Natural Language Classifier
Overview

Watson

November 2018

Contents

Every minute:

150K emails sent **450K tweets** **29M WhatsApp messages**

From messages to support tickets, text classification is at the core of natural language processing.

The key to understanding this data is natural language classification

Common Pain-points

Natural Language Text Processing

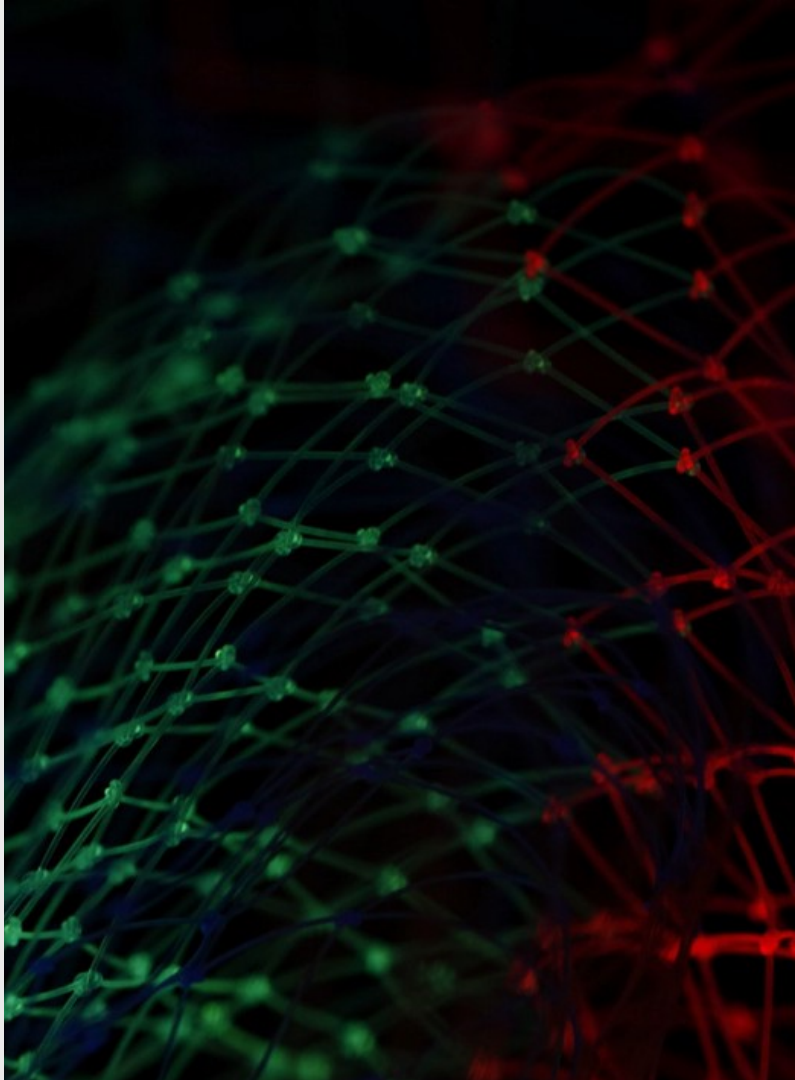
With the amount of data out there, is there a way I can categorize it all into classes

Customizable Classes

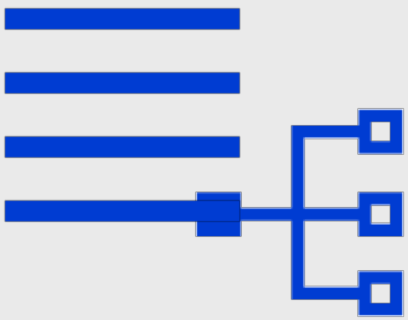
Current classification tools have predefined classes that I cannot customize

Scalable Models

There is too much data out there for my own model to handle. Do I have to be a developer?



Watson Natural Language Classifier is



A custom text classification service that enables users to quickly and accurately create, train, and evaluate text classifiers using machine learning.

Watson Natural Language Classifier provides

Tooling and API

Developers can leverage the NLC API for easy integration into applications in their preferred language.

Not a developer? No problem! Use the Watson Studio tooling to create, train, and evaluate your custom classifiers.

Customization and flexibility

Watson Natural Language Classifier allows you to train and build your own classifiers at scale.

Customize your models for various industry verticals from finance to social media.

Privacy & Security

IBM ensures that the data passed through our service remain secure on the IBM Cloud.

The IBM Cloud offers enterprise grade security, around the globe support, and multi-region deployments.

Natural Language Classifier features

Training

Build custom classifier models via tooling to categorize text phrases

Language Support

Create classifiers in: English, Arabic, French, German, Italian, Japanese, Korean, Portuguese (Brazilian), and Spanish

Deep Learning

Upload larger and **Train** faster datasets for training leveraging GPUs. Each .csv file can contain up to 20,000 rows

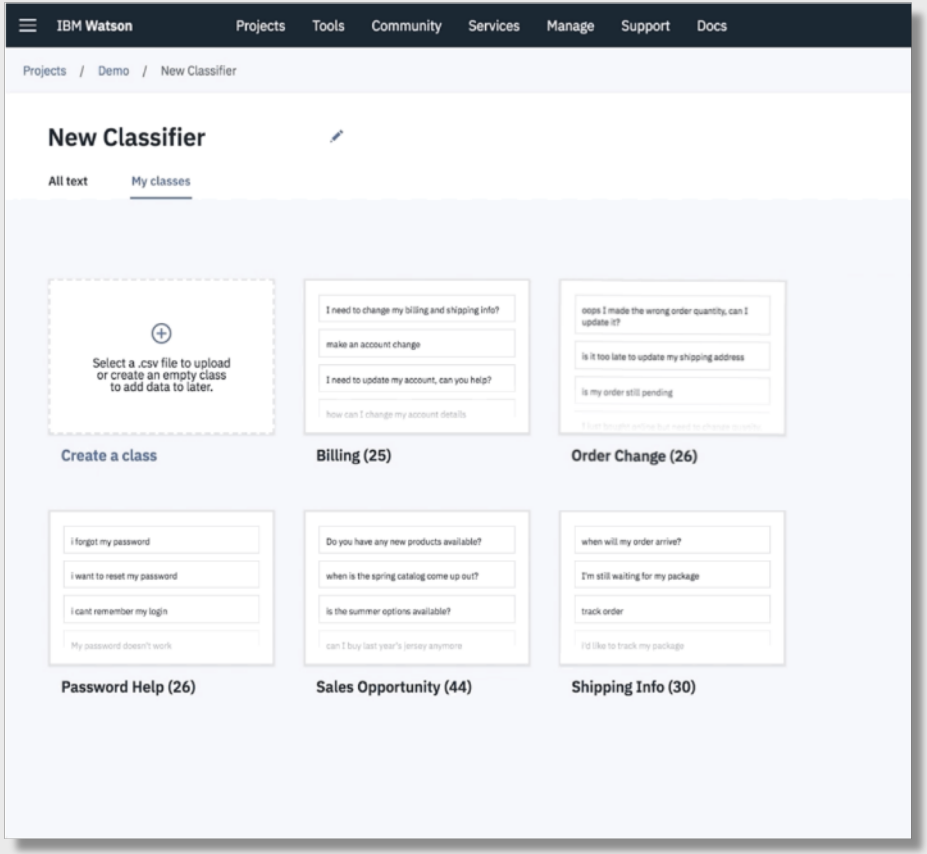
Multi-Phrase Classification

Simplify categorization by viewing confidence scores for multiple categories

Natural Language Classifier Tooling

Watson Studio

A suite of tools for data scientists, application developers and subject matter experts to collaboratively and easily work with data and use that data to build, train and deploy classifiers at scale



Learn how a subject matter expert can leverage the free Watson Studio tooling to train for a support ticket classification use case

YouTube Video | **IBM Webinar**

Industry Vertical Applications

Identify potential risks from online data sources

Simplify your workflow by classification of e-mails

Organize tweets, posts, and messages into categories

Analyze documents by categorizing them into user defined sections

Categorize support tickets and messages to identify issues

and many more...

Client Use Cases | IBM Talent: HR & Recruiting

Need

Derive soft skills & interpret phrases on resumes. Need to a way to search resumes for soft skills and key competencies not relating to specific terms

Solution

Used NLC to parse resumes. Input phrases from a resume and classify them into categories: Collaborative, Engaging, etc.

Benefit

Further automate resume scanning with the ability to search and extract key soft skills and competencies.

Free recruiters time to allocate toward interviews

Client Use Cases | Influential Network

Need

Firms that market using social media influencers can't guarantee demographic targeting and timed messaging. Influential sought a solution that could help agencies address such issues.

Solution

A demographic analysis & mobile solution helping influencers authentically amplify their marketing message throughout various social media campaigns.

Benefit

Analyze 5B follower accounts with ease, while streamlining daily workflows

Ensure influencers social voice and followers maximize marketing results

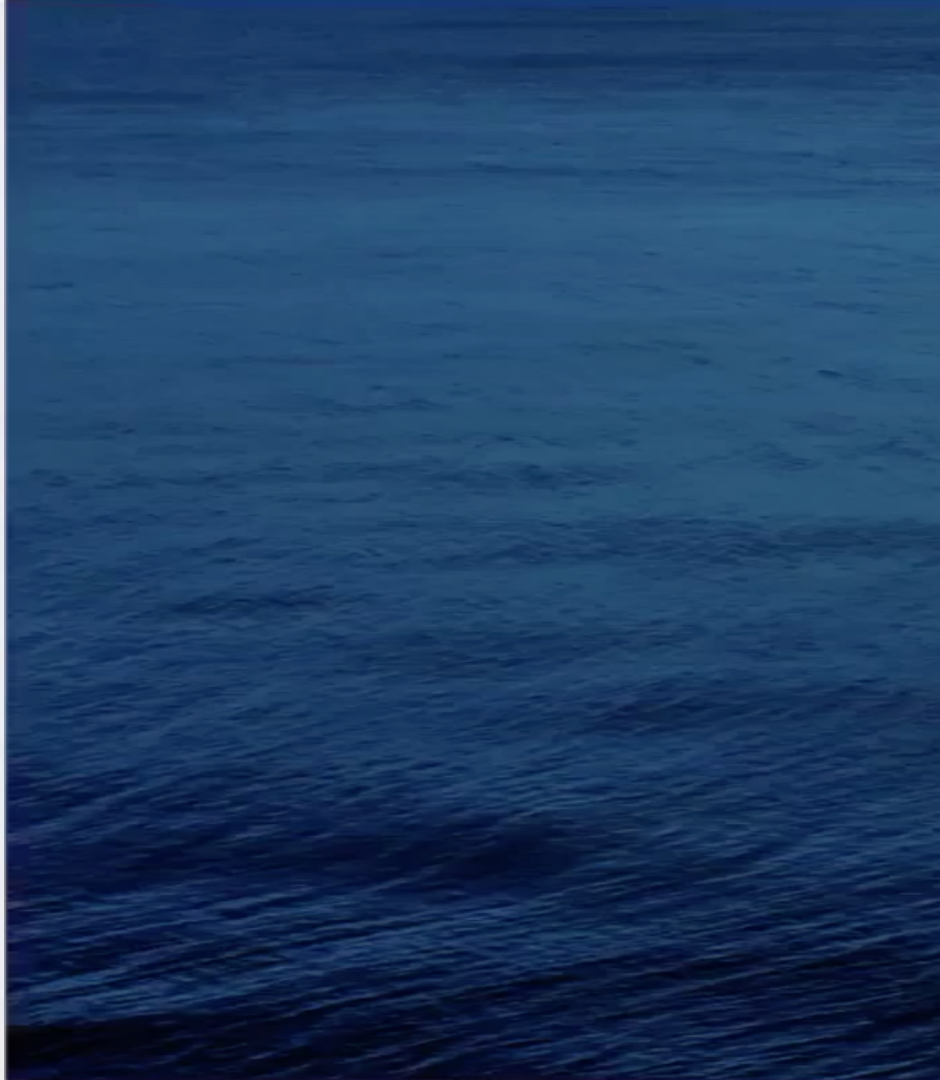
Deep Learning

Word Vectors

Background word embedding's on large general-domain corpus (Wikipedia) through unsupervised training (word2vec)

Adapt

Word embedding's with in-domain data (customer-provided) through online incremental training



Core NLC Models

Convolutional neural network (CNN)

Text classification with word embedding as input along with other ensemble models

IDFSVM

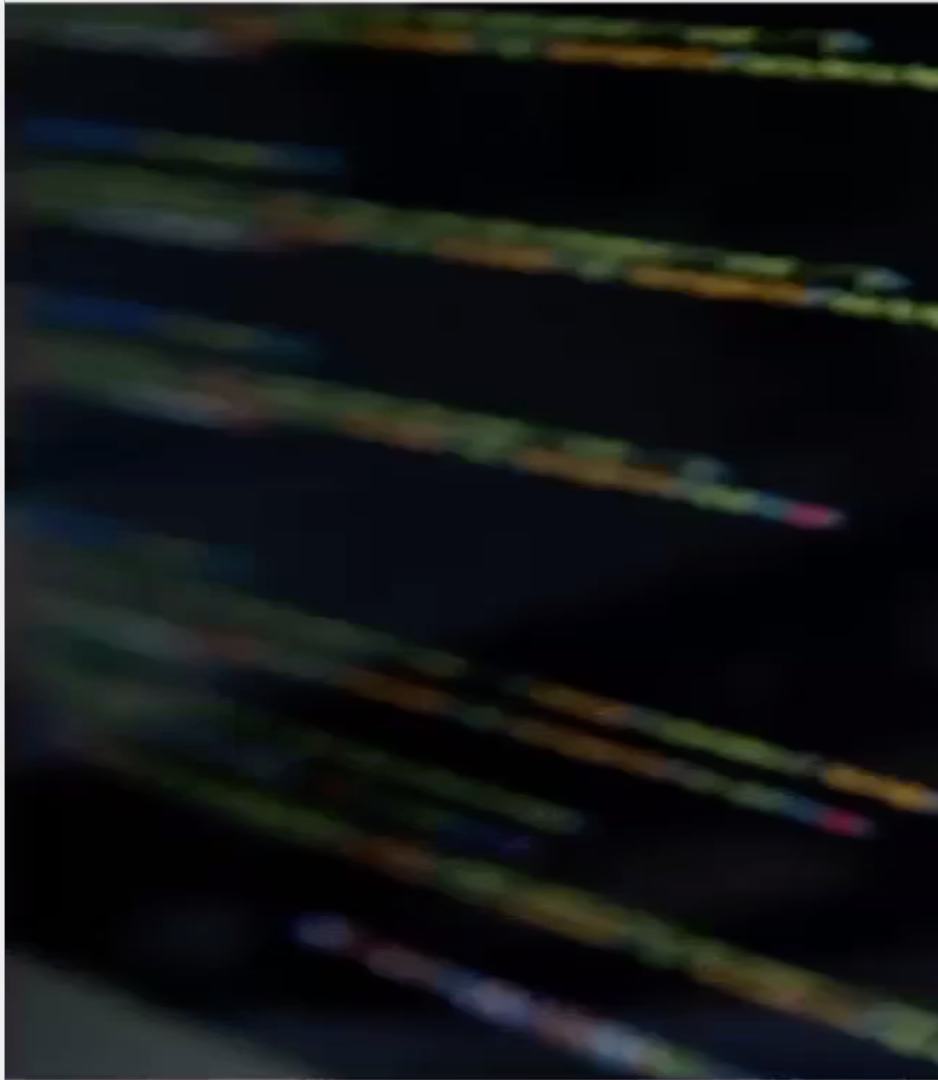
SVM-based classifier that uses TF-IDF features

BOWSVM

SVM-based classifier that uses bag-of-words features

MB

Learning to rank classifier with word features



CNNs for Sentence Classification

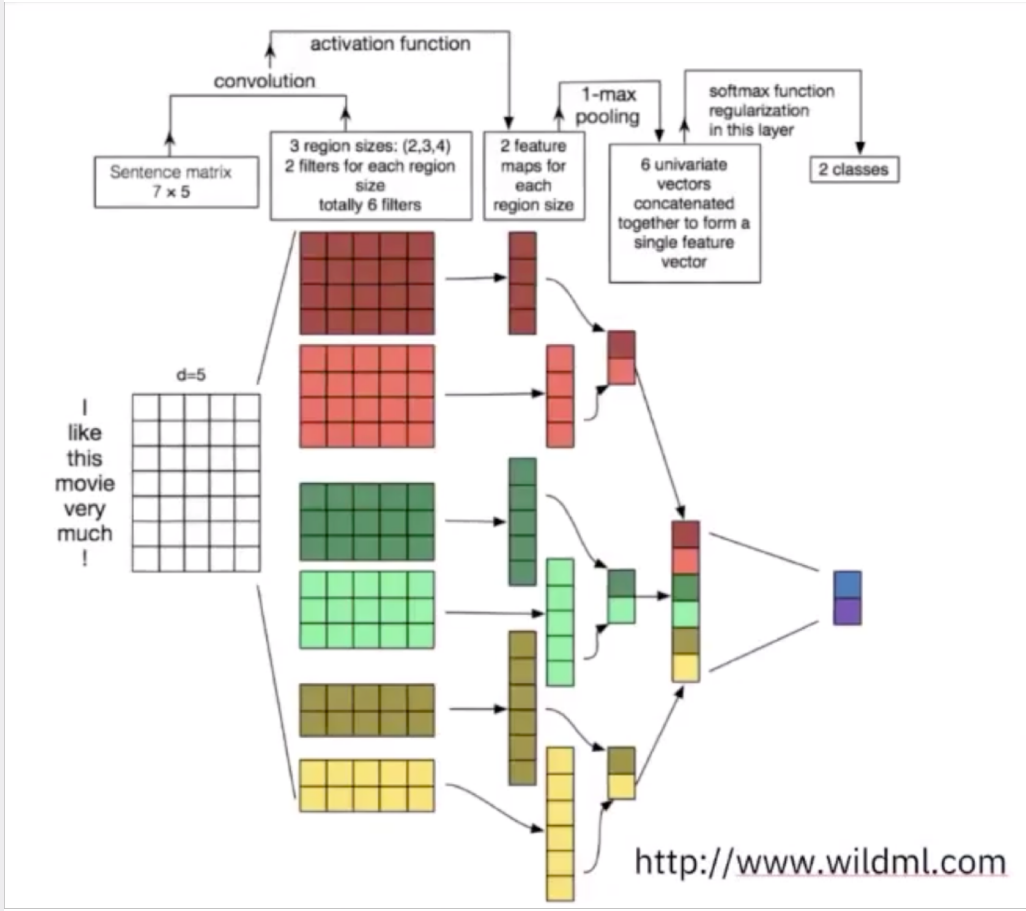
Commonly used for Computer Vision

CNNs work quite well for NLP tasks while providing state of the art performance in product.

Often slow to train with large amounts of training data without accelerators (GPUs).

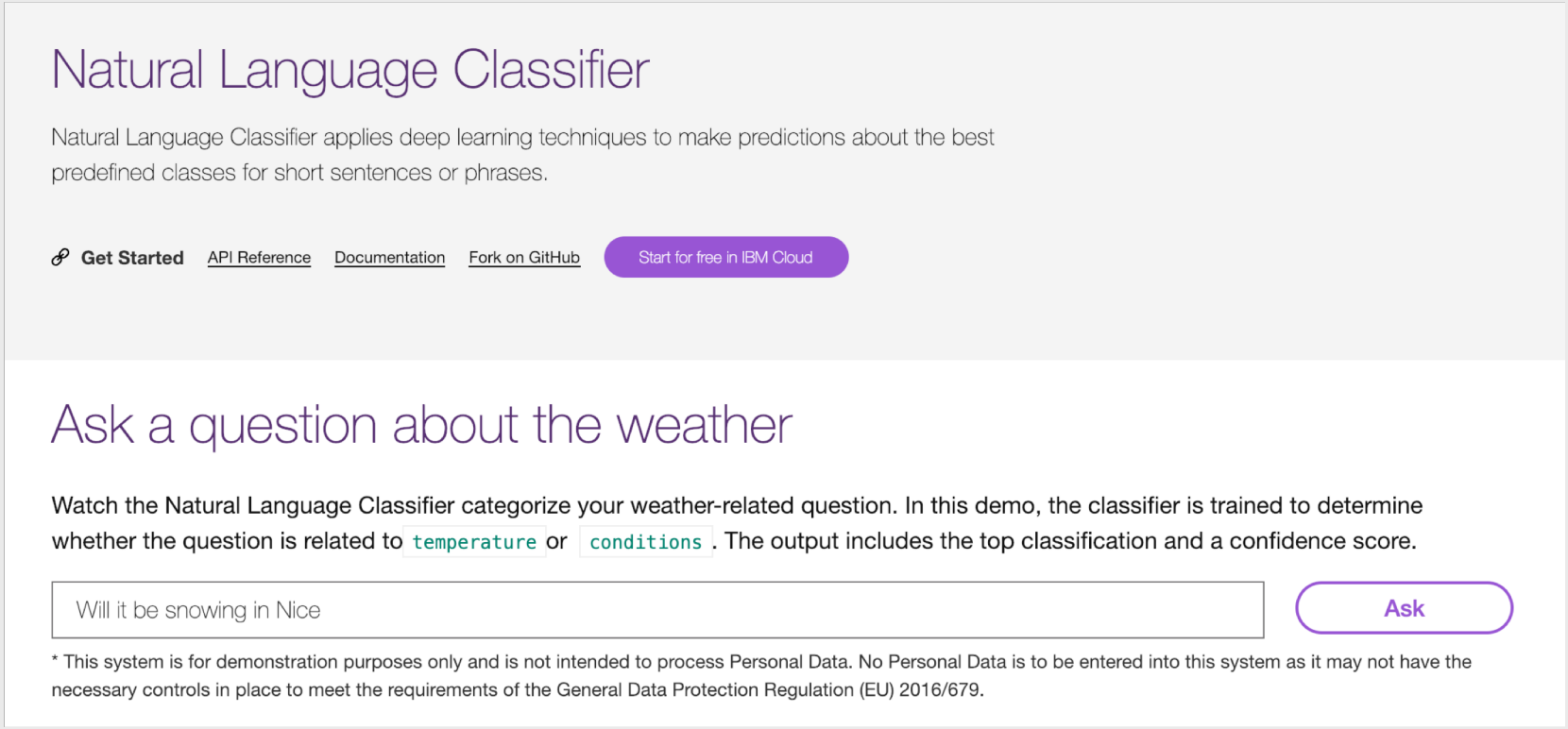
Convolution operations

Implemented at hardware-level on GPUs consumed through new IBM DLaaS service.

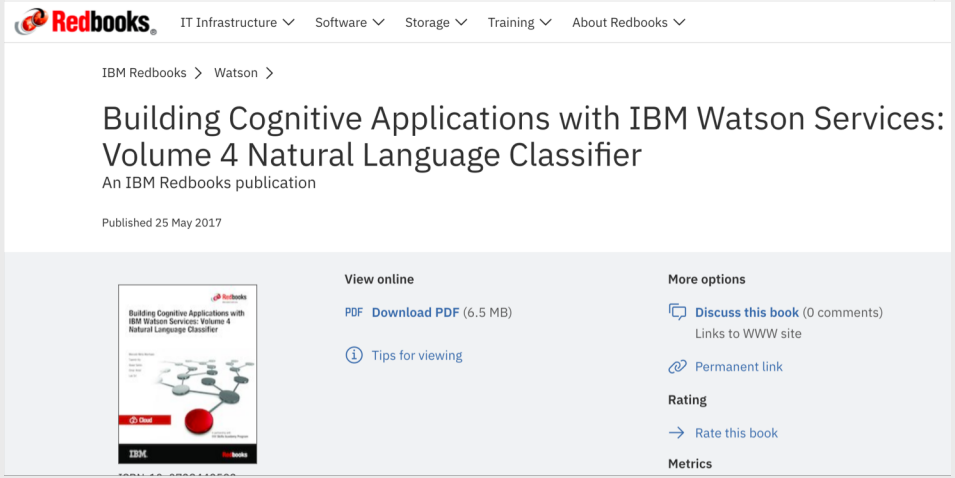


Demo

https://natural-language-classifier-demo.ng.bluemix.net/



Resources



http://www.redbooks.ibm.com/redbooks/pdfs/sg248391.pdf