Scalable Text Understanding

Multilingual sentence embeddings

Andrew Yeager
GTC - 2019

MEDALLIA

Today's Discussion

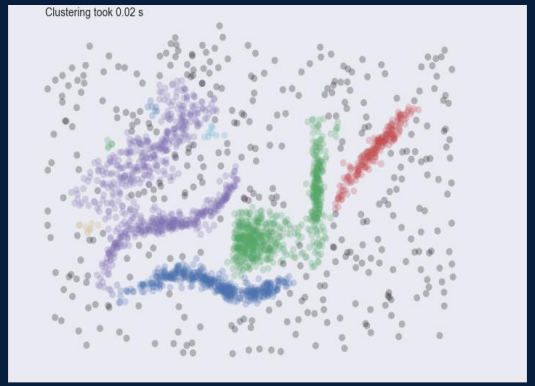
1. Why (multilingual) embeddings
2. The Common Techniques
3. Measuring Performance
4. Our Approach + Roadmap
5. Hands-On Learnings

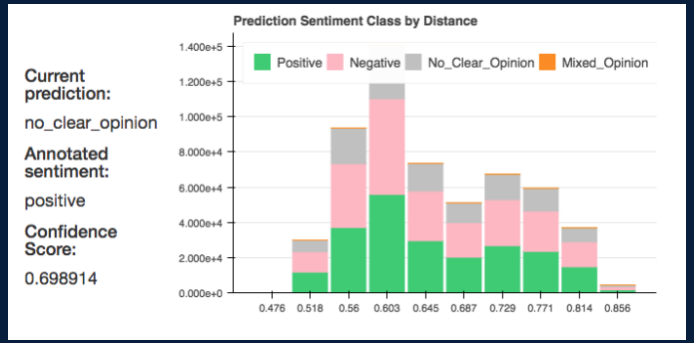


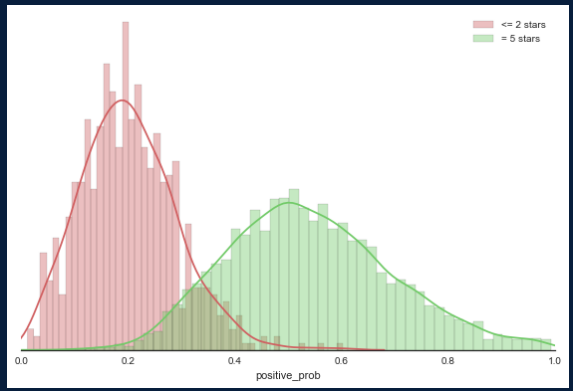





What is Medallia?

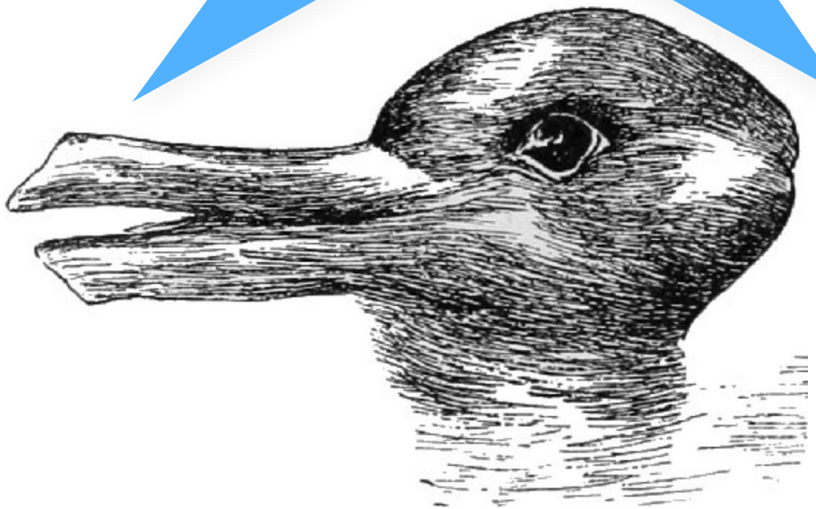




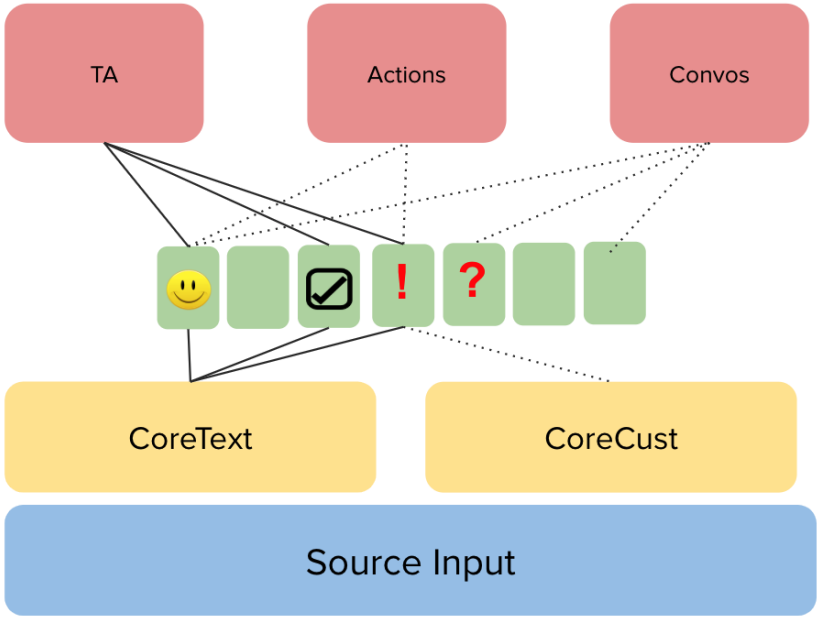


The Challenges

It crashed every day



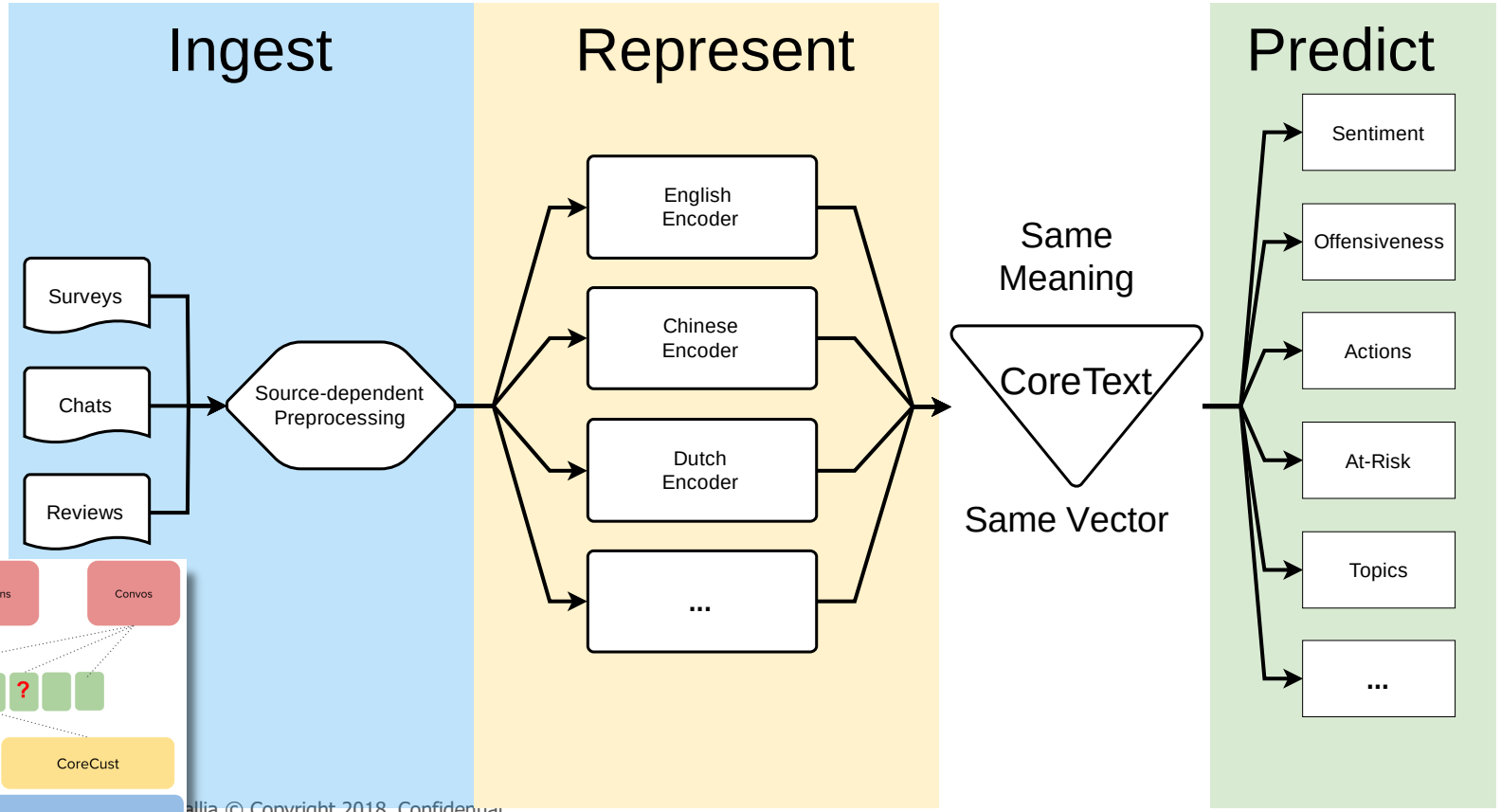
Athena

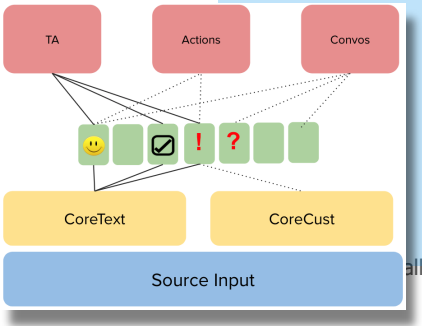


- Core Representations do heavy lifting
- Task models are lightweight
- Task models focus on task - not NLP

How do we deliver AI-based features fast?

CoreText





CoreText in Action

Finding Suggestions

 # Comments to Find 10 Suggestions:

English : ~~166~~ **25**
Portuguese : ~~200~~ **13**

2000 ***Portuguese*** Suggestions
\+ **CoreText** **+** Linear Model

15x reduction in Portuguese
7x reduction in English

Benefit from any annotation in any language

CoVe

The Embedding Space

Skip Thoughts

The Embedding Space

salesforce

CoVe

fast.ai

ULMFiT

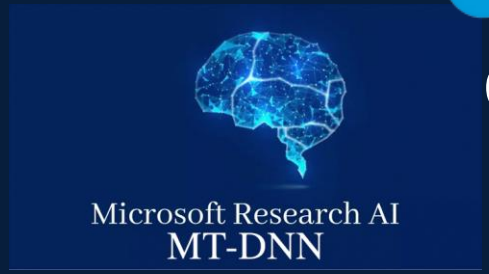
facebook research

InferSent

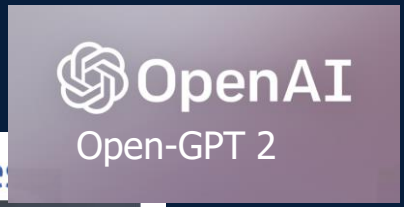
Skip Thoughts

AllenNLP

ELMo



CoVe



The Embedding Space

Skip Thoughts

The Recipe

Word Vectors

GloVe
Numberbatch

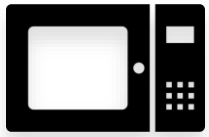
Add an Encoder

BiLSTM
Transformer
CNN + Attention

Mix in Training Task

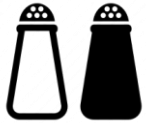
Language Model
Entailment
Translation
Others...

Cook on High



Place on GPU and train until ready, stirring data occasionally

Season to Taste



Fine-Tune

Performance Metrics

What's our Objective?

Support all current and **future** models

All models automatically support same set of languages

Models can be built using low volumes of annotated data

Performance Metrics

Traditional - Research Focus

- SentEval
- GLUE

Medallia's Metrics - Business Focus

- Multilingual Performance
- Task Coverage
- Learning Efficiency

Task	Type	#train	#test	needs_train	set_classifier
MR	movie review	11k	11k	1	1
CR	product review	4k	4k	1	1
SUBJ	subjectivity status	10k	10k	1	1
MPQA	opinion-polarity	11k	11k	1	1
SST	binary sentiment analysis	67k	1.8k	1	1
SST	fine-grained sentiment analysis	8.5k	2.2k	1	1
TREC	question-type classification	6k	0.5k	1	1
SICK-E	natural language inference	4.5k	4.9k	1	1
SNLI	natural language inference	550k	9.8k	1	1

MEDALLIA

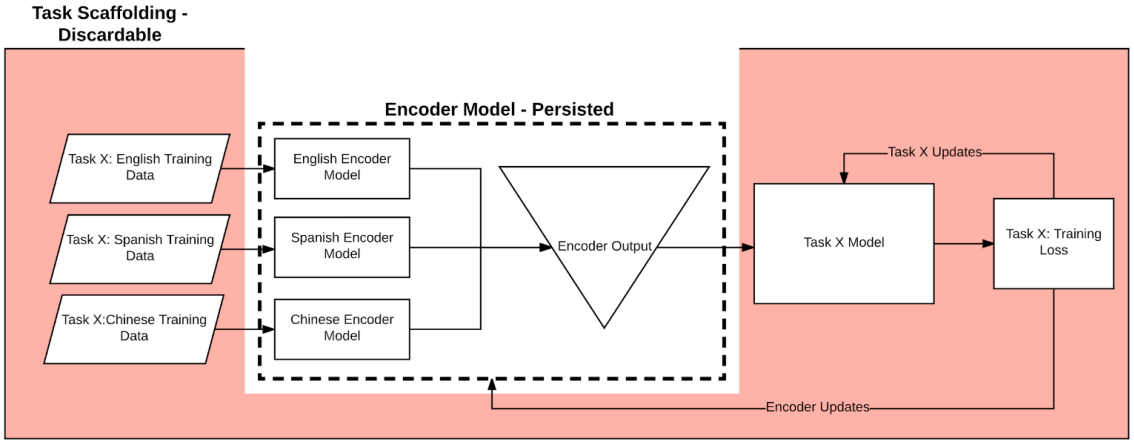
Medallia’s Strategy

Multi-task Learning with Composable Models

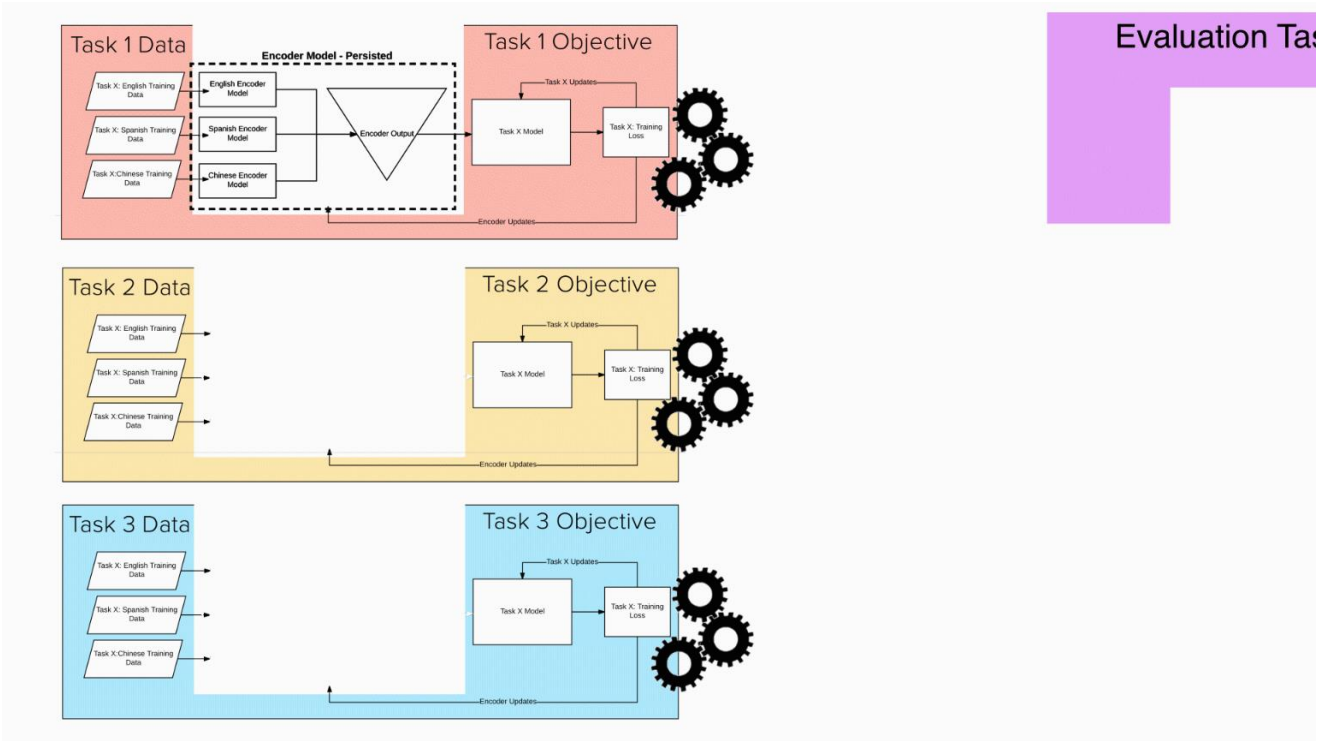
Abstraction of tasks and encoders

Allow a single encoder to be shared and updated by many tasks

Easily experiment with new combinations of tasks and encoder styles



Training Cycle



Results

MEDALLIA

CoreText

8 Languages where Sentiment Model Improved

+18% Japanese Accuracy

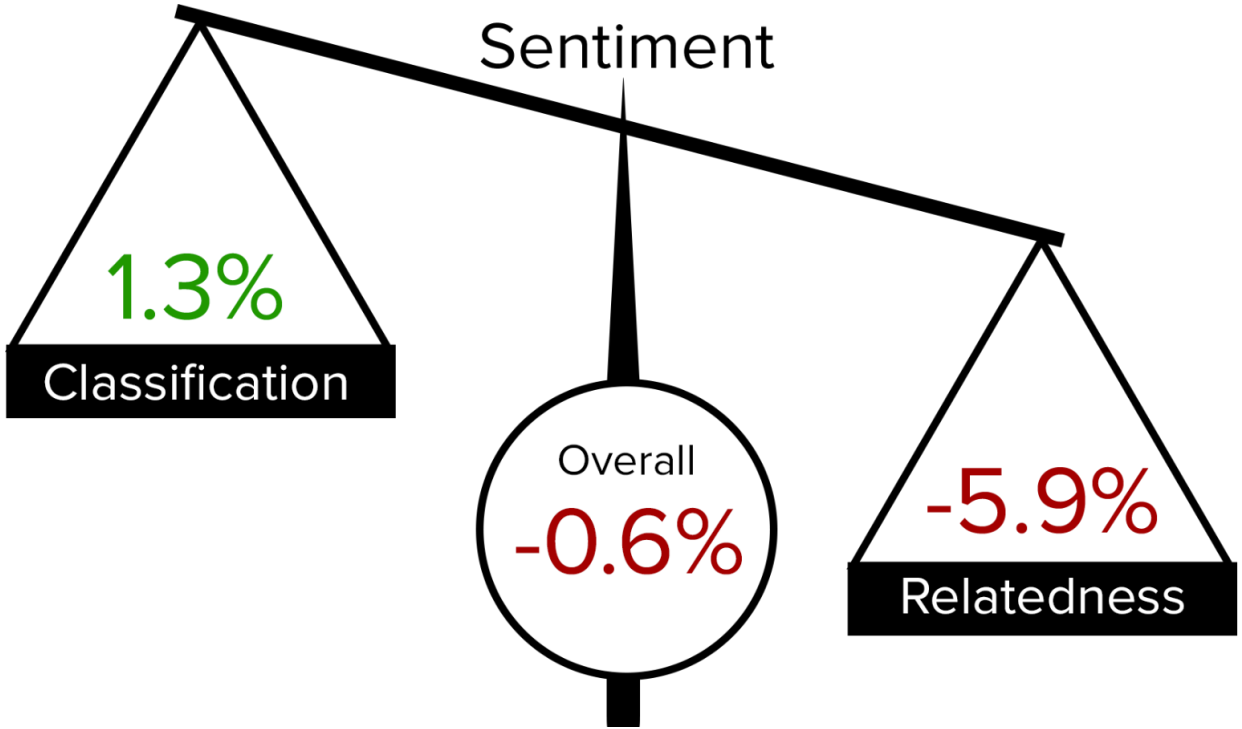
+4% English Accuracy

3 Production-Ready Models

0 Annotation Projects Needed

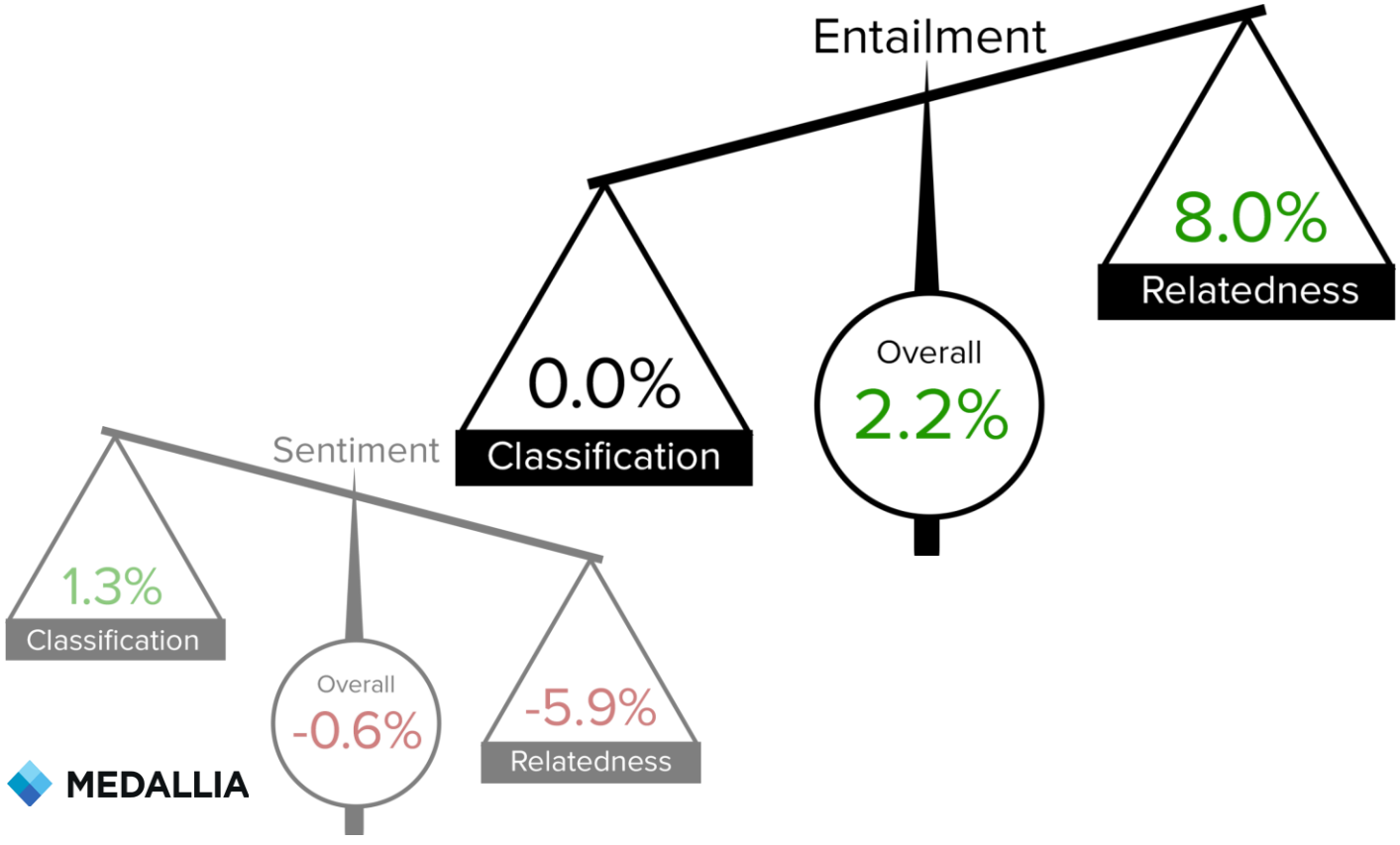
Validating Multitask Learning

Difference from Mean on SentEval Tasks



Validating Multitask Learning

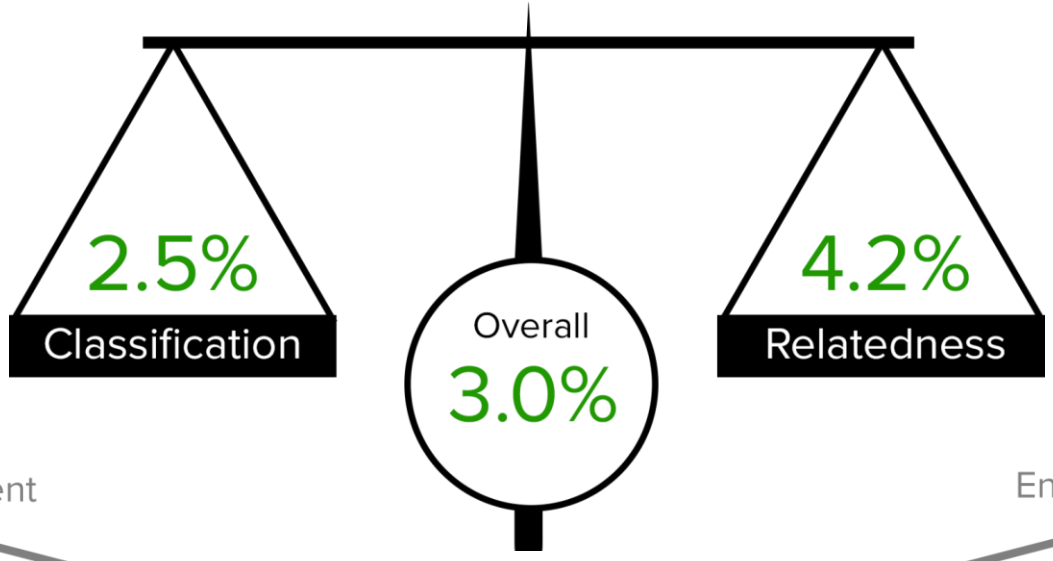
Difference from Mean on SentEval Tasks

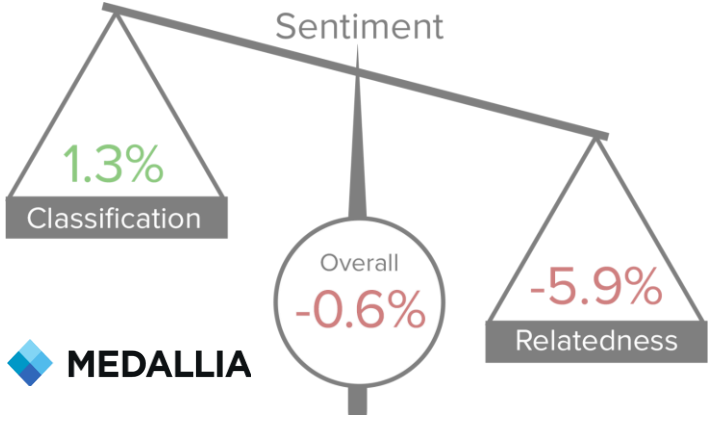


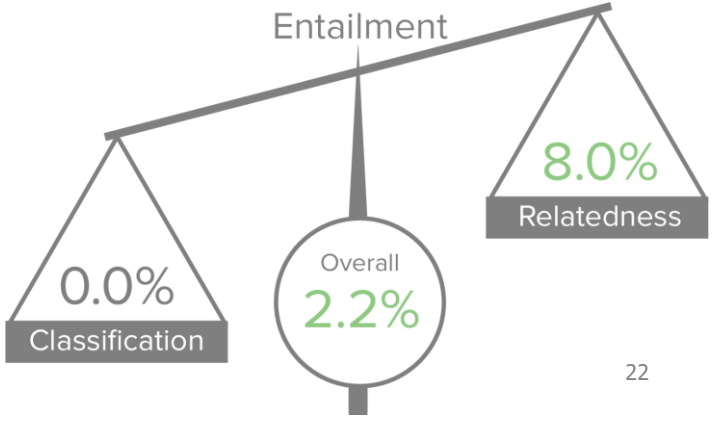
MEDALLIA

Validating Multitask Learning

Difference from Mean on SentEval Tasks





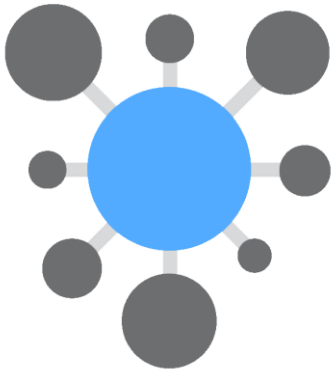


MEDALLIA

Next Steps

Experiment, Experiment, Experiment

Model Feedback Loop



Leverage Pre-trained Embeddings

Integration with Context

First-Hand Learnings

Multi-task Multi-lingual Approach

- Multilingual embeddings can replace model-per-language approach
- Introducing data in one language can improve performance in other languages, even when the task in those languages is also different

Design Decisions

- Sentiment prediction as a training task can work
- Multi-task learning by swapping is simple and effective
- Efficient experimentation is essential
- Latest-and-greatest not always best for you
- Encoder per language is a GPU memory hog
- We're getting near 100% GPU utilization with shallow LSTMs

Our GPU vs CPU Experience

- 14x training speedup for LSTM-based Encoder on single V100
- 20x Inference speed (6000 phrases/sec)

Please leave Feedback Through App

To learn more about Medallia visit

www.medallia.com