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

Tribe scrapes posts, and currently **uses a classification model** to detect whether a post talk about a specific brand and about what product. **(OUR FOCUS AREA)**

Based on the **social media engagement** (likes, comments and shares) on these posts, Tribe calculates a proprietary **EMV** metric – offers to clients.

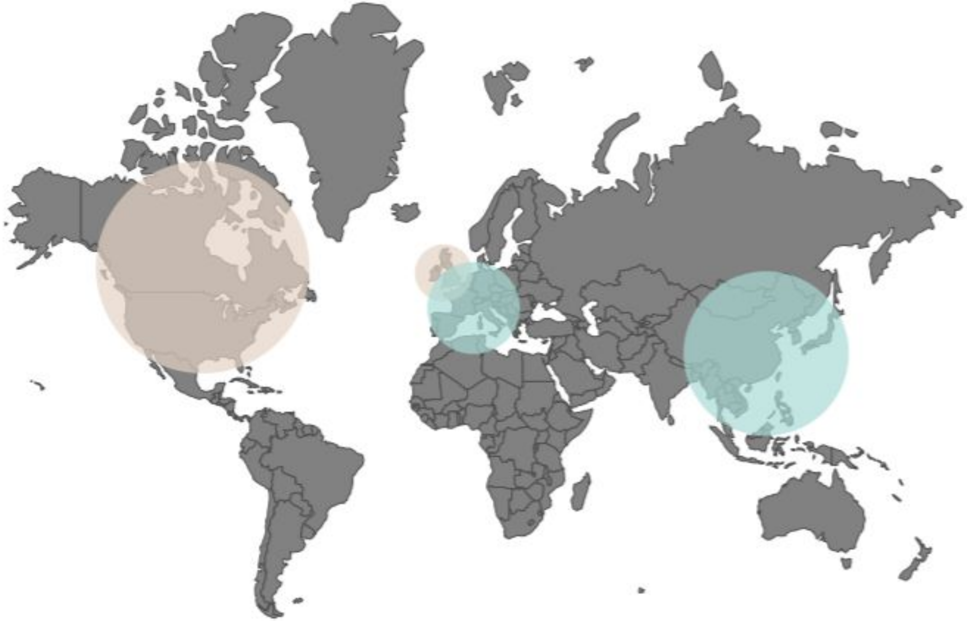
today.mims.com/malaysia-s-moh-tightening-governance-over-online-cosmetics-businesses

Operations

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

Tribe's current models do not have top performance in language understanding in non-English languages

Our team's goal is to build a **model that performs posts' classification 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 found that the results are fairly consistent across datasets.

English has more data points and dataset significantly imbalanced in all languages (measure AUC instead of accuracy)

AUC(ROC) N-grams LR model

	Tribe	**Our Team**
English :	**0.95**	**0.96**
Non English:	**0.76**	**0.82**

Observations and First Model Proposal

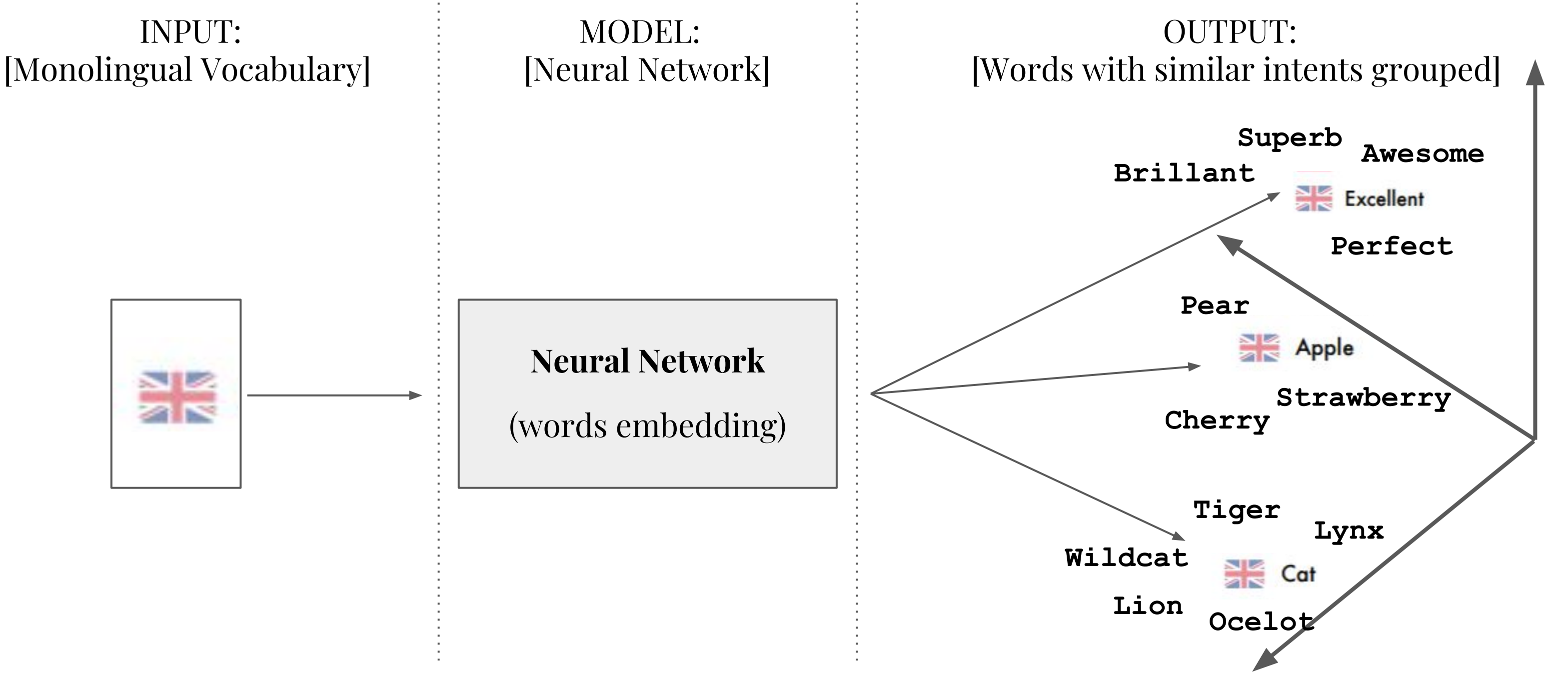
Issue with N-grams:

1. N-grams are **not scalable** (200+ languages)
2. **Context** and word **similarity** not encoded

Models proposed:

→ **Neural Nets - Word Embeddings** trained on each vocabulary (unsupervised ML)

Mono-Lingual Word Embeddings

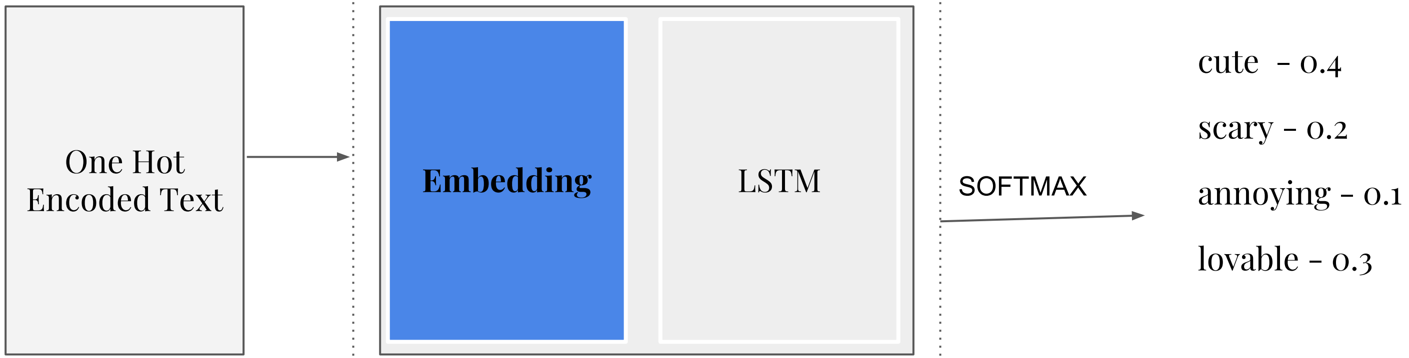


Word Embeddings Training : LM

INPUT: [Dogs, are] [Cats, are]

MODEL: [Neural Network]

SAMPLE OUTPUTS:



- Unsupervised.
- Context encoded in sentence.
- Similar words : Similar successors.

Proposal I : Static and Dynamic Channel Embeddings

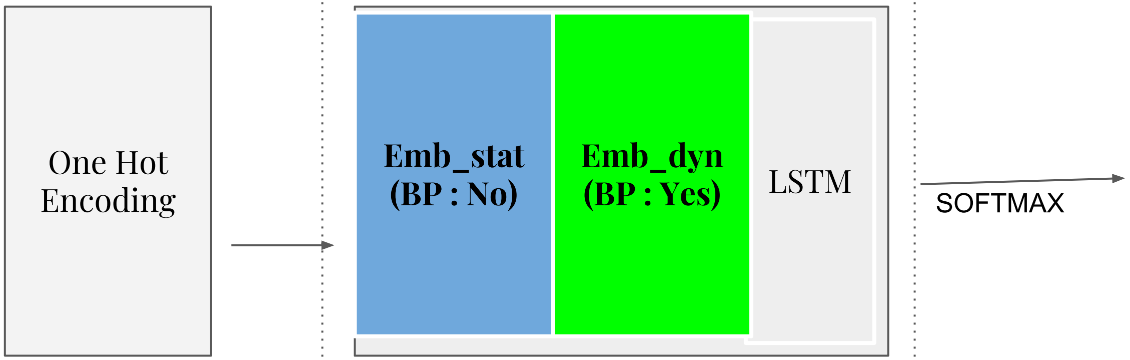
INPUT:

[**Eyeliners**, are]

MODEL

[Neural Network]

SAMPLE OUTPUTS



cute - 0.4

scary - 0.2

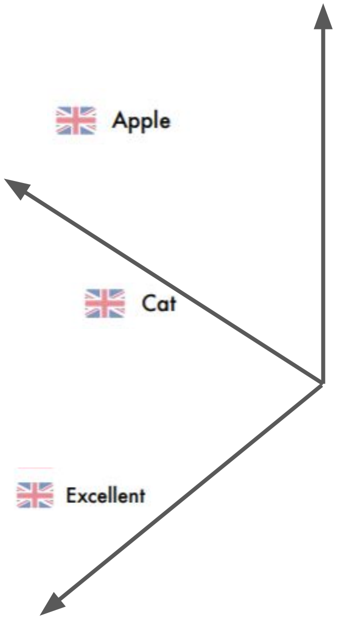
annoying - 0.1

lovable - 0.3

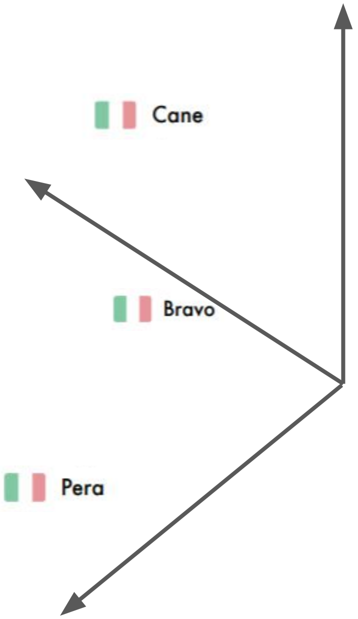
MOTIVATION : Standard knowledge(Wiki) - Static, Fashion Knowledge - Dynamic

Cross-Lingual Aligned Word Embeddings

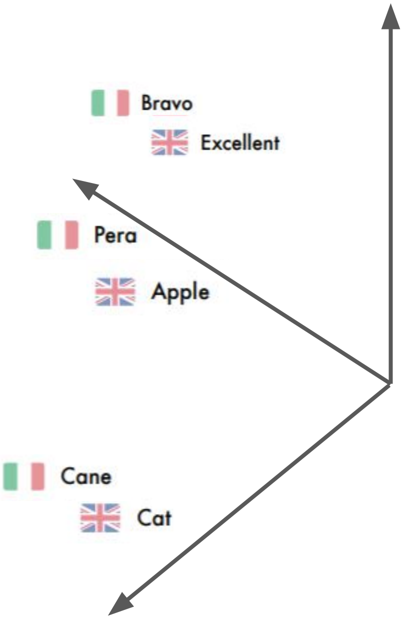
EMBEDDING_1:
[English Vocabulary]



EMBEDDING_2:
[Italian Vocabulary]



ALIGNMENT OUTPUT:
[Words with similar intents grouped]



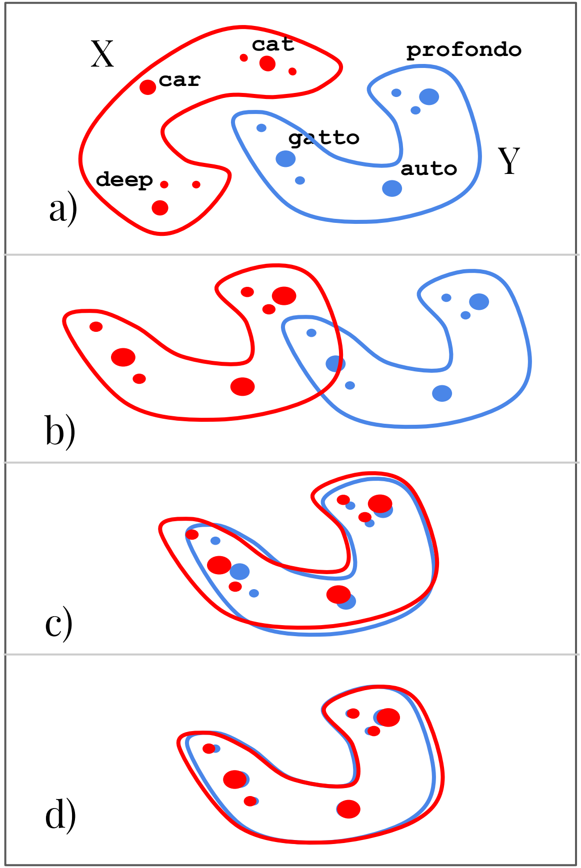
Alignment

Find the transformations that align dictionary X and Y

a) rotation
b) translation
c) stretch

Using few anchor words, find matrix operations that maximize cosine similarity between two embeddings.

Alignment is beneficial because only one model must be embedded in vector space.



Practical Discussion on Aligned Embeddings

1. **One-time cost** - New language train and align embeddings once. Retrain infrequent.
2. **Knowledge Transfer -** Happens at the embedding level.
3. **Singular Model -** One single model across languages .
4. **Generalizable -** Works with any classification model on embedded space.

   Eg; Logistic Regression, Feedforward nets, ConvNets etc.
5. **Data Boost -** Number of effective data points for model training increases.

Proposal 2 : Mixture Model

Why do we need two models ?

Deep Learning Training is harder. Simpler model with an alternate philosophy.

What if ?

One model in English. Translate all other languages to English.

Advantages : Maintenance, Core Strengths, Lesser n-grams, Lesser data costs in non-English

Issues

- Sentence Level : Inaccurate, Costly
- Word Level : Disambiguation

Word Sense Disambiguation

Key Idea : Context drives translation of words across languages.

EXAMPLE

Dove Sei (Not brand) - Where are you ?

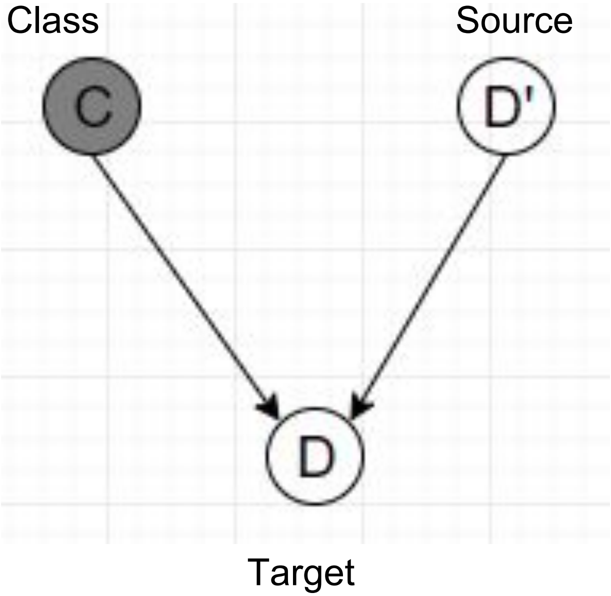
Amore Dove(brand) - Love Dove.

MAIN IDEA

- Consider each word has a mixture of translations.
- Infer the right translation through context during training.

Conditional Translations - PGM

- Generative model
- Document - Bag of words.
- P(D|D') :   "**amore Dove"** - "Love where", "Love Dove"(50%)
- P(D|D',C=brand) : "**amore Dove"** - "Love Dove"(>>50%)
- Needs bilingual lexicon and EM.



Source Classification and Class Inference

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

Conditional Translation

Classifier in source(Softmax)

IDEA

- Target class maximizes likelihood of generation.
- Could be viewed as weighted ensemble prediction by different translations to source language.

Preliminary Results

(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

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. Ground Truth - False