Indrasen Poola

Data Scientist & Artificial Intelligence -Industry Consultant

Fellow Member of International Association for Engineering and Technology
 Email me on :indrasen.research@gmail.com

Total of 13+ years of experience into IT -Artificial Intelligence ,Data Scientist, Machine Learning, Business Intelligence and Analytics and Natural Language Processing Specialist and Software Engineer with a unique combination of solid algorithm design skills and research acumen. Experience includes:

• Authorship of scholarly articles of **15 publications** in the prestigious International Journals on Artificial Intelligence.
• applying neural networks in NLP
• implementing Deep Learning nets for face recognition
• using Machine Learning methods for classification
• using 'Big Data" technologies like Apache Spark with Scala, Hadoop and Cascading
• expert programming in python (using Jupyter notebooks), C/C++ and Java
• staying abreast of the latest developments in ML and NLP.
• Graph mining in social network analysis
• Performance analysis and optimization

WORK EXPERIENCE

Data Scientist & AI Machine Learning Consultant

Cisco -April ,16 to Present

Natural Language Processing
Working on a new sentiment analysis algorithm for Cisco as well as an algorithm to extract information from email threads related to customer care (the cause and resolution of customer complaints) for another Fortune 10 company.

Implementing a cascaded convolutional network for use in a system for face detection. The network is intended to reproduce results from a recent research paper, using the Torch Deep Learning framework. Other frameworks like Theano/Lasagne were also evaluated. Studied a sequence of papers on other deep neural network architectures as well as OpenFace (CMU) and FaceNet (Google)..

Data Scientist
EA   July'15 to April'16

Enabling job seekers to find jobs relevant to the intent of their queries by understanding job descriptions - classifying them, tagging them, and finding the dominant "topics" in them.

• Used word2vec , a neural embedding algorithm, on millions of job descriptions along with graph clustering algorithms to assign "signatures" to them for job retrieval. Developed heuristics for word-sense disambiguation and for automatically determining term specificity.
• Tried a community detection approach to document clustering.
• Applied probabilistic topic modeling techniques such as LDA (Latent Dirichlet Allocation) and HDP (Hierararchical Dirichlet Processes) available in the gensim package to find the major themes in a large job

description corpus and use the model for Information Retrieval. Also experimented with methods that combine LDA with word2vec (Topical Word Embeddings,

• Above computations were done with Spark/Scala and PySpark in Databricks notebooks and AWS S3. Gained experience with Spark DataFrames, RDDs and Spark SQL.

• Extensively used python machine learning and NLP stacks (scikit-learn, nltk, scipy, numpy as well as newer libraries like spaCy and chainer - a python neural network library with CUDA and GPU computation support) plus open source Java libraries like OpenNLP, Stanford Core NLP and GATE.

• Developed a gold standard of responses to a carefully engineered set of queries and a random sample of job descriptions to evaluate search engine versions rapidly and without expensive and time-consuming A/B testing.

• Tried developing folksonomy-style tagging methods for documents. In this context, experimented with keyword extraction techniques (Kea, Maui-indexer and KP-Miner).

• Correlated click-through data with presented jobs and combined this with clustering of word neighborhood graphs to find jobs likely to be clicked on.

Skills Used: research aptitude, machine learning algorithms, document clustering, text classification, graph clustering, neural networks, word2vec, lda2vec, spaCy, chainer, pyLDAvis, NP-MSSG, statistical NLP, python, nltk, scikit-learn, numpy, scipy, Spark, Scala, Spark MLLib, Databricks, Spark Data Frames and Datasets, SQL, MySql, AWS, parquet files, gensim package, WordNet, Stanford Core NLP, OpenNLP, Solr, LDA, HDP, IR, Information Retrieval

Senior Software Engineer

Cisco   September 2014 to June 2015

Online Search
Searching for products that are relevant to a customer by looking at product descriptions in natural language in addition to structured data about them.

• Applied recent research on neural-net generated distributed, dense vector representations of words and phrases in experiments to understand the context and intent of a user query by mining a hitherto unexploited corpus of descriptions of ~1M products sold online by Cisco.

• Used word2vec to overcome vocabulary mismatch by suggesting related search terms with the objective of improving online customer experience on cisco.com and increasing conversion rates by an order of magnitude.

• Devised and selected algorithms that scale to millions of product descriptions.

• Categorized and provided insight into the reasons for "No Results Found" pages by mining query logs containing tens of millions of unique queries. Assessed the potential impacts of better spellchecking, model number recognition, automatic rephrasing of queries on the customer's experience and conversion rate.

• Evaluated spell checkers like aspell (with Metaphone 3), hunspell, LingPipe (based on the noisy channel model), and homegrown hybrids thereof to correct spelling errors taking phonetics and context into account and using custom dictionaries.

• Discovered a way to use word2vec for correcting spelling errors in O(1) time.

Technologies Used: Python, Java, C/C++, bash, Linux, cygwin, awk, sed, Maven, Ant, ontologies, OWL, RDF, Protégé, OpenRDF, WordNet, neural networks, word2vec, clustering, k-means, kNN, R, Dragon Toolkit, aspell, Hunspell, Jazzy, LingPipe, ARK TurboParser Dependency Parser, Stanford NLP, GATE, OpenNLP, Named Entity Recognition, Statistical NLP, TF-IDF, Jaro-Winkler, Levenshtein distance, fuzzy search algorithms, recommendation systems, collaborative filtering, Named Entity Recognition (NER), POS tagging.

Senior Software Engineer
Health Data Vision   Nov'13 to Aug'14

Social Network Analytics
Analyzing local neighborhood structure of social network nodes in a graph-theoretic way to discover and quantify similarities between them.

• Developed a highly scalable and fast technique for analyzing and characterizing roles of individuals within large social networks, by importing ideas stemming from the analysis of protein interaction networks in bioinformatics. This innovative application of graphlets to social networks with ~105 edges is able to precisely identify in a matter of seconds individuals who play similar roles to a single exemplar. It made a US Navy project for identifying potential terrorist threats in a large social network enormously successful and is now part of the core IP of 21CT.
• Employed R packages for principal components analysis, k-means clustering and decision trees to analyze results of using graphlet methods on Facebook100, a complete set of Facebook friendship data from 100 American Universities in 2005.
• Implemented the graphlet application in C++ as well as Java for incorporation into company codebase as a Maven project.
• Participated in a project to study collective entity resolution by fusing network data coming from sources in different modalities. System is aimed at coalescing multiple monikers belonging to the same individual.
• Gained experience working on DoD SBIR research projects with tight deadlines.
• Created a small OWL ontology with RDF n-triples using Protégé and Sesame. Experimented with Rya, a distributed RDF repository on top of the Accumulo key-value store. Generated and ran SPARQL queries against the repository.
• Converted a group detection algorithm to MapReduce, using the Cascading abstraction layer on top of Hadoop.
• Worked with several Python scripts and libraries as well as R packages for classification, clustering, principal components analysis and visualization.
Tools Used: Terrorism Intelligence Analytics, DoD contracts, Social Network Analysis, Java, Maven. C+ +, NoSQL, Accumulo, R, principal components analysis (PCA), machine learning, Python, iPython, Scipy, Numpy, Eclipse, Netbeans, Cytoscape, graphlets, graph mining, RDF, SPARQL, OpenRdf, ontologies, OWL, Protégé, Sesame, Hadoop, Cascading, MapReduce, Big Data, Cloud, Linux, cygwin, bash, sed, awk, svn, Agile, SCRUM, software integration..

Senior Software Engineer
Applied Materials - July'12 to Oct'13

Information Retrieval from free text databases
Retrieving legal documents relevant to a litigation with high precision and recall expanding queries where needed and dealing with "vocabulary mismatch",

• Researched and implemented some of the latest IR techniques for query suggestion, relevance feedback and ranked retrieval to modernize and differentiate the company's two main products in the eDiscovery marketplace.
• Experimented with Latent Semantic Indexing (LSI) as implemented in the "semantic vectors" package to creates models which represent collections of documents in terms of underlying concepts.
• Enhanced components which are written in Java, Ruby and C#, use MongoDB, MySQL and SQL Server databases and communicate via SOAP/REST web services. Technologies employed include Apache Lucene and Solr (for free text search), JBoss, Spring and Maven
Skills Used: C++, Boost, g++, cygwin, Visual C++, Java, JBoss, Maven, Spring, Svn, QuickBuild, web services, SOAP, REST, XML, Big Data, SQL, NoSQL, MongoDB, Agile, SCRUM, Rally, Applied Research in Information retrieval (IR), TF-IDF, machine learning, algorithm design and implementation, universal hash functions, Bloom filters, performance analysis and optimization, Eclipse, Mockito, Junit, document classification.

Consulting Software Engineer

Accenture -   Oct'09 to June'12

VoIP Phones

Refactored and completely re-implemented the Qt 4.6/C++-based Network Access Controller eliminating critical bugs and memory leaks in next-generation phones.

• PayPal, Austin, TX, Infrastructural Software

Modified enterprise-wide C++ software to use a standard version of the Xerces XML Parser. Led a pilot project to prevent buffer overflow, code injection and other vulnerabilities in PayPal software by introducing Fortify, a static analysis tool into the development process.

• Advanced Micro Devices, Austin, TX, CPU Diagnostics

Developed a remote diagnostics tool using the XML-RPC protocol.

• IBM, Austin, TX, AIX Kernel Technical Support

Interfaced with IBM customers worldwide as well as AIX kernel developers to resolve code defects in the loader and linker.

Tools and Skills: C, C++, Qt 4.6, Ubuntu Linux, Embedded Linux, CentOs Linux, Visual Studio 2008, Eclipse IDE, KDevelop, Subversion, CVS, Perforce, Rational ClearCase/ClearQuest, Jira, XML, Xerces DOM Parser, Agile methodology, SCRUM, MVC architecture, design patterns, multi-threaded systems, compiler front-end, XML-RPC protocol, http, ftp, AIX, bash, ksh, gcc, gdb, Oracle VirtualBox, network programming, sockets, client interaction

Engineer

Capgemini   July06 to Sep'09

Embedded compiler in C and message-oriented middleware in C++ for a memory tester

• Facilitated the use of the M1000 high-end memory tester by defining a custom language and implementing an embedded compiler (target - PPC405) for it using flex, bison and Gnu crosstool on Linux.

• Developed and deployed multi-threaded middleware in Visual C++ for a distributed system with socket communication between modules.

Skills: C, gcc, g++, flex, bison, compilers, GNU crosstool, embedded Linux, cygwin, Visual C++, sockets, multi-threaded programming, UML, use-case scenarios, SQL, Database Template Library (DTL)

EDUCATION

(PhD) in Artificial Intelligence & Deep Learning

M.Tech in Computer Science / AI

Institute for Advanced Analytics , USA

B Tech. in Computer Science

University of JNU , India