

# Study on Recent Developments on Deep Learning (Word Embeddings)

Aiden (Yansen) HAN

*aiden.han@connect.ust.hk*

- 1 PART I: The Reimplements of Classical Word Embeddings Models
  - The Architecture of PART I
  - Data Source
  - Configuration
  - Goolge Official Implementation of GloVe Model
  - Comparison of Goolge's GloVe Embedding with Our GloVe Model
  - Comparison of Goolge's GloVe Embedding with Our GloVe Model
  - Results Visualizatон
- 2 PART II: Warm ChatBot
  - Data Source for Warm ChatBot
  - Experiments Results
- 3 PART III: What Kind of Space should Our Languages In?
  - Papers Review

# The Architecture of PART I

Our project architecture can be divided into three parts, models reimpliments, data reduction and data visualization.

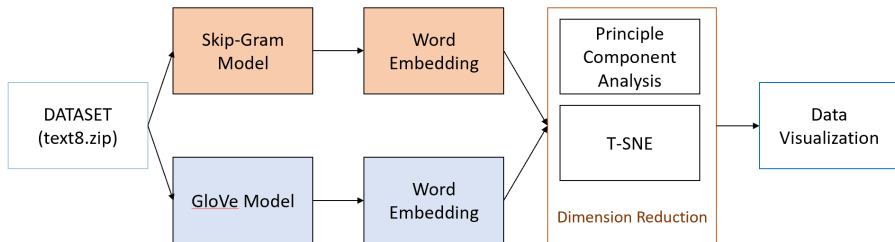


Figure: Project Architecture

# Data Source

We use a typical data source, text8, in which all of the punctuation has been eliminated.

anarchism originated as a term of abuse first used against early working class radicals including the diggers social relations based upon voluntary association of autonomous individuals mutual aid and self governance while state its intervention and regimentation and proclaimed the sovereignty of the moral law of the individual then tedly stated that he is an anarchist and so are all his ancestors in one seven nine three in the thick of the property where owners have complete rights to use and abuse their property as they wish such as exploiting worker his lifetime and there are differing interpretations of some of his ideas for more detailed discussion see his diverse american individualist anarchism benjamin tucker in one eight two five josiah warren had participated in to be the finest individualist anarchist periodical ever issued in the english language tucker's conception of anarchism later in one eight six four the international workingmen's association sometimes called the first international horitarian and predicted that if a marxist party gained to power its leaders would end up as bad as the ruling the worker has a right to but to the satisfaction of his or her needs whatever may be their nature he announced kin and publishing his work described communist anarchism as pseudo anarchism propaganda of the deed johann mos a for example wrote of violence as a necessary and sometimes desirable force in revolutionary settings but at the french anarchism reemerged influencing the bourses de travaux of autonomous workers groups and trade unions ying a spirit of resistance that has inspired many anglophone syndicalists cannot propaganda from april two zero z workers association an anarcho syndicalist successor to the first international contemporary anarcho syndicalists calists in general uphold principles of workers solidarity direct action and self management the russian revolution their experiences in russia aiming to expose the reality of bolshevik control for them bakunin's predictions about platform continues to inspire some contemporary anarchist groups who believe in an anarchist movement organized ething different fascism is not just another form of government which like all others uses violence it is the most collectivized the land but even before the eventual fascist victory in one nine three nine the anarchists were a combination of religious social conscience historical religiosity amongst oppressed social classes and the communist exhibits many anarchistic tendencies such as communal goods and wealth by aiming to obey utterly certain of it o writes extensively about both spirituality and activism anarchism and feminism emma goldman early french feminist of civilization and therefore consider primitivism to also be an anarchist school of thought that addresses fears exist and that it is highly desirable it should in the spanish civil war an anarcho feminist group free womanesis of classical liberalism and austrian economics was germinal for the development of contemporary anarcho communist anglophone and european countries have been taking action for the natural environment eco anarchists or green anarchism primitivism developed in the context of the reclaim the streets earth first and the earth liberation front movement to contrary leftist movements and single issue causes anti war anti nuclear etc it calls for a synthesis o use of the term however it has taken on a life of its own and a wide range of ideas including autonomism post modernistreich and alfredo m bonanno author of works including armed joy and the anarchist tension this tendency is el and crucially the rejection of any idea that the end justifies the means let alone that the business of a re

Figure: Data Source

# Configuration

- $\text{dimWordEmbedding} = 100 \text{ and } 300$
- $\text{numVocab} = 10000$
- $\text{lenWindow} = 3$
- $\text{learningRate} = 0.1$
- $\text{batchSize} = 128$
- $\text{numIteration} = 50000$
- $\text{numSamples} = 64$

# Google Official Implementation of GloVe Model

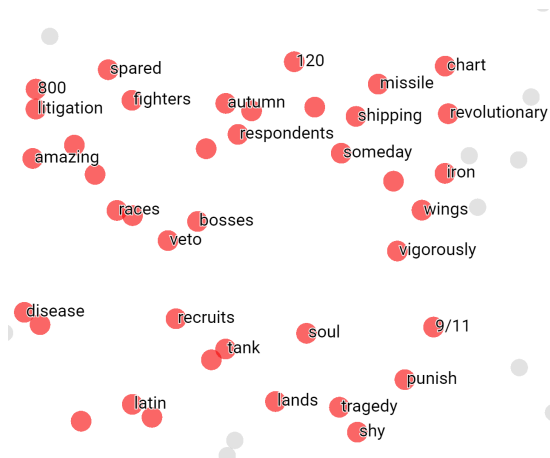


Figure: A Part of Visualization Result of Google's GloVe Implementation

# Comparison of Google's GloVe Embedding with Our GloVe Model

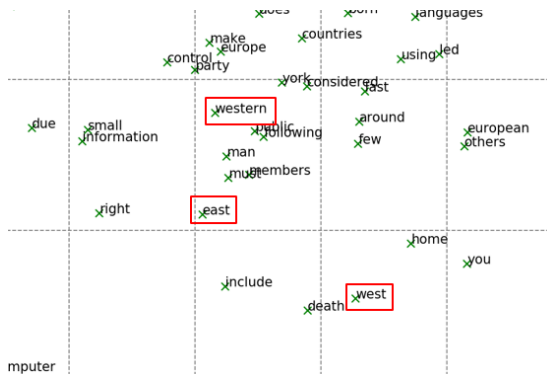


Figure: The Words centered on "east" in Our GloVe Model

# Comparison of Google's GloVe Embedding with Our GloVe Model

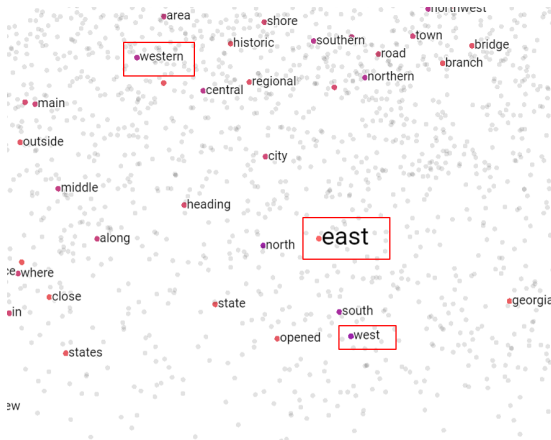


Figure: The Words centered on "east" in Google's GloVe Model



## Results Visualization after PCA

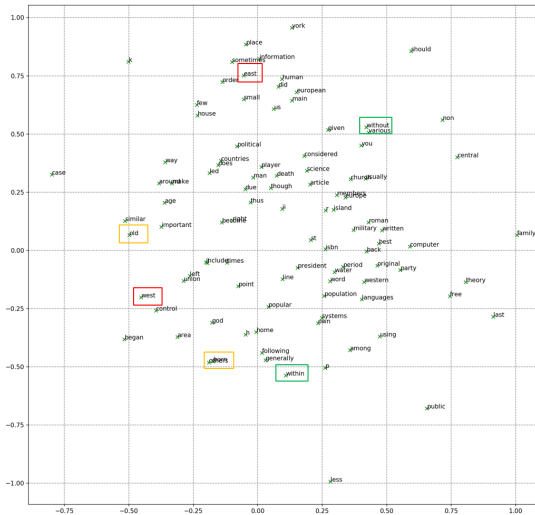


Figure: Skip gram Model, PCA

# Results Visualization after PCA

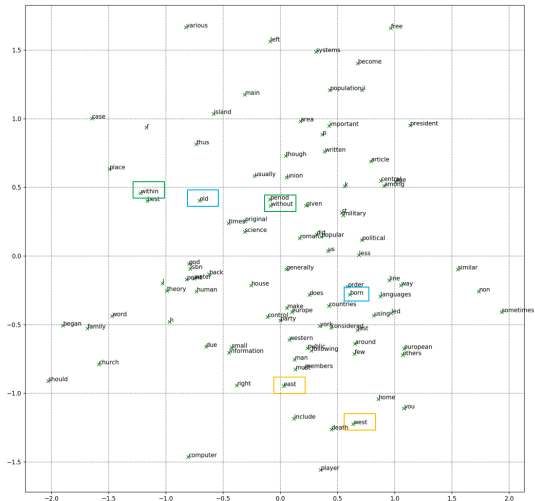


Figure: GloVe Model, PCA

# Results Visualization after TSNE

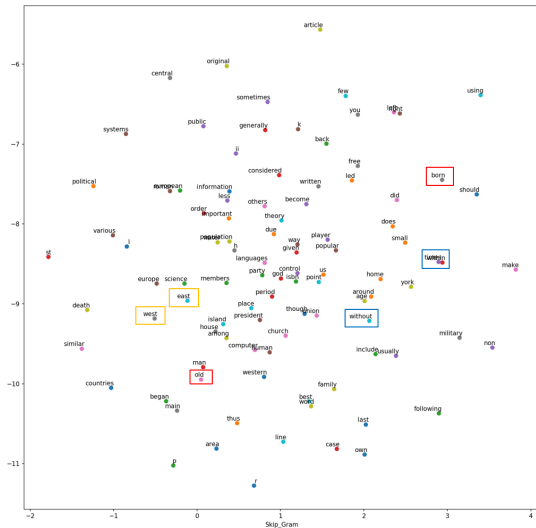


Figure: Skip gram Model, TSNE

## Results Visualization after TSNE

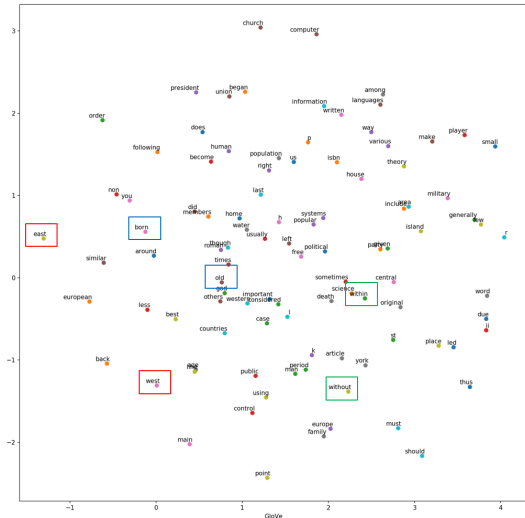


Figure: GloVe Model, TSNE

# Data Source for Warm ChatBot

The training dataset is called STC3, which is about 600 MB.

The dataset is constructed from Weibo posts and replies/comments. More than 1 million Weibo post-response pairs will be provided to participants for training their models.

The test dataset consists of about 5000 posts while 100-200 of the posts will be manually assessed, and for each post, at most 3 emotion classes will be manually specified to indicate the emotion class of a generated response.

The dataset will include labels of each post and response. These labels are obtained by a simple classifier that is based on a bidirectional LSTM model.

The correspondence between the label and the emotion class can be seen as follows:

0: Other, **1: Like**, 2: Sadness, 3: Disgust, 4: Anger, **5: Happiness**

## Experiments Results

```

> 你好呀
Bot: [嘻嘻]
> 这么开心的吗
Bot: [嘻嘻]
> 你可以说点别的吗
Bot: [嘻嘻]
> .....
Bot: [嘻嘻]
> aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa! 我很难受
Bot: [嘻嘻] 啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊
> 哈哈, 你终于变啦
Bot: [嘻嘻]
> .....
Bot: [嘻嘻]
> ni hao ya
Bot: [偷笑] 了了
> ni ke yi shuo dian bie de ma
Bot: [偷笑] 了了了
> aaaaaaaaaaaaaaaaaaaa nan shou xiang ku
Bot: 我也有啊
> ni you shen me
Bot: [吃惊] 了? 了了了
> shen me gui a
Bot: 我也是啊
> 你不是
Bot: [嘻嘻]
> 拜拜
Bot: [嘻嘻]
> 88
Bot: [嘻嘻]
> byebye
Bot: 你去哪里?????????
> Redesign you. smile
Bot: [吃惊]

```

# What Kind of Space should Our Languages In?

Nowadays, almost all of the wide-used models are defined in Euclid space, in which the addition operation is **commutative and associative**. In effect, our languages, however, don't obey those great properties. In the following, I will illustrate those properties one by one.

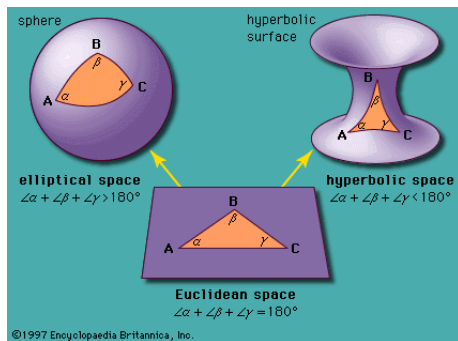
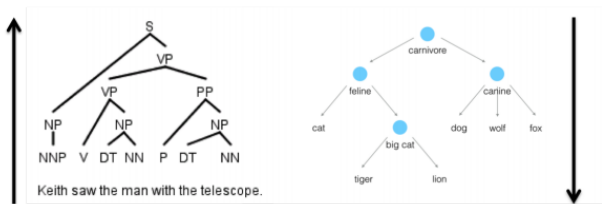


Figure: Topological Space

# What Kind of Space should Our Languages In?

**Topological space:** some researchers consider the hierarchy structure of the word relations, for example, 'music' should be more abstract than 'musician'. Therefore, they proposed **poincare embeddings**(2017) which is defined on hyperbolic space. This definition is actually a partial order relationship, like the '<', '>' in the natural number set. Thus, to keep this hierarchy property, recent researches take advantage of space structure to encode this kind of hierarchy relationship. Therefore, **we can consider our languages lying in hyperbolic topology space, otherwise, we need to find some methods to encode this kind of structure.**





# What Kind of Space should Our Languages In?

- **Operator–Communication:** In a latest paper (CMOW, 2019), researchers improve CBOW by converting word into matrix, because they find that word order is greatly important to exactly express the meaning of one sentence. For example, if we define a communicative operator  $op$  to measure the sentences' polarity. It is obviously that  $op(\text{"The movie was not awful, it was rather great."}) \neq op(\text{"The movie was not great, it was rather awful."})$ . Therefore, **the operator should not be commutative.**
- **Operator–Association:** Besides, some researchers proposed that the operator should also be non-associative, i.e.  $op(\text{"it is not only interesting, but also very comforting."})$ . It is obviously that  $op(\text{"not"}) \cdot op(\text{"only interesting"}) \neq op(\text{"not only"}) \cdot op(\text{"interesting"})$ . Therefore, **the operator should not be associative.**

# What Kind of Space should Our Languages In?

- **Possible Solutions:** I have searched for some materials about non-communicative and non-associative geometry and algebra. I find that Lie algebra meets this requirement.
- **Definition:** A Lie algebra is a vector space  $g$  over some field  $\mathcal{F}$  together with a binary operator  $[\cdot, \cdot] : g \times g \rightarrow g$  called the Lie bracket that satisfies the following axioms:
  - Bilinearity:

$$[ax + by, z] = a[x, z] + b[y, z]$$

$$[z, ax + by] = a[z, x] + b[z, y]$$

- Alternativity:  $[x, x] = 0$
- The Jacobi identity:  $[x, [y, z]] + [z, [x, y]] + [y, [z, x]] = 0$
- Anticommutativity (implied):  $[x, y] = -[y, x]$

**Thank You! QA**