Automated Essay Scoring

Synthesis Lectures on Human Language Technologies

Natural Language Processing for Social Media, Third Edition
Anna Atefeh Farzindar and Diana Inkpen
2020

Statistical Significance Testing for Natural Language Processing
Rotem Dror, Lotem Peled, Segev Shlomov, and Roi Reichart
2020

Deep Learning Approaches to Text Production
Shashi Narayan and Claire Gardent
2020

Linguistic Fundamentals for Natural Language Processing II: 100 Essentials from Semantics and Pragmatics
Emily M. Bender and Alex Lascarides
2019

Cross-Lingual Word Embeddings
Anders Søgaard, Ivan Vulić, Sebastian Ruder, Manaal Faruqui
2019

Bayesian Analysis in Natural Language Processing, Second Edition
Shay Cohen
2019

Argumentation Mining
Manfred Stede and Jodi Schneider
2018

Quality Estimation for Machine Translation
Lucia Specia, Carolina Scarton, and Gustavo Henrique Paetzold
2018

Natural Language Processing for Social Media, Second Edition
Atefeh Farzindar and Diana Inkpen
2017

Automatic Text Simplification
Horacio Saggion
2017

Neural Network Methods for Natural Language Processing
Yoav Goldberg
2017

Syntax-based Statistical Machine Translation
Philip Williams, Rico Sennrich, Matt Post, and Philipp Koehn
2016

Domain-Sensitive Temporal Tagging
Jannik Strötgen and Michael Gertz
2016

Linked Lexical Knowledge Bases: Foundations and Applications
Iryna Gurevych, Judith Eckle-Kohler, and Michael Matuschek
2016

Bayesian Analysis in Natural Language Processing
Shay Cohen
2016

Metaphor: A Computational Perspective
Tony Veale, Ekaterina Shutova, and Beata Beigman Klebanov
2016

Grammatical Inference for Computational Linguistics
Jeffrey Heinz, Colin de la Higuera, and Menno van Zaanen
2015

Automatic Detection of Verbal Deception
Eileen Fitzpatrick, Joan Bachenko, and Tommaso Fornaciari
2015

Natural Language Processing for Social Media
Atefeh Farzindar and Diana Inkpen
2015

Semantic Similarity from Natural Language and Ontology Analysis
Sébastien Harispe, Sylvie Ranwez, Stefan Janaqi, and Jacky Montmain
2015

Learning to Rank for Information Retrieval and Natural Language Processing, Second Edition
Hang Li
2014

Ontology-Based Interpretation of Natural Language
Philipp Cimiano, Christina Unger, and John McCrae
2014

Automated Grammatical Error Detection for Language Learners, Second Edition
Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel Tetreault
2014

Web Corpus Construction
Roland Schäfer and Felix Bildhauer
2013

Recognizing Textual Entailment: Models and Applications
Ido Dagan, Dan Roth, Mark Sammons, and Fabio Massimo Zanzotto
2013

Linguistic Fundamentals for Natural Language Processing: 100 Essentials from Morphology and Syntax
Emily M. Bender
2013

Semi-Supervised Learning and Domain Adaptation in Natural Language Processing
Anders Søgaard
2013

Semantic Relations Between Nominals
Vivi Nastase, Preslav Nakov, Diarmuid Ó Séaghdha, and Stan Szpakowicz
2013

Computational Modeling of Narrative
Inderjeet Mani
2012

Natural Language Processing for Historical Texts
Michael Piotrowski
2012

Sentiment Analysis and Opinion Mining
Bing Liu
2012

Discourse Processing
Manfred Stede
2011

Bitext Alignment
Jörg Tiedemann
2011

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

Automated Essay Scoring

Beata Beigman Klebanov and Nitin Madnani

ISBN: 978-3-031-01054-5 paperback
ISBN: 978-3-031-02182-4 PDF
ISBN: 978-3-031-00193-2 hardcover

DOI 10.1007/978-3-031-02182-4

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

Lecture #52
Series Editor: Graeme Hirst, *University of Toronto*
Series ISSN
Print 1947-4040 Electronic 1947-4059

Automated Essay Scoring

Beata Beigman Klebanov and Nitin Madnani
Educational Testing Service

SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES #52

ABSTRACT

This book discusses the state of the art of automated essay scoring, its challenges and its potential. One of the earliest applications of artificial intelligence to language data (along with machine translation and speech recognition), automated essay scoring has evolved to become both a revenue-generating industry and a vast field of research, with many subfields and connections to other NLP tasks. In this book, we review the developments in this field against the backdrop of Elias Page's seminal 1966 paper titled "The Imminence of Grading Essays by Computer".

Part I establishes what automated essay scoring is about, why it exists, where the technology stands, and what are some of the main issues. In Part II, the book presents guided exercises to illustrate how one would go about building and evaluating a simple automated scoring system, while Part III offers readers a survey of the literature on different types of scoring models, the aspects of essay quality studied in prior research, and the implementation and evaluation of a scoring engine. Part IV offers a broader view of the field inclusive of some neighboring areas, and Part V closes with summary and discussion.

This book grew out of a week-long course on automated evaluation of language production at the North American Summer School for Logic, Language, and Information (NASSLLI), attended by advanced undergraduates and early-stage graduate students from a variety of disciplines. Teachers of natural language processing, in particular, will find that the book offers a useful foundation for a supplemental module on automated scoring. Professionals and students in linguistics, applied linguistics, educational technology, and other related disciplines will also find the material here useful.

KEYWORDS

automated essay scoring, automated writing evaluation, natural language processing, educational technology, artificial intelligence, AES, AWE, NLP, EdTech, AI

To my Dad, a physicist who likes a good story, and
To my Mom, a doctor who likes a well-written one,
With love,
Beata

To my family for their enduring support … and patience!
Nitin

Contents

Preface

This book grew out of a week-long course on automated evaluation of language production that we gave at the North American Summer School for Logic, Language, and Information in 2016. As anyone who has given a natural language processing (NLP) course for NASSLLI or ESSLLI knows, one generally expects early- or middle-stage graduate students from a variety of disciplines, or advanced undergraduates; some students are looking to merely familiarize themselves with a sub-field, others want to understand how things are done in the sub-field in a hands-on fashion, and still others who already have substantial computer science training want some help in understanding the current research frontier.

In an attempt to cater to all these different types of audiences, we wrote a book that can be read in different ways by different readers. In Part I, we provide an introduction addressing the first set of questions on anybody's mind when discovering a new field of endeavor—What is it about? Why does it exist? Where does it stand? What are some of the main issues? We aimed for an introduction that isn't too technical for people without much technical background, yet that would offer a reader with such background a new perspective that would not be commonly taught in a computer science, linguistics, education, or natural language processing course—a historical perspective that highlights the hopes and promise of this line of endeavor, along with long-known and more recently discovered challenges.

The paths of different readers through the book might diverge from this point on. In Part II, we address the eager newcomer who is ready to give this sub-field some more attention; we provide a guided set of exercises to illustrate how one would go about building and evaluating an automated scoring system, and the kinds of issues one is likely to encounter (Chapter 2), followed by a broader discussion of lessons learned (Chapter 3). A reader who wanted a glimpse and a taste, without having background in either computer science or linguistics, and without the intention of engaging with the research frontier in the area, might be best served by skimming Part III (Chapters 4–8) and focusing next on Part IV, in order to get a broader view of the field—inclusive of some neighboring areas—and moving to Part V for the overall summary and discussion.

If you are a reader who is already reasonably well versed in the standard machine learning material, is perhaps taking advanced undergraduate- and graduate-level courses in NLP, and is looking for a more technical orientation into the state of the field and the relevant literature—Part III is written for you. We consider different types of scoring models (Chapter 4), describe in detail the specific aspects of essay quality that have been studied in prior research—both those that cover a relatively generic construct of a good quality essay (Chapter 5) and those that were built to address more specific requirements of task or genre (Chapter 6). Chapters 7

and 8 discuss the implementation of a scoring engine and the evaluation of its performance, respectively. The subsequent Part IV, where a variety of neighboring fields are discussed briefly, might be familiar to this type of reader, and so, perhaps, are to be skimmed on the way to the summary and discussion in Part V.

This book is not intended as a stand-alone textbook that would take the reader with no knowledge of computer science, linguistics, NLP and psychometrics to the cutting edge of current research. We believe that some of the perspectives offered in this book would help teachers of courses in natural language processing who consider including automated scoring as one of the modules/exercises to supplement instruction. We also hope that teachers of courses in linguistics, applied linguistics, educational technology, and other related disciplines would find the material here useful for supplementing their instruction. The large majority of the work discussed in this book has been done on essays in English; we provide some pointers to work on automated essay scoring in other languages in Part V.

Finally, we want to say a few words on the typographical emphases we will use in the book. We will use small caps—like THIS—when an important concept is most fully explained; these locations are linked from the Definitions-in-Context section. You can track mentions of these concepts throughout the book using the index. We will use italics for local emphasis, and boldface to draw the reader's attention to an important statement or finding, as well as for tracking some chapter-level organizational elements.

Beata Beigman Klebanov and Nitin Madnani
August 2021