Semantic Relations Between Nominals

Synthesis Lectures on Human Language Technologies

Linguistic Structure Prediction
Noah A. Smith
2011

Learning to Rank for Information Retrieval and Natural Language Processing
Hang Li
2011

Computational Modeling of Human Language Acquisition
Afra Alishahi
2010

Introduction to Arabic Natural Language Processing
Nizar Y. Habash
2010

Cross-Language Information Retrieval
Jian-Yun Nie
2010

Automated Grammatical Error Detection for Language Learners
Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel Tetreault
2010

Data-Intensive Text Processing with MapReduce
Jimmy Lin and Chris Dyer
2010

Semantic Role Labeling
Martha Palmer, Daniel Gildea, and Nianwen Xue
2010

Spoken Dialogue Systems
Kristiina Jokinen and Michael McTear
2009

Introduction to Chinese Natural Language Processing
Kam-Fai Wong, Wenjie Li, Ruifeng Xu, and Zheng-sheng Zhang
2009

Introduction to Linguistic Annotation and Text Analytics
Graham Wilcock
2009

Dependency Parsing
Sandra Kübler, Ryan McDonald, and Joakim Nivre
2009

Statistical Language Models for Information Retrieval
ChengXiang Zhai
2008

Semantic Relations Between Nominals

Vivi Nastase, Preslav Nakov, Diarmuid Ó Séaghdha, and Stan Szpakowicz

ISBN: 978-3-031-01020-0 paperback
ISBN: 978-3-031-02148-0 ebook

DOI 10.1007/978-3-031-02148-0

A Publication in the Springer series
SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES

Lecture #19
Series Editor: Graeme Hirst, *University of Toronto*
Series ISSN
Synthesis Lectures on Human Language Technologies
Print 1947-4040 Electronic 1947-4059

Semantic Relations Between Nominals

Vivi Nastase
FBK, Trento

Preslav Nakov
QCRI, Qatar Foundation

Diarmuid Ó Séaghdha
Computer Laboratory, University of Cambridge

Stan Szpakowicz
EECS, University of Ottawa
ICS, Polish Academy of Sciences

SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES #19

ABSTRACT

People make sense of a text by identifying the semantic relations which connect the entities or concepts described by that text. A system which aspires to human-like performance must also be equipped to identify, and learn from, semantic relations in the texts it processes. Understanding even a simple sentence such as "Opportunity and Curiosity find similar rocks on Mars" requires recognizing relations (*rocks* **are located on** *Mars*, signalled by the word *on*) and drawing on already known relations (*Opportunity* and *Curiosity* **are instances of the class of** *Mars rovers*). A language-understanding system should be able to find such relations in documents and progressively build a knowledge base or even an ontology. Resources of this kind assist continuous learning and other advanced language-processing tasks such as text summarization, question answering and machine translation.

The book discusses the recognition in text of semantic relations which capture interactions between base noun phrases. After a brief historical background, we introduce a range of relation inventories of varying granularity, which have been proposed by computational linguists. There is also variation in the scale at which systems operate, from snippets all the way to the whole Web, and in the techniques of recognizing relations in texts, from full supervision through weak or distant supervision to self-supervised or completely unsupervised methods. A discussion of supervised learning covers available datasets, feature sets which describe relation instances, and successful algorithms. An overview of weakly supervised and unsupervised learning zooms in on the acquisition of relations from large corpora with hardly any annotated data. We show how bootstrapping from seed examples or patterns scales up to very large text collections on the Web. We also present machine learning techniques in which data redundancy and variability lead to fast and reliable relation extraction.

KEYWORDS

natural language processing, computational linguistics, lexical semantics, semantic relations, nominals, noun compounds, information extraction

Contents

Preface

Relations and texts

Every non-trivial text describes interactions and relations: between people, between other entities or concepts, between events. What we know about the world consists in large part of similar relations between concepts representing people, other entities, events, and so on. Such knowledge contributes to the understanding of relations which occur in texts. Newly found relations can in turn become part of the knowledge we store.

For an automatic system to grasp a text's semantic content, it must be able to recognize and reason about relations in texts, possibly by applying and updating previously acquired knowledge. We focus here in particular on semantic relations which describe the interactions among nouns and compact noun phrases, and we present such relations from both a theoretical and a practical perspective. The theoretical exploration shows the historical path which brings us to the current interpretation and view of semantic relations, and the wide range of proposals of relation inventories; such inventories vary according to domain, granularity and suitability for downstream applications.

On the practical side, we investigate the recognition and acquisition of relations from texts. We look at supervised learning methods. We present available datasets, discuss the variety of features which can describe relation instances, and learning algorithms used successfully thus far. The overview of weakly supervised and unsupervised learning looks in detail at problems and solutions related to the acquisition of relations from large corpora with little or no previously annotated data. We show how enduring the bootstrapping algorithm based on seed examples or patterns has proved to be, and how it has been adapted to tackle Web-scale text collections. We also present a few machine learning techniques which can take advantage of data redundancy and variability for fast and reliable relation extraction.

Semantic relations play a fundamental role in ontology-based learning and information extraction from documents. They can also provide valuable information for higher-level language-processing tasks, including summarization, question answering and machine translation.

The audience

This book will appeal to graduate students, researchers and practitioners interested in computational semantics, information extraction and, more generally, modern natural language processing technology. We have tried to make the presentation broadly accessible to anyone with a little background in artificial intelligence. Even so, it helps to have some familiarity with computational lin-

guistics and a modicum of tolerance for mathematical formulae. A basic understanding of machine learning is useful but not strictly necessary.

Organization of the book

A brief chapter 1 explains why we found it worthwhile to write a book about recognizing semantic relations between nominals, what applications those recognized relations can facilitate, what the book does and what it does not discuss.

Chapter 2 offers a brief overview of how relations between nominals have emerged from combined sources—people's propensity for categorization and the organization of knowledge, and for expressing these (and more) in language. It then continues with examples of lists of relations, and a discussion of the different dimensions along which relations can vary. These differences are reflected in how relation analysis is performed, how they behave in a corpus, and what kind of evidence or features we can use to describe them.

Chapters 3 and 4 contain an overview of the various methods of obtaining or learning relations between nominals. Among the many ways in which this topic can be structured, we chose to organize our survey according to the data support. Chapter 3 is focused on methods of learning relations which come from (comparatively) small, predefined datasets. Chapter 4 describes methods which work on large collections of unstructured texts or directly on the Web.

A very brief chapter 5 recaps our main observations, draws conclusions about the work on semantic relations between nominals covered in this book, and sums up the lessons learned.

Vivi Nastase, Preslav Nakov, Diarmuid Ó Séaghdha, and Stan Szpakowicz
April 2013