Who is Tribe Dynamics?

Tribe is a **consulting firm** for brands operating in the cosmetics industry.

Tribe advises cosmetic brands ' on their **social media marketing strategies**.

Tribe **measures brands' engagement** on social media contents to quantify campaigns' success.

4horsemedia.com/product/professional-plan-social-media/

Tribe Dynamics

Operations

Trybe currently **uses a classification model** to detect whether a post talk about a specific brand and about what product.

Tribe **measures number of appearances** of brands and products on social media to quantify the success of brand’s campaign.

Tribe **uses the Earned Media Value metric** that provide its clients with a weighted average of likes, comments, and posts share for each brand.

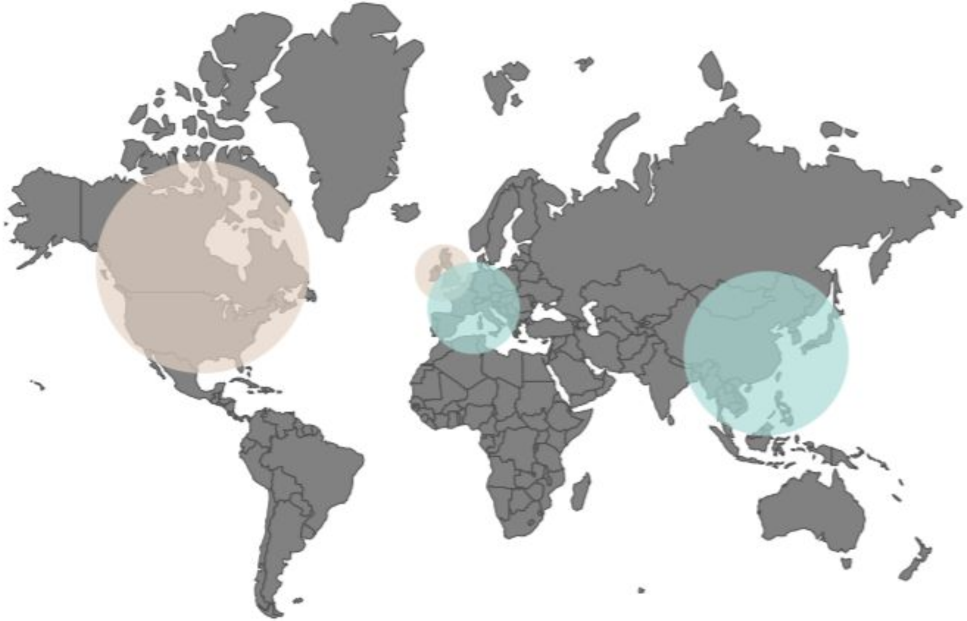
Tribe Dynamics

Operations

Tribe desires to **expand its operations** to European and Asian market.

Tribe's current classification model has **low accuracy for non-English** written posts.

Our team's goal is to build a **model that detects brands and scale well across languages**.



Classification Task Performance

Trybe currently **uses a N-grams Logistic Regression model** on a multilingual-vocabulary**.**

Our team **replicated Tribe's model** and we tested it on a different datasets (i.e. Dove).

Our results confirms that classification model achieves:

a) higher accuracy on English vocabulary compare to others

b) higher accuracy due to class imbalance and greater training data

Accuracy N-grams LR model

	Tribe	Our Team
English :	**0.92**	**0.91**
Non English:	**0.84**	**0.96***

*0.95 Class Imbalance Ratio

Observations and Models Proposal

Issue with N-grams:

1. N-grams are **not scalable**
2. Requires **many labels**
3. Words' **meaning disambiguation** (e.g. "Dove")

Models proposed:

Word Embeddings
trained on each vocabulary
(unsupervised ML)

Mixture Model
trained on class probability distribution
(generative model)

Tribe Dynamics

1. Word Embeddings

Across the languages there are words with similar intent

Model learning

Clustering words with similar intent together

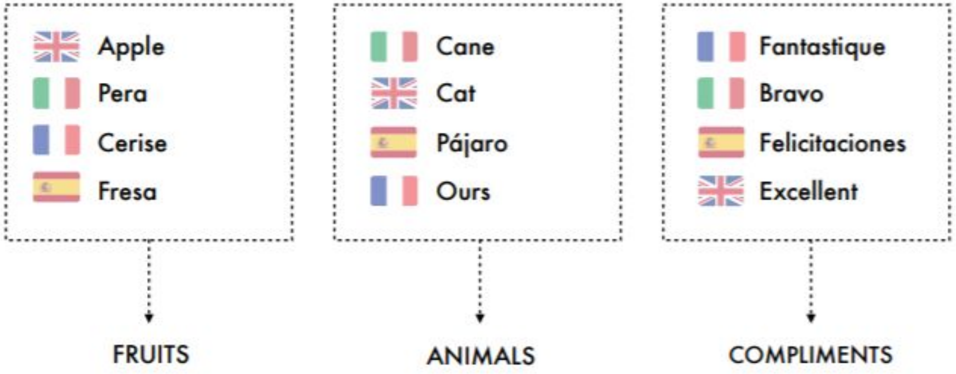
[one model for all languages]

Words are clustered by intent

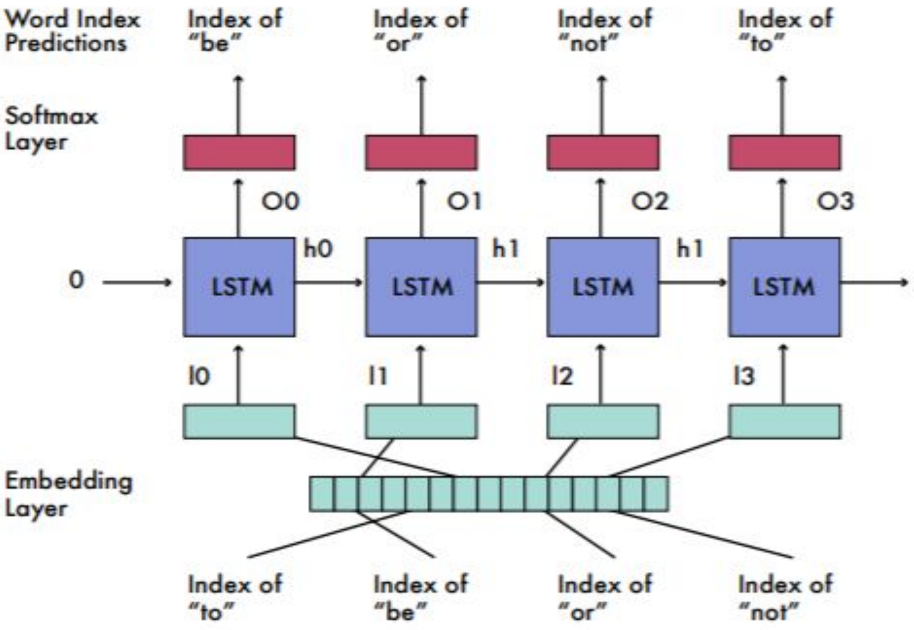
Using:

NEURAL NETWORKS

Functional form of similarity, end-to-end learning



Language models and NN architecture



Part of the Neural Network architecture
http://colah.github.io/posts/2014-07-NLP-RNNs-Representations/

LANGUAGE MODEL

- Predict next word given the current sequence history.
- Embeddings -> Transformation on the initial word vectors
- End-to-end backprop. Learn functional form without any domain knowledge.
- Key idea : Similar words in context lead to similar word successors.

Language models and NN architecture

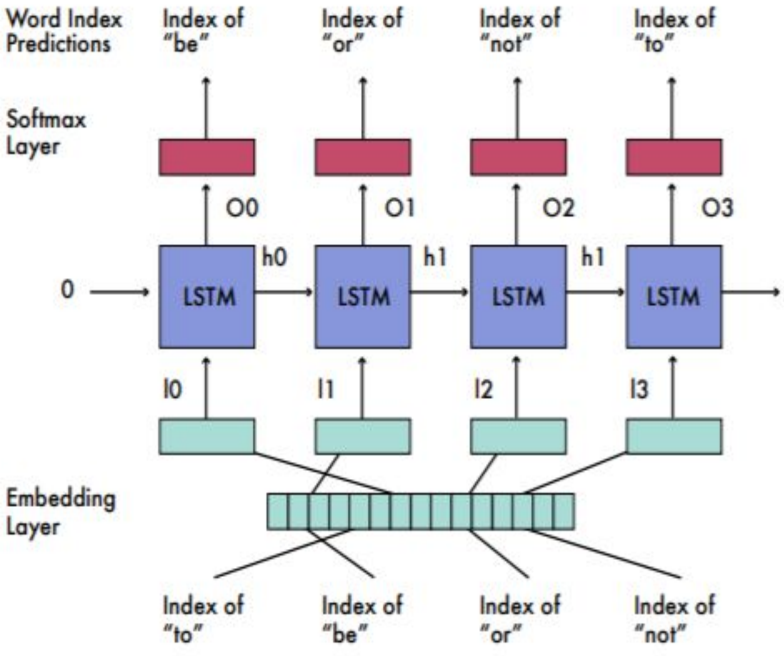
Pre trained embeddings was not completely useful to be used out of the box (fashion vocabulary is different).

<u>Dynamic Embeddings</u>

1) Increase vocabulary to add new fashion-specific words

2) Multichannel embeddings - Have two channels of embeddings(static and dynamic) and backpropagate only on the dynamic channel.

- Retain Facebook’s MUSE(FastText) embeddings knowledge and add new knowledge that is trained to task.



Part of the Neural Network architecture
http://colah.github.io/posts/2014-07-NLP-RNNs-Representations/

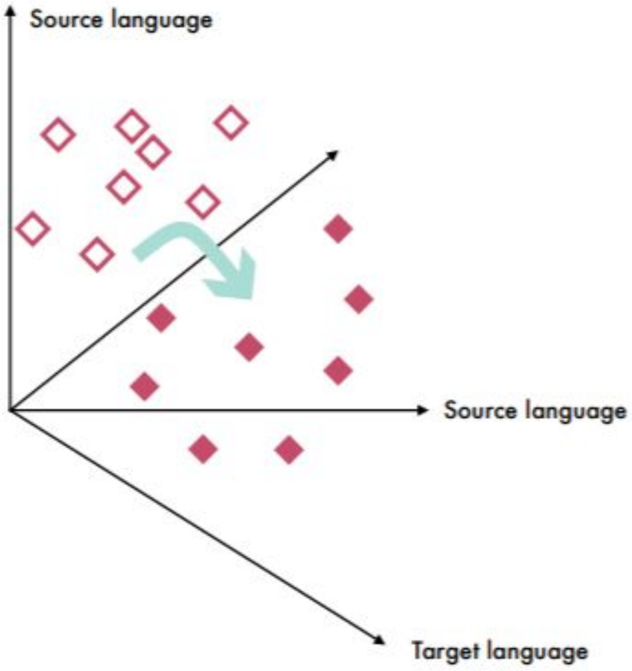
Alignment

1) Use existing alignment – Facebook Muse (supervised and unsupervised alignment using adversarial training). Complicated method.

2) Find transformation (rotation, translation and stretch using few anchor words) – find matrix operations that maximize cosine similarity between two embeddings.

<u>Why alignment beneficial?</u>

Because one single model in embedded space



http://colah.github.io/posts/2014-07-NLP-RNNs-Representations/

Mixture Model for Cross-Language Classification

Motivation

1) Multiplicative Cost of increasing bag of words and n-grams
2) Ground Truth unreliability in non-English languages (80% vs 97%)
3) Costs (MTurker, Initial data collection downtime)
4) Good solid model in English.

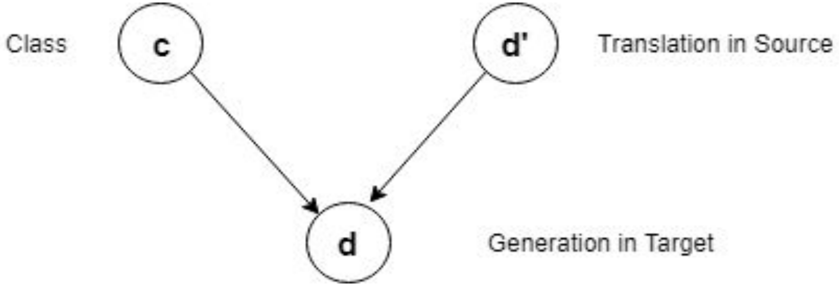
Question: What if we had a bilingual lexicon and translate words to english?

Issue: Word sense disambiguation

Dove Sei - Where are you ?

Io Amo **Dove** - I love Dove.

Model



Bag of Words Assumption

$$P(d) = \sum_{c} P(c) \sum_{d'} \prod_{i=1}^{l} P(w_i|w'_i, c)P(w'_i|c)$$

- E-step

$$\begin{aligned} P(w'c|w) &\leftarrow \frac{P(cw'w)}{P(w)} \\ &= \frac{P(w|w'c)P(w'c)}{\sum_c \sum_{w'} P(w|w'c)P(w'c)} \end{aligned} \quad (1)$$

- M-step

$$P(w|w'c) \leftarrow \frac{f(w)P(w'c|w)}{\sum_{w \in K} f(w)P(w'c|w)} \quad (2)$$

Model - Continued (Remember : We inferred P(W|W'))

Classifier in Source Language

$$P(c|d) = \frac{1}{Z(d)} \prod_{w \in V} e^{\lambda_w f(w,c)}$$

Max Entropy Classifier / Softmax Regression

Model Translation

$$\hat{c} = argmax_{c \in C} \prod_{i=1}^{v} \sum_{j=1}^{n_i} P(w_t^{ij} | w_s^i, c) e^{\lambda_{w_s^i} f(w_t^{ij}, c)}$$

Where: t-Target(Italian), s-Source(English)

Old Example : 1. Dove sei(Not brand), 2. Io Amo Dove(brand); Dove - where, Dove.

Source model high weights for Dove for brand classification.

1) High P mass on dove ➡ where (under class label not brand)

2) dove ➡ brand (under class label brand)

Results and Expected Outcomes

(Performance On Italian Data, 80-20 train-test)	**Baseline Model (**LR + all words BOW trained entirely)	**Mixture Model** (Under-constructed lexicon)
Accuracy	0.93	0.85
AUC	0.82	0.79
PR	0.62	0.52

Success example:

trattamento completo che consenta risultati migliori in modo decisivo e rapido, rallentando gli effetti del tempo e andando ad agire **dove** diete e palestra non hanno risultato.Frutto di ricerca scientifca e innovazione, Kiemè è una linea di cosmetici che garantisce la'

Class : Baseline - True, Mixture - False

Takeaways

1. Build Confidence `with labeled data in target language (semi-supervised finetune)
2. Good performance on low data regimes - Real Transfer
3. Small BOW - more scalable and tractable. Setting up lexicon is one-time cost
4. Single model on source language - easier to maintain