Transfer Learning in NLP

Helping Small Teams Account for Small Datasets

Ryan Smith
ryan@wootric.com

Transfer Learning in NLP

- What we'll cover
  - A look into a real problem involving NLP and Deep Learning
    - A brief discussion of the pros and cons of methods we tried
  - How Transfer Learning can help **small teams with less data** compete with established corporations
  - A look at our results from applying these methods

WOOTRIC

Wootric - What We Do

Collection

Track core CX metrics with micro-surveys that customers love

- Right question: NPS, CSAT, CES
- Right channel
- Right time

Analysis

Aggregate feedback from any source into one dashboard

- Autocategorization
- Sentiment analysis
- Algorithms tailored by industry

Action

Align the company around a single source of customer truth

- Deep & broad integrations
- Routing for action

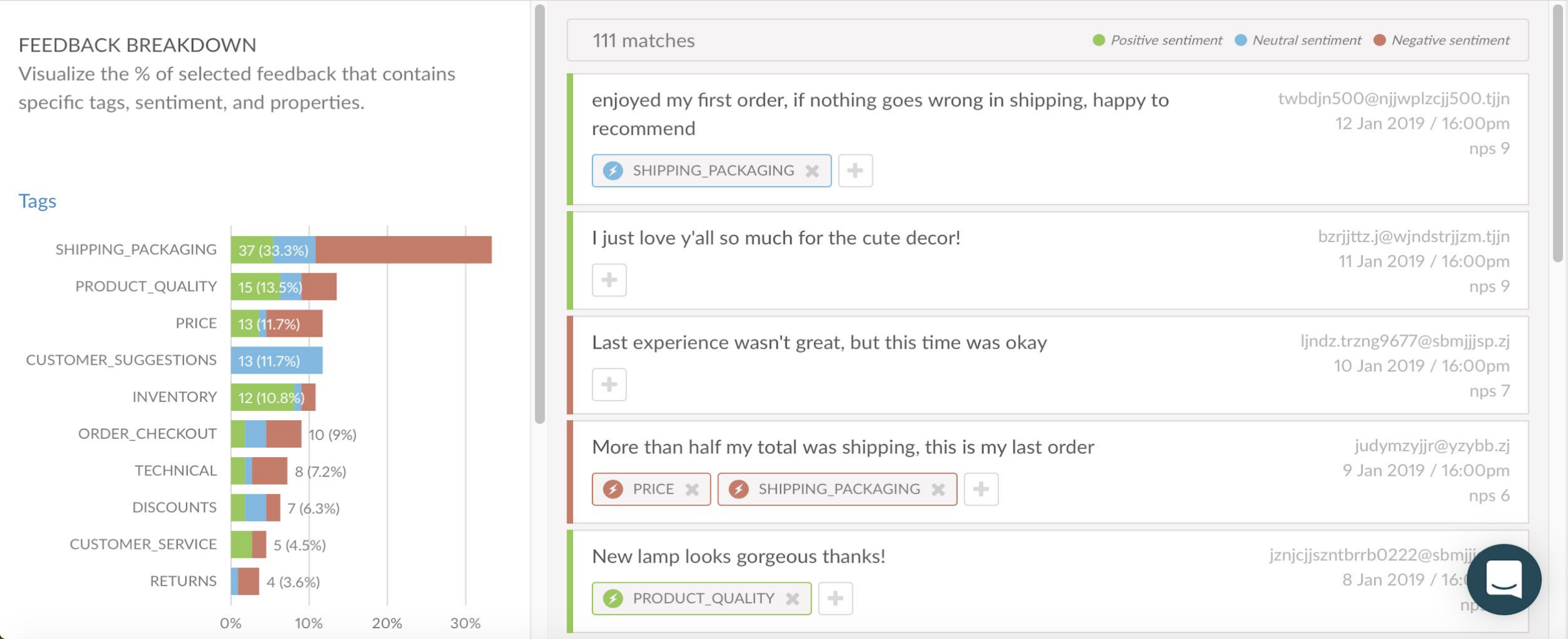
WOOTRIC

Wootric - Problem We Want Solved

- Survey collects a lot of feedback
  - What set of topics is the customer commenting on?
    - Multi-Label Classification
  - How does the customer feel about the product/service?
    - Sentiment Analysis

WOOTRIC

Wootric - Problem We Want Solved



Metrics to Evaluate

- Precision
  - Given we have “tagged” a piece of feedback, how often are we correct
- Recall
  - What percent of the feedback that we should tag are we actually tagging
- F1-Score
  - Combination of the two
  - F1-Score = 2 * Precision * Recall / (Precision + Recall)
  - We will report this for discussing model quality

WOOTRIC

Applying ML

- Formal Problem:
    - “Given this piece of feedback and its industry, what tags should be applied?”
        - Multi-Label Classification: Applying a set of binary labels
    - Metrics: Precision, Recall, F1-Score ***for each tag***
- For Business, it is nice to implement Low-Cost solutions first
    - A very basic model
    - An existing service

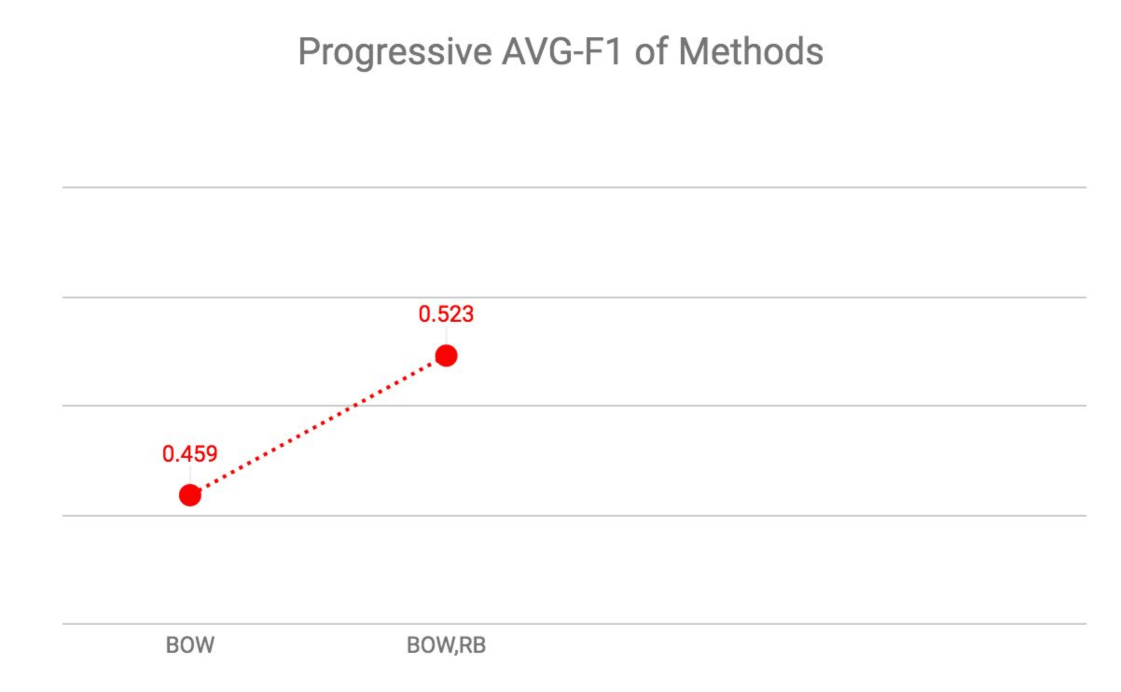
WOOTRIC

Using a Basic Model

- Models
  - Bag of Words
  - Rule Based
- Gives a good baseline
- Can keep iterating
- Requires that you have a production system in place

WOOTRIC

Using a Basic Model - Results



Using a Basic Model - Problems

- Language is hard to model
    - “The engineering ***cost*** to implement your product was too high”
        - Rule Based & BOW methods would tag as Price (incorrect)
    - “I really hate how much I love your product”
- Bag of Words and Rule Based approaches could be improved

WOOTRIC

Using an Existing Service

- Google Prediction API
  - Easy Interface
  - Had Binary or Multi-Class options
    - Used one classifier per tag, since our problem is Multi-Label
- Gave better results than BOW
  - Passed the baseline!

WOOTRIC

Problems

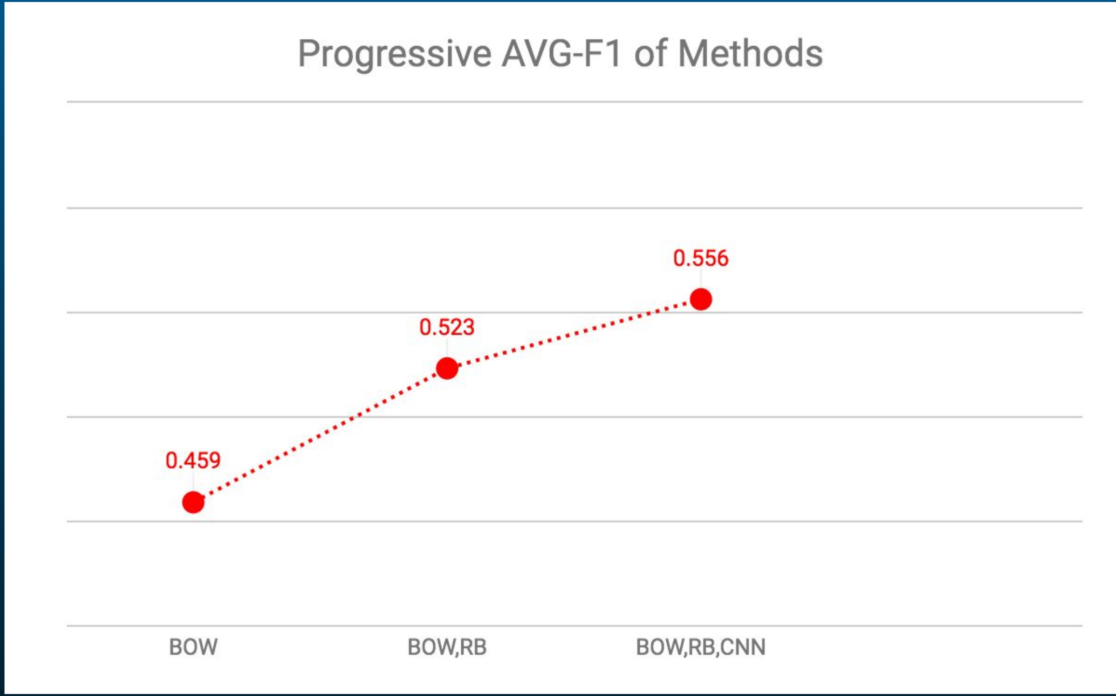
- Unfortunately, Prediction API began failing regression tests
  - Training process no longer gave good results
  - Google deprecated it soon after
    - AutoML did not come out until another year down the road
- Problem with black box systems: You have no control
- Now we only have basic methods, need better accuracy

Applying Deep Learning

- Deep learning is fun!
    - But (relatively) time consuming
    - Want to make sure it’s worth the time investment
- Used basic CNN and LSTM models
    - CNN did well
    - LSTM was not effective

WOOTRIC

Applying Deep Learning - Results



WOOTRIC

Problems - Small Training Set

- Have a lot of Feedback
  - Manually labeling is time consuming
- Class Imbalance Problem
  - Makes each additional chunk of labeled data less effective
- How can we learn from so few examples?
  - And still compete with models that use hundreds of thousands of training rows

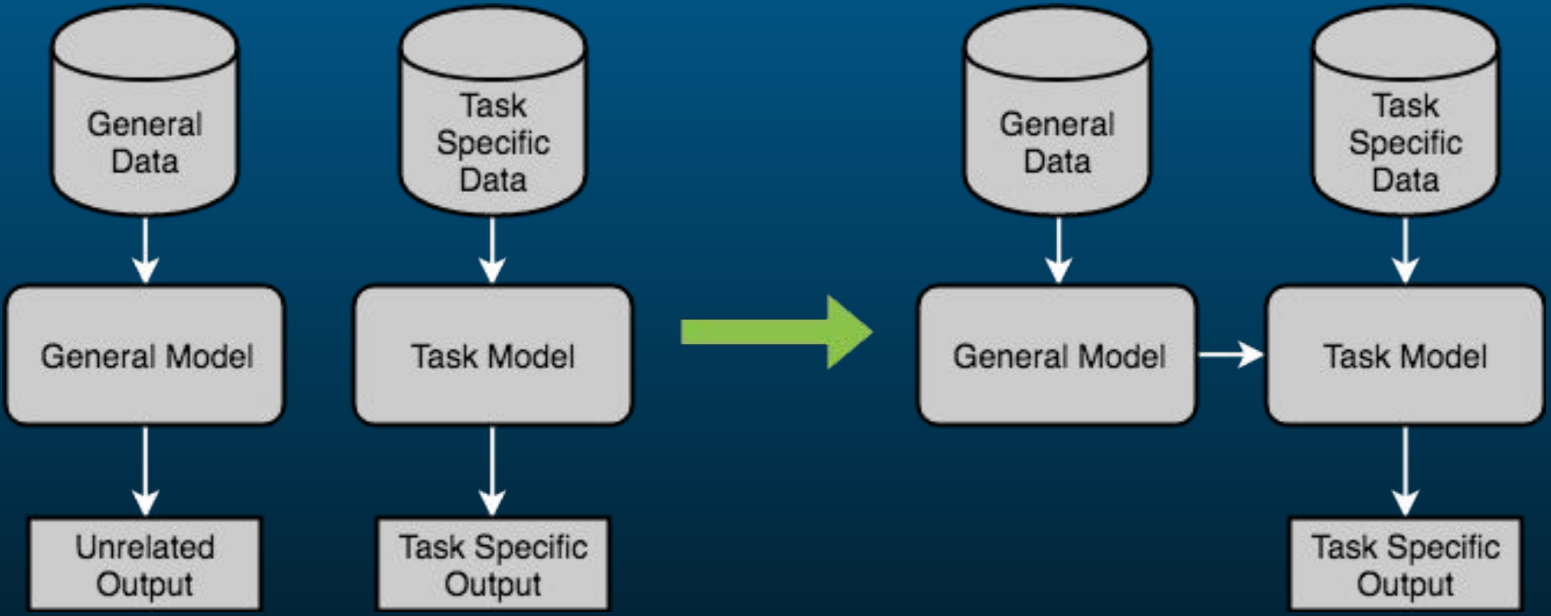
WOOTRIC

Transfer Learning

- Want to make use of as much data as possible
- A model trained on a separate domain **can still be useful**

WOOTRIC

Transfer Learning



WOOTRIC

Transfer Learning

- More Data is better but how do we utilize it?
- Common Techniques include
    - Using parts of ImageNet models
    - Prior distribution for Bayesian Analysis
    - Word Vectors
    - Language Models (Just Recently)

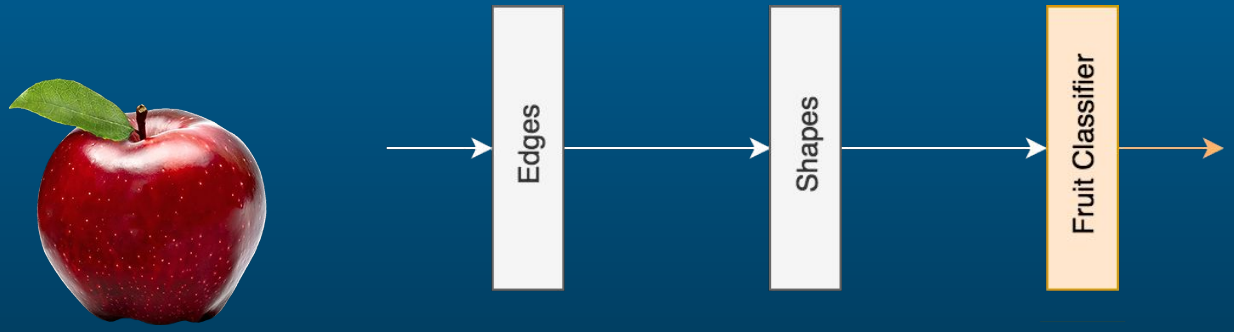
WOOTRIC

Transfer Learning in Computer Vision

- ImageNet
  - Learn low-level features from general data
    - Edges, shapes, colors, etc.
  - Build new classifiers on top for domain-specific tasks

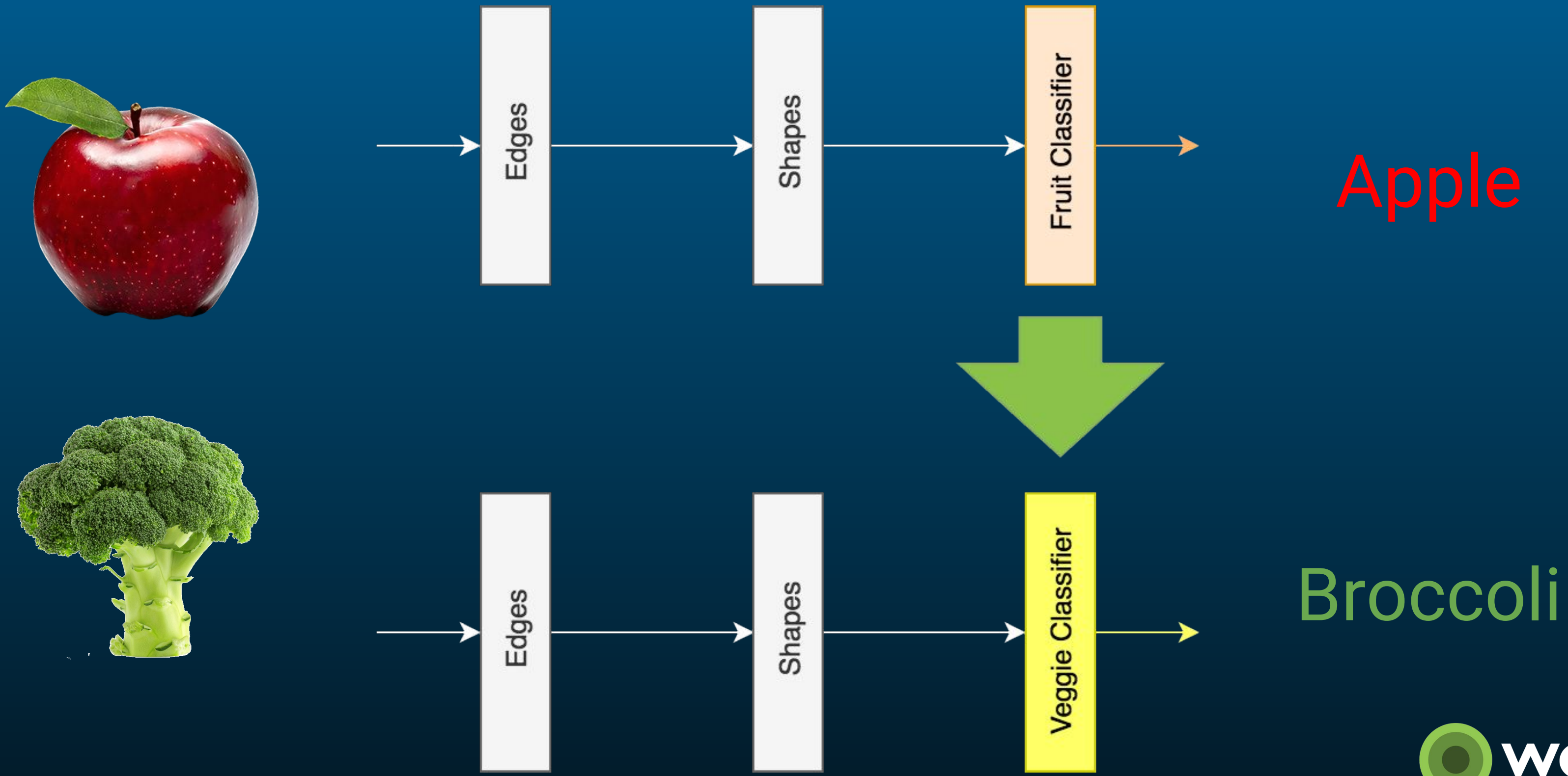
WOOTRIC

Transfer Learning in Computer Vision



Apple

Transfer Learning in Computer Vision



WOOTRIC

Transfer Learning in NLP

- Word Vectors
    - Huge stride in 2012
    - Learn **One Initial Layer** of a model
    - Only captures one aspect of language
    - Infamous GoogleNews generated word vectors

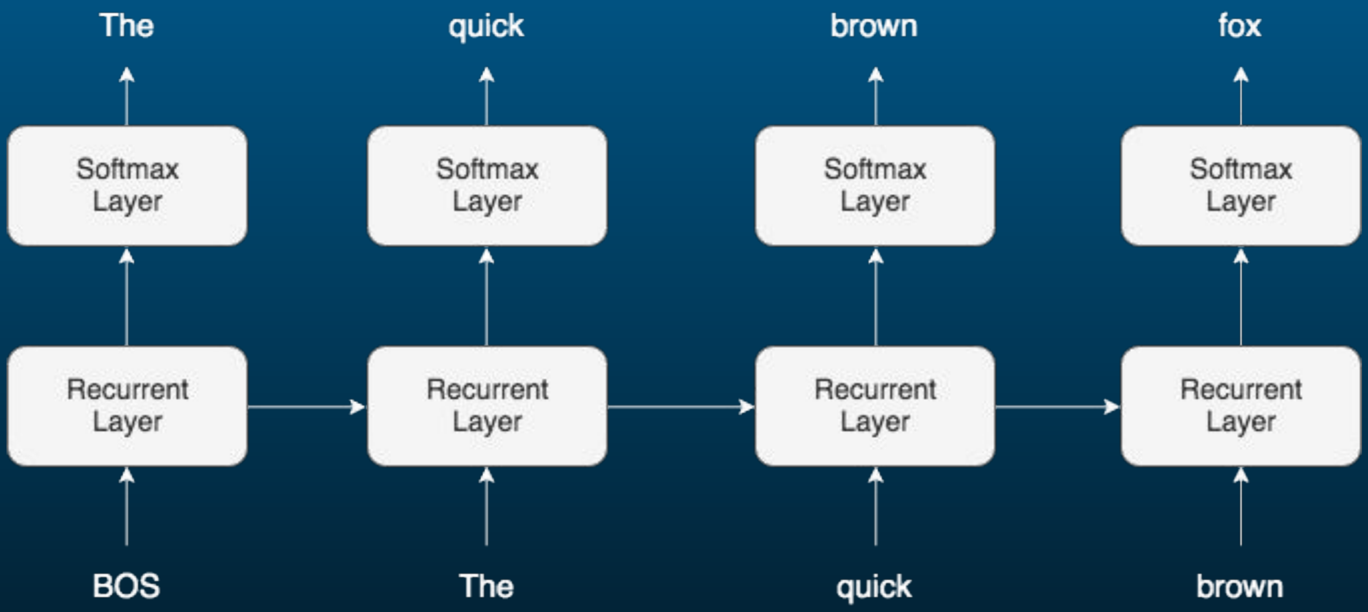
WOOTRIC

Transfer Learning in NLP

- Language Models
  - Learn **Multiple General Purpose Layers**
  - Trained to model language, not just words
    - A good Language Model will differentiate word sense
      - “I **hit** the ball”
      - “Our website got a lot of **hits**“
    - Order of words matters
  - **No labeled training data needed**

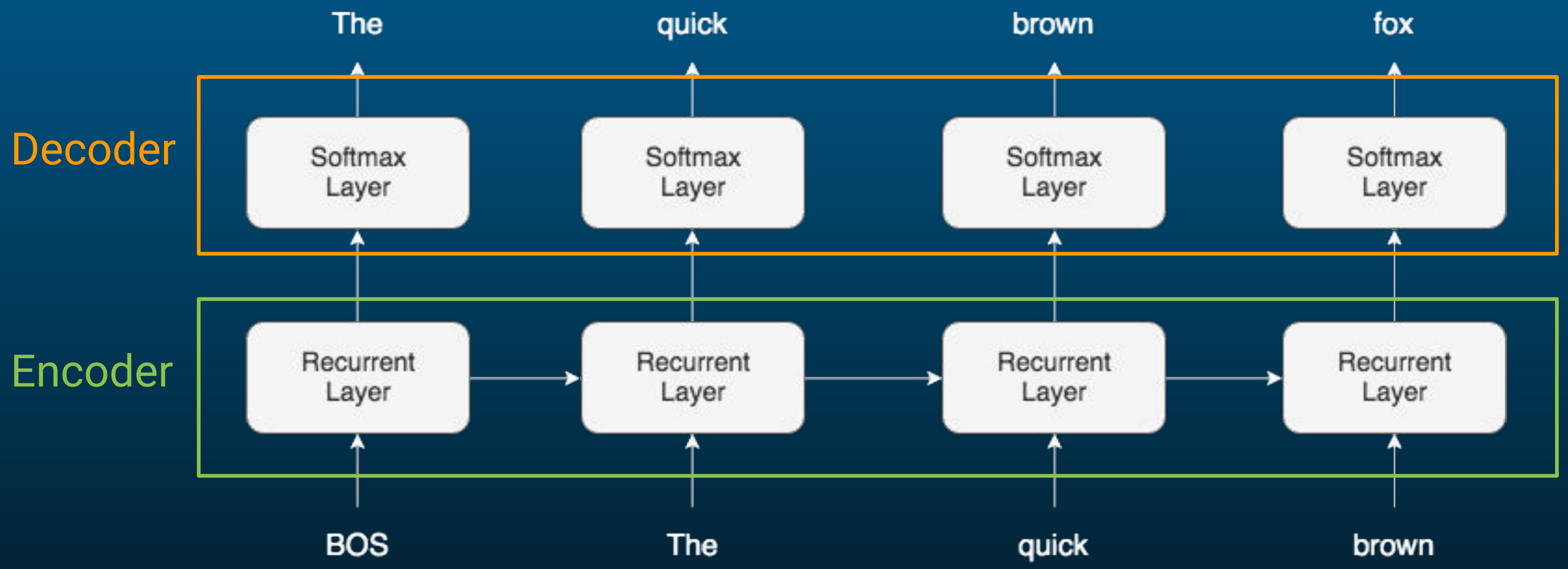
WOOTRIC

What is a Language Model?



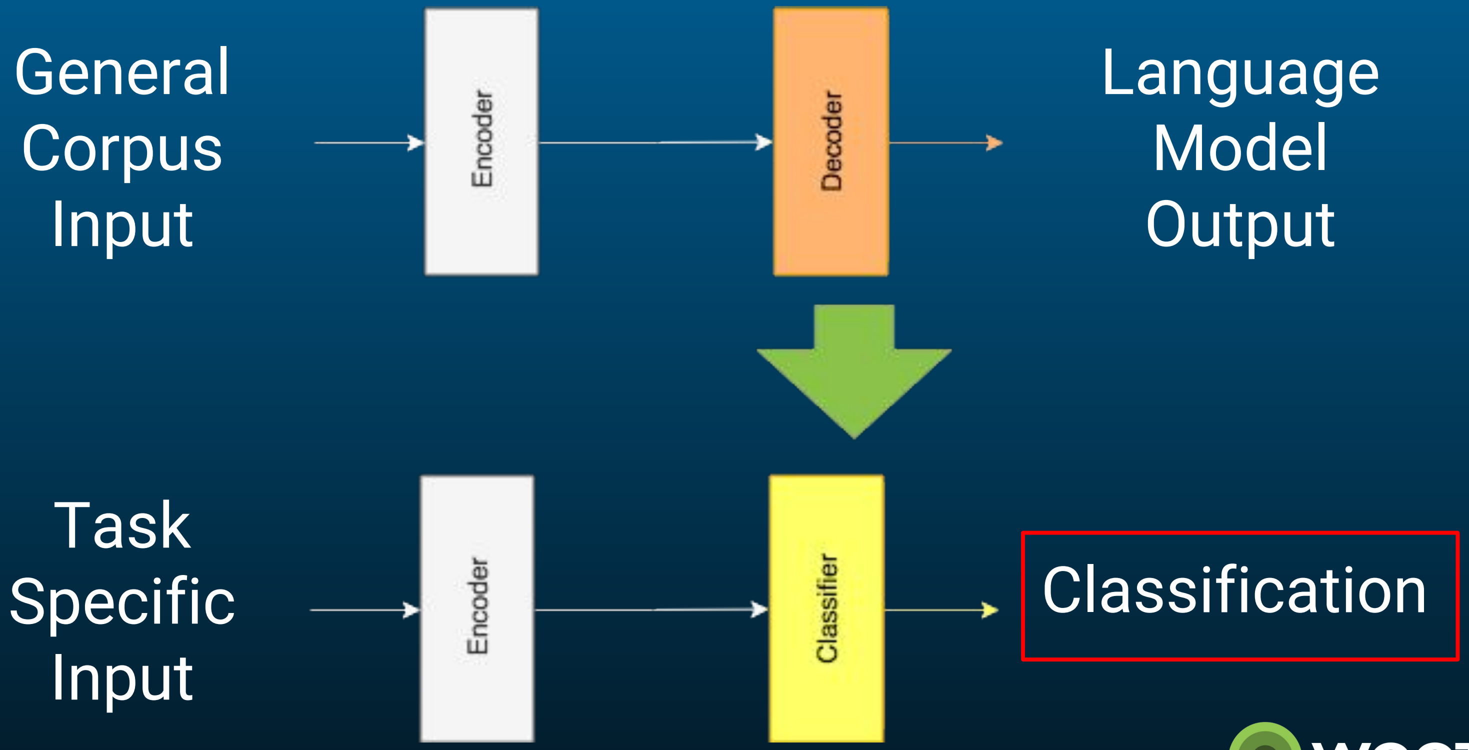
WOOTRIC

What is a Language Model?



WOOTRIC

Building from Language Models



WOOTRIC

Building from Language Models

- **Initialize Model State** for your next task with the **Encoder of the More General Task**
- Can iterate this process as much as necessary
    - Don't need to settle for one general purpose Language Model
    - Use progressively more relevant corpuses to fine tune the language you will see in your data
    - Add a classifier for the last step, on your labeled data

WOOTRIC

Transfer Learning in NLP

- "NLP's Imagenet Moment"
  - Finally, we can use Transfer Learning to quickly productize DL models for NLP
- Can make use of publicly available text (and models)
  - Wiki-Text
  - Penn TreeBank
  - Twitter Stream
  - Web Crawl

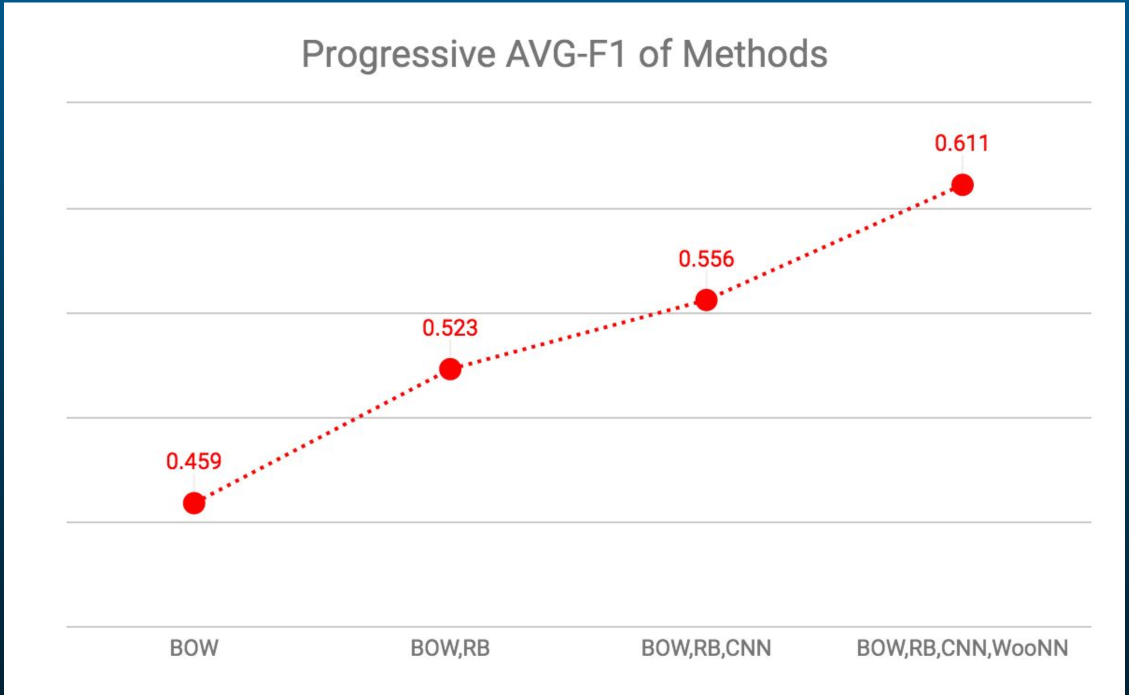
WOOTRIC

Our Transfer Learning Model

1. Language Model over WikiText-103
   - There are pre-existing versions of these
2. Refine the Language Model on our (unlabeled) corpus
   - Adapts to the Customer Feedback domain
3. Train on our specific labeled data

WOOTRIC

Our Transfer Learning Model - Results



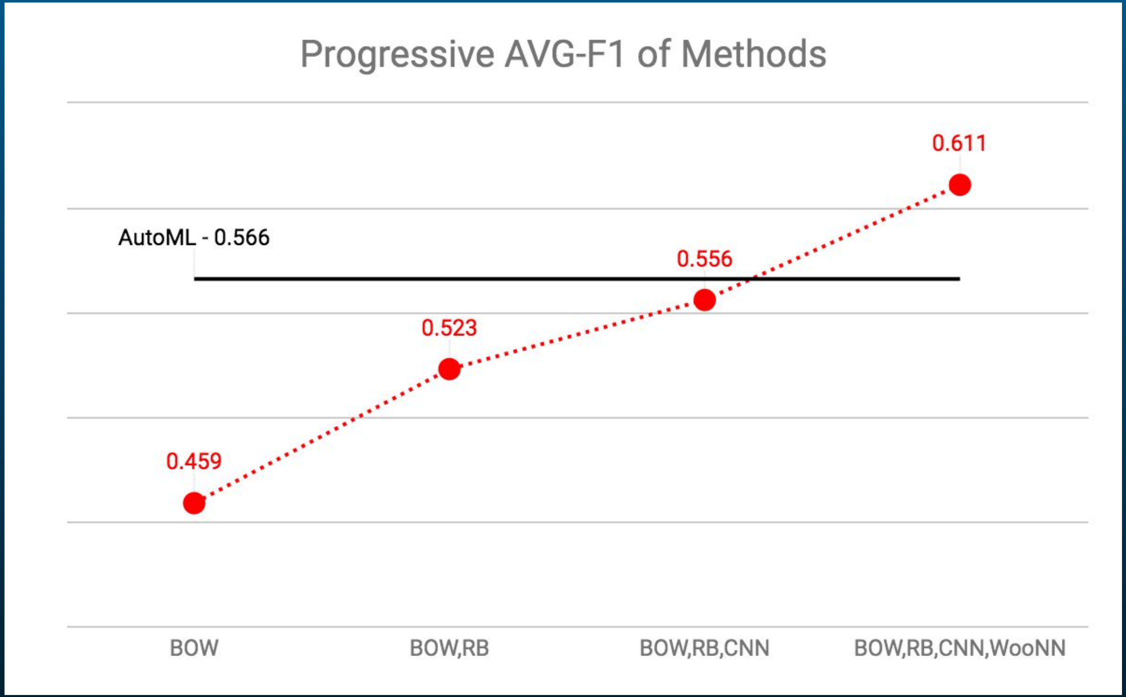
WOOTRIC

AutoML

- AutoML came out after we got used to not having the Google Prediction API anymore
- Needed to compare our own models to see how we did

WOOTRIC

AutoML - Results



WOOTRIC

Sentiment Model

- Sentiment model is also important
    - We use it in determining Aspect-Based Sentiment Analysis for our tags
    - Trained on our own (smaller) dataset
        - Language Model pre-trained with Customer Feedback corpus
- 96.5% accuracy (WooNN) vs. 92% accuracy (Google NL API Sentiment)
    - Just because Google is Google doesn't mean you can't beat them in your own domain

WOOTRIC

Conclusion

- Evaluate if you need DL/Transfer Learning first
- We often have access to general, unspecified data
    - Combine with small, specific data to succeed in your domain
- Make use of as many building blocks that can transfer as possible

References

- “NLP’s ImageNet Moment has Arrived” - Sebastian Ruder
  https://thegradient.pub/nlp-imagenet/

- “Universal Language Model Fine-tuning for Text Classification” - Jeremy Howard, Sebastian Ruder
  https://arxiv.org/abs/1801.06146

Questions