TF-IDF

Term frequency-inverse Document Frequency

Change Text to vector

Take into consideration the importance of a word in a sentence

Important word give any number expect zero

Not important word give it zero

Example

Sen1: the cat sat on my face

Sen2: the dog sat on my bed

First we calculate TF=

No.of rep in a sen / No.of words in a sen

Unique Words	Sen1	Sen2
The	1/6	1/6
Cat	1/6	0
Sat	1/6	1/6
On	1/6	1/6
My	1/6	1/6
Face	1/6	0
Dog	0	1/6

Bed	0	1/6

second we calculate IDF=

Log(No.of sen)/ No.of sen containing word

Unique word	IDF
The	LOG(2/2)=0
Cat	LOG(2/1)=0.3
Sat	LOG(2/2)=0
On	LOG(2/2)=0
My	LOG(2/2)=0
Face	LOG(2/1)=0.3
Dog	LOG(2/1)=0.3
Bed	LOG(2/1)=0.3

Last step we calculate =
TF * IDF

	The	Cat	Sat	On	My	Face	Dog	Bed
Sen1	0	0.115	0	0	0	0.115	0	0
Sen2	0	0	0	0	0	0	0.115	0.115