Learning and Interpreting STS with Structural Kernels

Alessandro Moschitti
Department of Information Engineering and Computer Science
University of Trento
Email: moschitti@disi.unitn.it

STS 13, 2012
CCLS, NY

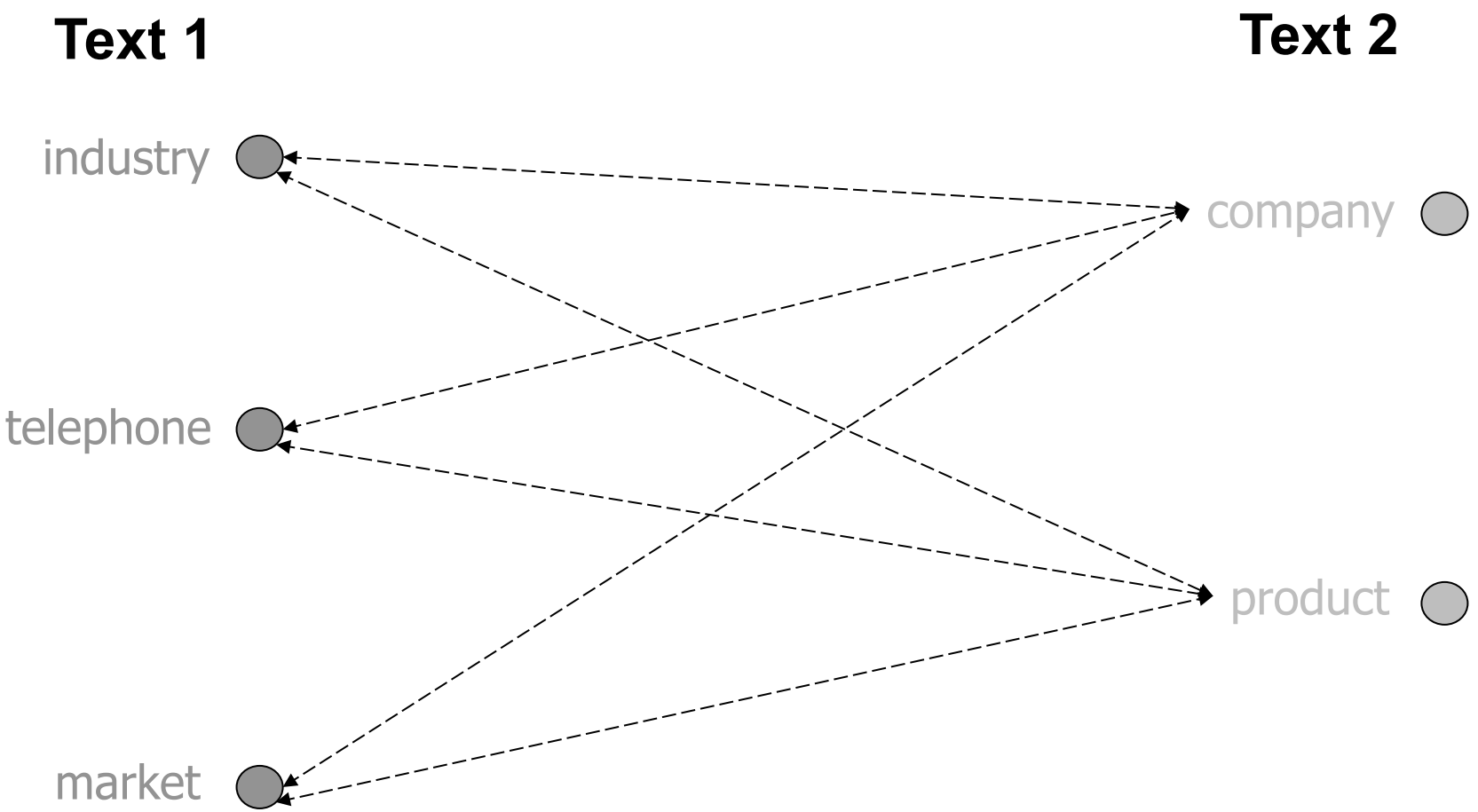
Motivations

- Learning STS automatically from sentence pairs
- Supervised Methods → Training Data
- Which features?
- What generalization?
- Which structures?
- What combination?
- Kernels can give a big help

Role of Kernels

- They can provide lexical similarities
- They can provide structural similarity
- They can also provide combined similarities
- Are they the similarity we want?
- **No!**
- They provide high level representation
- They are a big help to learn automatically sentence similarity that we want

Text Similarity



Lexical Semantic Kernel [CoNLL 2005]

- The text similarity is the K function:

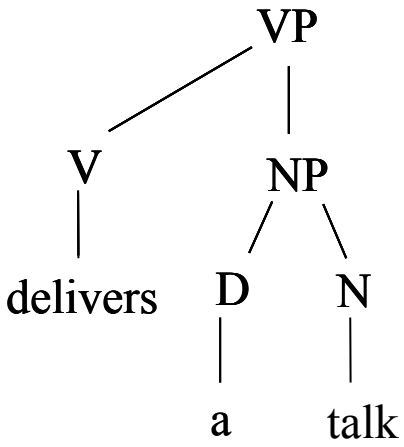
$$K(d_1, d_2) = \sum_{w_1 \in d_1, w_2 \in d_2} s(w_1, w_2)$$

- where *s* is any similarity function between words, e.g. WordNet [Basili et al.,2005] similarity or LSA [Cristianini et al., 2002]
- Good results when training data is small

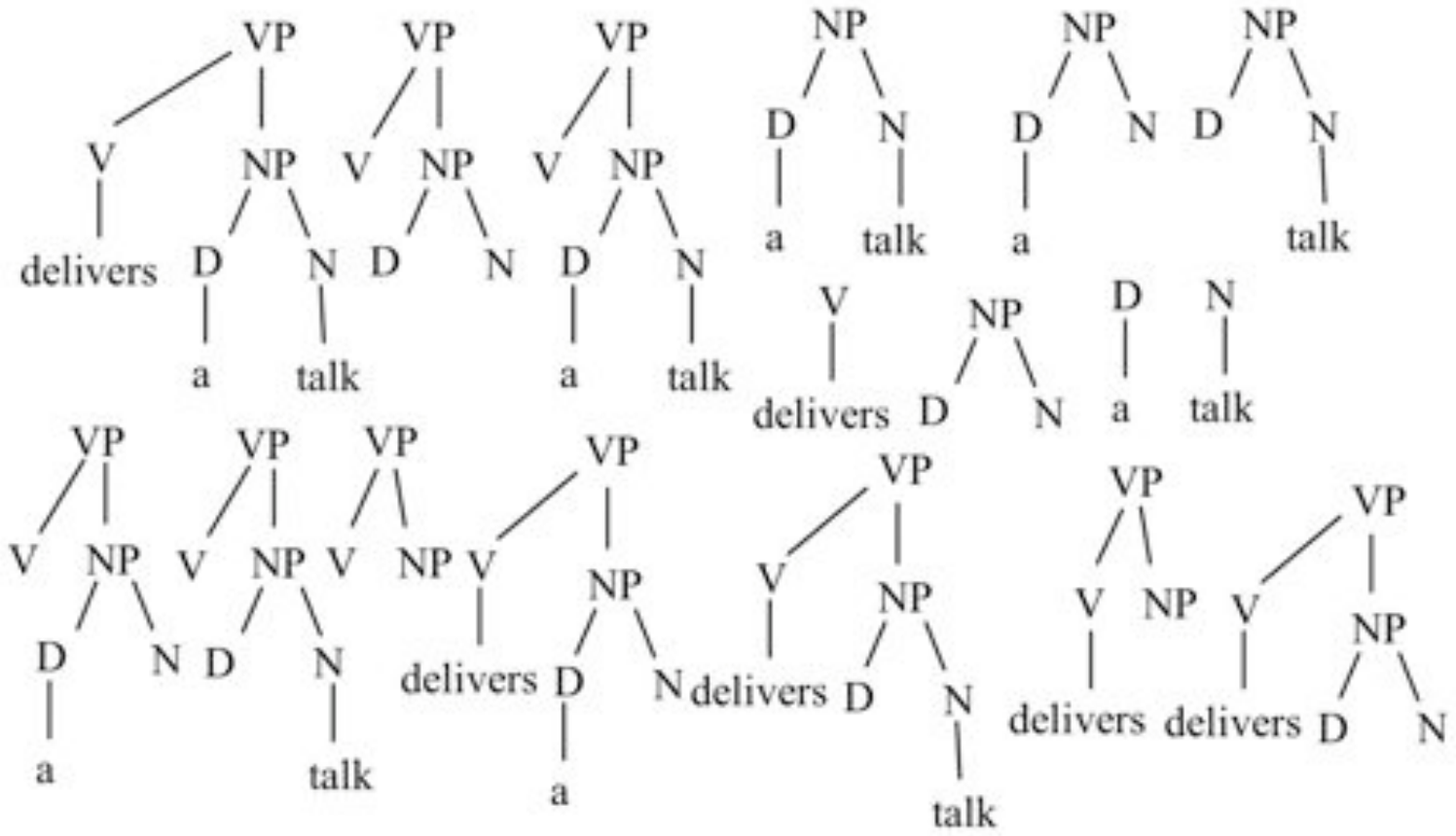
Sequence Similarity: Sequence Kernel

- *I am going to give a talk about structural kernels*
- *I give a talk on kernel methods*
- SK matches many subsequences:
- ***I give a talk kernels***, ***I talk kernels***, ***give kernels*** and all possible skip-grams

The Syntactic Tree Kernel (STK)
[Collins and Duffy, 2002]

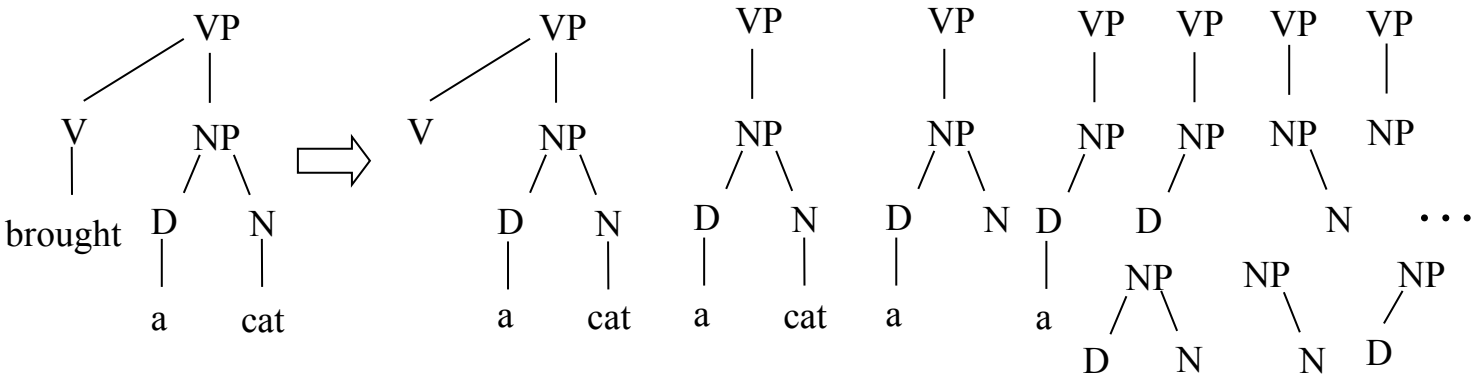


The overall fragment set

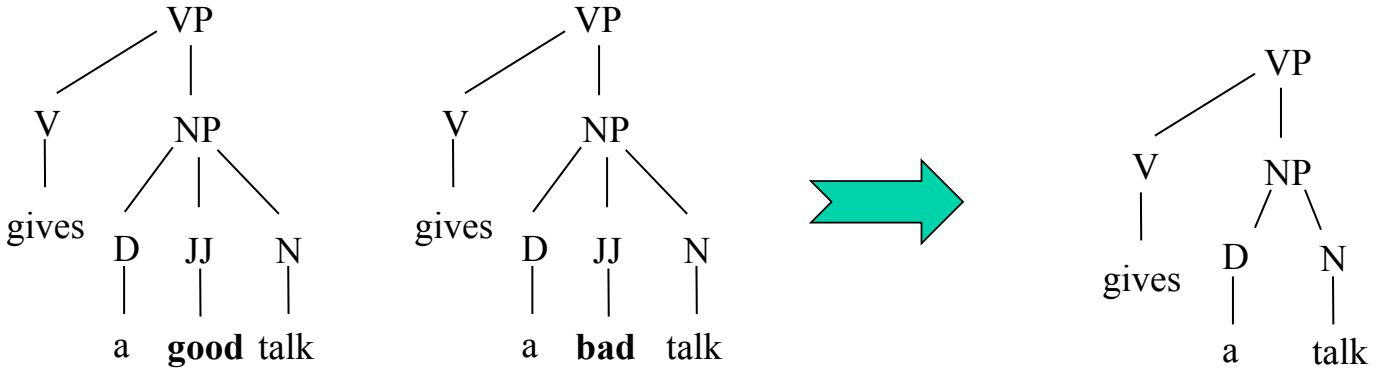


Partial Trees, [Moschitti, ECML 2006]

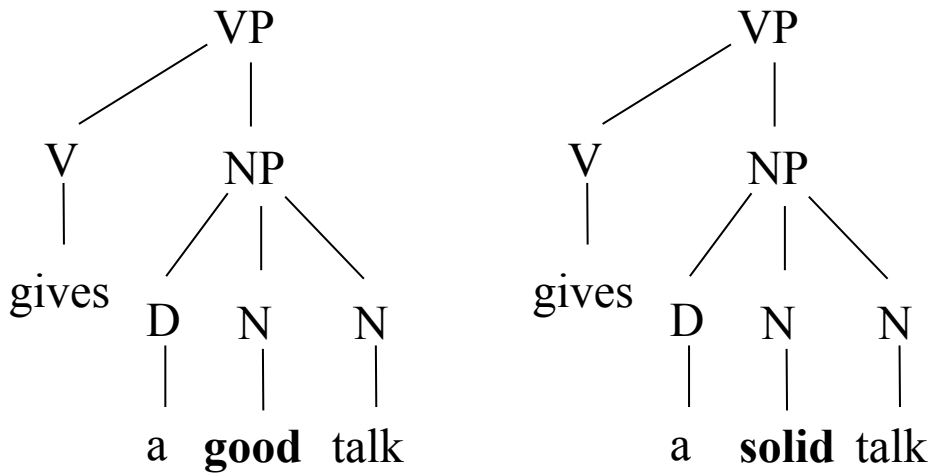
- STK + String Kernel with weighted gaps on Nodes' children



More and larger matches



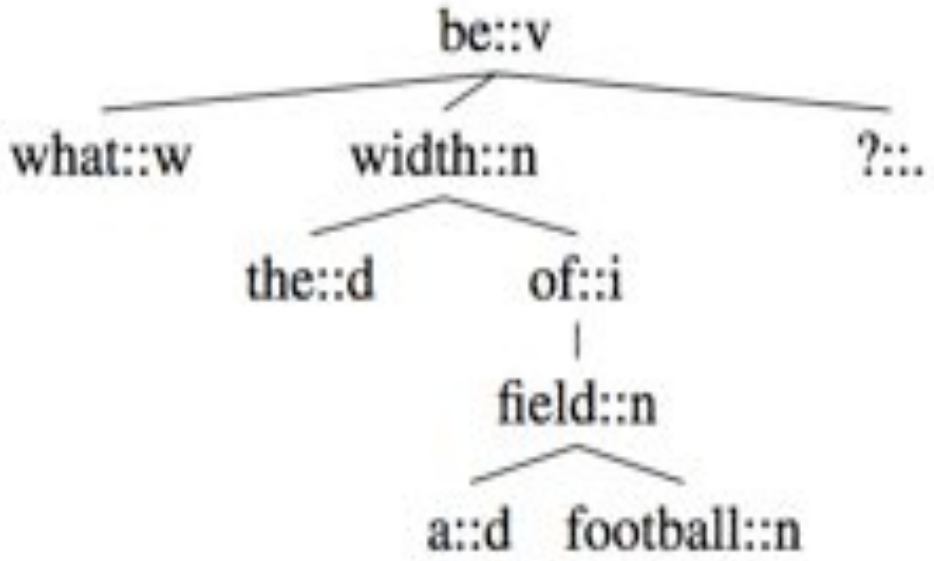
Syntactic/Semantic Tree Kernels
[Bloehdorn & Moschitti, ECIR 2007 & CIKM 2007]



- Similarity between the fragment leaves
  - Tree kernels + Lexical Similarity Kernel

Similarity on Dependency Trees

- What is the width of a football field?

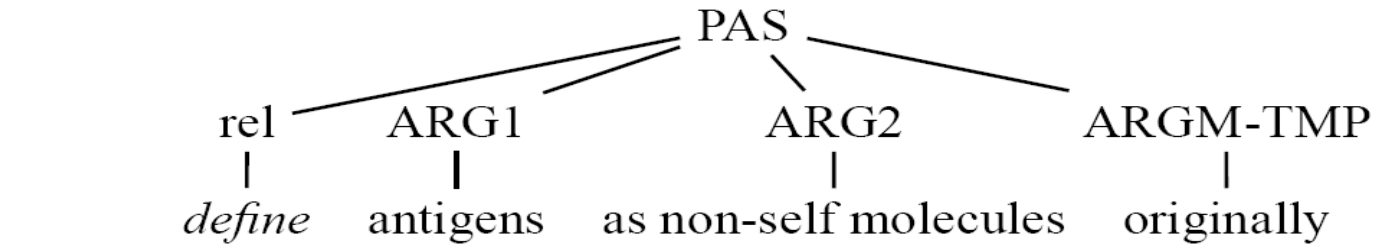


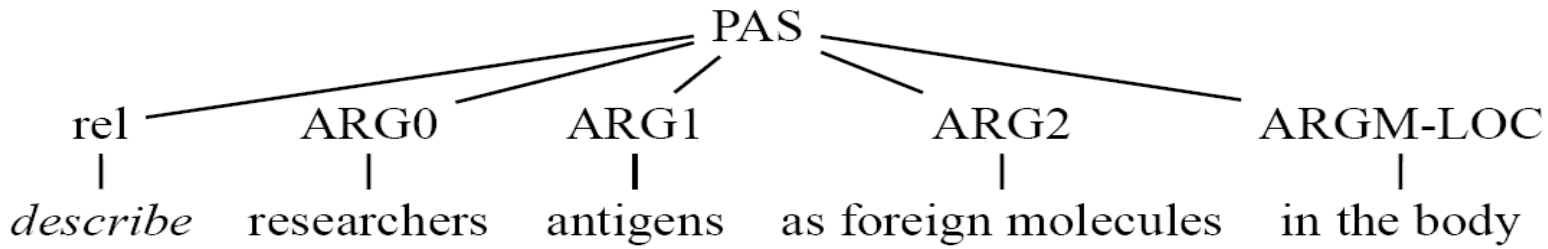
- Lexical similarity applied to any node of any substructure

- Word+generalized POS-tag

Predicate Argument Structure Similarity

- [**ARG1** Antigens] were [**AM–TMP** originally] [**rel** defined] [**ARG2** as non-self molecules].
- [**ARG0** Researchers] [**rel** describe] [**ARG1** antigens][**ARG2** as foreign molecules] [**ARGM–LOC** in the body]

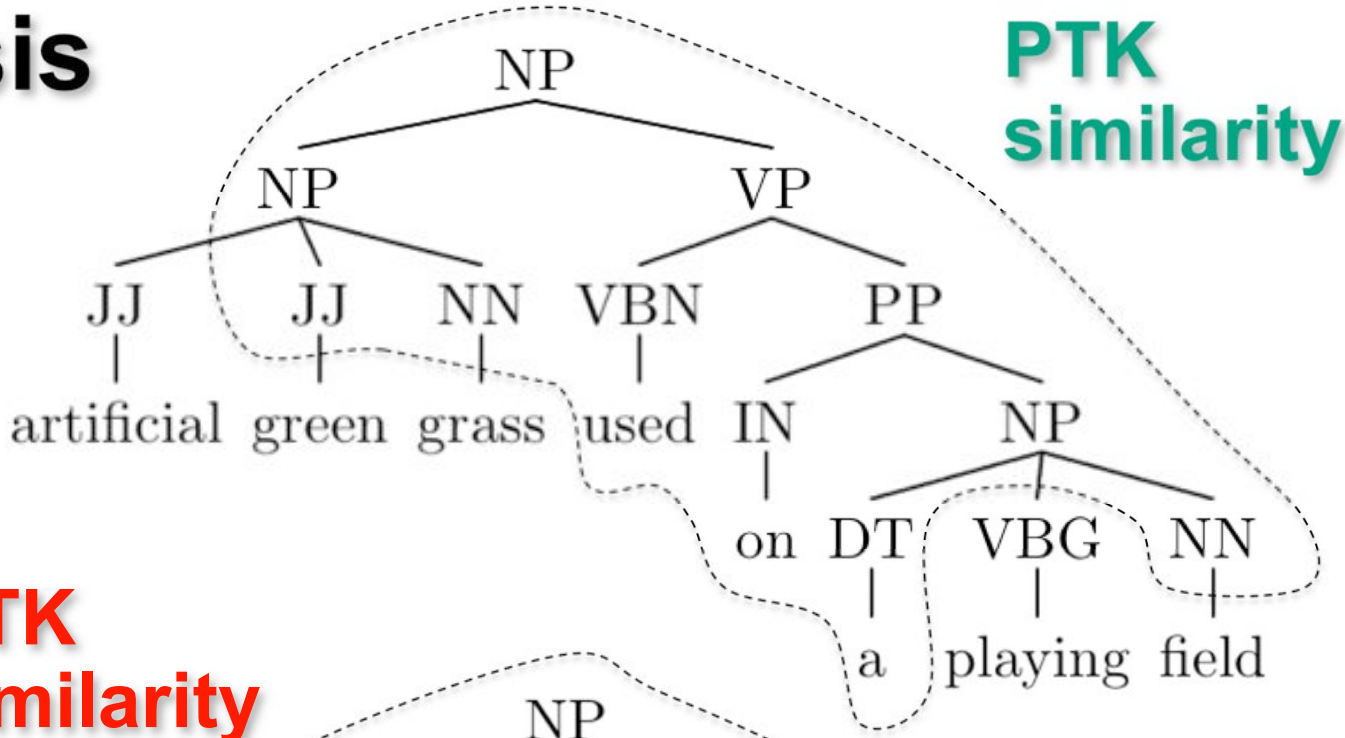




Error Analysis

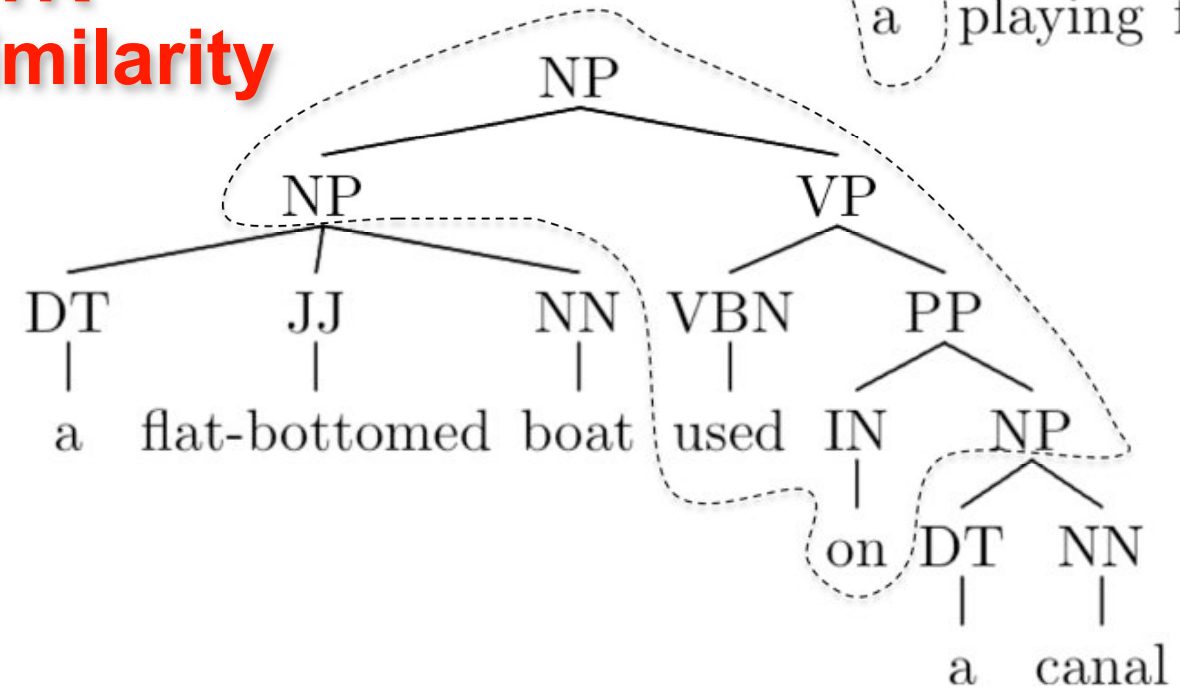
Test Example

- PTK ok
- STK not ok



STK similarity

Training Example

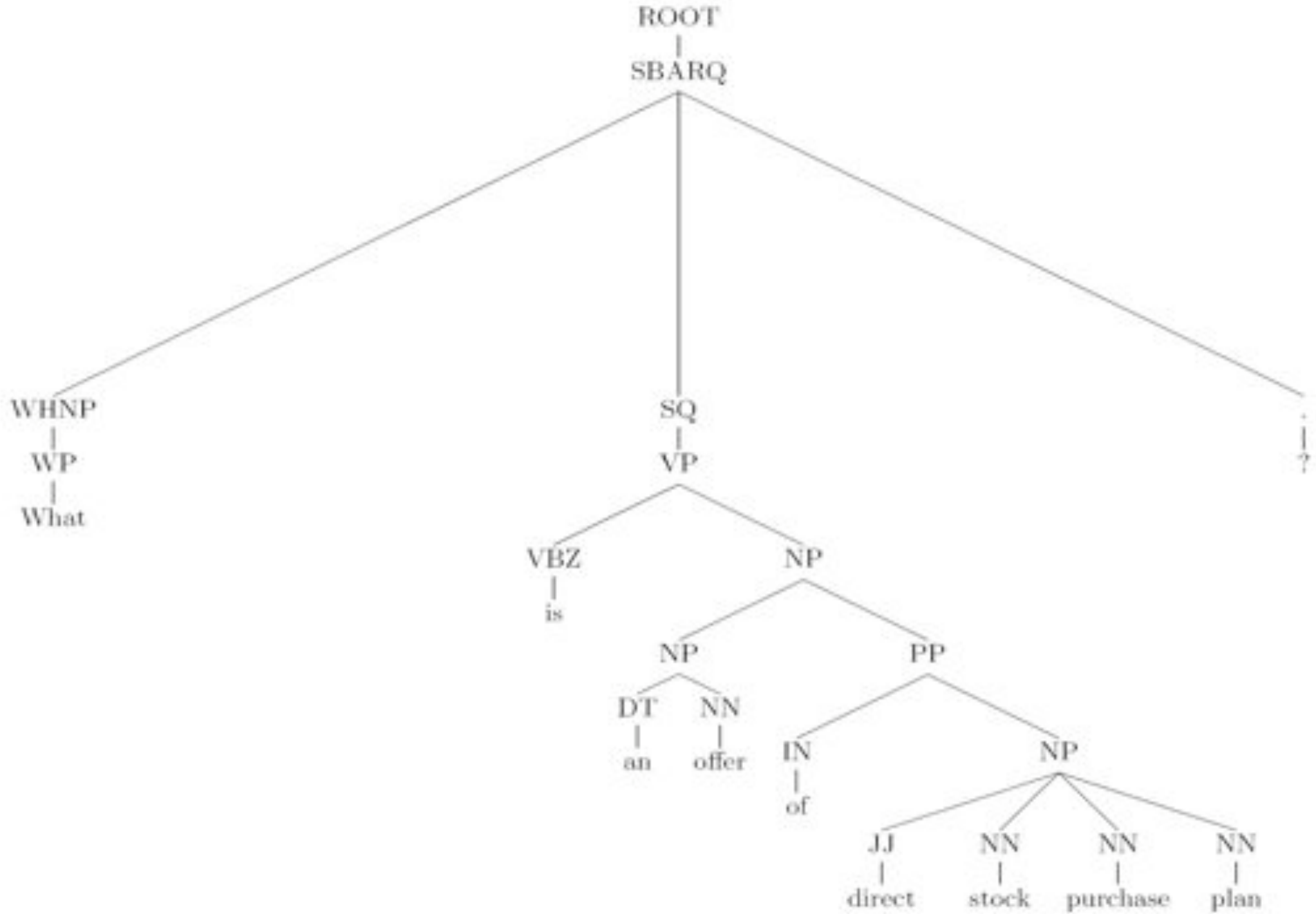


Objection: SVMs and Kernels are a black box

- SVMs provide models
  - Weight for each feature
  - We can watch the best features
- Not much meaningful, e.g., lexical features or string in isolation
- Do kernels make it worse?
- We can reverse engineering structural kernels!

Question Classification

- **Definition**: What does HTML stand for?
- **Description**: What's the final line in the Edgar Allan Poe poem "The Raven"?
- **Entity**: What foods can cause allergic reaction in people?
- **Human**: Who won the Nobel Peace Prize in 1992?
- **Location**: Where is the Statue of Liberty?
- **Manner**: How did Bob Marley die?
- **Numeric**: When was Martin Luther King Jr. born?
- **Organization**: What company makes Bentley cars?


ROOT
SBARQ
WHNP
WP
What
SQ
VP
VBZ
is
NP
NP
DT
an
NN
offer
PP
IN
of
NP
JJ
direct
NN
stock
NN
purchase
NN
plan
.
?

Interpretation (Abbreviation Class)

(NN(abbreviation))
(NP(DT)(NN(abbreviation)))
(NP(DT(the))(NN(abbreviation)))
(IN(for))
(VB(stand))
(VBZ(does))
(PP(IN))
(VP(VB(stand))(PP))
(NP(NP(DT)(NN(abbreviation)))(PP))
(SQ(VBZ)(NP)(VP(VB(stand))(PP)))
(SBARQ(WHNP)(SQ(VBZ)(NP)(VP(VB(stand))(PP)))(.))
(SQ(VBZ(does))(NP)(VP(VB(stand))(PP)))
(VP(VBZ)(NP(NP(DT)(NN(abbreviation)))(PP)))

Interpretation (Numeric Class)

(WRB(How))
(WHADVP(WRB(When)))
(WRB(When))
(JJ(many))
(NN(year))
(WHADJP(WRB)(JJ))
(NP(NN(year)))
(WHADJP(WRB(How))(JJ))
(NN(date))
(SBARQ(WHADVP(WRB(When)))(SQ)(.(?)))
(SBARQ(WHADVP(WRB(When)))(SQ)(.))
(NN(day))

Interpretation (Description Class)

(WRB(Why))

(WHADVP(WRB(Why)))

(WHADVP(WRB(How)))

(WHADVP(WRB))

(VB(mean))

(VBZ(causes))

(VB(do))

(SBARQ(WHADVP(WRB(How)))(SQ))

(WRB(How))

(SBARQ(WHADVP(WRB(How)))(SQ)(.))

(SBARQ(WHADVP(WRB(How)))(SQ)(.(?)))

Boundary Detection in SRL

(ADJP(RB-B)(VBN-P))
(NP(VBN-P)(NNS-B))
(S(NP-B)(VP))
(VP(VBD-P(said))(SBAR))
(VP(VB-P)(NP-B))
(NP(VBG-P)(NNS-B))
(VP(VBD-P)(NP-B))
(VP(VBG-P)(NP-B))
(VP(VBZ-P)(NP-B))
(VP(VBN-P)(NP-B))
(VP(VBP-P)(NP-B))
(NP(NP-B)(VP))
(NP(VBG-P)(NN-B))
(S(S(VP(VBG-P)))(NP-B))

Table 3: Best fragments for SRL BC.

Verb Class Classification

VerbNet class 13.5.1
(VP(VB(target))(NP))
(VP(VBG(target))(NP))
(VP(VBD(target))(NP))
(VP(TO)(VP(VB(target))(NP)))
(S(NP-SBJ)(VP(VBP(target))(NP)))
VerbNet class 60
(VBN(target))
(VP(VBD(target))(S))
(VP(VBZ(target))(S))
(VBP(target))
(VP(VBD(target))(NP-1)(S(NP-SBJ)(VP)))

Conclusions

- Learning STS with
  - Similarity functions (Kernel Methods)
  - Structural syntactic/semantic similarity
- Interpret the results to refine the representation

Future (on going work)

- Modeling more than one sentence with deeper structures: shallow semantics and *discourse*
- The objective is more compact and accurate models applicable to whole paragraphs.
- Use of reverse kernel engineering to study linguistic phenomena:
  - [Pighin&Moschitti, CoNLL2009, EMNLP2009, CoNLL2010]
  - To mine the most relevant fragments according to SVMs gradient
  - To use the linear space

Thank you

References

- Alessandro Moschitti' handouts http://disi.unitn.eu/~moschitt/teaching.html
- Alessandro Moschitti and Silvia Quarteroni, *Linguistic Kernels for Answer Re-ranking in Question Answering Systems,* Information and Processing Management, ELSEVIER, 2010.
- Yashar Mehdad, Alessandro Moschitti and Fabio Massimo Zanzotto. *Syntactic/ Semantic Structures for Textual Entailment Recognition*. Human Language Technology - North American chapter of the Association for Computational Linguistics (HLT-NAACL), 2010, Los Angeles, Calfornia.
- Daniele Pighin and Alessandro Moschitti. *On Reverse Feature Engineering of Syntactic Tree Kernels*. In Proceedings of the 2010 Conference on Natural Language Learning, Upsala, Sweden, July 2010. Association for Computational Linguistics.
- Thi Truc Vien Nguyen, Alessandro Moschitti and Giuseppe Riccardi. *Kernel-based Reranking for Entity Extraction.* In proceedings of the 23rd International Conference on Computational Linguistics (COLING), August 2010, Beijing, China.

References

- M. Dinarelli, A. Moschitti, and G. Riccardi. *Discriminative Reranking for Spoken Language Understanding*. IEEE Transaction on Audio, Speech and Language Processing, to appear in 2011.10.1109/TASL.2010 .2093520.
- Danilo Croce, Alessandro Moschitti, and Roberto Basili. *Structured lexical similarity via convolution kernels on dependency trees*. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pages 1034–1046, Edinburgh, Scotland, UK., July 2011. Association for Computational Linguistics.
- Aliaksei Severyn and Alessandro Moschitti. *Fast support vector machines for structural kernels*. In ECML-PKDD, 2011, Greece, 2011.
- Alessandro Moschitti, Jennifer Chu-carroll, Siddharth Patwardhan, James Fan, and Giuseppe Riccardi. *Using syntactic and semantic structural kernels for classifying definition questions in jeopardy!* In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pages 712–724, Edinburgh, Scotland, UK., July 2011. Association for Computational Linguistics.

References

- Alessandro Moschitti. *Syntactic and semantic kernels for short text pair categorization*. In Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009), pages 576–584, Athens, Greece, March 2009.
- Truc-Vien Nguyen, Alessandro Moschitti, and Giuseppe Riccardi. *Convolution kernels on constituent, dependency and sequential structures for relation extraction*. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pages 1378–1387, Singapore, August 2009.
- Marco Dinarelli, Alessandro Moschitti, and Giuseppe Riccardi. *Re-ranking models based-on small training data for spoken language understanding*. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pages 1076–1085, Singapore, August 2009.
- Alessandra Giordani and Alessandro Moschitti. *Syntactic Structural Kernels for Natural Language Interfaces to Databases*. In ECML/PKDD, pages 391–406, Bled, Slovenia, 2009.

References

- Alessandro Moschitti, Daniele Pighin and Roberto Basili. *Tree Kernels for Semantic Role Labeling,* Special Issue on Semantic Role Labeling, Computational Linguistics Journal. March 2008.
- Fabio Massimo Zanzotto, Marco Pennacchiotti and Alessandro Moschitti*, A Machine Learning Approach to Textual Entailment Recognition,* Special Issue on Textual Entailment Recognition, Natural Language Engineering, Cambridge University Press., 2008
- Mona Diab, Alessandro Moschitti, Daniele Pighin, *Semantic Role Labeling Systems for Arabic Language using Kernel Methods*. In proceedings of the 46th Conference of the Association for Computational Linguistics (ACL'08). Main Paper Section. Columbus, OH, USA, June 2008.
- Alessandro Moschitti, Silvia Quarteroni, Kernels on Linguistic Structures for Answer Extraction. In proceedings of the 46th Conference of the Association for Computational Linguistics (ACL'08). Short Paper Section. Columbus, OH, USA, June 2008.

References

- Yannick Versley, Simone Ponzetto, Massimo Poesio, Vladimir Eidelman, Alan Jern, Jason Smith, Xiaofeng Yang and Alessandro Moschitti, *BART: A Modular Toolkit for Coreference Resolution*, In Proceedings of the Conference on Language Resources and Evaluation, Marrakech, Marocco, 2008.
- Alessandro Moschitti, *Kernel Methods, Syntax and Semantics for Relational Text Categorization*. In proceeding of ACM 17th Conference on Information and Knowledge Management (CIKM). Napa Valley, California, 2008.
- Bonaventura Coppola, Alessandro Moschitti, and Giuseppe Riccardi. *Shallow semantic parsing for spoken language understanding*. In Proceedings of HLT-NAACL Short Papers, pages 85–88, Boulder, Colorado, June 2009. Association for Computational Linguistics.
- Alessandro Moschitti and Fabio Massimo Zanzotto, *Fast and Effective Kernels for Relational Learning from Texts*, Proceedings of The 24th Annual International Conference on Machine Learning (ICML 2007).

References

- Alessandro Moschitti, Silvia Quarteroni, Roberto Basili and Suresh Manandhar, *Exploiting Syntactic and Shallow Semantic Kernels for Question/Answer Classification*, Proceedings of the 45th Conference of the Association for Computational Linguistics (ACL), Prague, June 2007.
- Alessandro Moschitti and Fabio Massimo Zanzotto, *Fast and Effective Kernels for Relational Learning from Texts*, Proceedings of The 24th Annual International Conference on Machine Learning (ICML 2007), Corvallis, OR, USA.
- Daniele Pighin, Alessandro Moschitti and Roberto Basili, *RTV: Tree Kernels for Thematic Role Classification*, Proceedings of the 4th International Workshop on Semantic Evaluation (SemEval-4), English Semantic Labeling, Prague, June 2007.
- Stephan Bloehdorn and Alessandro Moschitti, *Combined Syntactic and Semanitc Kernels for Text Classification*, to appear in the 29th European Conference on Information Retrieval (ECIR), April 2007, Rome, Italy.
- Fabio Aiolli, Giovanni Da San Martino, Alessandro Sperduti, and Alessandro Moschitti, *Efficient Kernel-based Learning for Trees*, to appear in the IEEE Symposium on Computational Intelligence and Data Mining (CIDM), Honolulu, Hawaii, 2007

References

- Alessandro Moschitti, Silvia Quarteroni, Roberto Basili and Suresh Manandhar, *Exploiting Syntactic and Shallow Semantic Kernels for Question/Answer Classification*, Proceedings of the 45th Conference of the Association for Computational Linguistics (ACL), Prague, June 2007.
- Alessandro Moschitti, Giuseppe Riccardi, Christian Raymond, *Spoken Language Understanding with Kernels for Syntactic/Semantic Structures*, Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU2007), Kyoto, Japan, December 2007
- Stephan Bloehdorn and Alessandro Moschitti, *Combined Syntactic and Semantic Kernels for Text Classification*, to appear in the 29th European Conference on Information Retrieval (ECIR), April 2007, Rome, Italy.
- Stephan Bloehdorn, Alessandro Moschitti: Structure and semantics for expressive text kernels. In proceeding of ACM 16th Conference on Information and Knowledge Management (CIKM-short paper) 2007: 861-864, Portugal.

References

- Fabio Aiolli, Giovanni Da San Martino, Alessandro Sperduti, and Alessandro Moschitti, *Efficient Kernel-based Learning for Trees*, to appear in the IEEE Symposium on Computational Intelligence and Data Mining (CIDM), Honolulu, Hawaii, 2007.
- Alessandro Moschitti*, Efficient Convolution Kernels for Dependency and Constituent Syntactic Trees.* In Proceedings of the 17th European Conference on Machine Learning, Berlin, Germany, 2006.
- Fabio Aiolli, Giovanni Da San Martino, Alessandro Sperduti, and Alessandro Moschitti, *Fast On-line Kernel Learning for Trees*, International Conference on Data Mining (ICDM) 2006 (short paper).
- Stephan Bloehdorn, Roberto Basili, Marco Cammisa, Alessandro Moschitti, *Semantic Kernels for Text Classification based on Topological Measures of Feature Similarity*. In Proceedings of the 6th IEEE International Conference on Data Mining (ICDM 06), Hong Kong, 18-22 December 2006. (short paper).

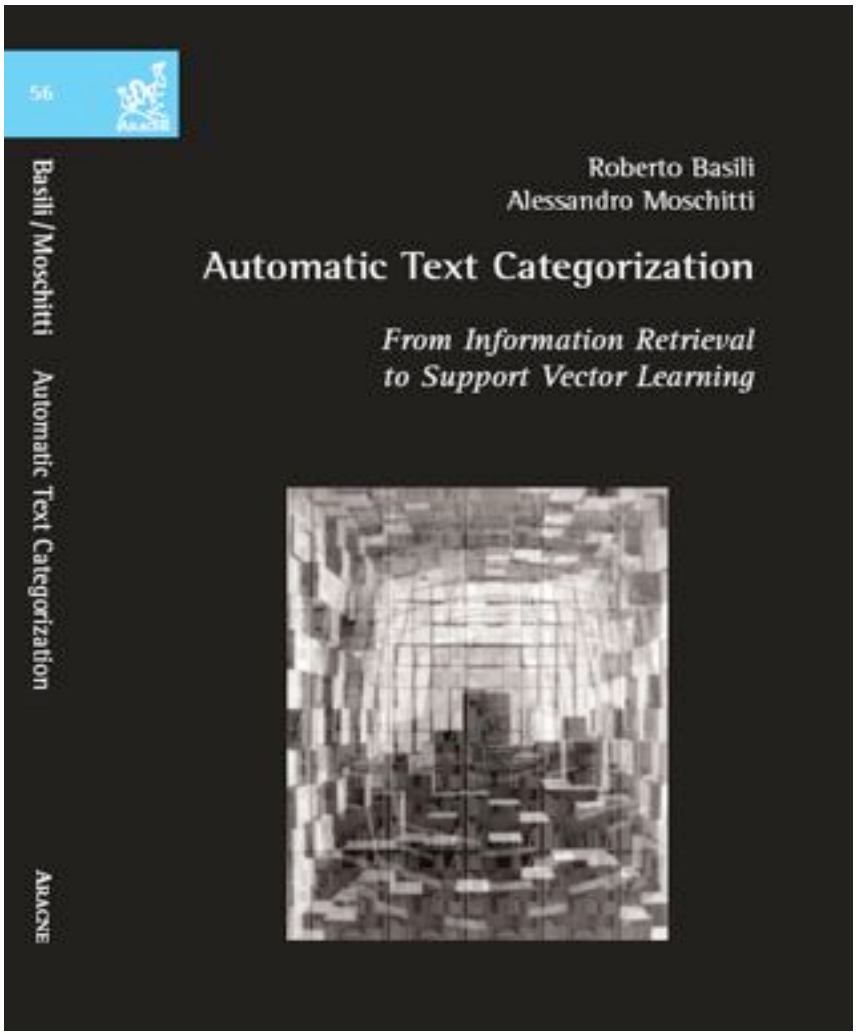
References

- Roberto Basili, Marco Cammisa and Alessandro Moschitti, *A Semantic Kernel to classify texts with very few training examples*, in Informatica, an international journal of Computing and Informatics, 2006.
- Fabio Massimo Zanzotto and Alessandro Moschitti, *Automatic learning of textual entailments with cross-pair similarities*. In Proceedings of COLING-ACL, Sydney, Australia, 2006.
- Ana-Maria Giuglea and Alessandro Moschitti*, Semantic Role Labeling via FrameNet, VerbNet and PropBank.* In Proceedings of COLING-ACL, Sydney, Australia, 2006.
- Alessandro Moschitti*, Making tree kernels practical for natural language learning.* In Proceedings of the Eleventh International Conference on European Association for Computational Linguistics, Trento, Italy, 2006.
- Alessandro Moschitti, Daniele Pighin and Roberto Basili. *Semantic Role Labeling via Tree Kernel joint inference*. In Proceedings of the 10th Conference on Computational Natural Language Learning, New York, USA, 2006.

References

- Roberto Basili, Marco Cammisa and Alessandro Moschitti, *Effective use of Wordnet semantics via kernel-based learning*. In Proceedings of the 9th Conference on Computational Natural Language Learning (CoNLL 2005), Ann Arbor (MI), USA, 2005
- Alessandro Moschitti, *A study on Convolution Kernel for Shallow Semantic Parsing*. In proceedings of the 42-th Conference on Association for Computational Linguistic (ACL-2004), Barcelona, Spain, 2004.
- Alessandro Moschitti and Cosmin Adrian Bejan, *A Semantic Kernel for Predicate Argument Classification.* In proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL-2004), Boston, MA, USA, 2004.

An introductory book on SVMs, Kernel methods and Text Categorization



Non-exhaustive reference list from other authors

- V. Vapnik. The Nature of Statistical Learning Theory. Springer, 1995.
- P. Bartlett and J. Shawe-Taylor, 1998. Advances in Kernel Methods - Support Vector Learning, chapter Generalization Performance of Support Vector Machines and other Pattern Classifiers. MIT Press.
- David Haussler. 1999. Convolution kernels on discrete structures. Technical report, Dept. of Computer Science, University of California at Santa Cruz.
- Lodhi, Huma, Craig Saunders, John Shawe Taylor, Nello Cristianini, and Chris Watkins. Text classification using string kernels. JMLR,2000
- Schölkopf, Bernhard and Alexander J. Smola. 2001. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, MA, USA.

Non-exhaustive reference list from other authors

- N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines (and other kernel-based learning methods)* Cambridge University Press, 2002
- M. Collins and N. Duffy, New ranking algorithms for parsing and tagging: Kernels over discrete structures, and the voted perceptron. In ACL02, 2002.
- Hisashi Kashima and Teruo Koyanagi. 2002. Kernels for semi-structured data. In Proceedings of ICML'02.
- S.V.N. Vishwanathan and A.J. Smola. Fast kernels on strings and trees. In Proceedings of NIPS, 2002.
- Nicola Cancedda, Eric Gaussier, Cyril Goutte, and Jean Michel Renders. 2003. Word sequence kernels. Journal of Machine Learning Research, 3:1059–1082. D. Zelenko, C. Aone, and A. Richardella. Kernel methods for relation extraction. JMLR, 3:1083–1106, 2003.

Non-exhaustive reference list from other authors

- Taku Kudo and Yuji Matsumoto. 2003. Fast methods for kernel-based text analysis. In Proceedings of ACL'03.
- Dell Zhang and Wee Sun Lee. 2003. Question classification using support vector machines. In Proceedings of SIGIR'03, pages 26–32.
- Libin Shen, Anoop Sarkar, and Aravind k. Joshi. Using LTAG Based Features in Parse Reranking. In Proceedings of EMNLP'03, 2003
- C. Cumby and D. Roth. Kernel Methods for Relational Learning. In Proceedings of ICML 2003, pages 107–114, Washington, DC, USA, 2003.
- J. Shawe-Taylor and N. Cristianini. Kernel Methods for Pattern Analysis. Cambridge University Press, 2004.
- A. Culotta and J. Sorensen. Dependency tree kernels for relation extraction. In Proceedings of the 42nd Annual Meeting on ACL, Barcelona, Spain, 2004.

Non-exhaustive reference list from other authors

- Kristina Toutanova, Penka Markova, and Christopher Manning. The Leaf Path Projection View of Parse Trees: Exploring String Kernels for HPSG Parse Selection. In Proceedings of EMNLP 2004.
- Jun Suzuki and Hideki Isozaki. 2005. Sequence and Tree Kernels with Statistical Feature Mining. In Proceedings of NIPS'05.
- Taku Kudo, Jun Suzuki, and Hideki Isozaki. 2005. Boosting based parse reranking with subtree features. In Proceedings of ACL'05.
- R. C. Bunescu and R. J. Mooney. Subsequence kernels for relation extraction. In Proceedings of NIPS, 2005.
- R. C. Bunescu and R. J. Mooney. A shortest path dependency kernel for relation extraction. In Proceedings of EMNLP, pages 724–731, 2005.
- S. Zhao and R. Grishman. Extracting relations with integrated information using kernel methods. In Proceedings of the 43rd Meeting of the ACL, pages 419–426, Ann Arbor, Michigan, USA, 2005.

Non-exhaustive reference list from other authors

- J. Kazama and K. Torisawa. Speeding up Training with Tree Kernels for Node Relation Labeling. In Proceedings of EMNLP 2005, pages 137–144, Toronto, Canada, 2005.
- M. Zhang, J. Zhang, J. Su, , and G. Zhou. A composite kernel to extract relations between entities with both flat and structured features. In Proceedings of COLING-ACL 2006, pages 825–832, 2006.
- M. Zhang, G. Zhou, and A. Aw. Exploring syntactic structured features over parse trees for relation extraction using kernel methods. Information Processing and Management, 44(2):825–832, 2006.
- G. Zhou, M. Zhang, D. Ji, and Q. Zhu. Tree kernel-based relation extraction with context-sensitive structured parse tree information. In Proceedings of EMNLP-CoNLL 2007, pages 728–736, 2007.

Non-exhaustive reference list from other authors

- Ivan Titov and James Henderson. Porting statistical parsers with data-defined kernels. In Proceedings of CoNLL-X, 2006
- Min Zhang, Jie Zhang, and Jian Su. 2006. Exploring Syntactic Features for Relation Extraction using a Convolution tree kernel. In Proceedings of NAACL.
- M. Wang. A re-examination of dependency path kernels for relation extraction. In Proceedings of the 3rd International Joint Conference on Natural Language Processing-IJCNLP, 2008.