Learning To Classify Text Using Support Vector Machines The Springer International Series In Engineering And Computer Science

Text classification is becoming a crucial task to analysts in different areas. In the last few decades, the production of textual documents in digital form has increased exponentially. Their applications range from web pages to scientific documents, including emails, news and books. Despite the widespread use of digital texts, handling them is inherently difficult - the large amount of data necessary to represent them and the subjectivity of classification complicate matters. This book gives a concise view on how to use kernel approaches for inductive inference in large scale text classification; it presents a series of new techniques to enhance, scale and distribute text classification tasks. It is not intended to be a comprehensive survey of the state-of-the-art of the whole field of text classification. Its purpose is less ambitious and more practical: to explain and illustrate some of the important methods used in this field, in particular kernel approaches and techniques.

As today's world continues to advance, Artificial Intelligence (AI) is a field that has become a staple of technological development and led to the advancement of numerous professional industries. An application within AI that has gained attention is machine learning. Machine learning uses statistical techniques and algorithms to give computer systems the ability to understand and its popularity has circulated through many trades. Understanding this technology and its countless implementations is pivotal for scientists and researchers across the world. The Handbook of Research on Emerging Trends and Applications of Machine Learning provides a high-level understanding of various machine learning algorithms along with modern tools and techniques using Artificial Intelligence. In addition, this book explores the critical role that machine learning plays in a variety of professional fields including healthcare, business, and computer science. While highlighting topics including image processing, predictive analytics, and smart grid management, this book is ideally designed for developers, data scientists, business analysts, information architects, finance agents, healthcare professionals, researchers, retail traders, professors, and graduate students seeking current research on the benefits, implementations, and trends of machine learning.

Text Mining: Applications and Theory presents the state-of-the-art algorithms for text mining from both the academic and industrial perspectives. The contributors span several countries and scientific domains: universities, industrial corporations, and government laboratories, and demonstrate the use of techniques from machine learning, knowledge discovery, natural language processing and information retrieval to design computational models for automated text analysis and mining. This volume demonstrates how advancements in the fields of applied mathematics, computer science, machine learning, and natural language processing can collectively capture, classify, and interpret words and their contexts. As suggested in the preface, text mining is needed when "words are not enough." This book: Provides state-of-the-art algorithms and techniques for critical tasks in text mining applications, such as clustering, classification, anomaly and trend detection, and stream analysis. Presents a survey of text visualization techniques and looks at the multilingual text classification problem. Discusses the issue of cybercrime associated with chatrooms. Features advances in visual analytics and machine learning along with illustrative examples. Is accompanied by a supporting website featuring datasets. Applied mathematicians, statisticians, practitioners and students in computer science, bioinformatics and engineering will find this book extremely useful.

Learn to build expert NLP and machine learning projects using NLTK and other Python libraries About This Book Break text down into its component parts for spelling correction, feature extraction, and phrase transformation Work through NLP concepts with simple and easy-to-follow programming recipes Gain insights into the current and budding research topics of NLP Who This Book Is For If you are an NLP or machine learning enthusiast and an intermediate Python programmer who wants to quickly master NLTK for natural language processing, then this Learning Path will do you a lot of good. Students of linguistics and semantic/sentiment analysis professionals will find it invaluable. What You Will Learn The scope of natural language complexity and how they are processed by machines Clean and wrangle text using tokenization and chunking to help you process data better Tokenize text into sentences and sentences into words Classify text and perform sentiment analysis Implement string matching algorithms and normalization techniques Understand and implement the concepts of information retrieval and text summarization Find out how to implement various NLP tasks in Python In Detail Natural Language Processing is a field of computational linguistics and artificial intelligence that deals with human-computer interaction. It provides a seamless interaction between computers and human beings and gives computers the ability to understand human speech with the help of machine learning. The number of human-computer interaction instances are increasing so it's becoming imperative that computers comprehend all major natural languages. The first NLTK Essentials module is an introduction on how to build systems around NLP, with a focus on how to create a customized tokenizer and parser from scratch. You will learn essential concepts of NLP, be given practical insight into open source tool and libraries available in Python, shown how to analyze social media sites, and be given tools to deal with large scale text. This module also provides a workaround using some of the amazing capabilities of Python libraries such as NLTK, scikit-learn, pandas, and NumPy. The second Python 3 Text Processing with NLTK 3 Cookbook module teaches you the essential techniques of text and language processing with simple, straightforward examples. This includes organizing text corpora, creating your own custom corpus, text classification with a focus on sentiment analysis, and distributed text processing methods. The third Mastering Natural Language Processing with Python module will help you become an expert and assist you in creating your own NLP projects using NLTK. You will be guided through model development with machine learning tools, shown how to create training data, and given insight into the best practices for designing and building NLP-based applications using Python. This Learning Path combines some of the best that Packt has to offer in one complete, curated package and is designed to help you quickly learn text processing with Python and NLTK. It includes content from the following Packt products: NTLK essentials by Nitin Hardeniya Python 3 Text Processing with NLTK 3 Cookbook by Jacob Perkins Mastering Natural Language Processing with Python by Deepti Chopra, Nisheeth Joshi, and Iti Mathur Style and approach This comprehensive course creates a smooth learning path that teaches you how to get started with Natural Language Processing using Python and NLTK. You'll learn to create effective NLP and machine learning projects using Python and NLTK.

Much of the data available today is unstructured and text-heavy, making it challenging for analysts to apply their usual data wrangling and visualization tools. With this practical book, you'll explore text-mining techniques with tidytext, a package that authors Julia Silge and David Robinson developed using the tidy principles behind R packages like ggraph and dplyr. You'll learn how tidytext and other tidy tools in R can make text analysis easier and more effective. The authors demonstrate how treating text

as data frames enables you to manipulate, summarize, and visualize characteristics of text. You'll also learn how to integrate natural language processing (NLP) into effective workflows. Practical code examples and data explorations will help you generate real insights from literature, news, and social media. Learn how to apply the tidy text format to NLP Use sentiment analysis to mine the emotional content of text Identify a document's most important terms with frequency measurements Explore relationships and connections between words with the ggraph and widyr packages Convert back and forth between R's tidy and non-tidy text formats Use topic modeling to classify document collections into natural groups Examine case studies that compare Twitter archives, dig into NASA metadata, and analyze thousands of Usenet messages

From news and speeches to informal chatter on social media, natural language is one of the richest and most underutilized sources of data. Not only does it come in a constant stream, always changing and adapting in context; it also contains information that is not conveyed by traditional data sources. The key to unlocking natural language is through the creative application of text analytics. This practical book presents a data scientist's approach to building language-aware products with applied machine learning. You'll learn robust, repeatable, and scalable techniques for text analysis with Python, including contextual and linguistic feature engineering, vectorization, classification, topic modeling, entity resolution, graph analysis, and visual steering. By the end of the book, you'll be equipped with practical methods to solve any number of complex real-world problems. Preprocess and vectorize text into high-dimensional feature representations Perform document classification and topic modeling Steer the model selection process with visual diagnostics Extract key phrases, named entities, and graph structures to reason about data in text Build a dialog framework to enable chatbots and language-driven interaction Use Spark to scale processing power and neural networks to scale model complexity

Leverage Natural Language Processing (NLP) in Python and learn how to set up your own robust environment for performing text analytics. This second edition has gone through a major revamp and introduces several significant changes and new topics based on the recent trends in NLP. You'll see how to use the latest state-of-the-art frameworks in NLP, coupled with machine learning and deep learning models for supervised sentiment analysis powered by Python to solve actual case studies. Start by reviewing Python for NLP fundamentals on strings and text data and move on to engineering representation methods for text data, including both traditional statistical models and newer deep learning-based embedding models. Improved techniques and new methods around parsing and processing text are discussed as well. Text summarization and topic models have been overhauled so the book showcases how to build, tune, and interpret topic models in the context of an interest dataset on NIPS conference papers. Additionally, the book covers text similarity techniques with a real-world example of movie recommenders, along with sentiment analysis using supervised and unsupervised techniques. There is also a chapter dedicated to semantic analysis where you'll see how to build your own named entity recognition (NER) system from scratch. While the overall structure of the book remains the same, the entire code base, modules, and chapters has been updated to the latest Python 3.x release. What You'll Learn • Understand NLP and text syntax, semantics and structure• Discover text cleaning and feature engineering• Review text classification and text clustering • Assess text summarization and topic models• Study deep learning for NLP Who This Book Is For IT professionals, data analysts, developers, linguistic experts, data scientists and engineers and basically anyone with a keen interest in linguistics, analytics and generating insights from textual data.

The contributions in this volume represent the latest research results in the field of Classification, Clustering, and Data Analysis. Besides the theoretical analysis, papers focus on various application fields as Archaeology, Astronomy, Bio-Sciences, Business, Electronic Data and Web, Finance and Insurance, Library Science and Linguistics, Marketing, Music Science, and Quality Assurance.

Most data scientists and engineers today rely on quality labeled data to train machine learning models. But building a training set manually is time-consuming and expensive, leaving many companies with unfinished ML projects. There's a more practical approach. In this book, Wee Hyong Tok, Amit Bahree, and Senja Filipi show you how to create products using weakly supervised learning models. You'll learn how to build natural language processing and computer vision projects using weakly labeled datasets from Snorkel, a spin-off from the Stanford AI Lab. Because so many companies have pursued ML projects that never go beyond their labs, this book also provides a guide on how to ship the deep learning models you build. Get up to speed on the field of weak supervision, including ways to use it as part of the data science process Use Snorkel AI for weak supervision and data programming Get code examples for using Snorkel to label text and image datasets Use a weakly labeled dataset for text and image classification Learn practical considerations for using Snorkel with large datasets and using Spark clusters to scale labeling Summary Deep Learning with Python introduces the field of deep learning using the Python language and the powerful Keras library. Written by Keras creator and Google AI researcher François Chollet, this book builds your understanding through intuitive explanations and practical examples. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Machine learning has made remarkable progress in recent years. We went from near-unusable speech and image recognition, to near-human accuracy. We went from machines that couldn't beat a serious Go player, to defeating a world champion. Behind this progress is deep learning—a combination of engineering advances, best practices, and theory that enables a wealth of previously impossible smart applications. About the Book Deep Learning with Python introduces the field of deep learning using the Python language and the powerful Keras library. Written by Keras creator and Google AI researcher François Chollet, this book builds your understanding through intuitive explanations and practical examples. You'll explore challenging concepts and practice with applications in computer vision, natural-language processing, and generative models. By the time you finish, you'll have the knowledge and hands-on skills to apply deep learning in your own projects. What's Inside Deep learning from first principles Setting up your own deep-learning environment Image-classification models Deep learning for text and sequences Neural style transfer, text generation, and image generation About the Reader Readers need intermediate Python skills. No previous experience with Keras, TensorFlow, or machine learning is required. About the Author François Chollet works on deep learning at Google in Mountain View, CA. He is the creator of the Keras deep-learning library, as well as a contributor to the TensorFlow machine-learning framework. He also does deep-learning research, with a focus on computer vision and the application of machine learning to formal reasoning. His papers have been published at major conferences in the field, including the Conference on Computer Vision and Pattern Recognition (CVPR), the Conference and Workshop on Neural Information Processing Systems (NIPS), the International Conference on Learning Representations (ICLR), and others. Table of Contents PART 1 - FUNDAMENTALS OF DEEP LEARNING What is deep learning? Before we begin: the mathematical building blocks of neural networks Getting started with neural networks Fundamentals of machine learning PART 2 -

DEEP LEARNING IN PRACTICE Deep learning for computer vision Deep learning for text and sequences Advanced deep-learning best practices Generative deep learning Conclusions appendix A - Installing Keras and its dependencies on Ubuntu appendix B - Running Jupyter notebooks on an EC2 GPU instance

An introduction to a broad range of topics in deep learning, covering mathematical and conceptual background, deep learning techniques used in industry, and research perspectives. "Written by three experts in the field, Deep Learning is the only comprehensive book on the subject." —Elon Musk, cochair of OpenAI; cofounder and CEO of Tesla and SpaceX Deep learning is a form of machine learning that enables computers to learn from experience and understand the world in terms of a hierarchy of concepts. Because the computer gathers knowledge from experience, there is no need for a human computer operator to formally specify all the knowledge that the computer needs. The hierarchy of concepts allows the computer to learn complicated concepts by building them out of simpler ones; a graph of these hierarchies would be many layers deep. This book introduces a broad range of topics in deep learning. The text offers mathematical and conceptual background, covering relevant concepts in linear algebra, probability theory and information theory, numerical computation, and machine learning. It describes deep learning techniques used by practitioners in industry, including deep feedforward networks, regularization, optimization algorithms, convolutional networks, sequence modeling, and practical methodology; and it surveys such applications as natural language processing, speech recognition, computer vision, online recommendation systems, bioinformatics, and videogames. Finally, the book offers research perspectives, covering such theoretical topics as linear factor models, autoencoders, representation learning, structured probabilistic models, Monte Carlo methods, the partition function, approximate inference, and deep generative models. Deep Learning can be used by undergraduate or graduate students planning careers in either industry or research, and by software engineers who want to begin using deep learning in their products or platforms. A website offers supplementary material for both readers and instructors.

Based on ideas from Support Vector Machines (SVMs), Learning To Classify Text Using Support Vector Machines presents a new approach to generating text classifiers from examples. The approach combines high performance and efficiency with theoretical understanding and improved robustness. In particular, it is highly effective without greedy heuristic components. The SVM approach is computationally efficient in training and classification, and it comes with a learning theory that can guide real-world applications. Learning To Classify Text Using Support Vector Machines gives a complete and detailed description of the SVM approach to learning text classifiers, including training algorithms, transductive text classification, efficient performance estimation, and a statistical learning model of text classification. In addition, it includes an overview of the field of text classification, making it self-contained even for newcomers to the field. This book gives a concise introduction to SVMs for pattern recognition, and it includes a detailed description of how to formulate text-classification tasks for machine learning.

This book explains the detectionbased approach to investigating crosslinguistic influence and illustrates the value of the approach through a collection of five empirical studies that use the approach to quantify, evaluate, and isolate the subtle and complex influences of learners' nativelanguage backgrounds on their English writing.

Many books and courses tackle natural language processing (NLP) problems with toy use cases and well-defined datasets. But if you want to build, iterate, and scale NLP systems in a business setting and tailor them for particular industry verticals, this is your guide. Software engineers and data scientists will learn how to navigate the maze of options available at each step of the journey. Through the course of the book, authors Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, and Harshit Surana will guide you through the process of building real-world NLP solutions embedded in larger product setups. You'll learn how to adapt your solutions for different industry verticals such as healthcare, social media, and retail. With this book, you'll: Understand the wide spectrum of problem statements, tasks, and solution approaches within NLP Implement and evaluate different NLP applications using machine learning and deep learning methods Fine-tune your NLP solution based on your business problem and industry vertical Evaluate various algorithms and approaches for NLP product tasks, datasets, and stages Produce software solutions following best practices around release, deployment, and DevOps for NLP systems Understand best practices, opportunities, and the roadmap for NLP from a business and product leader's perspective

Learn Machine Learning!Machine learning is one of those topics that can be daunting at first blush. It's not clear where to start, what path someone should take and what APIs to learn in order to get started teaching machines how to learn.This is where Machine Learning by Tutorials comes in! In this book, we'll hold your hand through a number of tutorials, to get you started in the world of machine learning. We'll cover a wide range of popular topics in the field of machine learning, while developing apps that work on iOS devices.Who This Book Is ForThis books is for the intermediate iOS developer who already knows the basics of iOS and Swift development, but wants to understand how machine learning works.Topics covered in Machine Learning by TutorialsCoreML: Learn how to add a machine learning model to your iOS apps, and how to use iOS APIs to access it.Create ML: Learn how to create your own model using Apple's Create ML Tool.Turi Create and Keras: Learn how to tune parameters to improve your machine learning model using more advanced tools.Image Classification: Learn how to apply machine learning models to predict objects in an image.Convolutional Networks: Learn advanced machine learning techniques for predicting objects in an image with Convolutional Neural Networks (CNNs).Sequence Classification: Learn how you can use recurrent neural networks (RNNs) to classify motion from an iPhone's motion sensor.Text-to-text Transform: Learn how to use machine learning to convert bodies of text between two languages.By the end of this book, you'll have a firm understanding of what machine learning is, what it can and cannot do, and how you can use machine learning in your next app!

This book brings together scientists, researchers, practitioners, and students from academia and industry to present recent and ongoing research activities concerning the latest advances, techniques, and applications of natural language processing systems, and to promote the exchange of new ideas and lessons learned. Taken together, the chapters of this book provide a collection of high-quality research works that address broad challenges in both theoretical and applied aspects of intelligent natural language processing. The book presents the state-of-the-art in research on natural language processing, computational linguistics, applied Arabic linguistics and related areas. New trends in natural language processing systems are rapidly emerging – and finding application in various domains including education, travel and tourism, and healthcare, among others. Many issues encountered during the development of these applications can be resolved by incorporating language technology solutions. The topics covered by the book include: Character and Speech Recognition; Morphological, Syntactic, and Semantic Processing; Information Extraction; Information Retrieval and Question Answering; Text Classification and Text Mining; Text Summarization; Sentiment Analysis; Machine Translation Building and Evaluating Linguistic Resources; and Intelligent Language Tutoring Systems.

This book constitutes the refereed proceedings of the 33rd annual European Conference on Information Retrieval Research, ECIR 2011, held in Dublin, Ireland, in April 2010. The 45 revised full papers presented together with 24 poster papers, 17 short papers, and 6 tool demonstrations were carefully reviewed and selected from 223 full research paper submissions and 64 poster/demo submissions. The papers are organized in topical sections on text categorization, recommender systems, Web IR, IR evaluation, IR for Social Networks, cross-language IR, IR theory, multimedia IR, IR applications, interactive IR, and question answering /NLP.

Class-tested and coherent, this textbook teaches classical and web information retrieval, including web search and the related areas of text classification and text clustering from basic concepts. It gives an up-to-date treatment of all aspects of the design and implementation of systems for gathering, indexing, and searching documents; methods for evaluating systems; and an introduction to the use of machine learning methods on text collections. All the important ideas are explained using examples and figures, making it perfect for introductory courses in information retrieval for advanced undergraduates and graduate students in computer science. Based on feedback from extensive classroom experience, the book has been carefully structured in order to make teaching more natural and effective. Slides and additional exercises (with solutions for lecturers) are also available through the book's supporting website to help course instructors prepare their lectures.

This book constitutes the thoroughly refereed post-conference proceedings of the Second International Symposium on Intelligent Informatics (ISI 2013) held in Mysore, India during August 23-24, 2013. The 47 revised papers presented were carefully reviewed and selected from 126 initial submissions. The papers are organized in topical sections on pattern recognition, signal and image processing; data mining, clustering and intelligent information systems; multi agent systems; and computer networks and distributed systems. The book is directed to the researchers and scientists engaged in various fields of intelligent informatics.

Comprehensive Coverage of the Entire Area of Classification Research on the problem of classification tends to be fragmented across such areas as pattern recognition, database, data mining, and machine learning. Addressing the work of these different communities in a unified way, Data Classification: Algorithms and Applications explores the underlying algorithms of classification as well as applications of classification in a variety of problem domains, including text, multimedia, social network, and biological data. This comprehensive book focuses on three primary aspects of data classification: Methods-The book first describes common techniques used for classification, including probabilistic methods, decision trees, rule-based methods, instance-based methods, support vector machine methods, and neural networks. Domains-The book then examines specific methods used for data domains such as multimedia, text, time-series, network, discrete sequence, and uncertain data. It also covers large data sets and data streams due to the recent importance of the big data paradigm. Variations-The book concludes with insight on variations of the classification process. It discusses ensembles, rare-class learning, distance function learning, active learning, visual learning, transfer learning, and semi-supervised learning as well as evaluation aspects of classifiers.

Text mining applications have experienced tremendous advances because of web 2.0 and social networking applications. Recent advances in hardware and software technology have lead to a number of unique scenarios where text mining algorithms are learned. Mining Text Data introduces an important niche in the text analytics field, and is an edited volume contributed by leading international researchers and practitioners focused on social networks & data mining. This book contains a wide swath in topics across social networks & data mining. Each chapter contains a comprehensive survey including the key research content on the topic, and the future directions of research in the field. There is a special focus on Text Embedded with Heterogeneous and Multimedia Data which makes the mining process much more challenging. A number of methods have been designed such as transfer learning and cross-lingual mining for such cases. Mining Text Data simplifies the content, so that advanced-level students, practitioners and researchers in computer science can benefit from this book. Academic and corporate libraries, as well as ACM, IEEE, and Management Science focused on information security, electronic commerce, databases, data mining, machine learning, and statistics are the primary buyers for this reference book.

This comprehensive encyclopedia, in A-Z format, provides easy access to relevant information for those seeking entry into any aspect within the broad field of Machine Learning. Most of the entries in this preeminent work include useful literature references.

Step-by-step tutorials on deep learning neural networks for computer vision in python with Keras.

This book constitutes the refereed proceedings of the 8th International Conference on Advanced Data Mining and Applications, ADMA 2012, held in Nanjing, China, in December 2012. The 32 regular papers and 32 short papers presented in this volume were carefully reviewed and selected from 168 submissions. They are organized in topical sections named: social media mining; clustering; machine learning: algorithms and applications; classification; prediction, regression and recognition; optimization and approximation; mining time series and streaming data; Web mining and semantic analysis; data mining applications; search and retrieval; information recommendation and hiding; outlier detection; topic modeling; and data cube computing.

"The fox knows many things, but the hedgehog knows one big thing." This ancient Greek aphorism, preserved in a fragment from the poet Archilochus, describes the central thesis of Isaiah Berlin's masterly essay on Leo Tolstoy and the philosophy of history, the subject of the epilogue to War and Peace. Although there have been many interpretations of the adage, Berlin uses it to mark a fundamental distinction between human beings who are fascinated by the infinite variety of things and those who relate everything to a central, all-embracing system. Applied to Tolstoy, the saying illuminates a paradox that helps explain his philosophy of history: Tolstoy was a fox, but believed in being a hedgehog. One of Berlin's most celebrated works, this extraordinary essay offers profound insights about Tolstoy, historical understanding, and human psychology. This new edition features a revised text that supplants all previous versions, English translations of the many passages in foreign languages, a new foreword in which Berlin biographer Michael Ignatieff explains the enduring appeal of Berlin's essay, and a new appendix that provides rich context, including excerpts from reviews and Berlin's letters, as well as a startling new interpretation of Archilochus's epigram.

A guide on the use of SVMs in pattern classification, including a rigorous performance comparison of classifiers and regressors. The book presents architectures for multiclass classification and function approximation problems, as well as evaluation criteria for classifiers and regressors. Features: Clarifies the characteristics of two-class SVMs; Discusses kernel methods for improving the generalization ability of neural networks and fuzzy systems; Contains ample illustrations and examples; Includes performance evaluation using publicly available data sets; Examines Mahalanobis kernels, empirical feature space, and the effect of model selection by cross-validation; Covers sparse SVMs, learning using privileged information, semi-supervised learning, multiple classifier systems, and multiple kernel learning; Explores incremental training based batch training and active-set training methods, and decomposition techniques for linear programming SVMs; Discusses variable selection for support vector regressors.

This thesis seeks to establish if the use of negation in Inductive Rule Learning (IRL) for text classification is effective. Text classification is a widely research topic in the domain of data mining. There have been many techniques directed at text classification; one of them is IRL, widely chosen because of its simplicity, comprehensibility and interpretability by humans. IRL is a process whereby rules in the form of $antecedent -> conclusion$ are learnt to build a classifier. Thus, the learnt classifier comprises a set of rules, which are used to perform classification. To learn a rule, words from pre-labelled documents, known as features, are selected to be used as conjunctions in the rule antecedent. These rules typically do not include any negated features in their antecedent; although in some cases, as demonstrated in this thesis, the inclusion of negation is required and beneficial for the text classification task. With respect to the use of negation in IRL, two issues need to be addressed: (i) the identification of the features to be negated and (ii) the improvisation of rule refinement strategies to generate

rules both with and without negation. To address the first issue, feature space division is proposed, whereby the feature space containing features to be used for rule refinement is divided into three sub-spaces to facilitate the identification of the features which can be advantageously negated. To address the second issue, eight rule refinement strategies are proposed, which are able to generate both rules with and without negation. Typically, single keywords which are deemed significant to differentiate between classes are selected to be used in the text representation in the text classification task. Phrases have also been proposed because they are considered to be semantically richer than single keywords. Therefore, with respect to the work conducted in this thesis, three different types of phrases (n-gram phrases, keyphrases and fuzzy phrases) are extracted to be used as the text representation in addition to the use of single keywords. To establish the effectiveness of the use of negation in IRL, the eight proposed rule refinement strategies are compared with one another, using keywords and the three different types of phrases as the text representation, to determine whether the best strategy is one which generates rules with negation or without negation. Two types of classification tasks are conducted; binary classification and multi-class classification. The best strategy in the proposed IRL mechanism is compared to five existing text classification techniques with respect to binary classification: (i) the Sequential Minimal Optimization (SMO) algorithm, (ii) Naive Bayes (NB), (iii) JRip, (iv) OlexGreedy and (v) OlexGA from the Waikato Environment for Knowledge Analysis (WEKA) machine learning workbench. In the multi-class classification task, the proposed IRL mechanism is compared to the Total From Partial Classification (TFPC) algorithm. The datasets used in the experiments include three text datasets: 20 Newsgroups, Reuters-21578 and Small Animal Veterinary Surveillance Network (SAVSNET) datasets and five UCI Machine Learning Repository tabular datasets. The results obtained from the experiments showed that the strategies which generated rules with negation were more effective when the keyword representation was used and less prominent when the phrase representations were used. Strategies which generated rules with negation also performed better with respect to binary classification compared to multi-class classification. In comparison with the other machine learning techniques selected, the proposed IRL mechanism was shown to generally outperform all the compared techniques and was competitive with SMO.

This open access book constitutes the refereed proceedings of the 18th International Conference on String Processing and Information Retrieval, ICOST 2020, held in Hammamet, Tunisia, in June 2020.* The 17 full papers and 23 short papers presented in this volume were carefully reviewed and selected from 49 submissions. They cover topics such as: IoT and AI solutions for e-health; biomedical and health informatics; behavior and activity monitoring; behavior and activity monitoring; and wellbeing technology. *This conference was held virtually due to the COVID-19 pandemic.

"We live in the age of data. In the last few years, the methodology of extracting insights from data or "data science" has emerged as a discipline in its own right. The R programming language has become one-stop solution for all types of data analysis. The growing popularity of R is due its statistical roots and a vast open source package library. The goal of "Beginning Data Science with R" is to introduce the readers to some of the useful data science techniques and their implementation with the R programming language. The book attempts to strike a balance between the how: specific processes and methodologies, and understanding the why: going over the intuition behind how a particular technique works, so that the reader can apply it to the problem at hand. This book will be useful for readers who are not familiar with statistics and the R programming language.

Deep learning methods are achieving state-of-the-art results on challenging machine learning problems such as describing photos and translating text from one language to another. In this new laser-focused Ebook, finally cut through the math, research papers and patchwork descriptions about natural language processing. Using clear explanations, standard Python libraries and step-by-step tutorial lessons you will discover what natural language processing is, the promise of deep learning in the field, how to clean and prepare text data for modeling, and how to develop deep learning models for your own natural language processing projects.

This book constitutes the post-conference proceedings of the 5th International Conference on Machine Learning, Optimization, and Data Science, LOD 2019, held in Siena, Italy, in September 2019. The 54 full papers presented were carefully reviewed and selected from 158 submissions. The papers cover topics in the field of machine learning, artificial intelligence, reinforcement learning, computational optimization and data science presenting a substantial array of ideas, technologies, algorithms, methods and applications.

The 13th International Conference on Human–Computer Interaction, HCI Inter- tional 2009, was held in San Diego, California, USA, July 19–24, 2009, jointly with the Symposium on Human Interface (Japan) 2009, the 8th International Conference on Engineering Psychology and Cognitive Ergonomics, the 5th International Conference on Universal Access in Human-Computer Interaction, the Third International Conf- ence on Virtual and Mixed Reality, the Third International Conference on Internati- alization, Design and Global Development, the Third International Conference on Online Communities and Social Computing, the 5th International Conference on Augmented Cognition, the Second International Conference on Digital Human Mod- ing, and the First International Conference on Human Centered Design. A total of 4,348 individuals from academia, research institutes, industry and gove- mental agencies from 73 countries submitted contributions, and 1,425 papers that were judged to be of high scientific quality were included in the program. These papers - dress the latest research and development efforts and highlight the human aspects of the design and use of computing systems. The papers accepted for presentation thoroughly cover the entire field of human–computer interaction, addressing major advances in knowledge and effective use of computers in a variety of application areas.

This book offers a highly accessible introduction to natural language processing, the field that supports a variety of language technologies, from predictive text and email filtering to automatic summarization and translation. With it, you'll learn how to write Python programs that work with large collections of unstructured text. You'll access richly annotated datasets using a comprehensive range of linguistic data structures, and you'll understand the main algorithms for analyzing the content and structure of written communication. Packed with examples and exercises, Natural Language Processing with Python will help you: Extract information from unstructured text, either to guess the topic or identify "named entities" Analyze linguistic structure in text, including parsing and semantic analysis Access popular linguistic databases, including WordNet and treebanks Integrate techniques drawn from fields as diverse as linguistics and artificial intelligence This book will help you gain practical skills in natural language processing using the Python programming language and the Natural Language Toolkit (NLTK) open source library. If you're interested in developing web applications, analyzing multilingual news sources, or documenting endangered languages -- or if you're simply curious to have a programmer's perspective on how human language works -- you'll find Natural Language Processing with Python both fascinating and immensely useful.

Summary Deep Learning with R introduces the world of deep learning using the powerful Keras library and its R language interface. The book builds your understanding of deep learning through intuitive explanations and practical examples. Continue your journey into the world of deep learning with Deep Learning with R in Motion, a practical, hands-on video course available exclusively at Manning.com (www.manning.com/livevideo/deep-?learning-with-r-in-motion). Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Machine learning has made remarkable progress in recent years. Deep-learning systems now enable previously impossible smart applications, revolutionizing image recognition and natural-language processing, and identifying complex patterns in data. The Keras deep-learning library provides data scientists and developers working in R a state-of-the-art toolset for tackling deep-learning tasks. About the Book Deep Learning with R introduces the world of deep learning using the powerful Keras library and its R language interface. Initially written for Python as Deep Learning with Python by Keras creator and Google AI researcher François Chollet and adapted for R by RStudio founder J. J. Allaire, this book builds your understanding of deep learning through intuitive

explanations and practical examples. You'll practice your new skills with R-based applications in computer vision, natural-language processing, and generative models. What's Inside Deep learning from first principles Setting up your own deep-learning environment Image classification and generation Deep learning for text and sequences About the Reader You'll need intermediate R programming skills. No previous experience with machine learning or deep learning is assumed. About the Authors François Chollet is a deep-learning researcher at Google and the author of the Keras library. J.J. Allaire is the founder of RStudio and the author of the R interfaces to TensorFlow and Keras. Table of Contents PART 1 - FUNDAMENTALS OF DEEP LEARNING What is deep learning? Before we begin: the mathematical building blocks of neural networks Getting started with neural networks Fundamentals of machine learning PART 2 - DEEP LEARNING IN PRACTICE Deep learning for computer vision Deep learning for text and sequences Advanced deep-learning best practices Generative deep learning Conclusions

Implement natural language processing applications with Python using a problem-solution approach. This book has numerous coding exercises that will help you to quickly deploy natural language processing techniques, such as text classification, parts of speech identification, topic modeling, text summarization, text generation, entity extraction, and sentiment analysis. Natural Language Processing Recipes starts by offering solutions for cleaning and preprocessing text data and ways to analyze it with advanced algorithms. You'll see practical applications of the semantic as well as syntactic analysis of text, as well as complex natural language processing approaches that involve text normalization, advanced preprocessing, POS tagging, and sentiment analysis. You will also learn various applications of machine learning and deep learning in natural language processing. By using the recipes in this book, you will have a toolbox of solutions to apply to your own projects in the real world, making your development time quicker and more efficient. What You Will Learn Apply NLP techniques using Python libraries such as NLTK, TextBlob, spaCy, Stanford CoreNLP, and many more Implement the concepts of information retrieval, text summarization, sentiment analysis, and other advanced natural language processing techniques. Identify machine learning and deep learning techniques for natural language processing and natural language generation problems Who This Book Is ForData scientists who want to refresh and learn various concepts of natural language processing through coding exercises.

Learning to Classify Text Using Support Vector MachinesSpringer Science & Business Media

This three volume set (CCIS 1237-1239) constitutes the proceedings of the 18th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2020, in June 2020. The conference was scheduled to take place in Lisbon, Portugal, at University of Lisbon, but due to COVID-19 pandemic it was held virtually. The 173 papers were carefully reviewed and selected from 213 submissions. The papers are organized in topical sections: homage to Enrique Ruspini; invited talks; foundations and mathematics; decision making, preferences and votes; optimization and uncertainty; games; real world applications; knowledge processing and creation; machine learning I; machine learning II; XAI; image processing; temporal data processing; text analysis and processing; fuzzy interval analysis; theoretical and applied aspects of imprecise probabilities; similarities in artificial intelligence; belief function theory and its applications; aggregation: theory and practice; aggregation: pre-aggregation functions and other generalizations of monotonicity; aggregation: aggregation of different data structures; fuzzy methods in data mining and knowledge discovery; computational intelligence for logistics and transportation problems; fuzzy implication functions; soft methods in statistics and data analysis; image understanding and explainable AI; fuzzy and generalized quantifier theory; mathematical methods towards dealing with uncertainty in applied sciences; statistical image processing and analysis, with applications in neuroimaging; interval uncertainty; discrete models and computational intelligence; current techniques to model, process and describe time series; mathematical fuzzy logic and graded reasoning models; formal concept analysis, rough sets, general operators and related topics; computational intelligence methods in information modelling, representation and processing.

The Definitive Resource on Text Mining Theory and Applications from Foremost Researchers in the Field Giving a broad perspective of the field from numerous vantage points, Text Mining: Classification, Clustering, and Applications focuses on statistical methods for text mining and analysis. It examines methods to automatically cluster and classify text documents and applies these methods in a variety of areas, including adaptive information filtering, information distillation, and text search. The book begins with chapters on the classification of documents into predefined categories. It presents state-of-the-art algorithms and their use in practice. The next chapters describe novel methods for clustering documents into groups that are not predefined. These methods seek to automatically determine topical structures that may exist in a document corpus. The book concludes by discussing various text mining applications that have significant implications for future research and industrial use. There is no doubt that text mining will continue to play a critical role in the development of future information systems and advances in research will be instrumental to their success. This book captures the technical depth and immense practical potential of text mining, guiding readers to a sound appreciation of this burgeoning field.

Comprehensive Coverage of the Entire Area of Classification Research on the problem of classification tends to be fragmented across such areas as pattern recognition, database, data mining, and machine learning. Addressing the work of these different communities in a unified way, Data Classification: Algorithms and Applications explores the underlying algorithms of classification as well as applications of classification in a variety of problem domains, including text, multimedia, social network, and biological data. This comprehensive book focuses on three primary aspects of data classification: Methods: The book first describes common techniques used for classification, including probabilistic methods, decision trees, rule-based methods, instance-based methods, support vector machine methods, and neural networks. Domains: The book then examines specific methods used for data domains such as multimedia, text, time-series, network, discrete sequence, and uncertain data. It also covers large data sets and data streams due to the recent importance of the big data paradigm. Variations: The book concludes with insight on variations of the classification process. It discusses ensembles, rare-class learning, distance function learning, active learning, visual learning, transfer learning, and semi-supervised learning as well as evaluation aspects of classifiers.