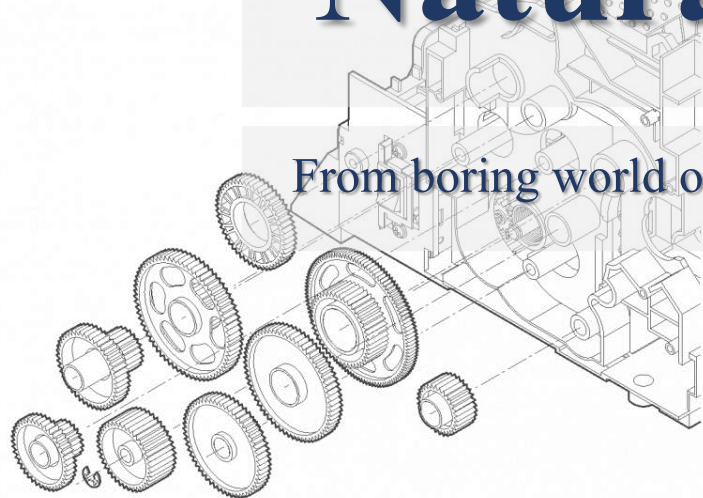


Natural Language Processing

From boring world of Tokens & Embeddings to fascinating world of knowledge processing.



$\Delta \Xi \delta \chi$
[A] [I] [E] [d] [X]

A I E d X

Case Study



28 Million views in 10 years!

Is the title correct?

Why is the martial arts called Kyudo?

Kyudo – The sound of an arrow being released from the bow

Problem 1 – Video Title

Point of Origin



3 days ago

This isn't the sound of an arrow being released. This is a whistle arrow. Basically there is a whistle on the tip of the arrow that makes that noise as air travels through the holes carved into it.



2.8K



REPLY

Summarization



Amardeep Singh Sidhu
@adssidhu86

Such rare and beautiful discussion is the real reason why working in NLP so much fun . Video shows a Whistling arrow making screaming sound. Martial arts is called Kyudo and arrow head makes a kyu sound. Question: is the title of video correct??

youtu.be/maMnRvJVih8



Problem 2 – Martial arts Kyudo

Point of Origin

New Concept



4 days ago

It's cool how the martial art is called "kyūdō" and the arrow literally goes "KYUUUUUUU"



434



REPLY



Hide 8 replies



3 days ago

True! But that's just a coincidence so keep that in mind! Kyudo (弓道) means "Way of the Bow", so just "Archery". The Kyuu part is just how the kanji for Bow is read in its on'yomi form. The other reading (Kun'yomi) is Yumi, which is why you maybe have heard of a japanese bow being called Yumi instead of Kyuu.



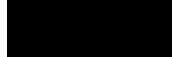
27



REPLY



Amardeep Singh Sidhu 3 days ago



Thanks for such an insightful comment. Made me so inquisitive to learn more.

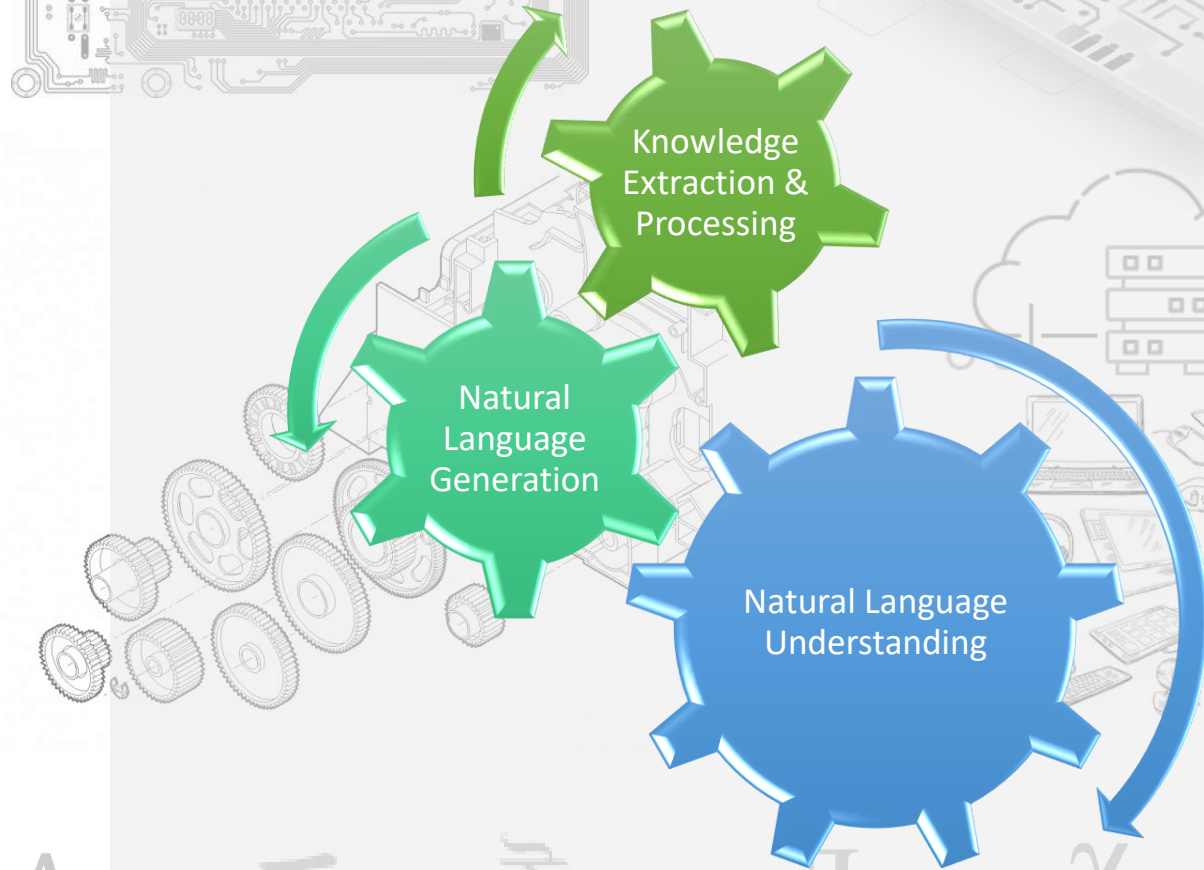


4



REPLY

What happened?



Sound of arrow is a physical concept and is most engaging part of the content. AI system should understand concepts instead of simply parsing and processing statements.

Visual Features such as traditional Japanese attire and large bow are also engaging part of the content.

Understanding word Kyudo leads to new knowledge such as there are 3 scripts in Japanese Kanji, Hiragana & Katakana. Japanese scripts can be written in two styles vertical and horizontal.

What can we do?

Parsing

Semantic Extraction

**Ontology based
method**

**Use unsupervised deep
learning models**

**Maybe a combination
of all of the above.**

This is too much work

- Break sentence into smaller structure, create a tree structure, resolve ambiguity in sentences.

- Named Entity Resolution, Sentiment Analysis, Summarization, Topic Modeling, Parts of speech tagging.

- Use ontology , graphs & Knowledge bases for information extraction.

- Use models like BERT pre trained on large datasets.(Tokens and embeddings)

- I don't know how!!!!

- Wait for a breakthrough in AI !!!!!



One Hot Vector vs Embeddings-2

- **Document 1:** Shelf life of fruits in refrigeration is described in this section. Banana 2 to 7 days, Watermelon 2 to 3 weeks, Jackfruit 1 to 2 days and Peach 5 to 6 days.
- **Document 2:** Maximum load capacity of vehicles is described in this section. Truck 18.5 tonne , Car 300 kg, Jeep 450 kg and Bus 12.5 tonnes

A

I

₹

Д

χ

₹

©

∠

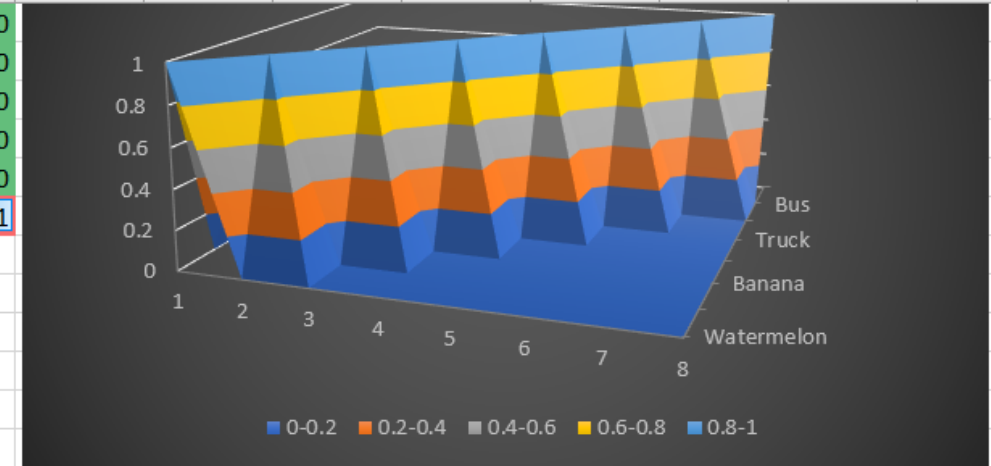
∠

X

One Hot Vector vs Embeddings-2

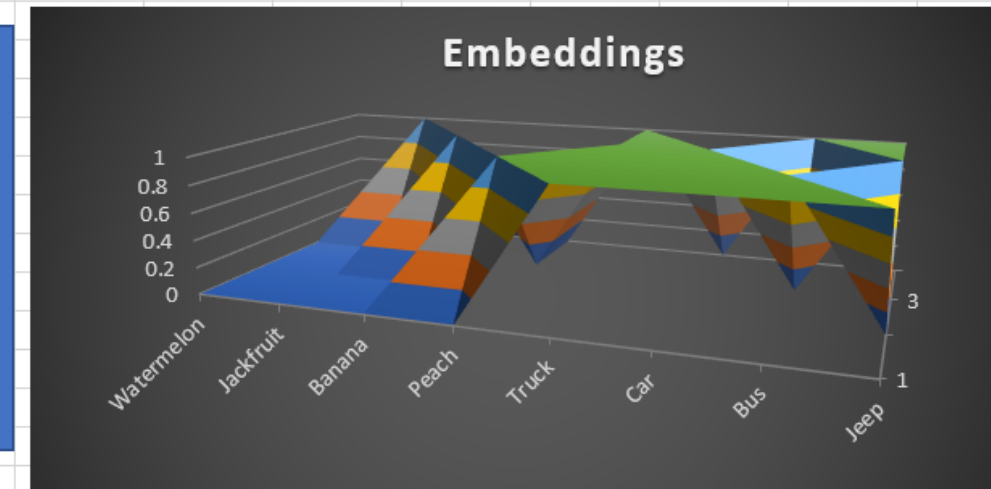
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Banana		0	0	1	0	0	0	0	0								
Peach		0	0	0	1	0	0	0	0								
Truck		0	0	0	0	1	0	0	0								
Car		0	0	0	0	0	1	0	0								
Bus		0	0	0	0	0	0	1	0								
Jeep		0	0	0	0	0	0	0	1								

Like roll number in class each tokens-ngrams is unique



	1	2	3	4
Watermelon	0	0	0	0
Jackfruit	0	0	0	1
Banana	0	0	1	0
Peach	0	1	0	0
Truck	1	1	1	1
Car	1	1	1	0
Bus	1	1	0	1
Jeep	1	0	1	1

Learning & representing information in vector form. Example: 0 in first column is for perishable items



Actual Simple Embedding under the hood (Demo)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
Group1	Group2	Group3	Group4		1	2	3	4	5	6	7	8	9	10	11	
Organic	Other	Other	Fruit	grapefruit	7.05E-02	-2.63E-02	-6.25E-02	5.55E-02	7.21E-03	4.35E-02	2.75E-02	-8.17E-02	6.83E-03	6.58E-03	-3.26E-02	7.1
Organic	Other	Other	Fruit	avocado	-1.04E-01	-1.94E-02	-9.62E-02	5.71E-02	-6.37E-02	2.85E-02	-4.08E-02	-9.16E-02	-4.57E-02	-7.91E-02	-1.22E-01	5.0
Organic	Other	Other	Fruit	watermelon	-2.02E-02	5.29E-03	-2.13E-02	6.57E-02	-5.53E-03	8.87E-03	2.33E-02	-4.89E-02	-2.02E-02	1.15E-02	-4.07E-02	2.5
Organic	Other	Other	Fruit	mango	-0.11804	-6.99E-03	-1.12E-01	2.91E-03	-9.00E-02	3.14E-02	6.31E-02	-8.16E-02	3.48E-02	2.70E-03	1.21E-02	4.2
Organic	Other	Other	Fruit	peach	-2.10E-02	5.39E-03	-1.57E-02	5.95E-02	-7.50E-02	-9.27E-02	-3.17E-03	-1.42E-01	5.61E-02	2.09E-02	-3.44E-02	-3.6
Organic	Other	Other	Fruit	Strawberry	-0.00657	-0.01804	-0.01887	0.009766	-0.06048	-0.0068	0.009378	-0.07132	-0.00566	-0.02071	-0.00583	0
Organic	Other	Other	Fruit	pear	0.048419	0.019115	-0.01299	0.1243	-0.09994	-0.03984	-0.01156	-0.11173	0.03379	-0.03111	-0.06619	-0.0
Organic	Other	Other	Fruit	guava	3.21E-03	-4.05E-02	-8.27E-02	4.23E-02	-1.35E-02	-1.95E-02	1.75E-02	-1.06E-01	-3.46E-02	-3.33E-02	-1.59E-02	1.6
Organic	Other	Other	Fruit	jackfruit	-8.17E-02	-4.29E-02	-3.10E-02	5.27E-02	-1.78E-02	-4.99E-02	-9.18E-04	-5.59E-02	-4.64E-03	-4.19E-02	1.60E-03	1.7
Organic	Other	Other	Vegetable	potato	-2.40E-02	4.34E-03	-9.57E-03	8.55E-02	-1.03E-03	-7.65E-02	-2.18E-02	-9.10E-02	2.60E-02	8.31E-03	-1.23E-01	1.6
Organic	Other	Other	Vegetable	peas	4.73E-02	6.38E-02	-3.86E-02	3.95E-02	-1.26E-02	-7.99E-02	-3.17E-02	-7.96E-02	-5.19E-02	1.79E-02	-7.19E-02	-7.9
Organic	Other	Other	Vegetable	cucumber	-0.08499	-0.0469	0.035246	0.092009	-0.00253	-0.00756	-0.03019	-0.09208	-0.05241	-0.00543	-0.0759	0.0
Organic	Other	Other	Vegetable	peanut	-6.26E-02	5.03E-02	1.37E-03	4.38E-02	-7.08E-02	9.05E-03	3.34E-02	-6.56E-02	-1.03E-01	6.82E-03	-6.01E-02	1.9
Organic	Other	Other	Vegetable	cauliflower	-4.85E-03	-9.76E-03	4.62E-02	7.89E-02	-1.07E-02	-5.52E-02	-4.75E-02	-4.14E-02	-4.39E-02	-2.12E-02	5.25E-03	-1.8
Inorganic	Other	Other	Other	glass	-3.01E-02	5.21E-03	4.72E-02	-1.17E-02	5.41E-02	-3.45E-02	5.68E-02	-1.09E-01	8.26E-05	7.77E-04	1.31E-02	-4.1
Inorganic	Other	Other	Other	ceramic	-6.13E-02	2.10E-02	4.15E-02	-6.61E-03	2.37E-02	-5.70E-03	-3.62E-03	-8.85E-02	1.52E-02	4.10E-02	-4.13E-02	-3.9
Other	Person	Other	Other	girl	2.24E-02	6.84E-02	-3.40E-02	8.11E-02	-7.27E-02	-4.84E-02	6.28E-02	-1.18E-01	5.52E-02	4.35E-02	1.01E-01	-3.0
Other	Person	Other	Other	boy	-4.78E-02	6.04E-02	-2.50E-02	6.29E-02	-9.60E-02	-9.06E-02	1.92E-02	-1.37E-01	6.53E-02	8.23E-02	2.25E-02	1.9
Other	Other	Measurement	Other	angle	1.40E-02	-9.01E-02	2.92E-02	2.11E-02	-7.36E-02	-5.64E-03	-7.23E-03	-1.00E-01	1.99E-02	-1.48E-03	3.83E-03	-4.1
Other	Other	Measurement	Other	radius	0.03492	-0.04718	0.010522	0.01585	-0.09144	0.012556	0.013075	-0.07767	0.018667	-0.05606	0.013646	-1

Note: These are Context Free Embeddings

Contextual Embeddings(Demo-Bertviz)

