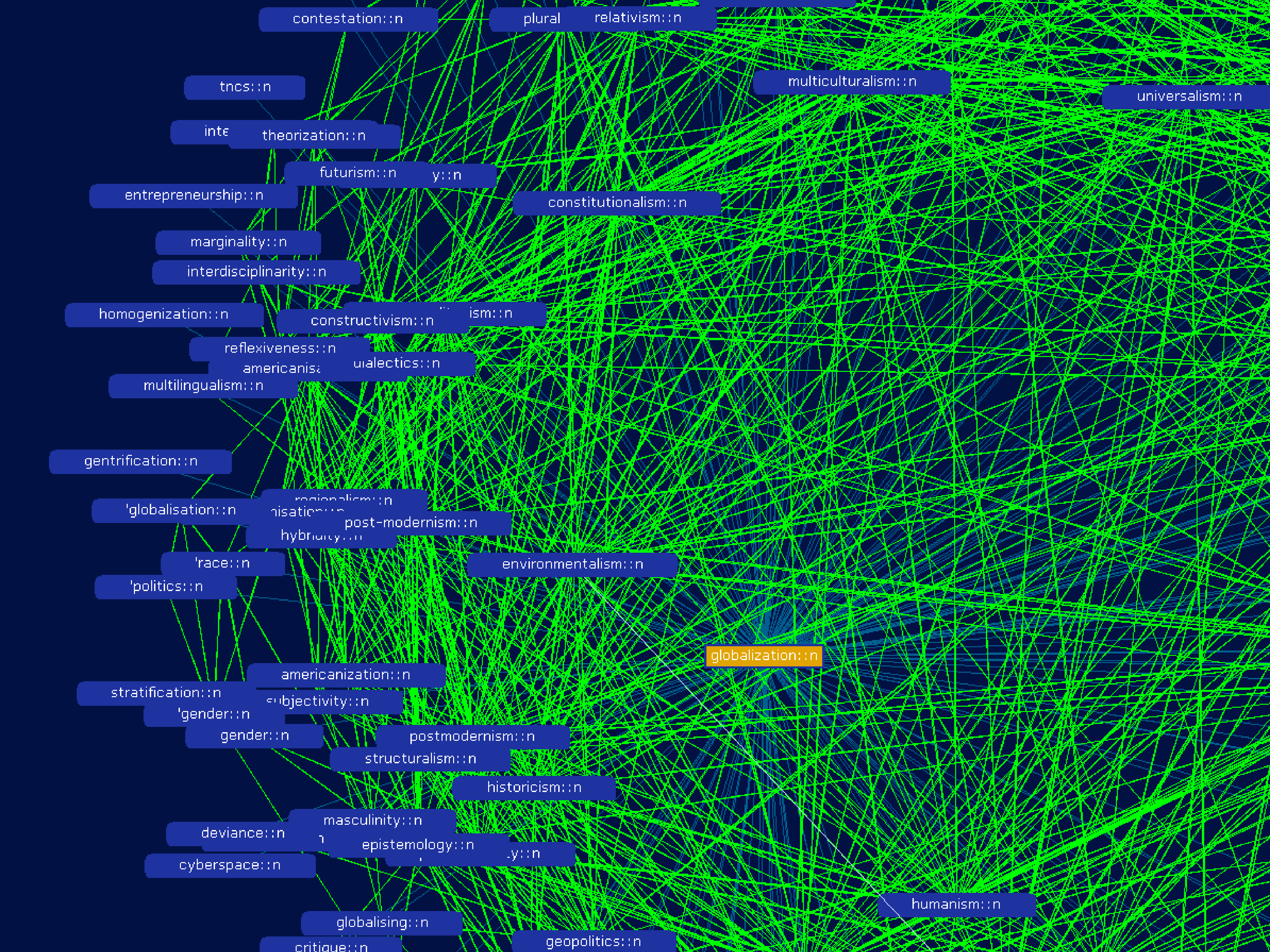
- If a *Kernel Function* ***k*** such that $k(x_i,x_j)=\phi(x_i)\cdot\phi(x_j)$ is available, there is no need to explicitly know the projection function ϕ [Cristianini et al., 2002]
- A **Structured Learning** paradigm can be adopted
  - Learning can be directly applied over (complex) structures
- A semantic similarity function ***k*** able to reflect lexical and syntactic aspects of linguistic examples is possible

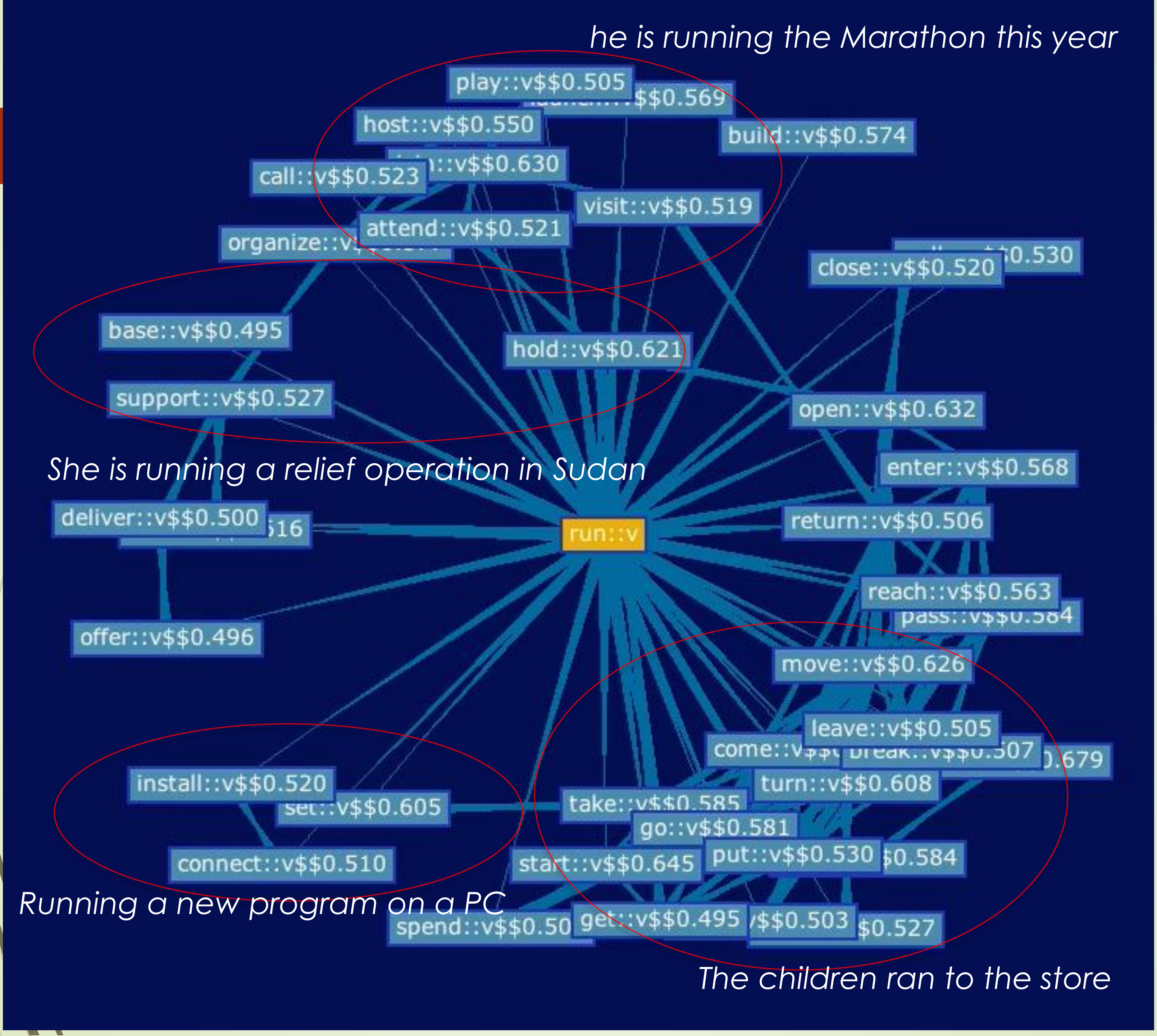
Learning NL Semantics

- Main perspective: the role of Semantic Compositionality
  - Frege's principle: "The meaning of a sentence must be derived by the composition of the meanings of its parts"
- Textual inference is based on the meaning of
  - single words
  - basic grammatical structures (i.e.V-Obj bigrams)
  - the overall interactions across the entire parse trees
- "… meaning of its parts" vs. "meaning as context"
  - Distributional Hypothesis [Harris, 1964] "words with similar meaning occur in similar contexts"
  - A geometrical space, a Word Space, can be acquired through statistical analysis of large corpora [Schutze,2001], [Sahlgren,2006][Baroni & Lenci, 2008], [Mikolov,2013]

Distributional Approaches to Lexical Semantics

- Vector spaces and Lexical Information
- Distributional approaches
  - Bow, the bayesian and IR tradition
  - Latent Semantic Spaces
  - HAL or counting-based wordspaces
  - Neural Language models
    - *Associative encoders* for Lexical Prediction (Word2Vect)
    - Continuous Probabilstic Language Models , Convolutional Neural Models


contestation::n
plural
relativism::n
tncs::n
multiculturalism::n
universalism::n
theorization::n
futurism::n
entrepreneurship::n
constitutionalism::n
marginality::n
interdisciplinarity::n
homogenization::n
constructivism::n
reflexiveness::n
multilingualism::n
gentrification::n
'globalisation::n
post-modernism::n
environmentalism::n
'race::n
'politics::n
globalization::n
americanization::n
stratification::n
'gender::n
gender::n
postmodernism::n
structuralism::n
historicism::n
masculinity::n
deviance::n
epistemology::n
cyberspace::n
humanism::n
globalising::n
geopolitics::n
critique::n

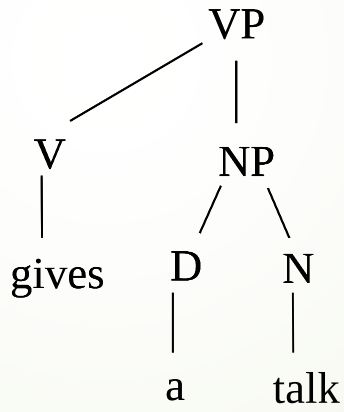

he is running the Marathon this year
play::v$$0.505
$$0.569
host::v$$0.550
build::v$$0.574
::v$$0.630
call::v$$0.523
visit::v$$0.519
attend::v$$0.521
organize::v
close::v$$0.520
$0.530
base::v$$0.495
hold::v$$0.621
support::v$$0.527
open::v$$0.632
She is running a relief operation in Sudan
enter::v$$0.568
deliver::v$$0.500
516
run::v
return::v$$0.506
reach::v$$0.563
offer::v$$0.496
move::v$$0.626
leave::v$$0.505
come::v
break::v$$0.507
.679
install::v$$0.520
turn::v$$0.608
set::v$$0.605
take::v$$0.585
go::v$$0.581
connect::v$$0.510
start::v$$0.645
put::v$$0.530
$0.584
Running a new program on a PC
spend::v$$0.50
get::v$$0.495
$$0.503
$0.527
The children ran to the store

The big issue

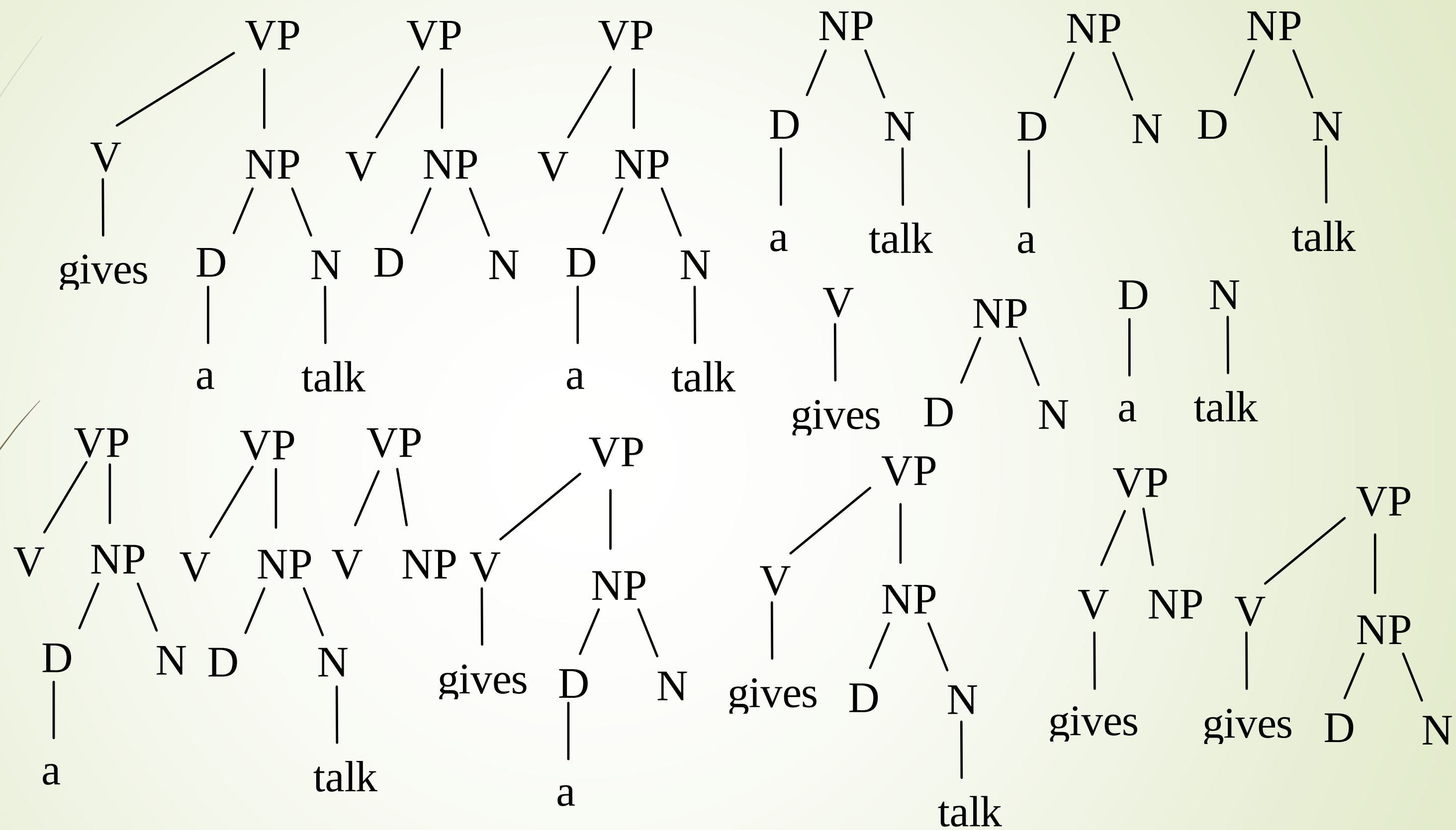
- *"How to combine word representations in order to characterize a model for sentence semantics?"*
- DM are typically focusing on isolated words
  - **Distributional Compositional Semantic** (DCS) models aim at capturing the meaning of phrases (i.e. bi-gram)…
  - …but they should be also sensitive to the full syntactic structure!
- IDEA: *Convolution Kernels* (Haussler, 1999) are well-known similarity functions among such complex structures (see also Zanzotto et al, 2013 CL paper)

TKs, PTKs and their limitations

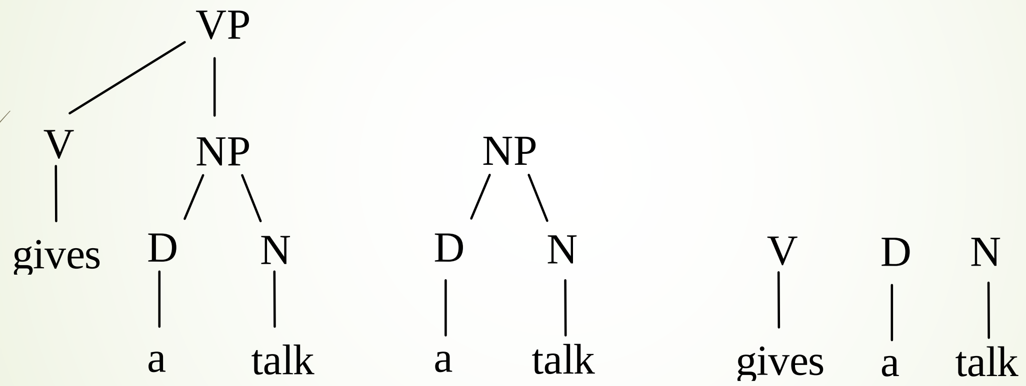
- The Collins and Duffy's Tree Kernel (called SST in [Vishwanathan and Smola, 2002])



The overall fragment set



SubTree (ST) Kernel [Vishwanathan and Smola, 2002]



Evaluation

- Given the equation for the SST kernel

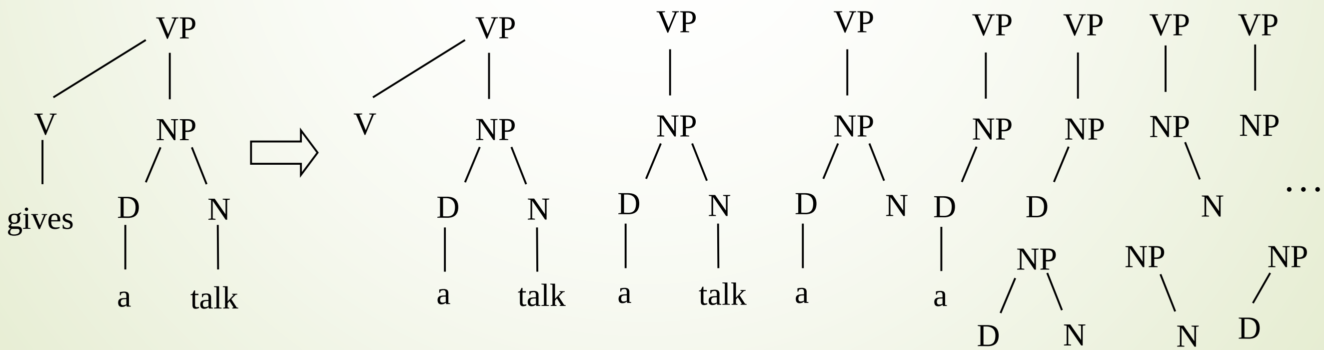
$$\Delta(n_1, n_2) = 0, \text{ if the productions are different else}$$

$$\Delta(n_1, n_2) = 1, \text{ if pre-terminals else}$$

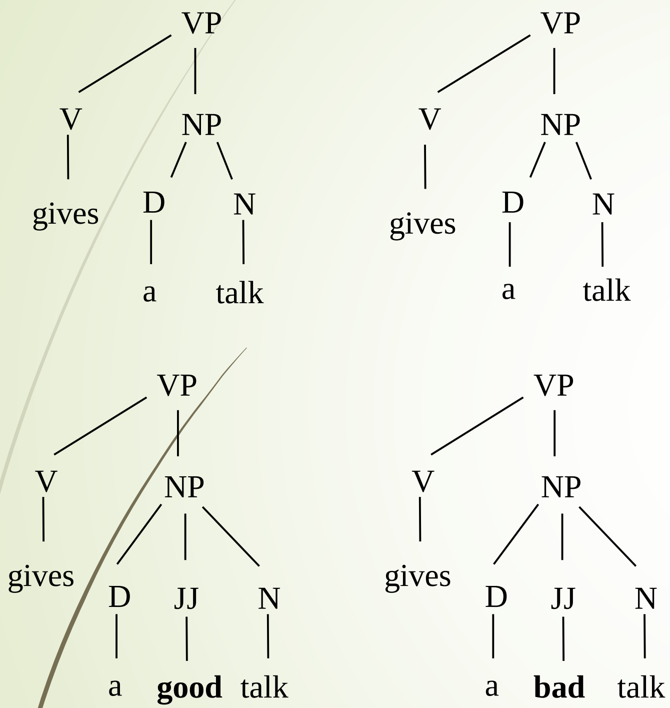
$$\Delta(n_1, n_2) = \prod_{j=1}^{nc(n_1)} (1 + \Delta(ch(n_1, j), ch(n_2, j)))$$

Labeled Ordered Tree Kernel

- SST satisfies the constraint "remove 0 or all children at a time".
- If we relax such constraint we get more general substructures [Kashima and Koyanagi, 2002]



Weighting Problems



- Both matched pairs give the same contribution.
- Gap based weighting is needed.
- A novel efficient evaluation has to be defined

Partial Tree Kernel

- if the node labels of n_1 and n_2 are different then $\Delta(n_1, n_2) = 0$;

- else

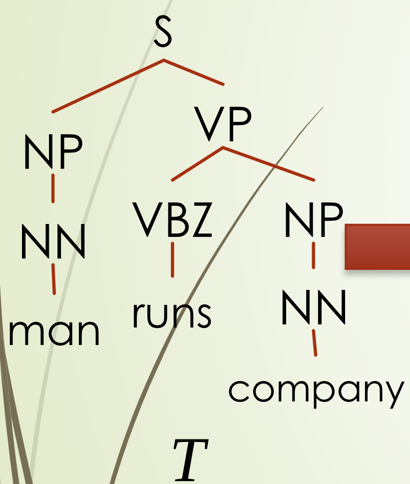
$$\Delta(n_1, n_2) = 1 + \sum_{\vec{J}_1, \vec{J}_2, l(\vec{J}_1)=l(\vec{J}_2)} \prod_{i=1}^{l(\vec{J}_1)} \Delta(c_{n_1}[\vec{J}_{1i}], c_{n_2}[\vec{J}_{2i}])$$

- By adding two decay factors we obtain:

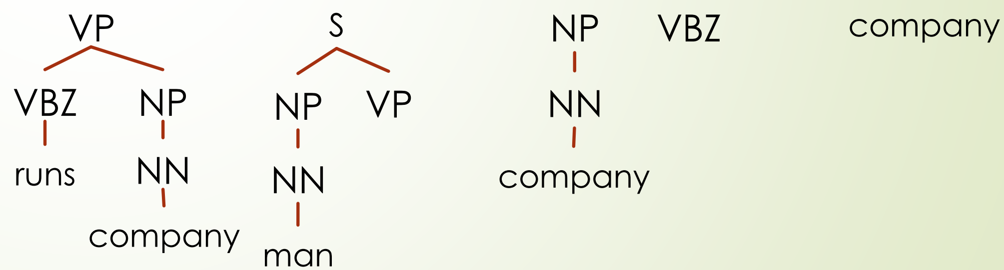
$$\mu\left(\lambda^2 + \sum_{\vec{J}_1, \vec{J}_2, l(\vec{J}_1)=l(\vec{J}_2)} \lambda^{d(\vec{J}_1)+d(\vec{J}_2)} \prod_{i=1}^{l(\vec{J}_1)} \Delta(c_{n_1}[\vec{J}_{1i}], c_{n_2}[\vec{J}_{2i}])\right)$$

Applying DCS to complex syntactic structures

- Tree Kernels [Collins and Duffy, 2003] account for structural analogies between syntactic parse trees

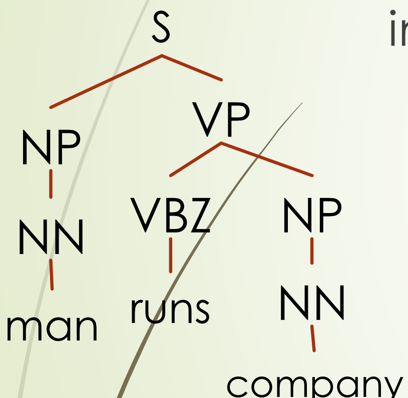


$$\phi(T) = (0,\ldots,1,\ldots,0,\ldots,1,\ldots,0,\ldots,1,\ldots,1,\ldots,0,\ldots,1)$$



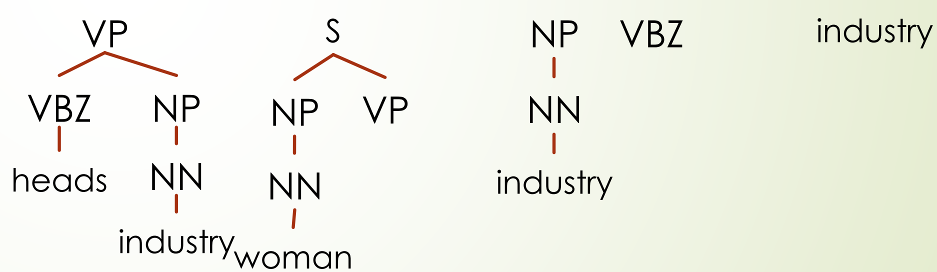
Applying DCS to complex syntactic structures

- Tree Kernels [Collins and Duffy, 2003] account for structural analogies between syntactic parse trees
- Smoothed Partial Tree Kernels (SPTKs) [Croce, 2011] introduce lexical semantic similarity within Tree Kernel

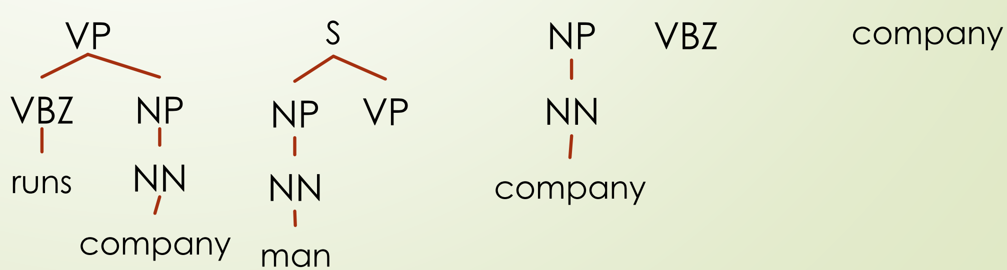


T

$$\phi(T_1) = (0,0,\ldots,1,\ldots,0,\ldots,1,\ldots,0,\ldots,1,\ldots,1,\ldots,0,\ldots,1)$$

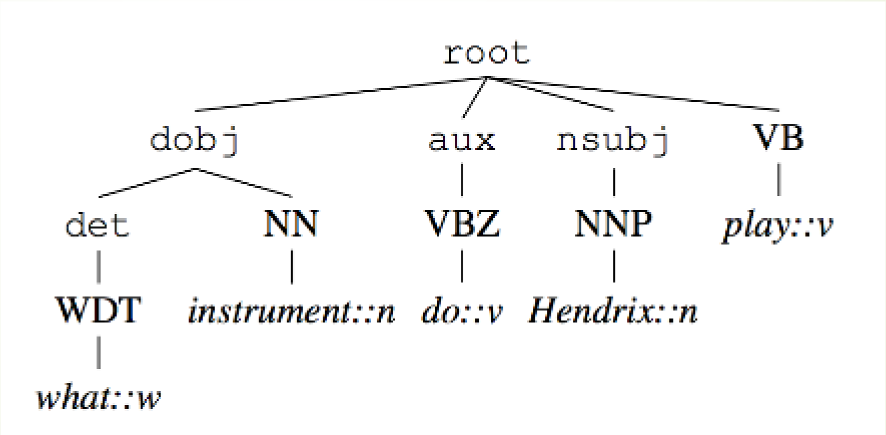


$$\phi(T_2) = (0,\ldots,1,\ldots,0,\ldots,1,\ldots,0,\ldots,1,\ldots,1,\ldots,0,\ldots,1)$$



Compositionally Smoothed Partial Tree Kernels

- Dependency trees include nodes expressing
  - Lexical information (e.g. verbs and nouns)
  - Grammatical and morphosyntactic information
    - Dependency relations
    - POS tags



Grammatical Relation Centered Tree (GRCT)

SPTK: Formal definition

- Given two trees T1 and T2
  - If n1 and n2 are leaves then

$$\Delta_\sigma(n_1, n_2) = \mu\lambda\sigma_\tau(n_1, n_2)$$

  - else

$$\Delta_\sigma(n_1, n_2) = \mu\sigma_\tau(n_1, n_2) \times \Big(\lambda^2 + \sum_{\vec{I}_1, \vec{I}_2, l(\vec{I}_1)=l(\vec{I}_2)} \lambda^{d(\vec{I}_1)+d(\vec{I}_2)} \prod_{j=1}^{l(\vec{I}_1)} \Delta_\sigma(c_{n_1}(\vec{I}_{1j}), c_{n_2}(\vec{I}_{2j}))\Big)$$

$\sigma_\tau(n_1,n_2)$ is a similarity function among the tree nodes depending on their linguistic type τ

MAIN LIMITATION:

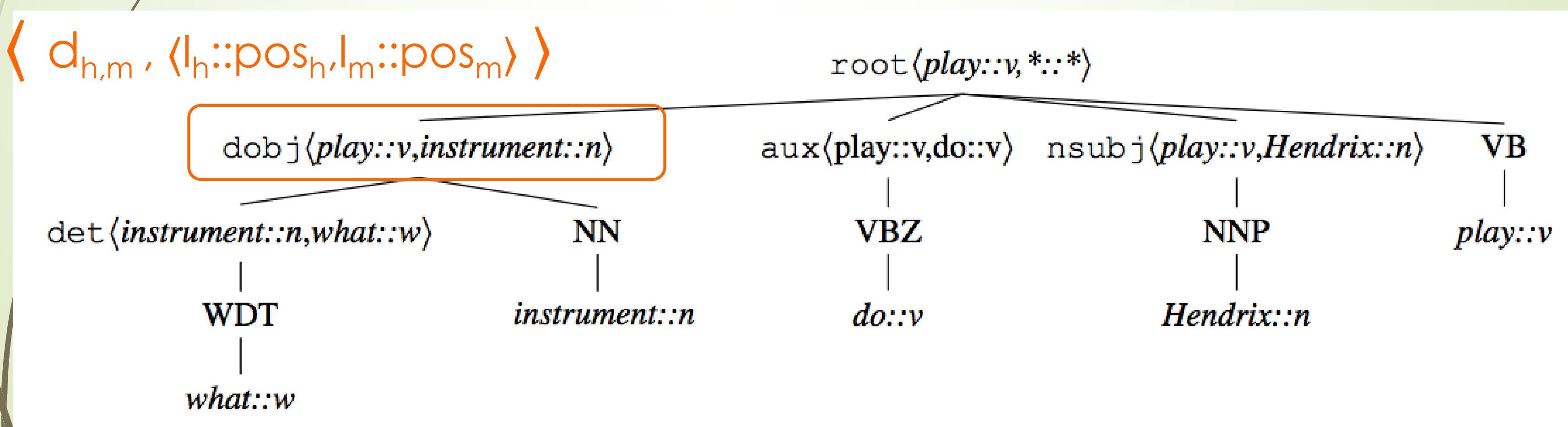
- *Again*, word similarity is still computed in **isolation**…
- How can we correctly handle a lexical node like *run* in all the **possible senses**?

Algorithm 1 $\sigma_\tau(n_1, n_2, lw)$

$\sigma_\tau \leftarrow 0$;
if $\tau(n_1) = \tau(n_2) = \text{SYNT} \wedge label(n_1) = label(n_2)$ **then**
$\quad\sigma_\tau \leftarrow 1$
end if
if $\tau(n_1) = \tau(n_2) = \text{POS} \wedge label(n_1) = label(n_2)$ **then**
$\quad\sigma_\tau \leftarrow 1$
end if
if $\tau(n_1) = \tau(n_2) = \text{LEX} \wedge pos(n_1) = pos(n_2)$ **then**
$\quad\sigma_\tau \leftarrow \sigma_{LEX}(n_1, n_2)$
end if
if leaf$(n_1) \wedge$ leaf(n_2) **then**
$\quad\sigma_\tau \leftarrow \sigma_\tau \times lw$
end if
return σ_τ

Compositionally Smoothed Partial Tree Kernels

- CSPTK is a novel kernel function that exploits Compositional Semantics within Tree Kernels
  - Compositionally labeled Tree: Compositional information over an entire parse tree is made explicit
  - Node similarity of the SPTK can be extended to host a DCS operator



Grammatical Relation Centered Tree (GRCT)

Compositionally labeled GRCT (CGRCT)

Similarity and DCS approaches

- Main idea: words in a composition influence each other's interpretation
- From individual concepts (word vectors) u and v, to the concept u·v for their appropriate composition, e.g.
  - Algebraic operators, e.g. sum, product or dilation [Mitchell & Lapata, 2008]
  - Regressor functions [Baroni, 2010], [Guevara, 2010] [Zanzotto et el, 2010]

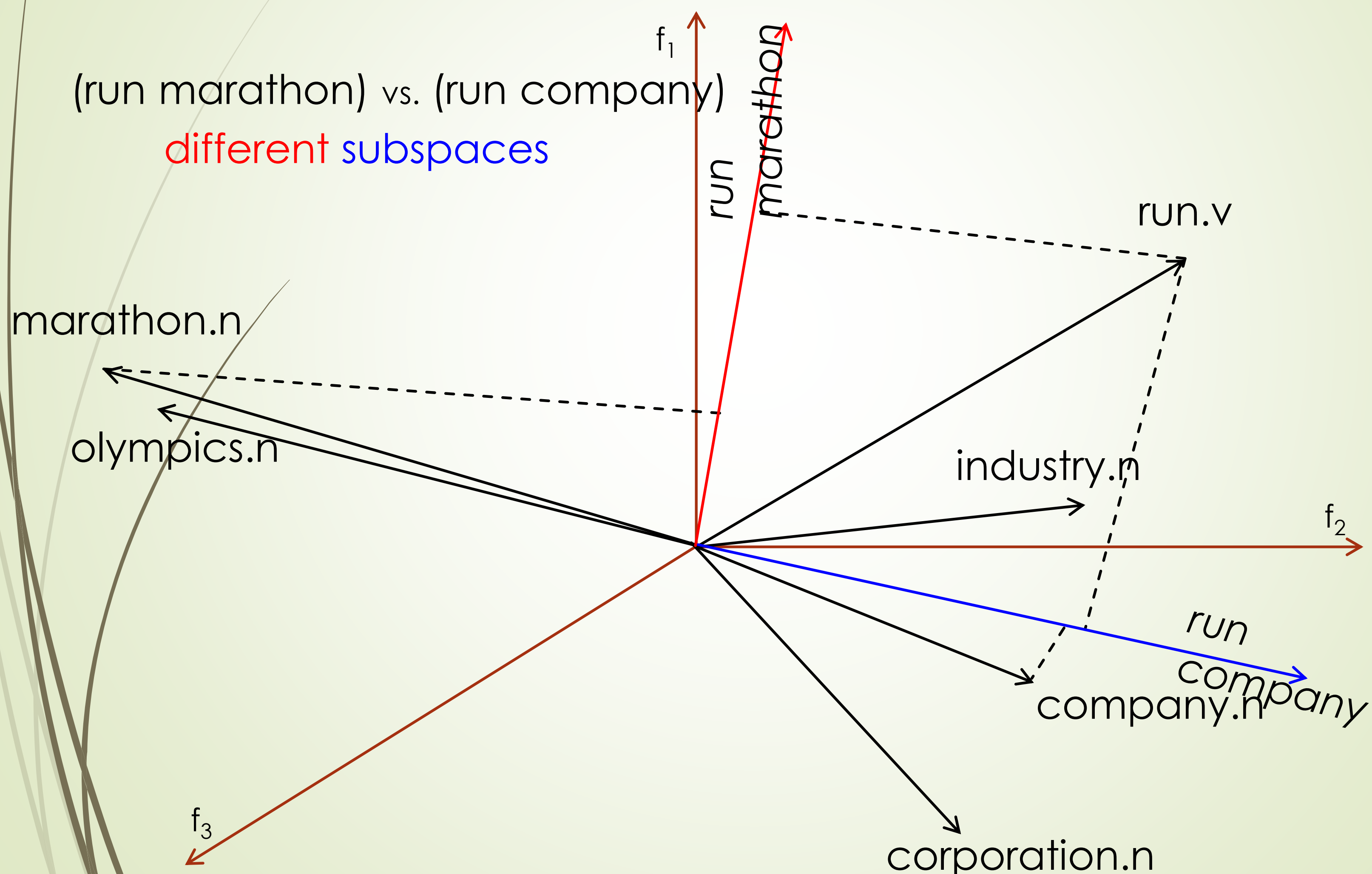
Similarity and DCS approaches (2)

- How to emphasize lexical composition through lexical vectors
- Intuition: word bi-grams can be represented into subspaces
  - By defining a projection function to identify common semantic features
  - Each subspace expresses properties shared by the specific sense of compounds
  - The resulting subspace is called Support Subspace [Annesi et al, 2012]

Support Subspaces: The underlying idea

(run marathon) vs. (run company)

different subspaces



Compositionality in Support Subspaces

Buy-Car	Buy-Time
cheap::Adj	*consume::V*
insurance::N	*enough::Adj*
rent::V	*waste::V*
lease::V	*save::In*
dealer::N	*permit::N*
motorcycle::N	*stressful::Adj*
hire::V	*spare::Adj*
auto::N	*save::V*
california::Adj	*warner::N*
tesco::N	*expensive::Adj*

Support Subspaces (Annesi et al, 2012)

- k-dimensional support subspaces for a pair *(h,m)*
  - the k indexes $I^k(h,m) = \{i_1, \ldots, i_k\}$ maximizing $\sum_{t=1}^{n} h_{i_t} \cdot m_{i_t}$
- **Projection matrix**

$$(M_{hm}^k)_{ij} = \begin{cases} 1 & iff \quad i = j \in I^k(h,m) \\ 0 & otherwise \end{cases}$$

- **Projected vectors**

$$\tilde{h} = M_{hm}^k \vec{h} \qquad \tilde{m} = M_{hm}^k \vec{m}$$

- A **compositional similarity between phrases** :

$$\sigma_{comp}\big((h,m),(h',m')\big) = \big[(M_1 \vec{h} \cdot M_2 \vec{h}') \circ (M_1 \vec{m} \cdot M_2 \vec{m}')\big]$$

CSPTK: Full definition

- Starting from SPTKs formulation
- New estimation of σ
  - The same for lexical nodes and pre-terminals
  - The DCS operator is introduced for **non-terminal nodes**

Compositional operator

Algorithm 1 $\sigma_\tau(n_x, n_y, lw)$ Compositional estimation of the lexical contribution to semantic tree kernel

Algorithm 1 $\sigma_\tau(n_1, n_2, lw)$

$\sigma_\tau \leftarrow 0,$
if $\tau(n_1) = \tau(n_2) = \text{SYNT} \wedge label(n_1) = label(n_2)$ **then**
  $\sigma_\tau \leftarrow 1$
end if
if $\tau(n_1) = \tau(n_2) = \text{POS} \wedge label(n_1) = label(n_2)$ **then**
  $\sigma_\tau \leftarrow 1$
end if
if $\tau(n_1) = \tau(n_2) = \text{LEX} \wedge pos(n_1) = pos(n_2)$ **then**
  $\sigma_\tau \leftarrow \sigma_{LEX}(n_1, n_2)$
end if
if leaf$(n_1) \wedge$ leaf(n_2) **then**
  $\sigma_\tau \leftarrow \sigma_\tau \times lw$
end if
return σ_τ

*/*Matching between compositional nodes: one modifier is missing*/*
if $li_x = \langle h_x{::}pos_h \rangle$ **and** $li_y = \langle h_y{::}pos_h, m_y{::}pos_m \rangle$ **then**
  $\sigma_\tau \leftarrow \sigma_{Comp}\big((h_x, h_x), (h_y, m_y)\big)$
end if
*/*Matching between compositional nodes: the general case*/*
if $li_x = \langle h_x{::}pos_h, m_x{::}pos_m \rangle$ **and** $li_y = \langle h_y{::}pos_h, m_y{::}pos_m \rangle$ **then**
  $\sigma_\tau \leftarrow \sigma_{Comp}\big((h_x, m_x), (h_y, m_y)\big)$
end if
end if
return σ_τ

CSPTK: Experimental evaluation

- Tasks (see CIKM 2014 paper):
  - Argument Classification in Semantic Role Labeling:
  - Question Classification (QC) in Question Answering
  - Paraphrase Identification
  - Metaphor Detection
- Set-up:
  - Co-occurrence Word Space, acquired through the distributional analysis of the UkWaC [Baroni et al,2009]
  - Representation of the examples derived by dependency parse trees
    - for CSPTK we use the compositionally labeled variant

SMPTK for Argument Classification

SRL at RTV: Smoothed Partial Tree Kernels

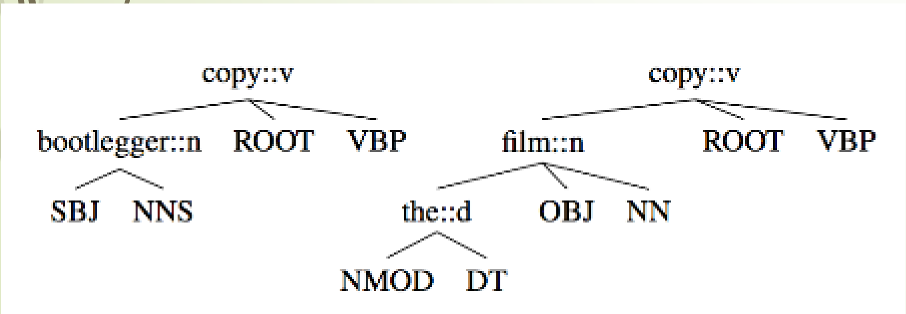
- Experimental Set-up (Croce et al., EMNLP 2011)

- FrameNet version: 1.3
- 271,560 training and 30,173 test examples respectively
- LTH dependency parser (Malt, Johansson & Nugues, 2007).
- Word space: LSA applied to the BNC corpus (about 10M words).

- Number of targeted frames: 648 frames
- Parse trees format: GRCT and LCT
- A total of 4,254 binary role classifiers (RC)

Argument Classification (Croce et al., 2013)

- UTV experimented with a FrameNet SRL classification (gold standard boundaries)
- We used the FrameNet version 1.3: 648 frames are considered
  - Training set: 271,560 arguments (90%)
  - Test set: 30,173 arguments (10%)
- [Bootleggers]$_{CREATOR}$, then copy [the film]$_{ORIGINAL}$ [onto hundreds of VHS tapes]$_{GOAL}$



Kernel	Accuracy
GRCT	87,60%
$GRCT_{LSA}$	88,61%
LCT	87,61%
LCT_{LSA}	88,74%
GRCT+LCT	87,99%
$GRCT_{LSA}+LCT_{LSA}$	**88,91%**

Question Classification: The task

- Reference corpus: UIUC dataset
  - Including
    - a training set of 5,452 questions and
    - a test set of 500 questions
  - Organized in six coarse-grained classes
    - ABBREVIATION abbreviation
    - ENTITY entities
    - DESCRIPTION description and abstract concepts
    - HUMAN human beings
    - LOCATION locations
    - NUMERIC numeric values

Examples

- `DESC:manner` How did serfdom develop in and then leave Russia ?
- `HUM:gr` What team did baseball 's St. Louis Browns become ?
- `ENTY:cremat` What films featured the character Popeye Doyle ?
- `DESC:manner` How can I find a list of celebrities ' real names ?
- `ENTY:animal` What fowl grabs the spotlight after the Chinese Year of the Monkey ?
- `ABBR:exp` What is the full form of .com ?
- `HUM:ind` What contemptible scoundrel stole the cork from my lunch ?

Question Classification: Results

Kernel	Accuracy	Std, Dev
BoW	86,3%	±0,3%
PTK_{LCT}	90,3%	±1,8%
$SPTK_{LCT}$	92,2%	±0,6%
$CSPTK^{+}_{CLCT}$	**95,6%**	**±0,6%**
$CSPTK^{cdot}_{CLCT}$	94,6%	±0,5%
$CSPTK^{d}_{CLCT}$	94,2%	±0,4%
$CSPTK^{ss}_{CLCT}$	93,3%	±0,7%
$CSPTK^{+}_{CGRCT}$	94,6%	±0,6%
$CSPTK^{cdot}_{CGRCT}$	94,1%	±0,6%
$CSPTK^{d}_{CGRCT}$	93,5%	±0,4%
$CSPTK^{SS}_{CGRCT}$	93,5%	±0,4%

Paraphrase Identification: The task

- Binary task: recognize if given a sentence pair, s1 and s2, they are in a paraphrase relation or not
  - MSRPC dataset: 5,801 sentence pairs.
- Given two sentence pairs (si1, si2) and (sj1, sj2), different kernels can be defined
  - We adopted a strategy similar to [Zanzotto&Moschitti, 2006] for Entailment
    - K1 = max{ k(si1,sj1)· k(si2,sj2, k(si1,sj2) · k(si2,sj1)}
    - K2 = k(si1, si2) · k(sj1, sj2)
    - **K = K1+ K2**

Paraphrase Identification: examples

Sentence1	Sentence2	is Paraphrase	Evaluation
Crews worked to install a new culvert and prepare the highway so motorists could use the east-bound lanes for travel as storm clouds threatened to dump more rain.	Crews worked to install a new culvert and repave the highway so motorists could use the east-bound lanes for travel.	false	true
Bethany Hamilton remained in stable condition Saturday after the attack Friday morning.	Bethany, who remained in stable condition after the attack Friday morning, talked of the attack Saturday.	false	true
Remaining shares will be held by QVC's management.	Members of the QVC management team hold the remaining shares.	true	false
Mr. Malik assured him that he would be considered a martyr if he did not return, the witness testified.	Mr. Malik assured him that he would be considered a martyr if anything happened to him as a result of his trip, the witness said.	true	false

Paraphrase Identification: Results

Kernel	Accuracy
baseline [Mihalcea et al, 2006]	65,40%
[Blacoe & Lapata, 2012]	73,00%
[Finch et al.,2005]	75,00%
[Srivastava et al., 2013]	72,00%
PTK_{LCT}	69,52%
$SPTK_{LCT}$	71,44%
$CSPTK^{+}_{CLCT}$	72,30%
$CSPTK^{+}_{CGRCT}$	72,20%
BoWK + PTK_{LCT}	74,96%
BoWK + $SPTK_{LCT}$	74,85%
BoWK + $CSPTK^{+}_{CLCT}$	**75,30%**

Metaphor Detection

- Task introduced in (Hovy and Shrivastava, 2013), http://www.edvisees.cs.cmu.edu/metaphordata.tar.gz
- The problem:
  - `yes 8 Stocks of California-based thrifts also were hard hit`

    implies

    "*hard hit*" corresponds to a metaphorical usage
- Previous work has applied
  - Walk-based kernels (Hovy et Srivastava, 2013)
- Experimental set-up:
  - 3872 sentences manually annotated
  - Manual splitting into training, dev, and test sets, using a 80-10-10 proportion

Metaphor Detection task

Kernel	Accuracy
Interannotator Agreement	57,0%
BoW	71,3
PTK_{LCT}	71,6%
$SPTK_{LCT}$	71,0%
$CSPTK^{+}_{CLCT}$	72,40%
$CSPTK^{ss}_{CLCT}$	**75,30%**
$CSPTK^{+}_{CGRCT}$	73,70%
$CSPTK^{ss}_{CGRCT}$	**74,50%**
[Hovy et al., 2013]	75,00%
[Srivastava et al., 2013]	**76,00%**

Conclusions

- Kernels allows to trigger a variety of very effective ML algorithms with a clear separation between the **induction** and **representation**
- They provide an expressive formalism for the optimization of NL semantics
  - ***Features as substructures***
  - *Complex convolutions are possible*
  - Optimization means maximization of linguistic resemblance (at different levels)
- Kernels can be combined to design very complex feature spaces
- Data-driven metrics are obtained by combining unsupervised feature modeling with supervised learning

Conclusions: advanced kernels & compositionality

- A Compositionality model (CSPTK) has been presented
- It combines the robustness of distributional models of the lexicons with grammatical information provided by the underlying tree kernel
- In this way *the full potential of unification-based formalisms (see AVG structures of LFGs) can be preserved*
- Advantages for a semantic task
  - Selective *sampling*: Automatic selection of suitable examples (i.e. the support vectors)
  - **Native Feature weighting according to the task**
  - **Efficient inference**

Conclusions: applications & perspectives

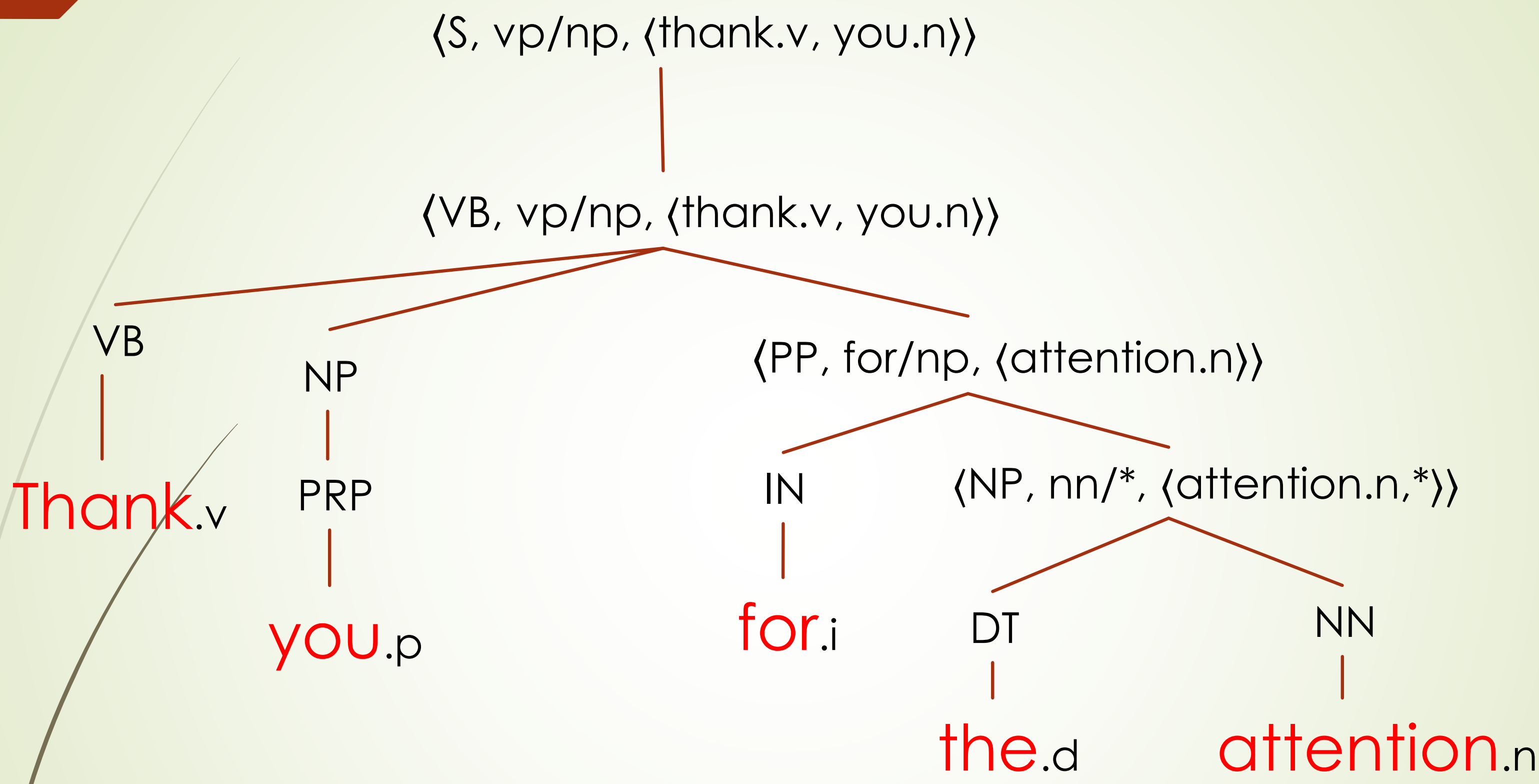
- Most applications (ranging from text classification, QA, parapharsing or sentiment analysis), benefit by the adoption of **CSPTK kernels**
- **No *ad-hoc* feature engineering is strictly required** thus improving
  - Design complexity
  - Data and Model Management
  - Time to market of applications
- Current work:
  - Extensive integration of neural word embedding information
  - Optimization of the tagging algorithm (see ECIR 2016 paper on *Nystrom linearization*)
  - *Adaptive on-line learning* in robotics (IJCAI 2016, accepted)

References

- Marco Pennacchiotti, Diego De Cao, Roberto Basili, Danilo Croce, Michael Roth, Automatic induction of FrameNet lexical units. EMNLP 2008: 457-465
- Alessandro Moschitti, Daniele Pighin, Roberto Basili, Tree Kernels for Semantic Role Labeling. Computational Linguistics 34(2): 193-224 (2008)
- Danilo Croce, Alessandro Moschitti, Roberto Basili, Structured Lexical Similarity via Convolution Kernels on Dependency Trees. EMNLP 2011: 1034-1046
- Danilo Croce, Alessandro Moschitti, Roberto Basili, Martha Palmer, Verb Classification using Distributional Similarity in Syntactic and Semantic Structures. ACL (1) 2012: 263-272
- Paolo Annesi, Valerio Storch, Roberto Basili, Space Projections as Distributional Models for Semantic Composition. CICLing (1) 2012: 323-335
- Danilo Croce, Simone Filice, Roberto Basili, Distributional Models and Lexical Semantics in Convolution Kernels. CICLing (1) 2012: 336-348
- Paolo Annesi, Danilo Croce, Roberto Basili, Semantic Compositionality in Tree Kernels. CIKM 2014: 1029-1038

Further Topics

- **Optimized kernel-based Learning**
  - Simone Filice, Danilo Croce, Roberto Basili, "*A Stratified Strategy for Efficient Kernel-Based Learning*". AAAI 2015: 2239-2245, 2015. On-Line & stratified Learning: AAAI2015
  - Danilo Croce, Roberto Basili "*Large-Scale Kernel-Based Language Learning Through the Ensemble Nystrom Methods*". ECIR 2016: 100-112, 2016.
- **Interactive Robotics**
  - Emanuele Bastianelli, Giuseppe Castellucci, Danilo Croce, Roberto Basili, Daniele Nardi, "*Effective and Robust Natural Language Understanding for Human-Robot Interaction*", Proc. of ECAI 2014, pp. 57-62, 18-22 August 2014, Prague, Czech Republic, 2014.
  - Emanuele Bastianelli, Danilo Croce, Roberto Basili, Daniele Nardi, "*Using semantic maps for robust natural language interaction with robots*", Proceedings of INTERSPEECH 2015, pp. 1393-1397, Dresden, Germany, September 6-10, 2015.
  - Emanuele Bastianelli, Danilo Croce, Andrea Vanzo, Roberto Basili, Daniele Nardi, A Discriminative Approach to Grounded Natural Language Learning in Interactive Robotics (accepted paper at IJCAI 2016)



⟨S, vp/np, ⟨thank.v, you.n⟩⟩
⟨VB, vp/np, ⟨thank.v, you.n⟩⟩
VB
Thank.v
NP
PRP
you.p
⟨PP, for/np, ⟨attention.n⟩⟩
IN
for.i
⟨NP, nn/*, ⟨attention.n,*⟩⟩
DT
the.d
NN
attention.n