Recognai

Natural Language Processing with neural nets

Paco Nathan

Daniel Vila Suero

Recognai

NLP concepts

NLP progress

Trends

Recognai

3 exercises

Data preparation for NN

NN from scratch

Modular NN

Software 2.0

Recognai

About membership

Medium

Sign in Get started

Andrej Karpathy Follow
Director of AI at Tesla. Previously Research Scientist at OpenAI and PhD student at Stanford. I like to train deep neural nets on large datasets.
Nov 11 · 7 min read

Software 2.0

I sometimes see people refer to neural networks as just “another tool in your machine learning toolbox”. They have some pros and cons, they work here or there, and sometimes you can use them to win Kaggle competitions. Unfortunately, this interpretation completely misses the forest for the trees. Neural networks are not just another classifier, they represent the beginning of a fundamental shift in how we write software. They are Software 2.0.

The “classical stack” of **Software 1.0** is what we’re all familiar with—it is written in languages such as Python, C++, etc. It consists of explicit instructions to the computer written by a programmer. By writing each line of code, the programmer is identifying a specific point in program space with some desirable behavior.

https://medium.com/@karpathy/software-2-0-a64152b37c35

Where we are

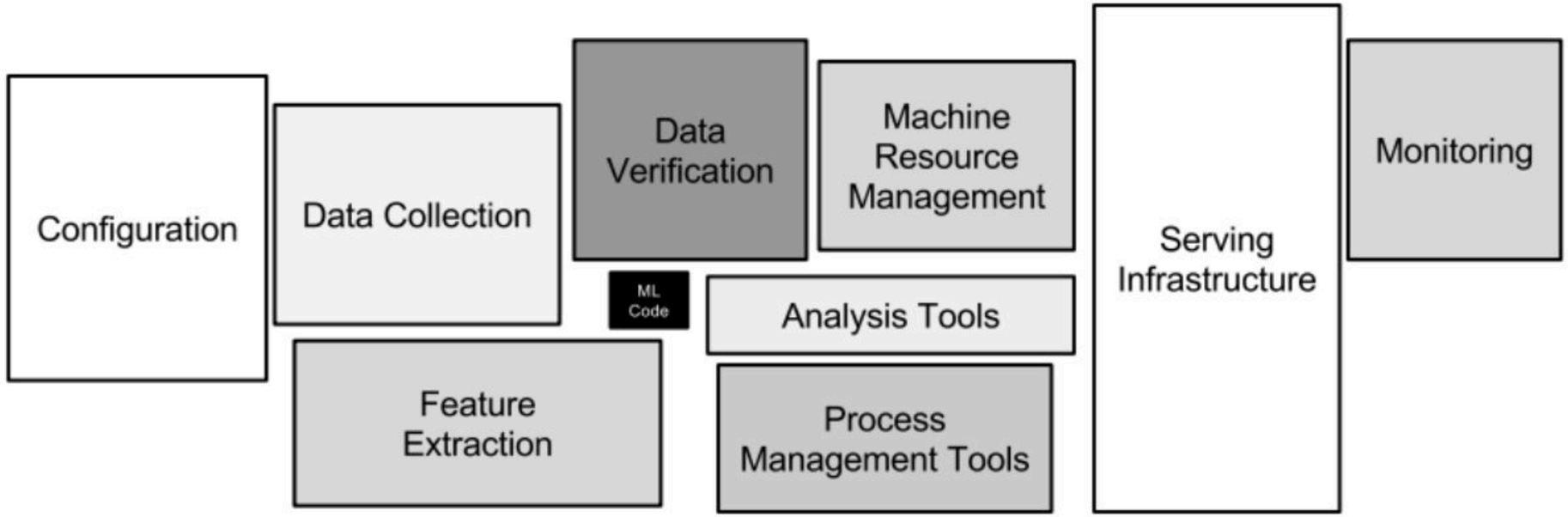
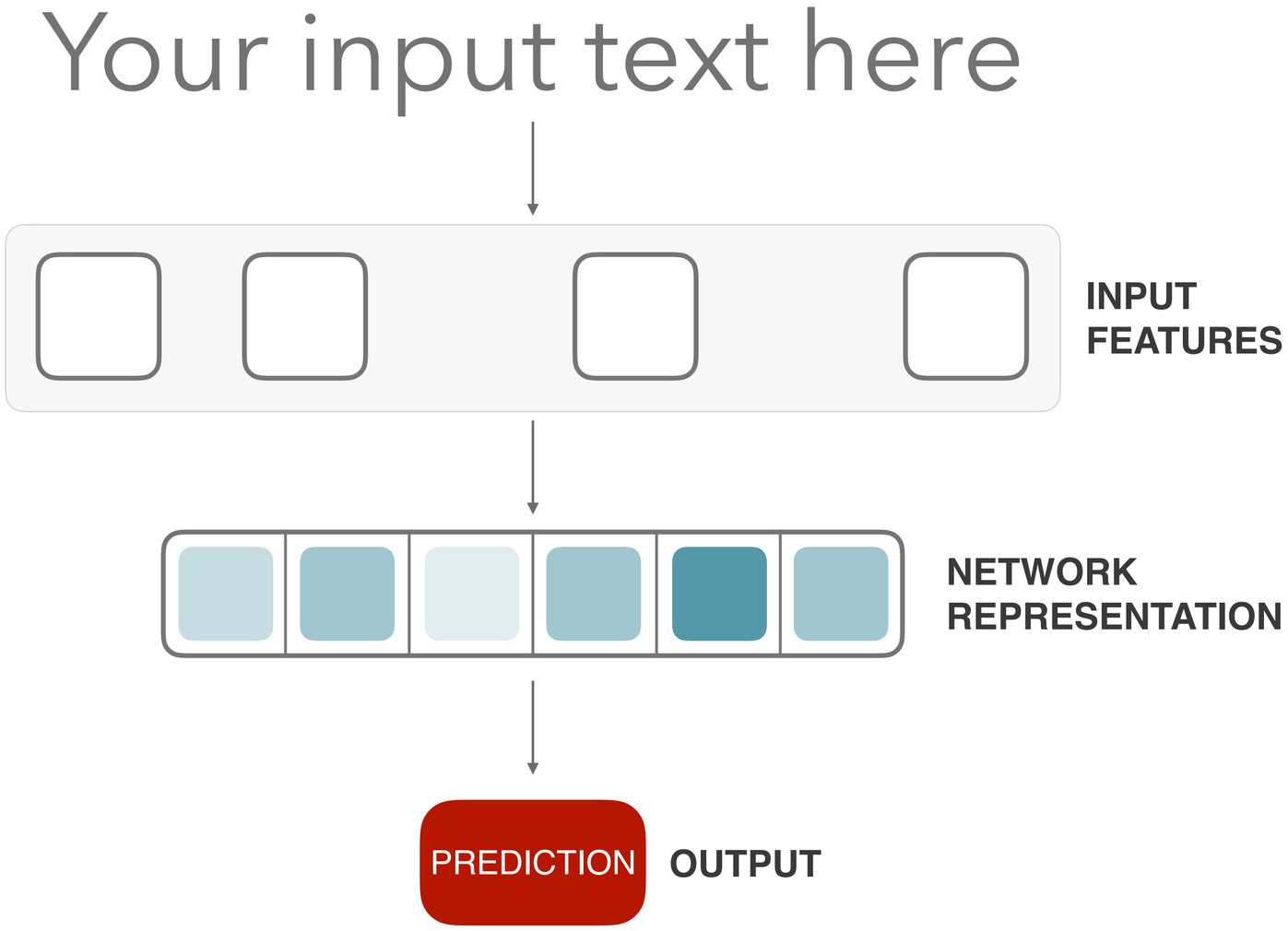
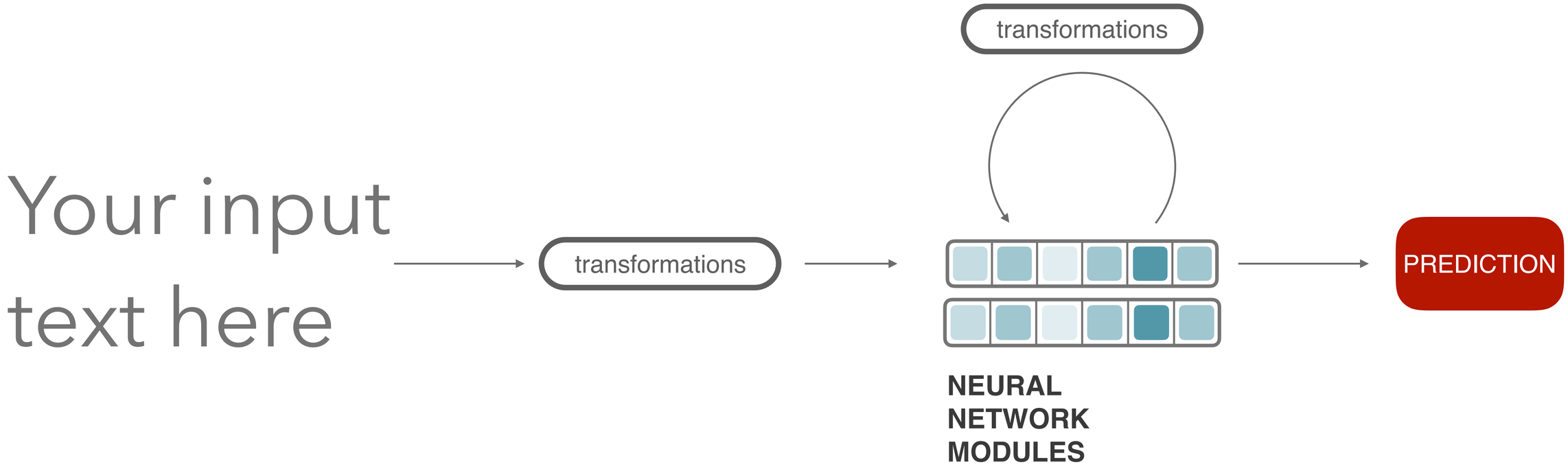


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

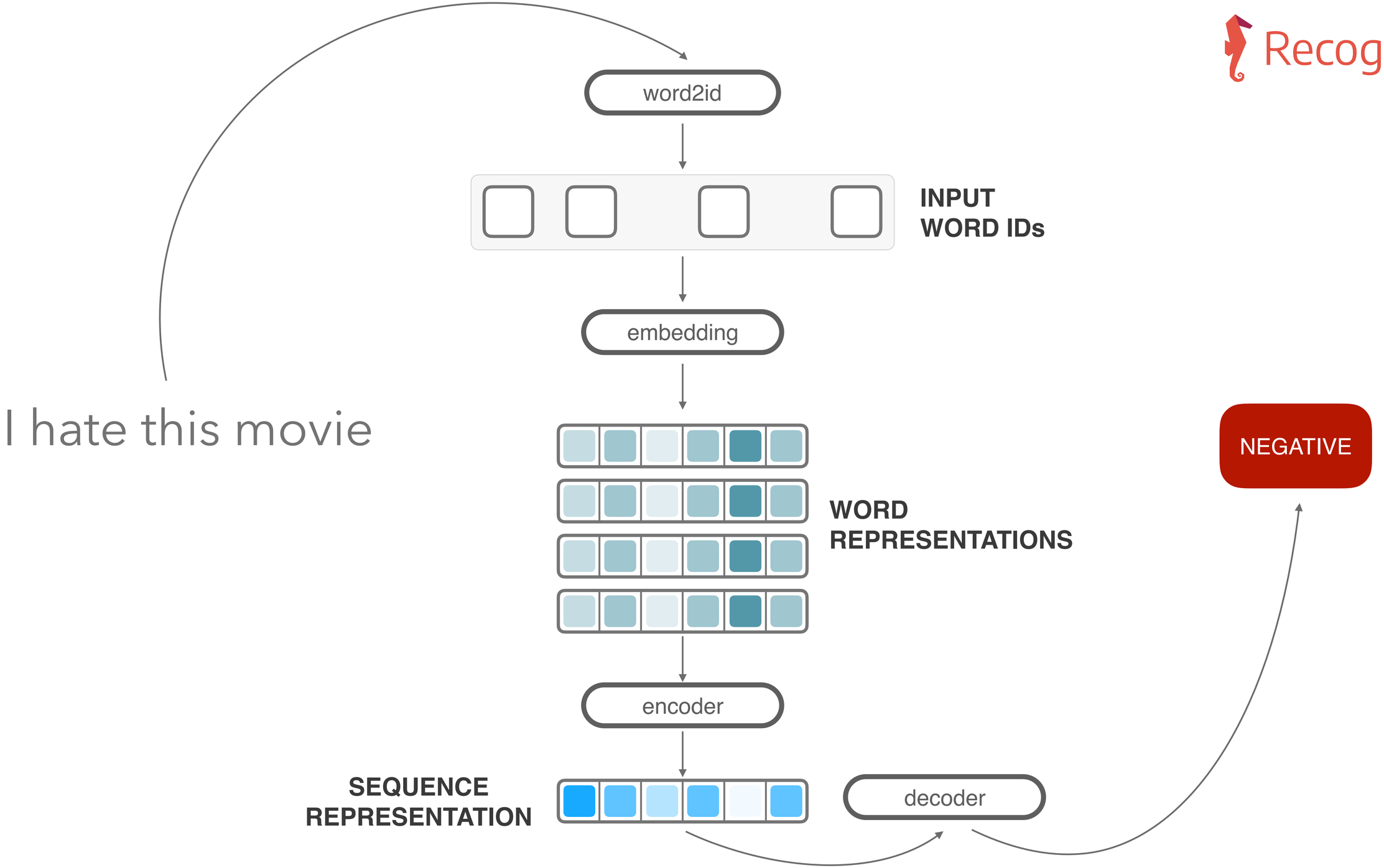
Basic flow


Your input text here
INPUT FEATURES
NETWORK REPRESENTATION
PREDICTION
OUTPUT

Another perspective



- **The goal is to learn useful features by transforming the input text**


Recognai
word2id
INPUT
WORD IDs
embedding
I hate this movie
WORD
REPRESENTATIONS
NEGATIVE
encoder
SEQUENCE
REPRESENTATION
decoder

What is our input?

Recognai

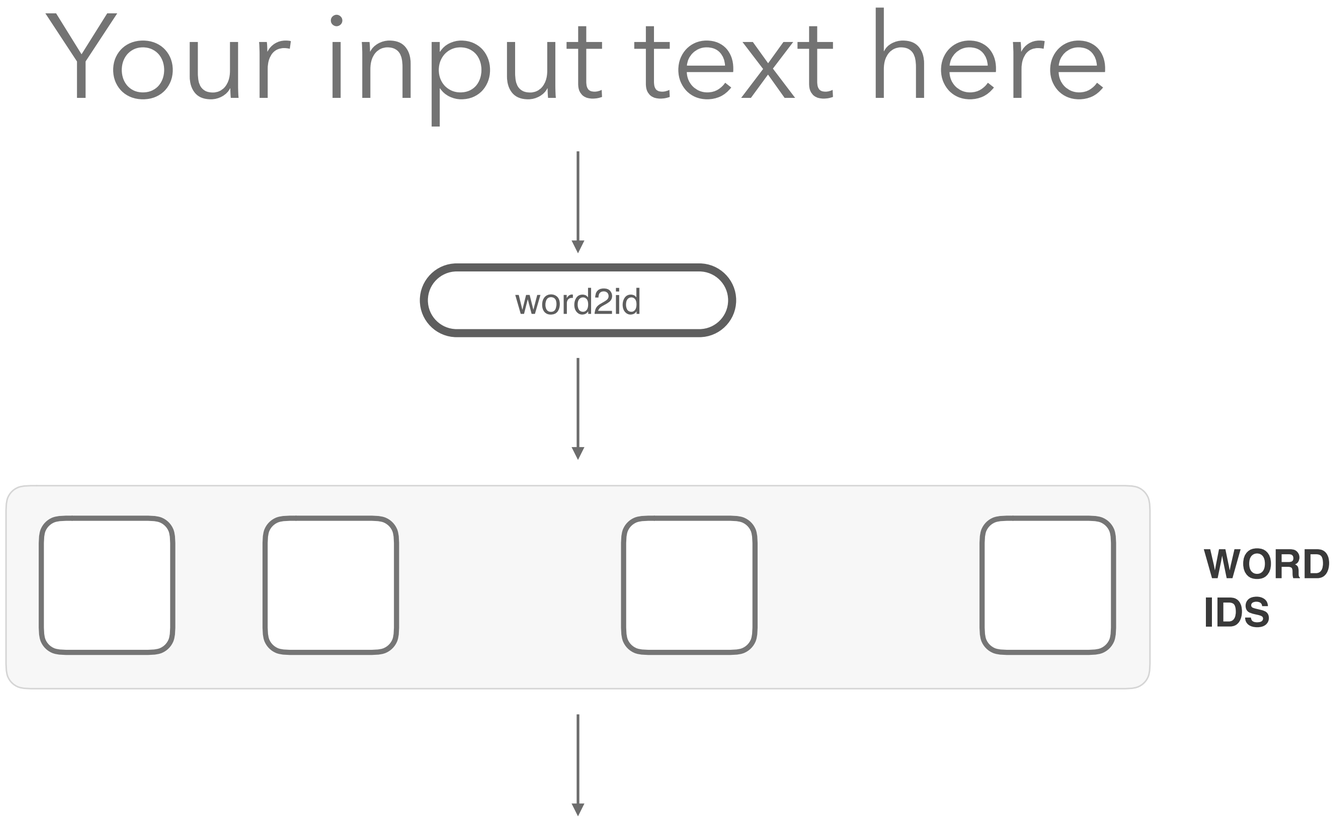
- **Input:** text as a sequence of words (symbols)

I hate this movie

INPUT

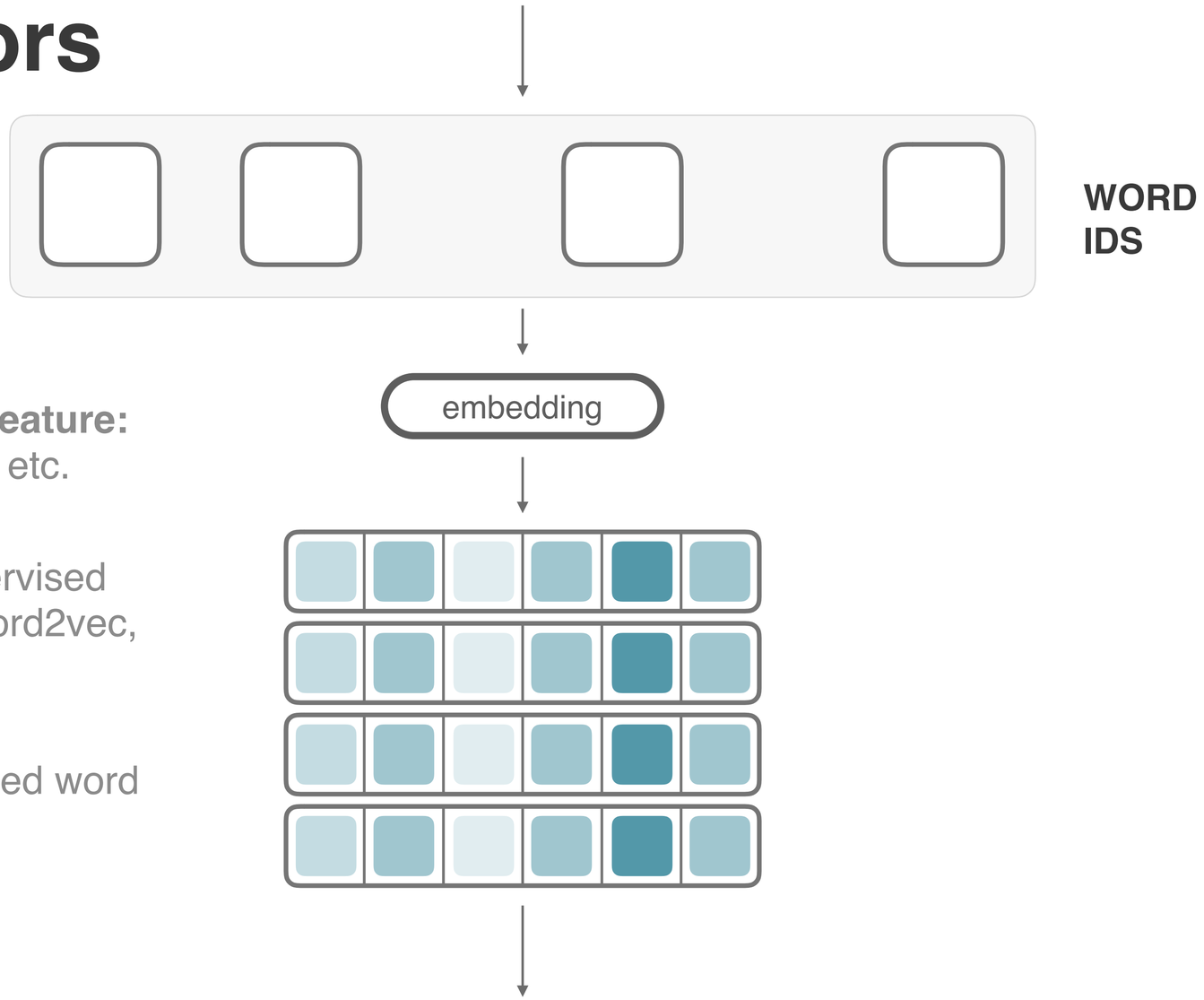
- **Examples:** Word IDs, Char IDs, POS tags, shapes

Transformations:
Text to IDs



- **Key concepts:** Tokenization and vocabulary for mapping words/chars/.. to IDs

Transformations:
IDs to vectors



- We can **embed any feature:** word id, shape, POS, etc.
- Different from unsupervised word embeddings (word2vec, GloVe)
- We can load pre-trained word embeddings.

Recognai

Exercise 1: Getting your data NN ready

Types of output

Recognai

- **Output:** a single label

I hate this movie

INPUT

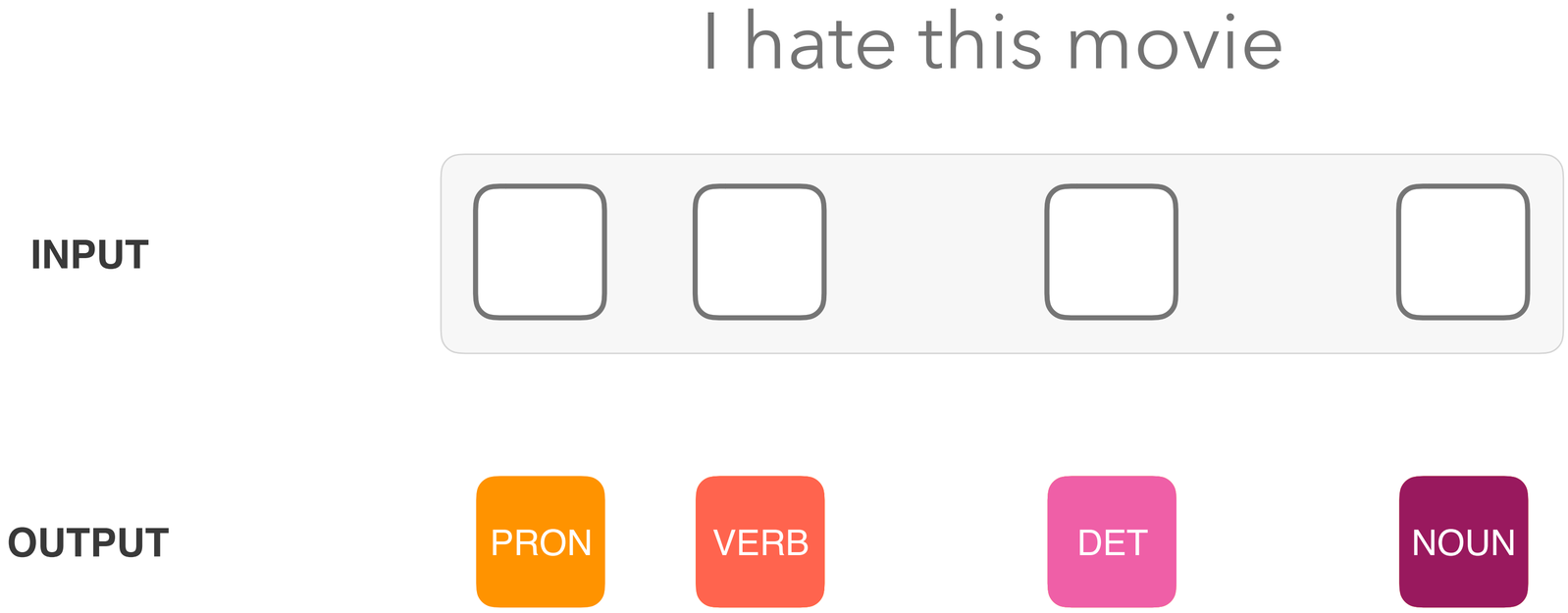
OUTPUT

NEGATIVE

- **Examples:** Sentiment, Topic, any classifier

Types of output

- **Output:** a sequence of tags



- **Examples:** Part of speech tagging, named entity recognition

Types of output

- **Output:** a sequence of words (symbols)

I hate this movie

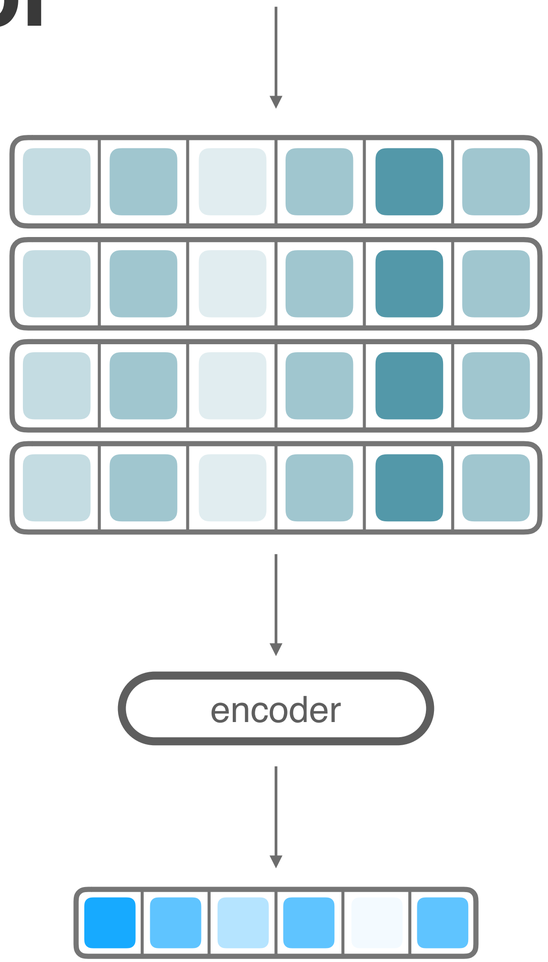
INPUT

OUTPUT

Odio esta película

- **Examples:** Machine translation, summarization, language models

Transformations:
Sequence to vector

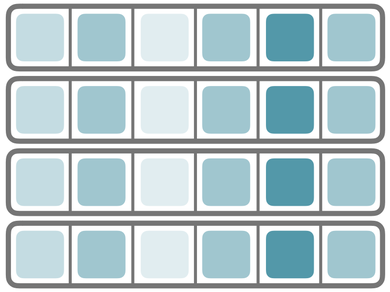


- **Hopefully encodes a good representation of the input sequence**

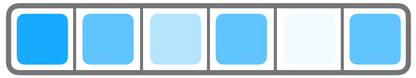
Transformations:
Sequence to vector encoders

Recognai

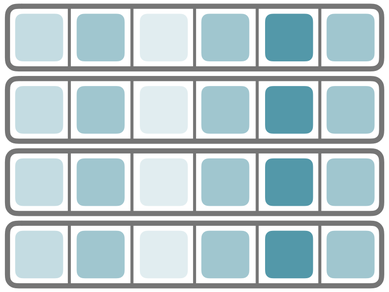
Bag of embeddings



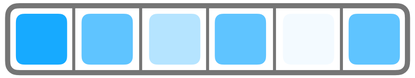
SUM/AVG



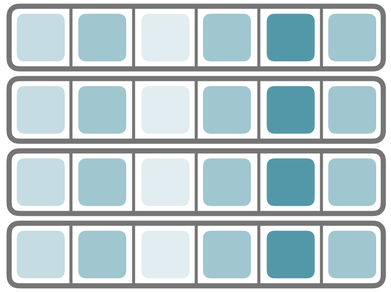
CNN encoder (N-gram detector)



Sliding window + POOLING



RNN encoder (recurrent)



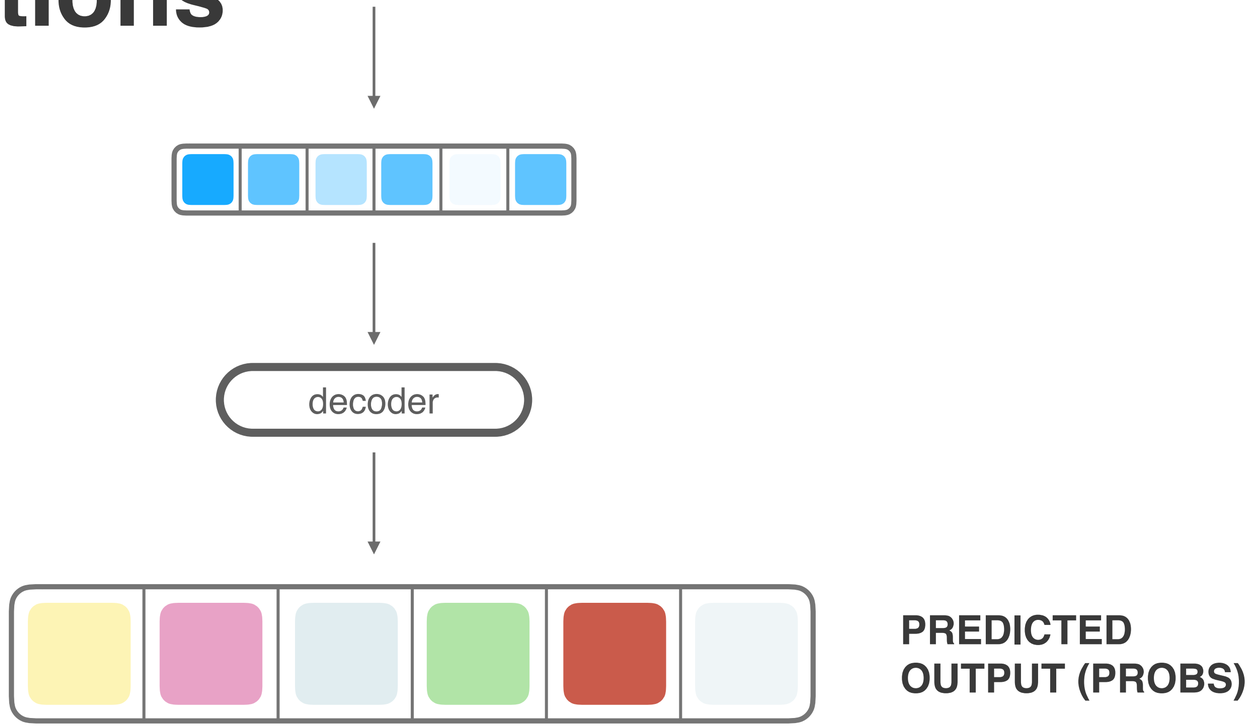
Hidden/output state

No order

Unlimited?

Ability to encode sequential ordering

Transformations:
Vector to predictions



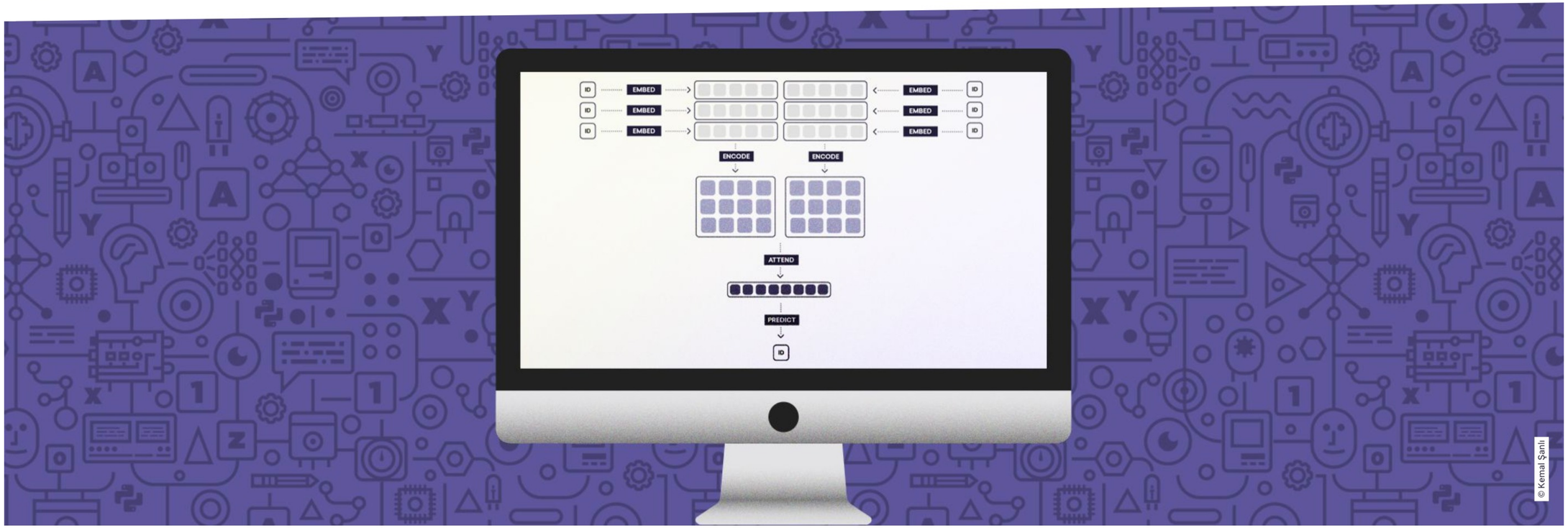
- **For classification: typically a linear layer + softmax**

Deep-learning formula

Recognai

EXPLOSION

Home About us **Blog** Request



Embed, encode, attend, predict: The new deep learning formula for state-of-the-art NLP models

BY MATTHEW HONNIBAL ON NOVEMBER 10, 2016

https://explosion.ai/blog/deep-learning-formula-nlp

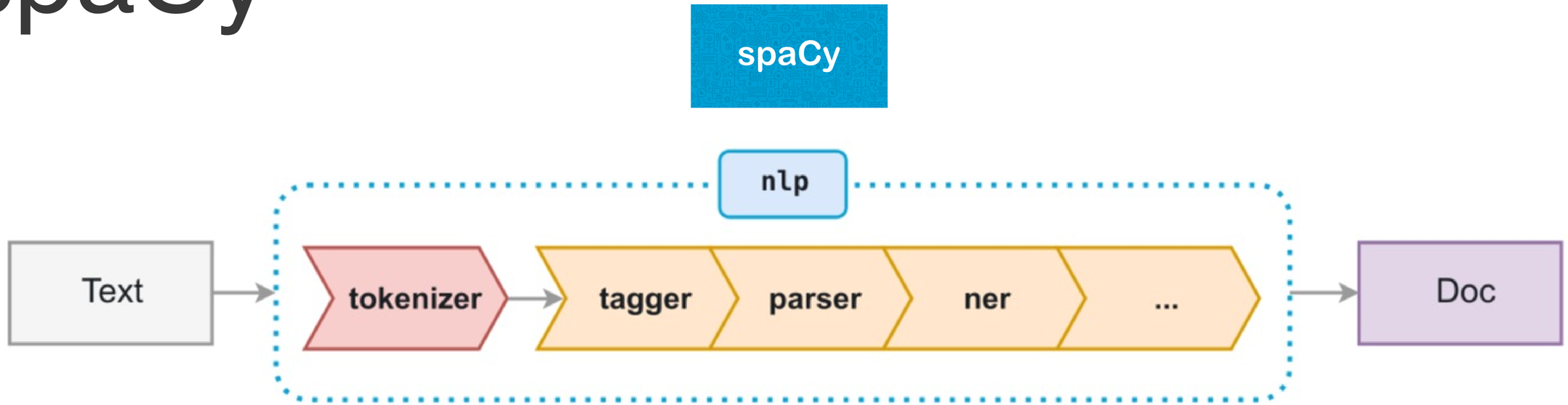
Recognai

Exercise 2: Our first NN in PyTorch

Recognai

Tooling

spaCy



- High quality **documentation**, focus on **production** (and research)
- Good integration with **deep learning libraries** (e.g., PyTorch, AllenNLP).
- **Training and fine-tuning** new models is a **not an after-thought.**

fast.ai v1

Recognai

fastai | Nav | GitHub | News | search...

- Training
- Applications
  - Overview
  - vision
  - text
    - Overview
    - text.learner
    - text.transform
    - text.data
    - text.models
  - tabular
  - collab
- Core
- Doc authoring
- Library development

text

Summary: Application to NLP, including ULMFiT fine-tuning

Table of Contents

- Quick Start: Training an IMDb sentiment model with ULMFiT
  - Reading and viewing the IMDb data
  - Getting your data ready for modeling
  - Fine-tuning a language model
  - Building a classifier

Text models, data, and training

The `text` module of the fastai library contains all the necessary functions to define a Dataset suitable for the various NLP (Natural Language Processing) tasks and quickly generate models you can use for them. Specifically:

- `text.transform` contains all the scripts to preprocess your data, from raw text to token ids,
- `text.data` contains the definition of `TextDataset`, which the main class you'll need in NLP,
- `text.learner` contains helper functions to quickly create a language model or an RNN classifier.

Have a look at the links above for full details of the API of each module, of read on for a quick overview.

Quick Start: Training an IMDb sentiment model with *ULMFiT*

Let's start with a quick end-to-end example of training a model. We'll train a sentiment classifier on a sample of the popular IMDb data, showing 4 steps:

AllenNLP

Recognai

AllenNLP

- Machine Comprehension
- Textual Entailment
- Semantic Role Labeling
- Coreference Resolution
- Named Entity Recognition
- Constituency Parsing
- Dependency Parsing
- **Open Information Extraction**
- Event2Mind
- Your model here!

Open Information Extraction

Given an input sentence, Open Information Extraction (Open IE) extracts a list of propositions, each composed of a single predicate and an arbitrary number of arguments. These often simplify syntactically complex sentences, and make their predicate-argument structure easily accessible for various downstream tasks. The AllenNLP toolkit provides the following Open IE visualization, which can be used for any Open IE model in AllenNLP. This page demonstrates a reimplementation of **a deep BiLSTM sequence prediction model (Stanovsky et al., 2018)**.

Enter text or Choose an example...

Sentence

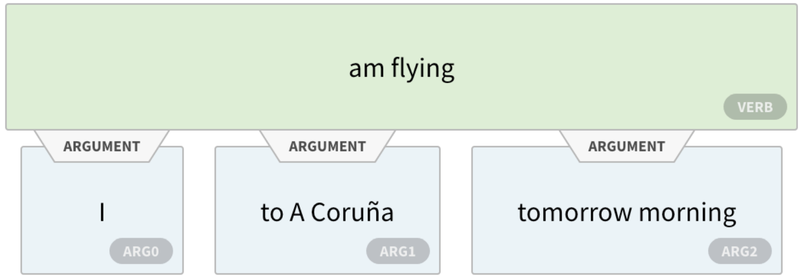
I am flying to A Coruña tomorrow morning

RUN

Tree Text

Verb 1 of 1: **am flying**

I am flying to A Coruña tomorrow morning



Recognai

flair

pypi package 0.2.1 | issues 24 open | contributions welcome | license MIT | build passing

A very simple framework for **state-of-the-art NLP**. Developed by Zalando Research

Flair is:

- **A powerful syntactic-semantic tagger / classifier.** Flair al
  entity recognition (NER), part-of-speech tagging (PoS), fra
  your text.
- **A text embedding library.** Flair has simple interfaces that
  embeddings. In particular, you can try out our proposed co
  art NLP methods.
- **A Pytorch NLP framework.** Our framework builds directly
  experiment with new approaches using Flair embeddings a

Now at version 0.2.0!

Comparison with State-of-the-Art

Flair outperforms the previous best methods on a range of NLP tasks:

Task	Dataset	Our Result	Previous best
Named Entity Recognition (English)	Conll-03	**93.09** (F1)	*92.22 (Peters et al., 2018)*
Named Entity Recognition (English)	Ontonotes	**89.71** (F1)	*86.28 (Chiu et al., 2016)*
Emerging Entity Detection (English)	WNUT-17	**50.20** (F1)	*45.55 (Aguilar et al., 2018)*
Named Entity Recognition (German)	Conll-03	**88.32** (F1)	*78.76 (Lample et al., 2016)*
Named Entity Recognition (German)	Germeval	**84.65** (F1)	*79.08 (Hänig et al, 2014)*
Part-of-Speech tagging	WSJ	**97.85**	*97.64 (Choi, 2016)*
Chunking	Conll-2000	**96.72** (F1)	*96.36 (Peters et al., 2017)*

Here's how to reproduce these numbers using Flair. You can also find a detailed evaluation and discu

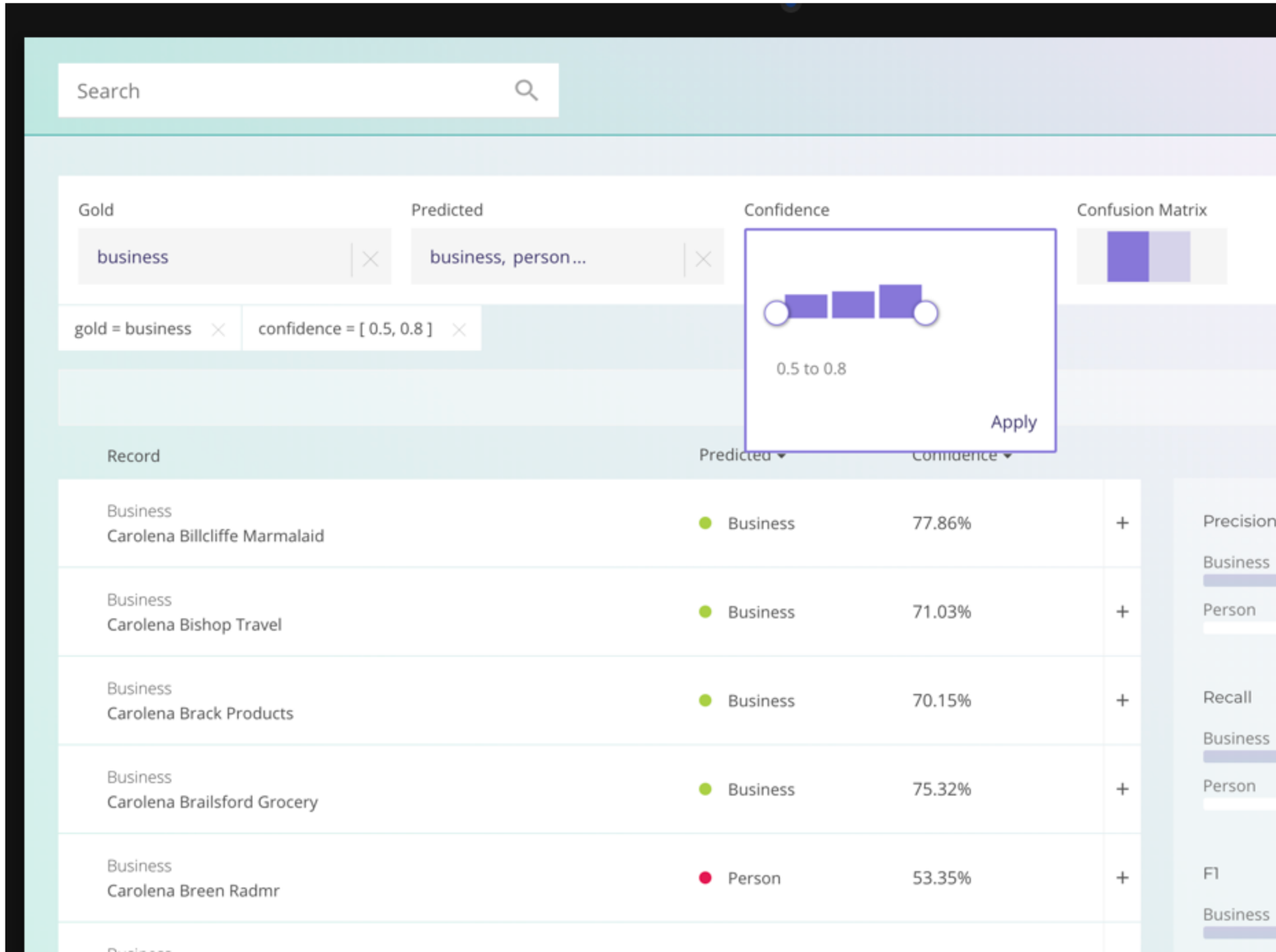
Comparison with State-of-the-Art

Flair outperforms the previous best methods on a range of NLP tasks:

biome

Recognai

- **Open**, **extensible** architecture
- **CI/CD approach** to NLP models and tasks
- **Guided** learning
- Focus **not only in data scientists**



Who is who

Recognai

spaCy

- **Text to input features**: tokenisation, POS tagging, NER..

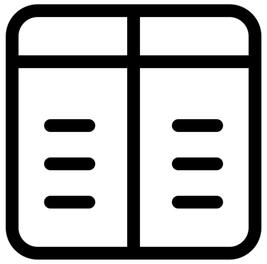
- **Flexible and dynamic deep learning framework:** core building blocks, autograd, loss functions.

AllenNLP

- **NLP Toolbox:** faster experimentation, best practices, plug and play higher-level modules (encoders..)

Working together

Recognai

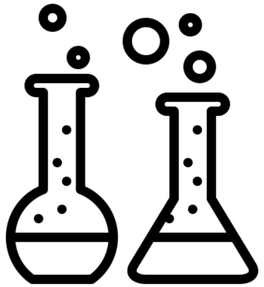


DATASETS

- **Dataset reader:**
  - **Type:** CSV, JSONL, write your own.
  - **Inputs and outputs:** where to find the inputs and outputs in the dataset.
  - **Tokenization:**
    - **Setup spaCy:** lang, features, etc.
    - Pipe different tokenisations: chars, words.

Working together

Recognai



EXPERIMENT DEFINITION

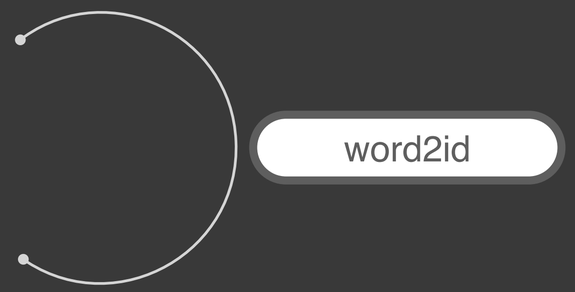
- **Model definition:**
  Combine modules, define new modules
- **Iterator:** Setup batching strategy/size
- **Trainer:**
  - **Optimizer:** Adam, SGD …
  - **Configuration:** Epochs, early stopping, etc.

AllenNLP

Experiment definition

RECOGNAI

```
{
  "dataset_reader": {
    "type": "jsonl_classification_reader",
    "input": "reviewText",
    "gold_label": "overall",
    "tokenizer": {
      "word_splitter": {
        "language": "es"
      }
    },
    "token_indexers": {
      "tokens": {
        "type": "single_id",
        "lowercase_tokens": true
      }
    }
  }
```



Experiment definition

RECOGNAI

```
{
  "model": {
    "type": "sequence_classifier",
    "text_field_embedder": {
          "tokens": {
            "type": "embedding",
            "embedding_dim": 300
          }
    },
    "encoder": {
          "type": "boe",
          "embedding_dim": 300
    }
  }
}
```

embedding

encoder

decoder

Recognai

NLP tasks and progress

nlpprogress.com

Recognai

View on GitHub

NLP-progress

Repository to track the progress in Natural Language Processing (NLP), including the datasets and the current state-of-the-art for the most common NLP tasks.

Tracking Progress in Natural Language Processing

Table of contents

- ASR
- CCG supertagging
- Chunking
- Constituency parsing
- Coreference resolution
- Dependency parsing
- Dialog
- Domain adaptation
- Entity Linking

Recognai

Language modeling

Language modeling is the task of predicting the next word or character in a document.

* Indicates models using dynamic evaluation.

Word Level Models

Penn Treebank

A common evaluation dataset for language modeling ist the Penn Treebank, as pre-processed by Mikolov et al. (2010). The dataset consists of 929k training words, 73k validation words, and 82k test words. As part of the pre-processing, words were lower-cased, numbers were replaced with N, newlines were replaced with `<eos>`, and all other punctuation was removed. The vocabulary is the most frequent 10k words with the rest of the tokens replaced by an `<unk>` token. Models are evaluated based on perplexity, which is the average per-word log-probability (lower is better).

Model	Validation perplexity	Test perplexity	Paper / Source	Code
AWD-LSTM-MoS + dynamic eval* by Yang et al. (2018)	48.33	47.69	Breaking the Softmax Bottleneck: A High-Rank RNN Language Model	
AWD-LSTM + dynamic eval* by Krause et al. (2017)	51.6	51.1	Dynamic Evaluation of Neural Sequence Models	
AWD-LSTM + continuous cache pointer* by Merity et al. (2017)	53.9	52.8	Regularizing and Optimizing LSTM Language Models	

Recognai

Text classification

Text classification is the task of assigning a sentence or document an appropriate category. The categories depend on the chosen dataset and can range from topics.

AG News

The AG News corpus consists of news articles from the AG's corpus of news articles on the web pertaining to the 4 largest classes. The dataset contains 30,000 training examples for each class 1,900 examples for each class for testing. Models are evaluated based on error rate (lower is better).

Model	Error	Paper / Source	Code
ULMFiT (Howard and Ruder, 2018)	5.01	Universal Language Model Fine-tuning for Text Classification	Official
CNN (Johnson and Zhang, 2016) *	6.57	Supervised and Semi-Supervised Text Categorization using LSTM for Region Embeddings	Official
DPCNN (Johnson and Zhang, 2017)	6.87	Deep Pyramid Convolutional Neural Networks for Text Categorization	Official
VDCN (Alexis et al., 2016)	8.67	Very Deep Convolutional Networks for Text Classification	Non Official
Char-level CNN (Zhang et al., 2015)	9.51	Character-level Convolutional Networks for Text Classification	Non Official

* Results reported in Johnson and Zhang, 2017

Sentiment analysis

Sentiment analysis is the task of classifying the polarity of a given text.

IMDb

The IMDb dataset is a binary sentiment analysis dataset consisting of 50,000 reviews from the Internet Movie Database (IMDb) labeled as positive or negative. The dataset contains an even number of positive and negative reviews. Only highly polarizing reviews are considered. A negative review has a score ≤ 4 out of 10, and a positive review has a score ≥ 7 out of 10. No more than 30 reviews are included per movie. Models are evaluated based on accuracy.

Model	Accuracy	Paper / Source
ULMFiT (Howard and Ruder, 2018)	95.4	Universal Language Model Fine-tuning for Text Classification
Block-sparse LSTM (Gray et al., 2017)	94.99	GPU Kernels for Block-Sparse Weights

SST

The Stanford Sentiment Treebank contains of 215,154 phrases with fine-grained sentiment labels in the parse trees of 11,855 sentences in movie reviews. Models are evaluated either on fine-grained (five-way) or binary classification based on accuracy.

Fine-grained classification (SST-5, 94,2k examples):

Model	Accuracy	Paper / Source
BCN+ELMo (Peters et al., 2018)	54.7	Deep contextualized word representations
BCN+Char+CoVe (McCann et al., 2017)	53.7	Learned in Translation: Contextualized Word Vectors

Binary classification (SST-2, 56.4k examples):

Named entity recognition

Named entity recognition (NER) is the task of tagging entities in text with their corresponding type. Approaches typically use BIO notation, which differentiates the beginning (B) and the inside (I) of entities. O is used for non-entity tokens.

Example:

Mark	Watney	visited	Mars
B-PER	I-PER	O	B-LOC

CoNLL 2003 (English)

The CoNLL 2003 NER task consists of newswire text from the Reuters RCV1 corpus tagged with four different entity types (PER, LOC, ORG, MISC). Models are evaluated based on span-based F1.

Model	F1	Paper / Source	Code
Flair embeddings (Akbik et al., 2018)	93.09	Contextual String Embeddings for Sequence Labeling	Flair framework
BiLSTM-CRF+ELMo (Peters et al., 2018)	92.22	Deep contextualized word representations	AllenNLP Project AllenNLP GitHub
Peters et al. (2017)	91.93	Semi-supervised sequence tagging with bidirectional language models	
LM-LSTM-CRF (Liu et al., 2018)	91.71	Empowering Character-aware Sequence Labeling with Task-	LM-LSTM-CRF

Semantic role labeling

Semantic role labeling aims to model the predicate-argument structure of a sentence and is often described as answering "Who did what to whom". BIO notation is typically used for semantic role labeling.

Example:

Housing	starts	are	expected	to	quicken	a	bit	from	August's	pace
B-ARG1	I-ARG1	O	O	O	V	B-ARG2	I-ARG2	B-ARG3	I-ARG3	I-ARG3

OntoNotes

Models are typically evaluated on the OntoNotes benchmark based on F1.

Model	F1	Paper / Source
He et al., (2018) + ELMO	85.5	Jointly Predicting Predicates and Arguments in Neural Semantic Role Labeling
(He et al., 2017) + ELMo (Peters et al., 2018)	84.6	Deep contextualized word representations
Tan et al. (2018)	82.7	Deep Semantic Role Labeling with Self-Attention
He et al. (2018)	82.1	Jointly Predicting Predicates and Arguments in Neural Semantic Role Labeling
He et al. (2017)	81.7	Deep Semantic Role Labeling: What Works and What's Next

Recognai

Trends

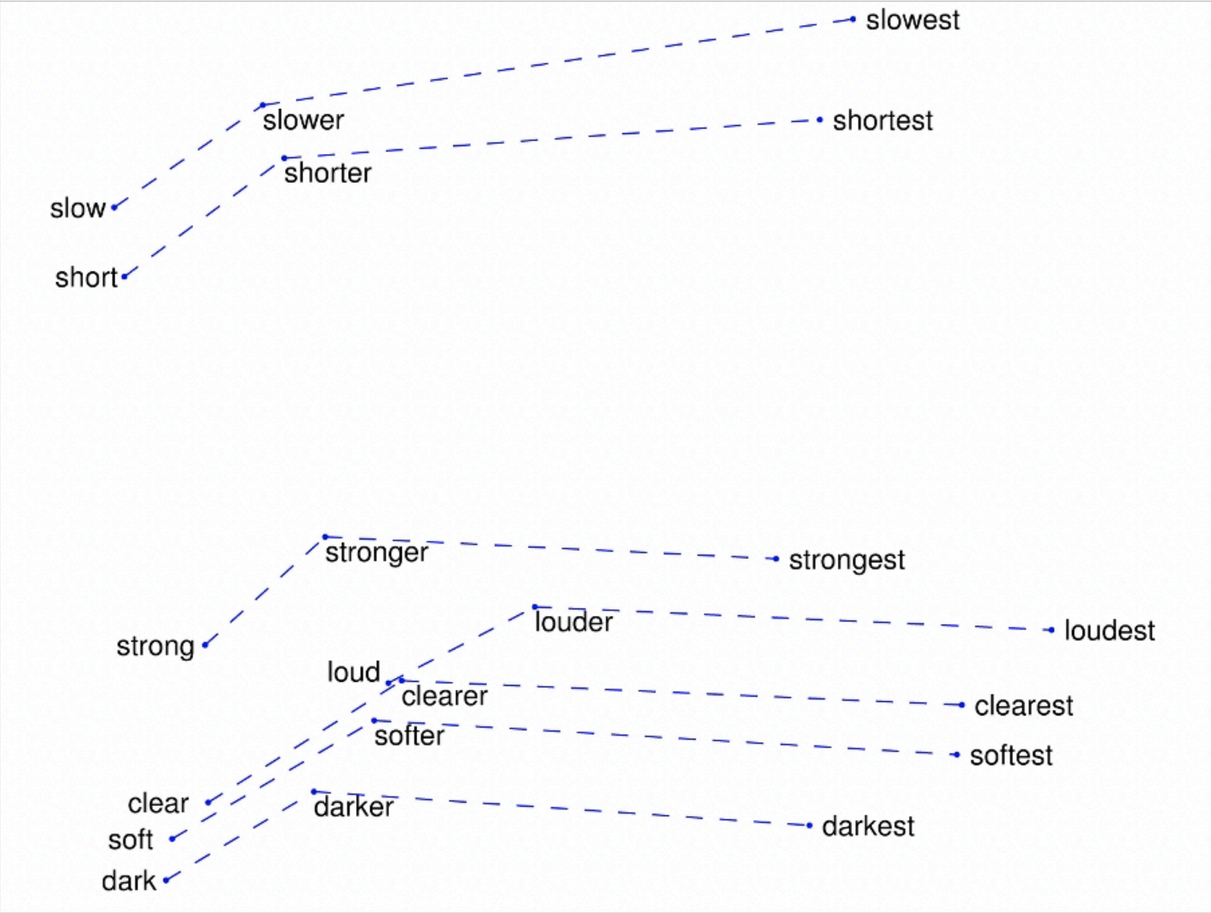
1. Encoders
2. Transfer learning
3. Knowledge graphs
4. Tooling
5. ….

Recognai

Encoders are “all” you need

Word vectors are dying

- Unsupervised learning of word representations (based on co-occurrences)
- Became standard in NLP very quickly.
- First form of (very) limited transfer learning (limited to 1st layer)
- They do not encode: context, word sense, word function
- We still need a lot of supervision examples



Comparative and superlative relationships (Pennington et al 2014).

Model words in context

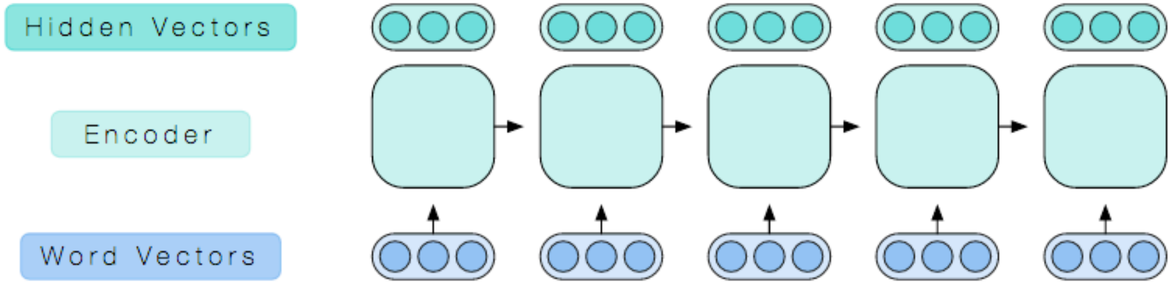
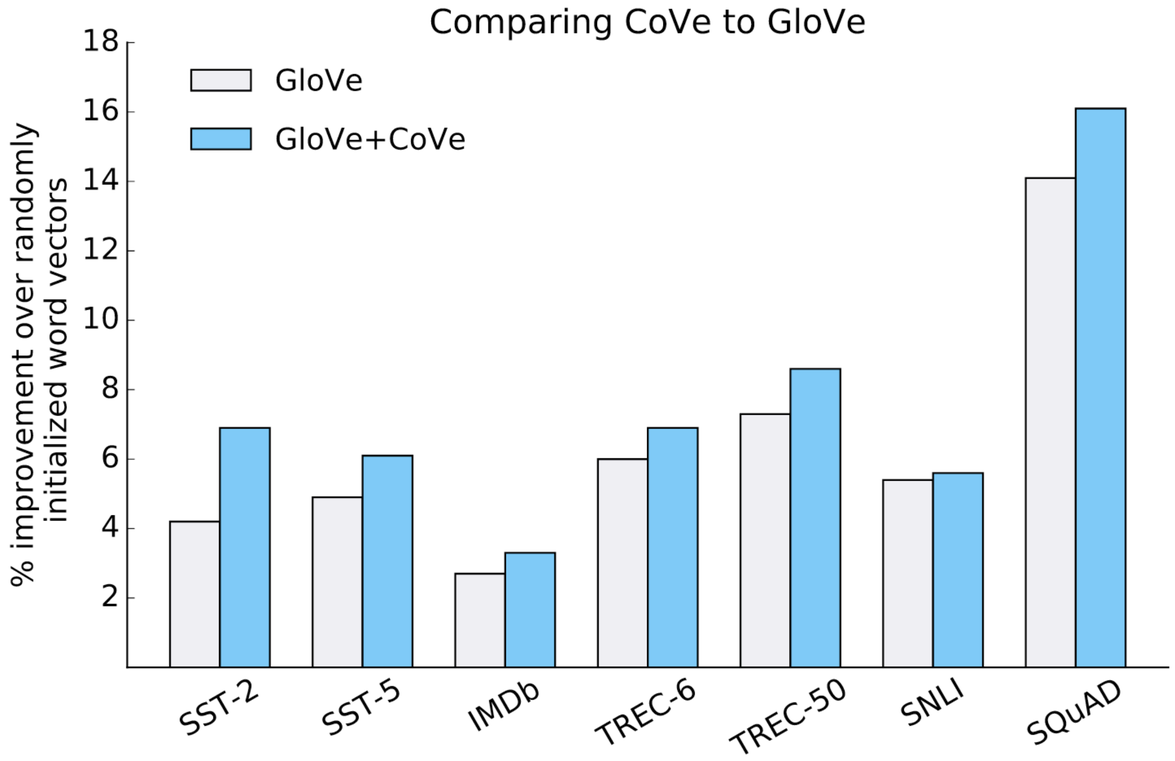


Figure 6: An LSTM encoder takes in a sequence of word vectors and outputs a sequence of hidden vectors.

Bryan McCann, James Bradbury, Caiming Xiong, and Richard Socher. 2017.
[Learned in Translation: Contextualized Word Vectors]

Learned in translation

- Learn to encode words by using NMT as an auxiliary task: German —> English
- Still combined with word vectors (Stanford GloVe)
- Large improvements several benchmarks



Bryan McCann, James Bradbury, Caiming Xiong, and Richard Socher. 2017.
Learned in Translation: Contextualized Word Vectors

Sentence encoders

Recognai

Universal Sentence Encoder

Daniel Cer[a], Yinfei Yang[a], Sheng-yi Kong[a], Nan Hua[a], Nicole Limtiaco[b], Rhomni St. John[a], Noah Constant[a], Mario Guajardo-Céspedes[a], Steve Yuan[c], Chris Tar[a], Yun-Hsuan Sung[a], Brian Strope[a], Ray Kurzweil[a]

[a]Google Research
Mountain View, CA

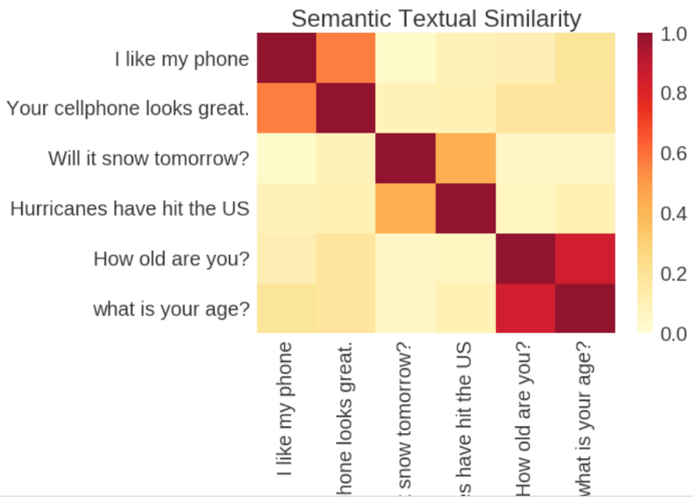
[b]Google Research
New York, NY

[c]Google
Cambridge, MA

12 Apr 2018

Abstract

We present models for encoding sentences into embedding vectors that specifically target transfer learning to other NLP tasks. The models are efficient and result in accurate performance on diverse transfer tasks. Two variants of the encoding models allow for trade-offs between accuracy and compute resources. For both variants, we investigate and report the rela-



Encoder Hubs

Recognai

TensorFlow Hub

USER GUIDE

Text
- Embedding

Image
- Classification
- Feature Vector
- Generator
- Other

Video
- Classification

Publishers
- Google
- DeepMind

Text embedding

universal-sentence-encoder-large By Google

text-embedding transformer en

Encoder of greater-than-word length text trained on a variety of data.

universal-sentence-encoder By Google

text-embedding dan en

Encoder of greater-than-word length text trained on a variety of data.

elmo By Google

text-embedding billion-word-benchmark elmo en

Embeddings from a language model trained on the 1 Billion Word Benchmark.

View more text embeddings

Image feature vectors

imagenet/inception_v3/feature_vector By Google

imagenet-ilsvrc-2012-cls image-feature-vector inception-v3

Feature vectors of images with Inception V3 trained on ImageNet (ILSVRC-2012-CLS).

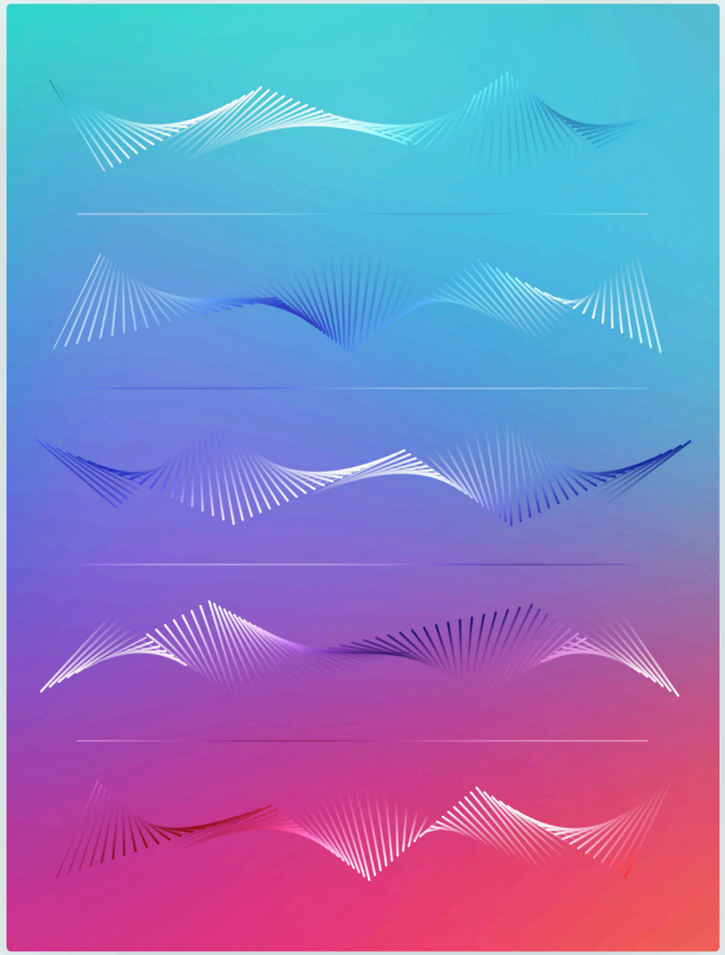
imagenet/resnet_v2_50/feature_vector By Google

imagenet-ilsvrc-2012-cls image-feature-vector resnet-v2-50

Sentiment neuron

Recognai

ABOUT BLOG

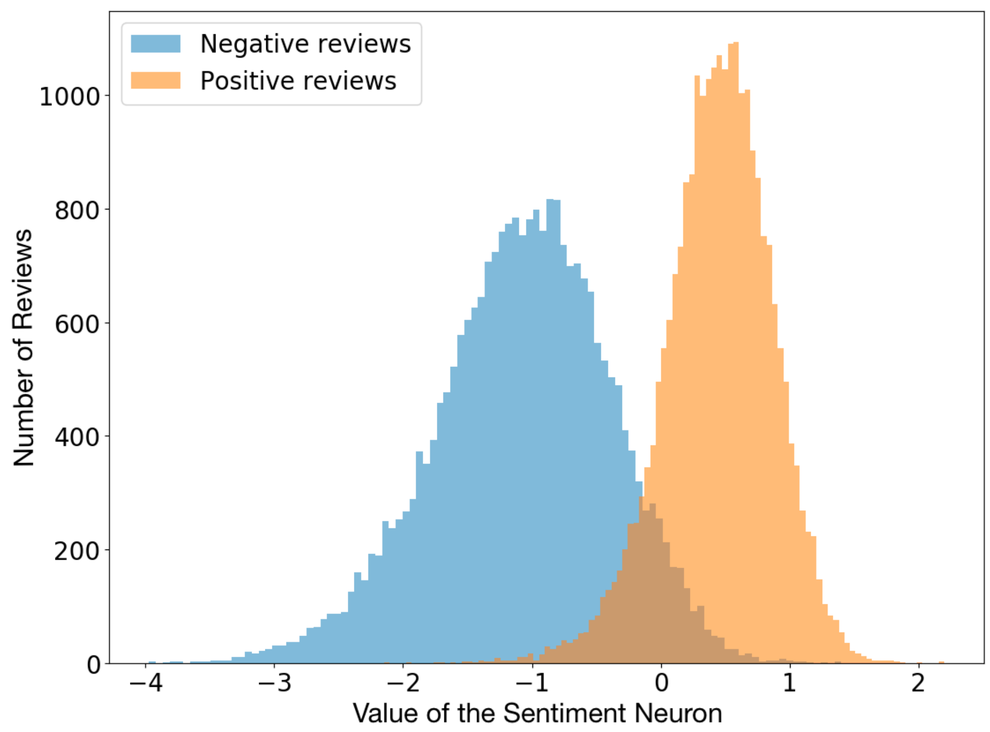


APRIL 6, 2017

Unsupervised Sentiment Neuron

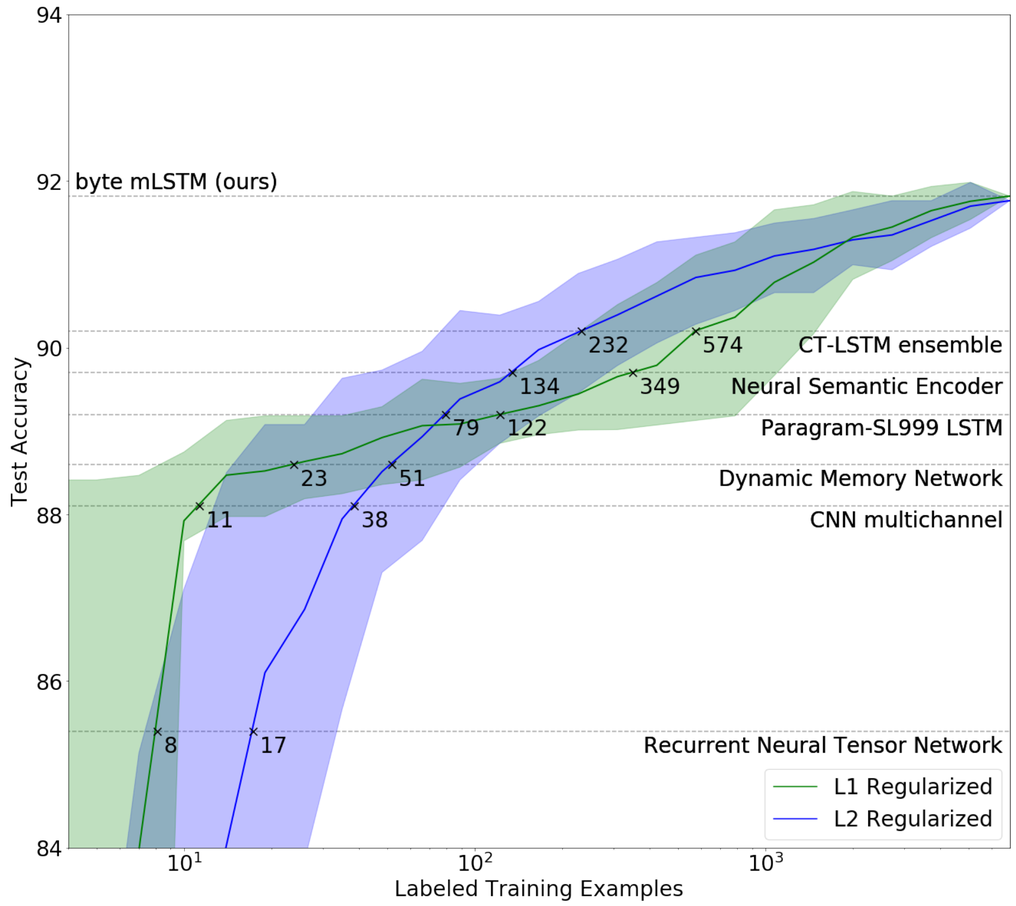
We've developed an unsupervised system which learns an excellent representation of sentiment, despite being trained only to predict the next character in the text of Amazon reviews.

A linear model using this representation achieves state-of-the-art sentiment analysis accuracy on a small but extensively-studied dataset, the Stanford Sentiment Treebank (we get 91.8% accuracy versus the previous best of 90.2%), and can match the performance of previous supervised systems using 30-100x fewer labeled examples. Our representation also contains a distinct "sentiment neuron" which contains almost all of the sentiment signal.



Sentiment neuron

- Char-level lang model on Amazon Reviews (82M)
- Took 1 month, 8GPUs
- Linear model using 4,096 learned units beats SoTA with fewer examples



Alec Radford, Rafal Jozefowicz, Ilya Sutskever. 2017
Learning to Generate Reviews and Discovering Sentiment

The power of language models

Recognai

This is one of Crichton's best books. The characters of Karen Ross, Peter Elliot, Munro, and Amy are beautifully developed and their interactions are exciting, complex, and fast-paced throughout this impressive novel. And about 99.8 percent of that got lost in the film. Seriously, the screenplay AND the directing were horrendous and clearly done by people who could not fathom what was good about the novel. I can't fault the actors because frankly, they never had a chance to make this turkey live up to Crichton's original work. I know good novels, especially those with a science fiction edge, are hard to bring to the screen in a way that lives up to the original. But this may be the absolute worst disparity in quality between novel and screen adaptation ever. The book is really, really good. The movie is just dreadful.

Alec Radford, Rafal Jozefowicz, Ilya Sutskever. 2017
Learning to Generate Reviews and Discovering Sentiment

Recognai

Contextual String Embeddings for Sequence Labeling

Alan Akbik
Zalando Research
Mühlenstraße 25
10243 Berlin

Duncan Blythe
Zalando Research
Mühlenstraße 25
10243 Berlin

Roland Vollgraf
Zalando Research
Mühlenstraße 25
10243 Berlin

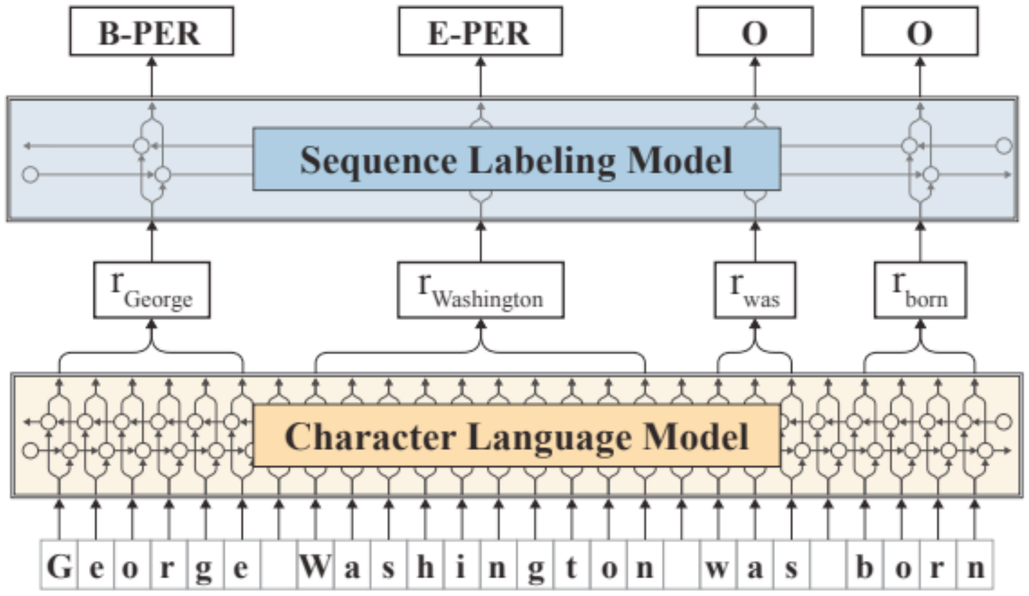
{firstname.lastname}@zalando.de

Abstract

Recent advances in language modeling using recurrent neural networks have made it viable to model language as distributions over characters. By learning to predict the next character on the basis of previous characters, such models have been shown to automatically internalize linguistic concepts such as words, sentences, subclauses and even sentiment. In this paper, we propose to leverage the internal states of a trained character language model to produce a novel type of word embedding which we refer to as *contextual string embeddings*. Our proposed embeddings have the distinct properties that they (a) are trained without any explicit notion of words and thus fundamentally model words as sequences of characters, and (b) are *contextualized* by their surrounding text, meaning that the same word will have different embeddings depending on its contextual use. We conduct a comparative evaluation against previous embeddings and find that our embeddings are highly useful for downstream tasks: across four classic sequence labeling tasks we consistently outperform the previous state-of-the-art. In particular, we significantly outperform previous work on English and German named entity recognition (NER), allowing us to report new state-of-the-art F1-scores on the CoNLL03 shared task.

We release all code and pre-trained language models in a search community, to enable reproduction of these experim embeddings to other tasks: https://github.com/za

flair

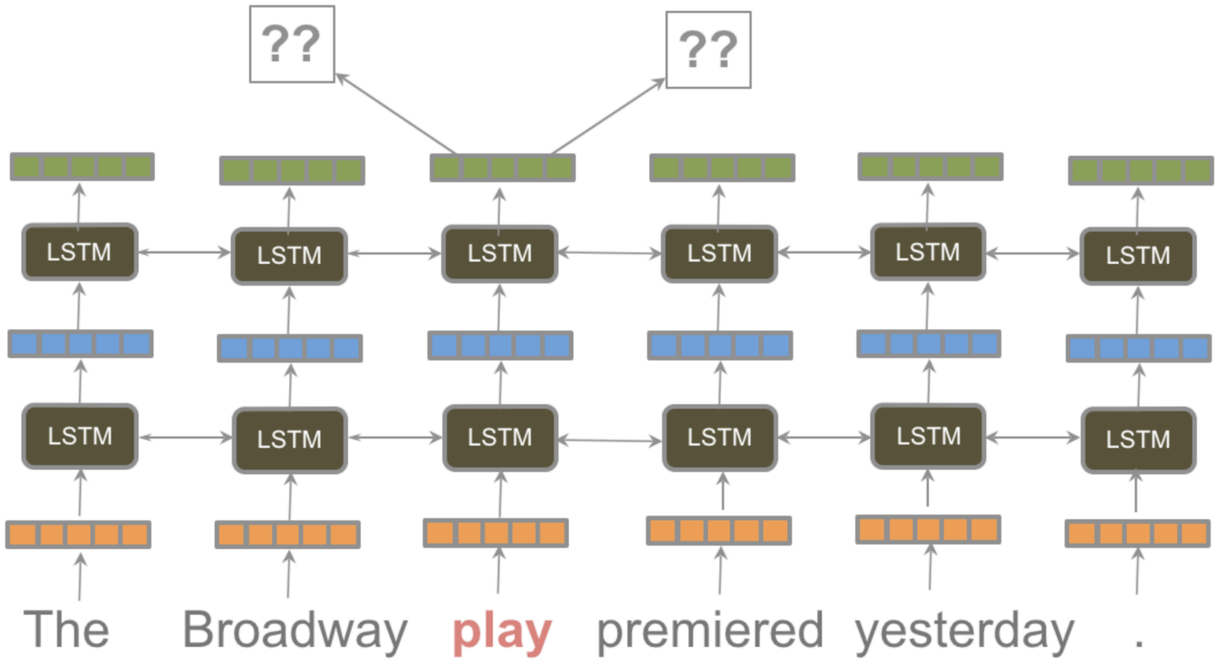


Recognai

Welcome transfer learning

The Gradient

HOME OVERVIEWS PERSPECTIVES ABOUT SUBSCRIBE



NLP's ImageNet moment has arrived

08.JUL.2018

https://thegradient.pub/nlp-imagenet/

Elmo

https://allennlp.org/elmo

Recognai

AllenNLP

Demo | Tutorials | Models | Elmo | Publications | Documentation | GitHub

ELMo

Deep contextualized word representations

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner,
Christopher Clark, Kenton Lee, Luke Zettlemoyer.
NAACL 2018.

Introduction

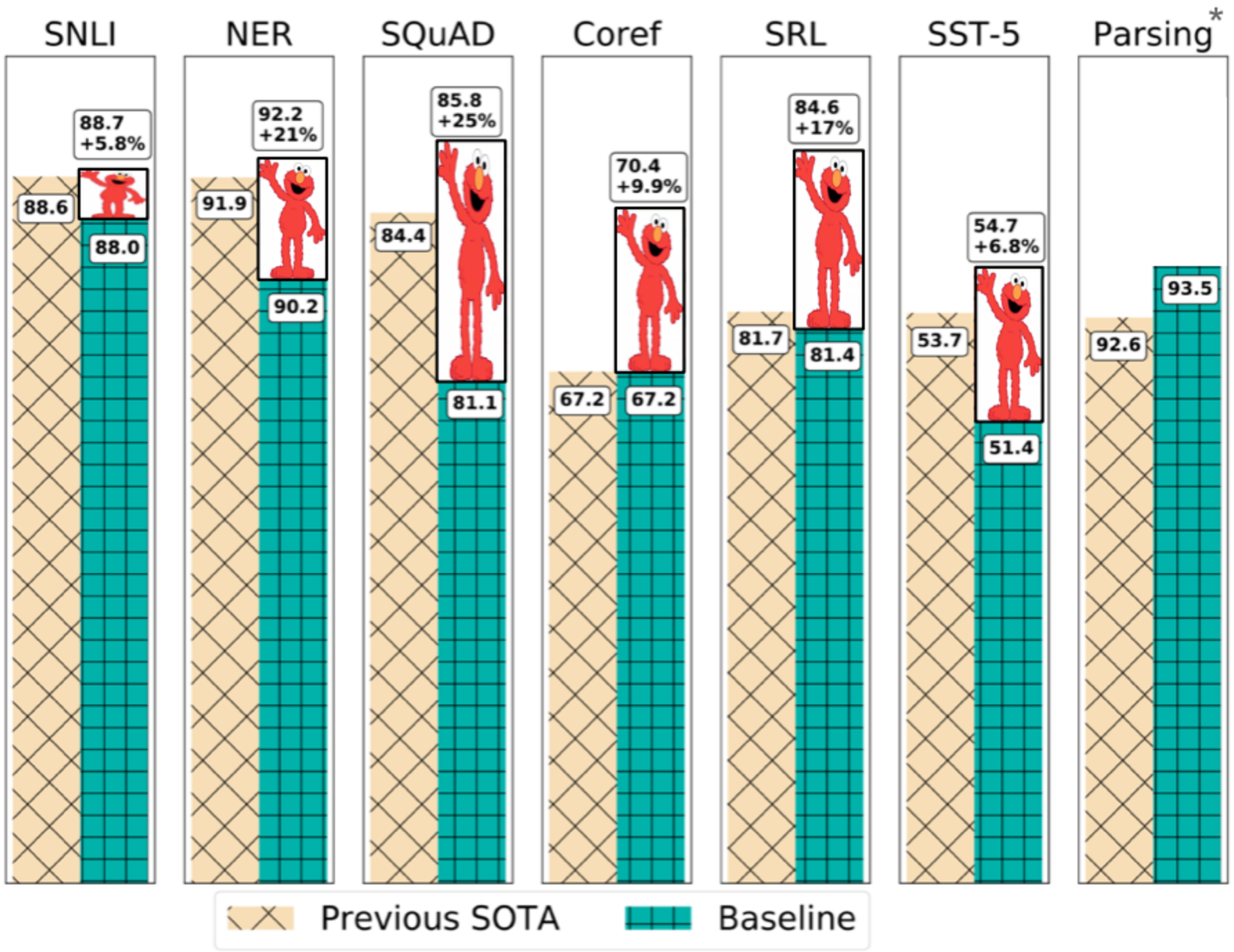
ELMo is a deep contextualized word representation that models both (1) complex characteristics of word use (e.g., syntax and semantics), and (2) how these uses vary across linguistic contexts (i.e., to model polysemy). These word vectors are learned functions of the internal states of a deep bidirectional language model (biLM), which is pre-trained on a large text corpus. They can be easily added to existing models and significantly improve the state of the art across a broad range of challenging NLP problems, including question answering, textual entailment and sentiment analysis.

Salient features

ELMo representations are:

- *Contextual*: The representation for each word depends on the entire context in which it is used.
- *Deep*: The word representations combine all layers of a deep pre-trained neural network.
- *Character based*: ELMo representations are purely character based, allowing the network to use morphological clues to form robust

Elmo



*Kitaev and Klein, ACL 2018 (see also Joshi et al., ACL 2018)

The improvements ELMo achieved on a wide range of NLP tasks. (Source: Matthew Peters)

fast.ai NLP

Projects from fast.ai researchers and collaborators

Home
About

Projects:

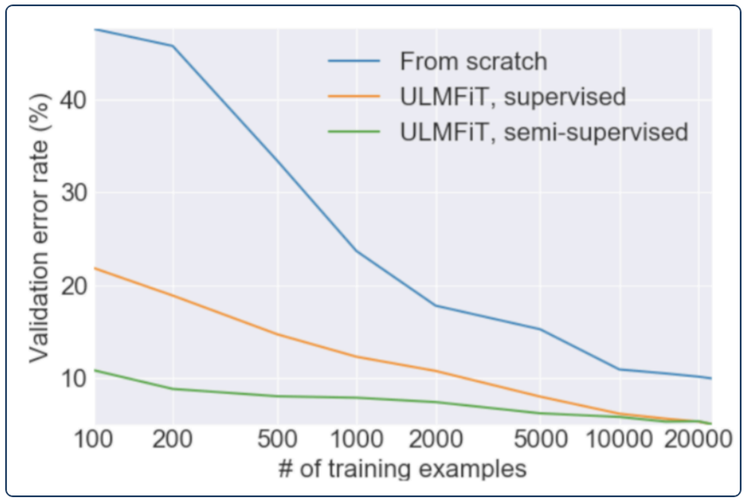
- Classification
- Model Zoo (soon)
- Seq2seq (soon)
- Sequence Labeling (soon)

© 2018. MIT License.

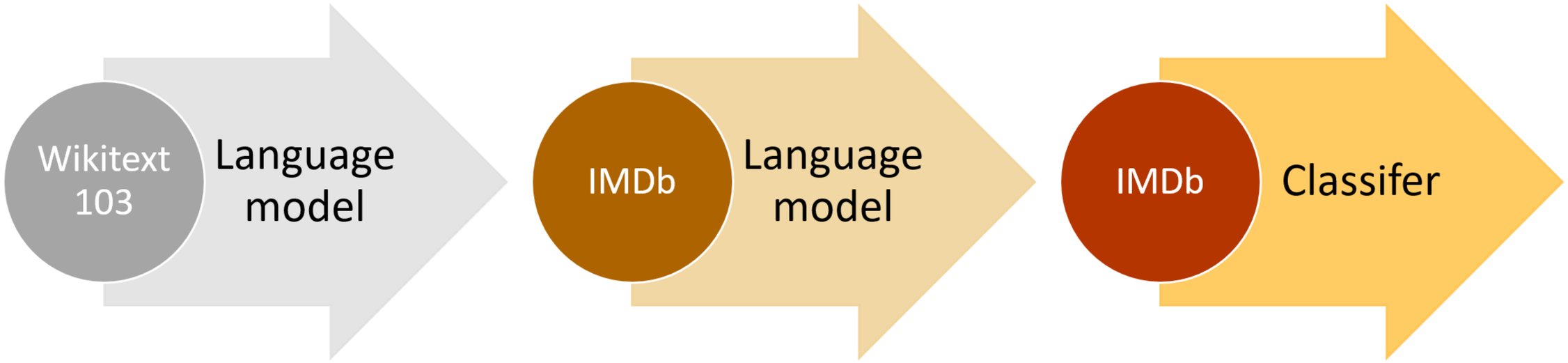
Today we're releasing our paper Universal Language Model Fine-tuning for Text Classification (ULMFiT), pre-trained models, and full source code in the Python programming language. The paper has been peer-reviewed and accepted for presentation at the Annual Meeting of the Association for Computational Linguistics (ACL 2018). For links to videos providing an in-depth walk-through of the approach, all the Python modules used, pre-trained models, and scripts for building your own models, see our NLP classification page.

This method dramatically improves over previous approaches to *text classification*, and the code and pre-trained models allow anyone to leverage this new approach to better solve problems such as:

- Finding documents relevant to a legal case;
- Identifying spam, bots, and offensive comments;
- Classifying positive and negative reviews of a product;
- Grouping articles by political orientation;
- ...and much more.



ULMfit

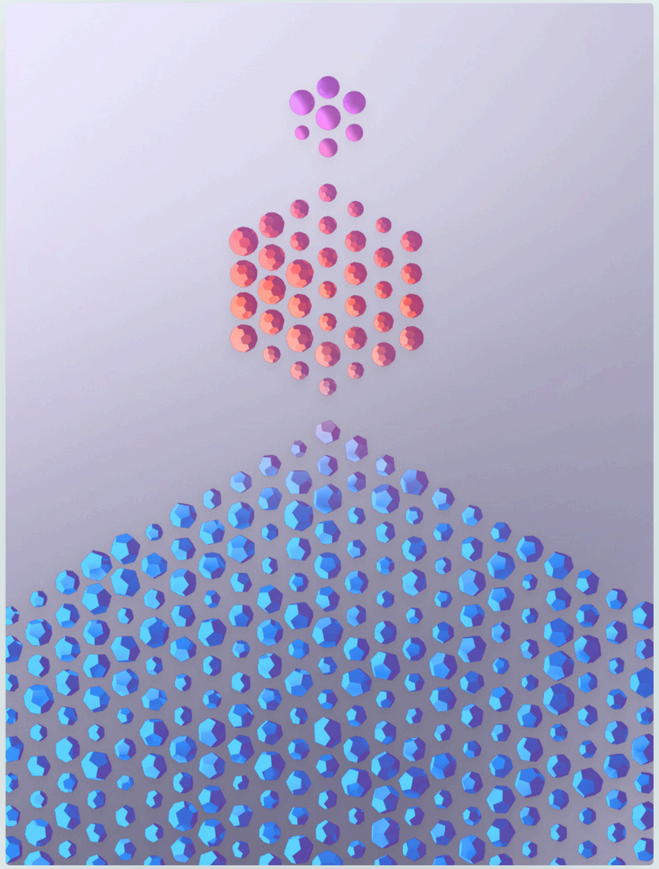


Jeremy Howard, Sebastian Ruder. ACL 2018
Universal Language Model Fine-tuning for Text Classification

Transformer

Recognai

RESEARCH SYSTEMS ABOUT BL



JUNE 11, 2018

Improving Language Understanding with Unsupervised Learning

We've obtained state-of-the-art results on a suite of diverse language tasks with a scalable, task-agnostic system, which we're also releasing. Our approach is a combination of two existing ideas: transformers and unsupervised pre-training. These results provide a convincing example that pairing supervised learning methods with unsupervised pre-training works very well; this is an idea that many have

Transformer

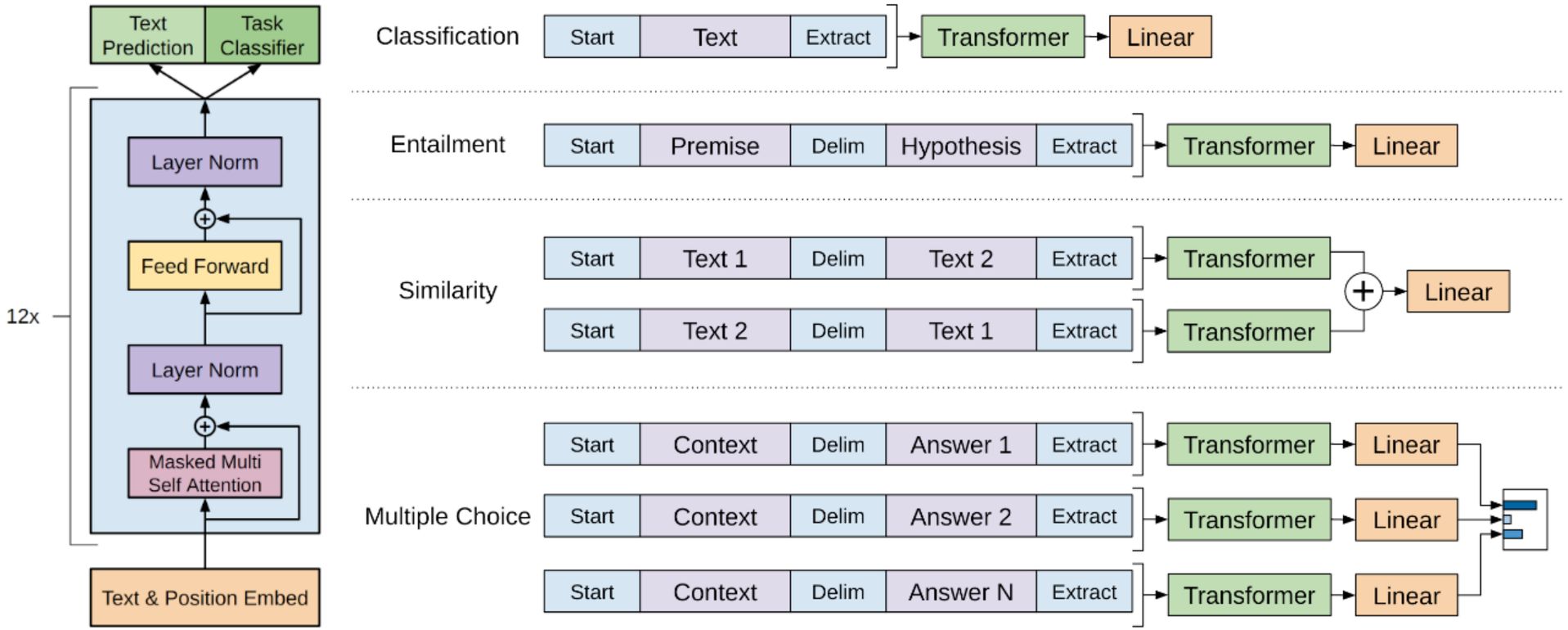


Figure 1: **(left)** Transformer architecture and training objectives used in this work. **(right)** Input transformations for fine-tuning on different tasks. We convert all structured inputs into token sequences to be processed by our pre-trained model, followed by a linear+softmax layer.

Google AI's BERT

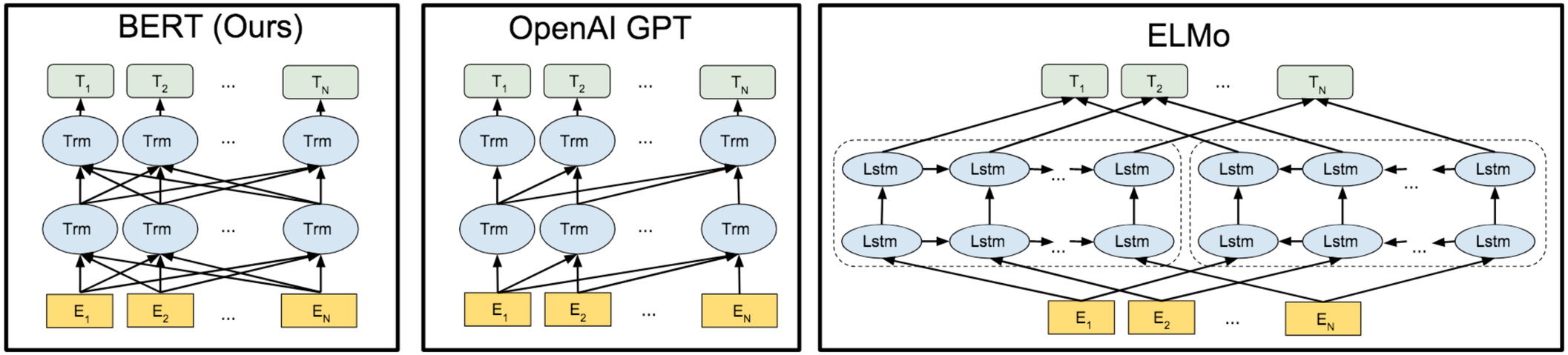
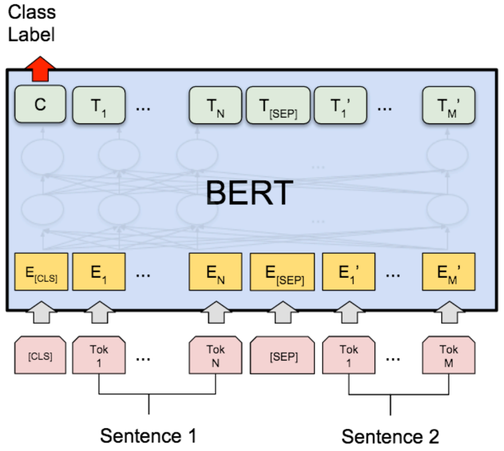
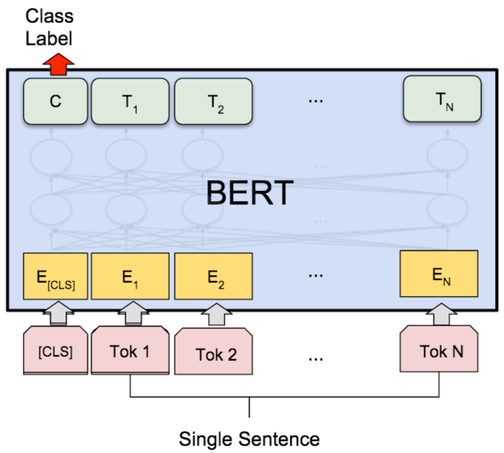


Figure 1: Differences in pre-training model architectures. BERT uses a bidirectional Transformer. OpenAI GPT uses a left-to-right Transformer. ELMo uses the concatenation of independently trained left-to-right and right-to-left LSTM to generate features for downstream tasks. Among three, only BERT representations are jointly conditioned on both left and right context in all layers.

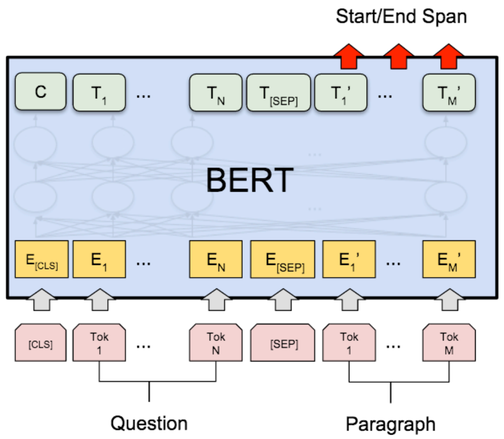
Google AI's BERT



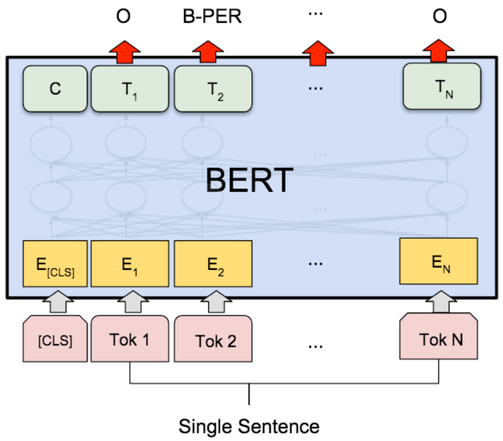
(a) Sentence Pair Classification Tasks: MNLI, QQP, QNLI, STS-B, MRPC, RTE, SWAG



(b) Single Sentence Classification Tasks: SST-2, CoLA



(c) Question Answering Tasks: SQuAD v1.1



(d) Single Sentence Tagging Tasks: CoNLL-2003 NER

Implications

Recognai

- **Compute**: Pre-training on large corpora is expensive, inference is also more expensive than before (deeper models).
- **Corpora**: Size? Domain? What about other languages?
- **Workflow and reproducibility**: Learning is not anymore a 1-step process. Model evolution with successive pre-training, fine-tuning steps

Recognai

Exercise 3:
Creating a modern, robust neural classifier

Recognai

Knowledge is a graph

What’s a knowledge graph

Recognai

Google | eduardo mendoza | Daniel Vila | 0 | + Compartir

Web | Imágenes | Maps | Shopping | Más | Herramientas de búsqueda

Aproximadamente 2.490.000 resultados (0,24 segundos)

Searching your Diigo library ... Try other search engines

Eduardo Mendoza - Wikipedia, la enciclopedia libre
es.wikipedia.org/wiki/**Eduardo_Mendoza**
Hijo de un fiscal, **Eduardo Mendoza** Arias-Carvajal, y una ama de casa, Cristina Garriga Alemany, estudió un año en una escuela de las monjas de Nuestra ...
Biografía - Obra - Adaptaciones cinematográficas - Premios

Pagina oficial de **Eduardo Mendoza**. Textos
www.clubcultura.com/clubliteratura/.../**mendoza**/
Pagina oficial del escritor **Eduardo Mendoza** donde se analizan sus obras, sus premios o una entrevista, ademas de informar de los trabajos que esta ...

Eduardo Mendoza - Libros y obras del autor, biografía y bibliografía ...
www.lecturalia.com/autor/25/**eduardo-mendoza**
Eduardo Mendoza nació el 11 de enero de 1943 en Barcelona. Se licenció en Derecho en 1966, tras lo que pasó a trabajar como asesor jurídico para un...

EDUARDO MENDOZA, Libros, obras, biografía y más en ...
www.casadellibro.com/.../**mendoza-eduardo/mendoza**32e...
Libros de **EDUARDO MENDOZA**, Biografía, obras y su último libro. Comprar libros de **EDUARDO MENDOZA** en tu librería online casadellibro.com.

Eduardo Mendoza | PlanetadeLibros.com
www.planetadelibros.com › Autores
Eduardo Mendoza. Barcelona, 1943. Nació en Barcelona en 1943. Ha publicado las

Eduardo Mendoza

Eduardo Mendoza Garriga es un escritor español. Su obra literaria, que inaugura con la publicación en 1975 de La verdad sobre el caso Savolta, mezcla un estilo sencillo y directo con el uso de cultismos ... Wikipedia

Fecha de nacimiento: 11 de enero de 1943 (edad 69), Barcelona

También se ha buscado

Juan Marsé | Manuel Vázquez Montalbán | Almudena Grandes | Javier Marías | Mario Camus

Knowledge Vault

Recognai

Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion

Download Search Copy Bibtex

Venue

The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014, pp. 601-610

Authors

Xin Luna Dong

Evgeniy Gabrilovich

Geremy Heitz

Abstract

Recent years have witnessed a proliferation of large-scale knowledge bases, including Wikipedia, Freebase, YAGO, Microsoft's Satori, and Google's Knowledge Graph. To increase the scale even further, we need to explore automatic methods for constructing knowledge bases. Previous approaches have primarily focused on text-based extraction, which can be very noisy. Here we introduce Knowledge Vault, a Web-scale probabilistic knowledge base that combines extractions from Web content (obtained via analysis of text, tabular data, page structure, and human annotations) with prior knowledge derived from existing knowledge repositories. We employ

Knowledge Vault

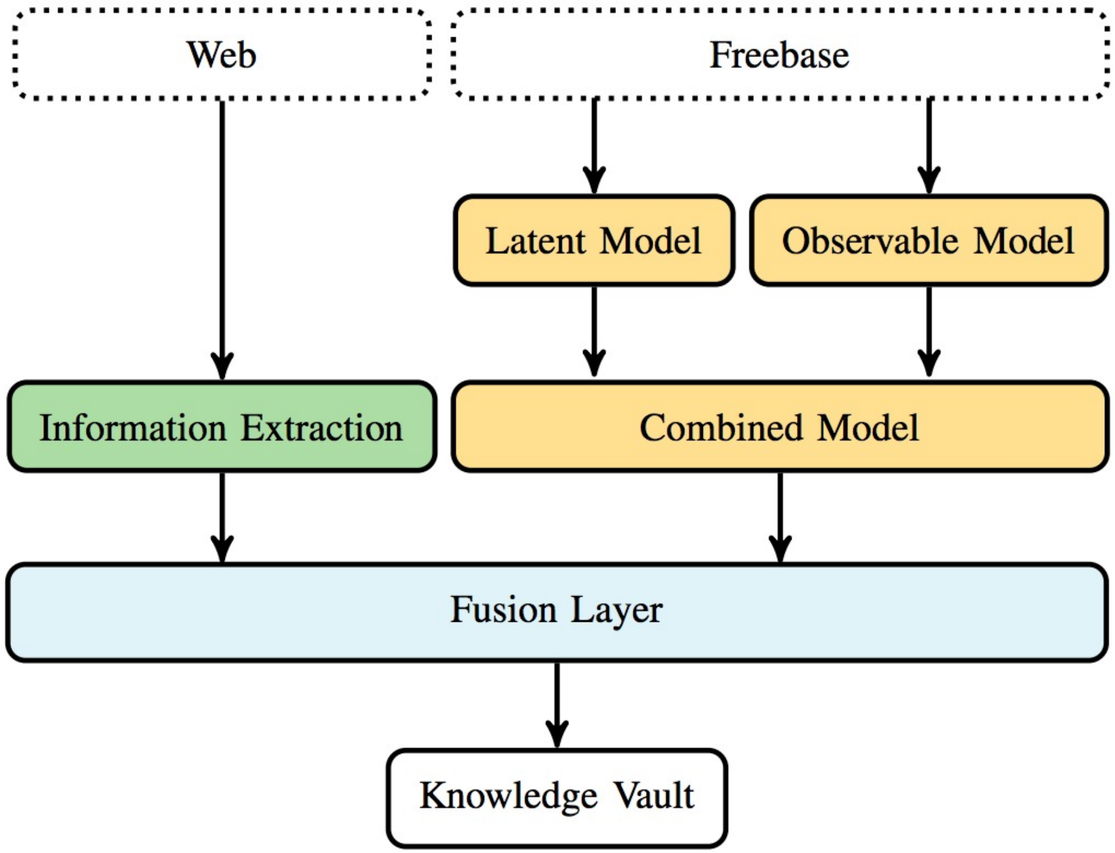


Fig. 7. Architecture of the Knowledge Vault.

AKBC

Recognai

AKBC 2019 | Call for Papers | Speakers | Dates | Workshops | Organization | Previous Workshops

AKBC 2019

1st Conference on Automated Knowledge Base Construction (AKBC) 2019

in Amherst, Massachusetts, May 20-22nd, 2019.

Tweet G+ Like 44 Share

Tweets about #AKBC2019 OR akbc.ws

Knowledge Base Construction

Knowledge gathering, representation, and reasoning are among the fundamental challenges of artificial intelligence. Large-scale repositories of knowledge about entities, relations, and their abstractions are known as “knowledge bases”. Most major technology companies now have substantial efforts in knowledge base construction, and related scholarly work spans many research areas, including machine learning, natural language processing, computer vision, information integration, databases, search, data mining, knowledge representation, human computation, human-computer interfaces, and fairness. The AKBC conference serves as a research forum for all these areas, in both academia and industry.

New Conference

Nearly ten years after the first AKBC workshop in Grenoble, France, AKBC is becoming a conference. Why a new stand-alone conference?

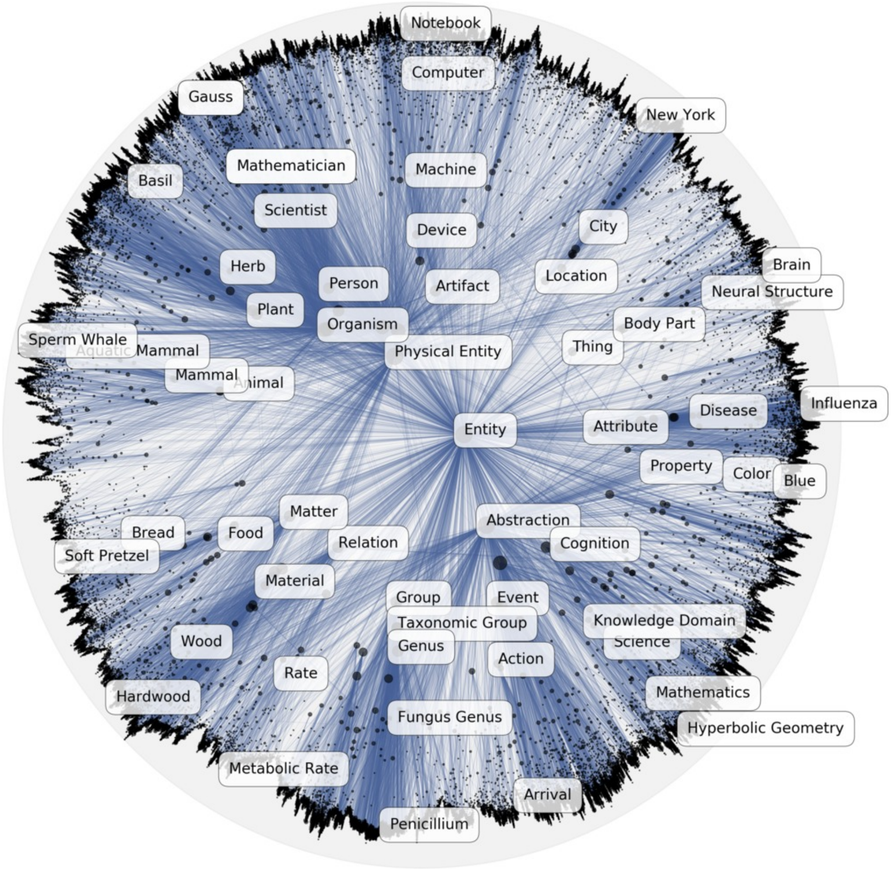
- Long-standing and growing interest in the area, now with too much material for a one-day workshop. We have sufficient material for a two-day conference plus topical workshops.
- We want to grow and connect the community beyond existing individual conference communities, bringing together ML, NLP, DB, IR,

KG embeddings

Recognai

MAXIMILIAN NICKEL

Home Projects Publications Workshops & Symposia



Geometric Representation Learning

Representation learning has become an invaluable approach in machine learning and artificial intelligence. For instance, word embeddings such as word2vec, GloVe and fastText are widely used for tasks rangingfrom machine translation to sentiment analysis. Similarly, embeddings of (multi-)graphs such as RESCAL and node2vec have found important applications for learning in semantic and social networks.

In this project, we study a fundamental aspect of representation learning, i.e., the **influence of the underlying geometry on embedding structured data**. This spans research areas such as *machine learning in non-Euclidean geometries*, *Riemannian geometry and optimization*, *graph theory*, as well as *representation learning for structured data*.

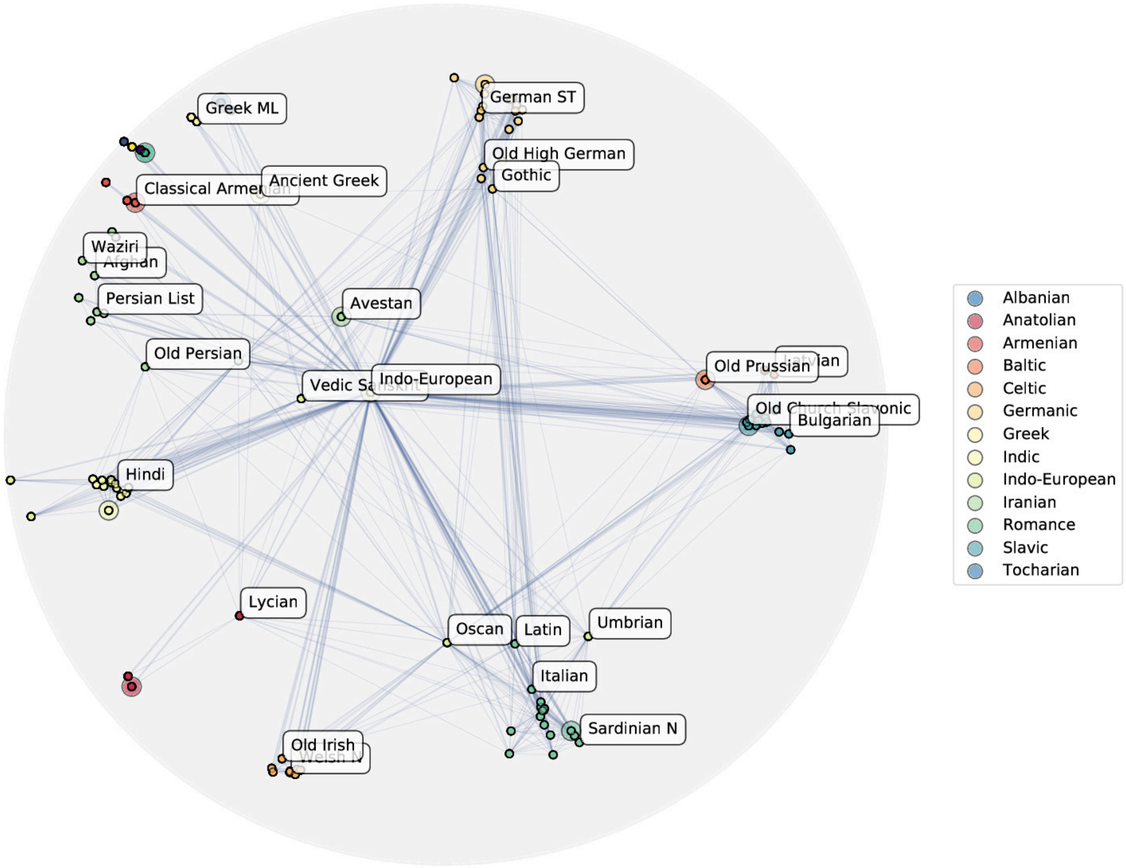
For instance, finding a way to represent hierarchies explicitly in an

KG embeddings

Recognai

MAXIMILIAN NICKEL

Home Projects Publications Workshops & Symposia



Geometric Representation Learning

Representation learning has become an invaluable approach in machine learning and artificial intelligence. For instance, word embeddings such as word2vec, GloVe and fastText are widely used for tasks rangingfrom machine translation to sentiment analysis. Similarly, embeddings of (multi-)graphs such as RESCAL and node2vec have found important applications for learning in semantic and social networks.

In this project, we study a fundamental aspect of representation learning, i.e., the **influence of the underlying geometry on embedding structured data**. This spans research areas such as *machine learning in non-Euclidean geometries*, *Riemannian geometry and optimization*, *graph theory*, as well as *representation learning for structured data*.

For instance, finding a way to represent hierarchies explicitly in an

Airbnb

Recognai

Airbnb Engineering & Data Science Follow

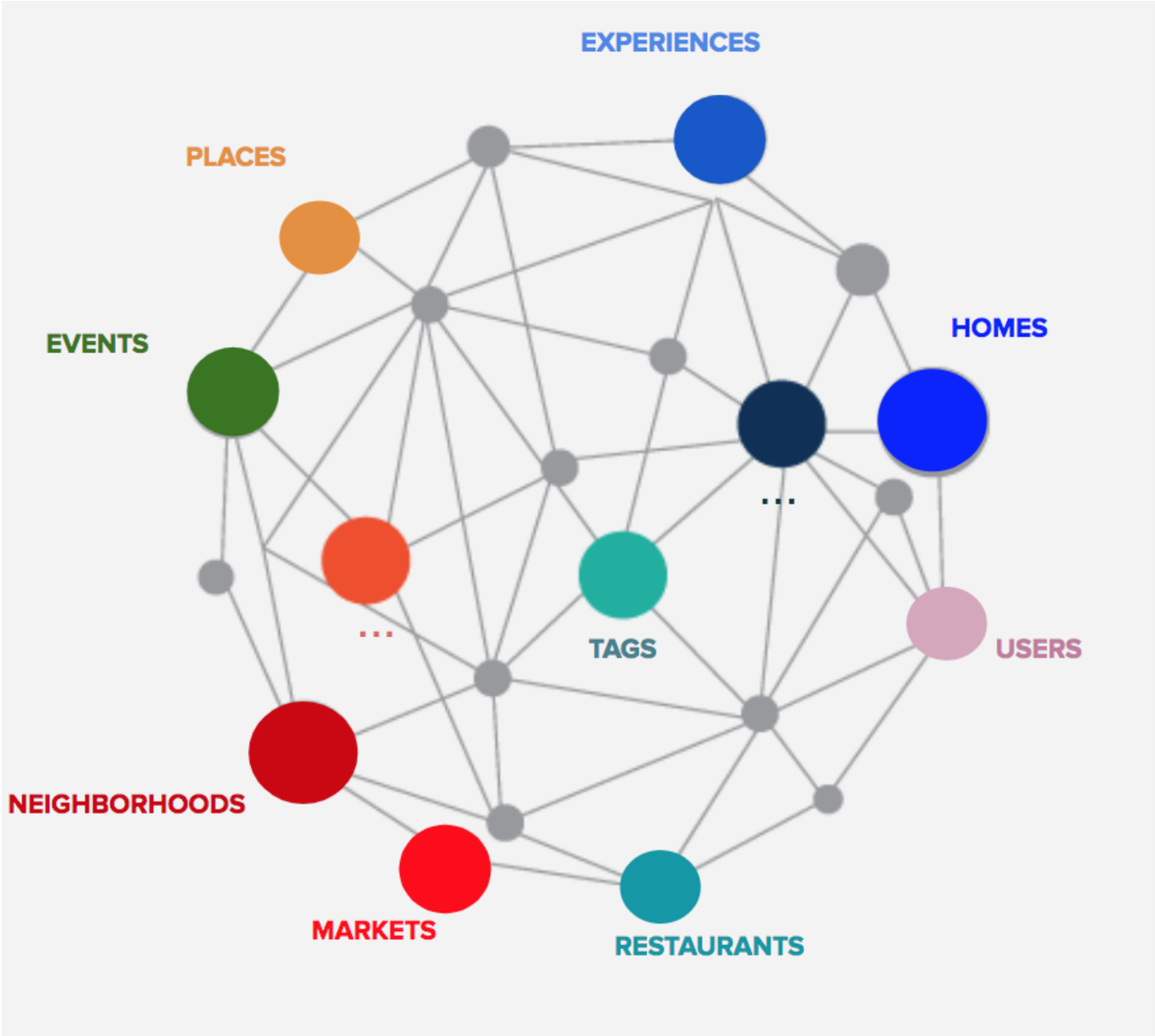
Sign in Get started

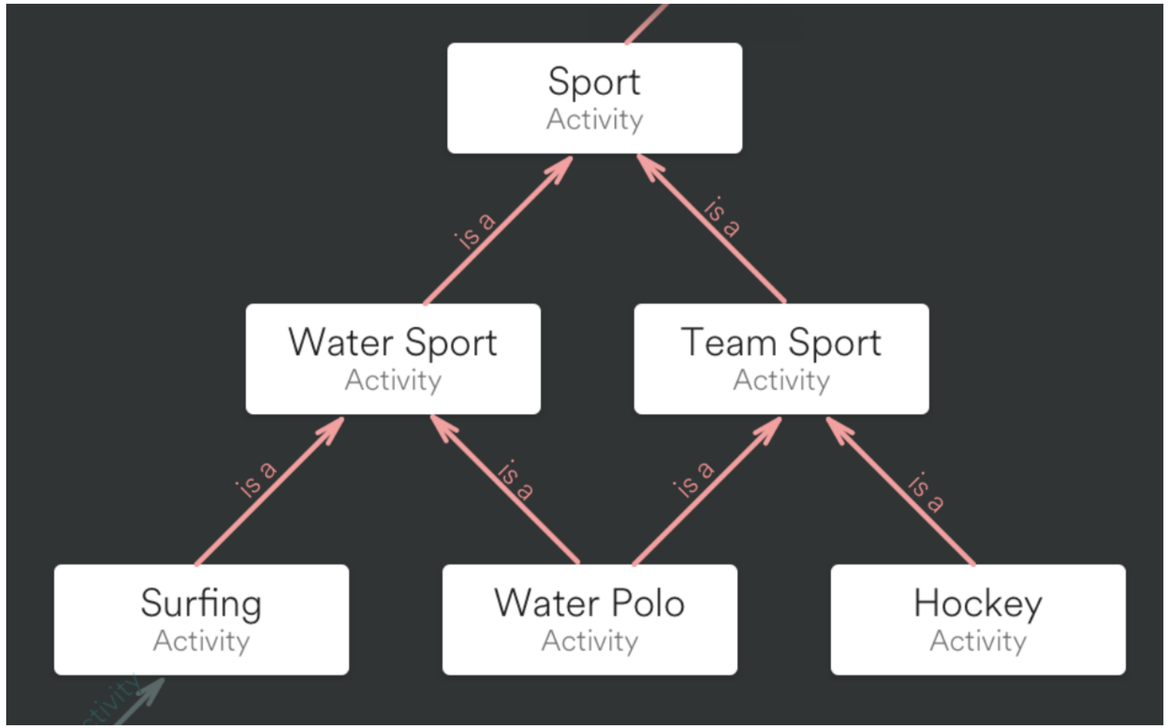
HOME AI BACKEND DATA INFRASTRUCTURE NATIVE WEB | OPEN SOURCE

Scaling Knowledge Access and Retrieval at Airbnb

Introducing our Knowledge Graph for encoding relationships and surfacing relevant information

Airbnb

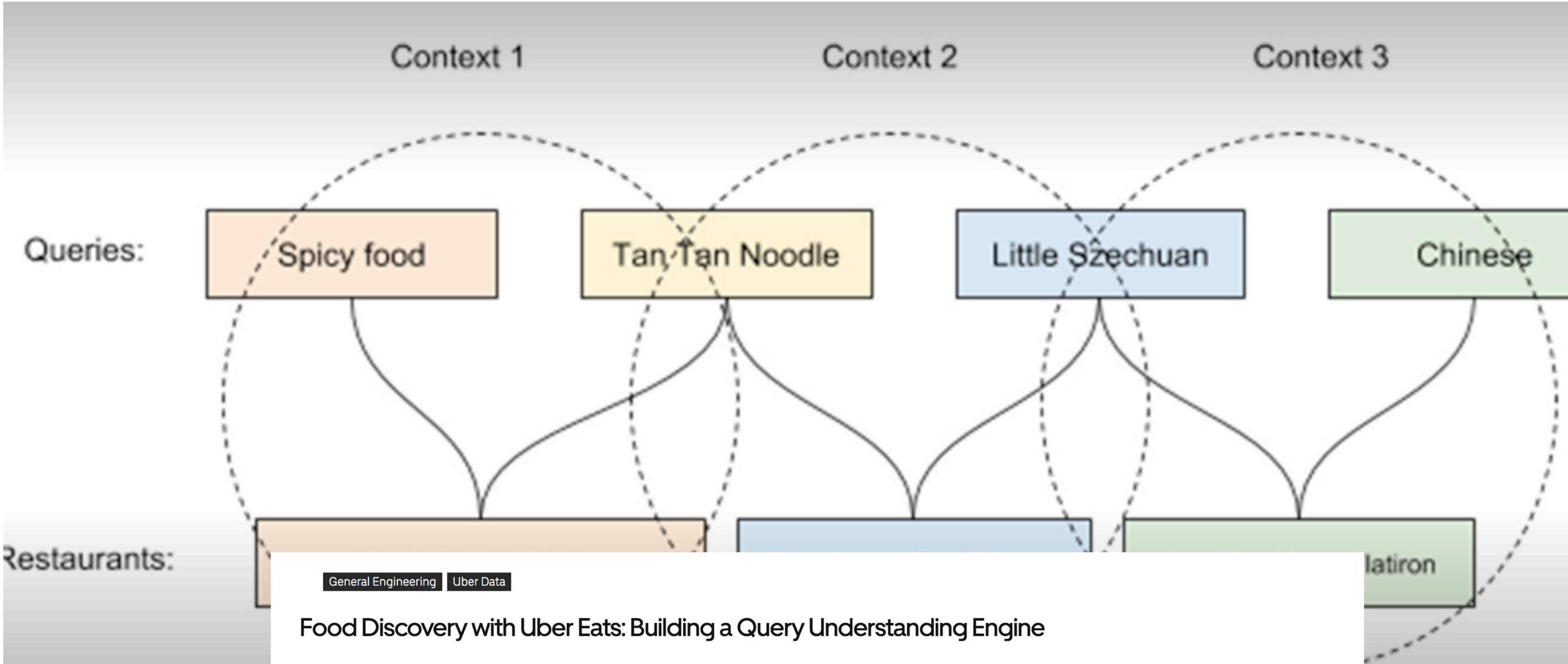

EXPERIENCES
PLACES
EVENTS
HOMES
...
...
TAGS
USERS
NEIGHBORHOODS
MARKETS
RESTAURANTS


Sport
Activity
is a
is a
Water Sport
Activity
Team Sport
Activity
is a
is a
is a
is a
Surfing
Activity
Water Polo
Activity
Hockey
Activity

UberEats

Uber Engineering

https://eng.uber.com/uber-eats-query-understanding/



General Engineering | Uber Data

Food Discovery with Uber Eats: Building a Query Understanding Engine

Recognai

Learn new things from your wild data.

Everyday.

REQUEST A DEMO

contact@recogn.ai