

N-GRAM MODELLING

Peyman Shobeiri

Supervisor : Dr.Moattar

Markov Chains – States

- A is a state.
- B is a state.

Markov Chains – States

- A is a state.
- B is a state.
- Probability of $A \rightarrow B$: 50%
- Probability of $A \rightarrow A$: 50%
- Probability of $B \rightarrow A$: 50%
- Probability of $B \rightarrow B$: 50%

Markov Chains – States

- A is a state.
- B is a state.

- Probability of $A \rightarrow B$: 50%
- Probability of $A \rightarrow A$: 50%
- Probability of $B \rightarrow A$: 50%
- Probability of $B \rightarrow B$: 50%

A

Markov Chains – States

- A is a state.
- B is a state.

- Probability of $A \rightarrow B$: 50%
- Probability of $A \rightarrow A$: 50%
- Probability of $B \rightarrow A$: 50%
- Probability of $B \rightarrow B$: 50%

AA

Markov Chains – States

- A is a state.
- B is a state.

- Probability of $A \rightarrow B$: 50%
- Probability of $A \rightarrow A$: 50%
- Probability of $B \rightarrow A$: 50%
- Probability of $B \rightarrow B$: 50%

AAB

Markov Chains – States

- A is a state.
- B is a state.

- Probability of $A \rightarrow B$: 50%
- Probability of $A \rightarrow A$: 50%
- Probability of $B \rightarrow A$: 50%
- Probability of $B \rightarrow B$: 50%

AABA

N-Gram Modelling – What are N-Grams?

“An N-gram is a contiguous **sequence** of n items from a given sample of text or speech” – Wikipedia

Items refer to states in Markov Chains

Items can be “**characters**”, “**words**”, “sentences” etc.

N-Gram Modelling – What are N-Grams?

$N = 2$, then bigrams

$N = 3$, then trigrams

and so on

N-Gram Modelling – Character Grams

Characters are the states of Markov Chains

N-Gram Modelling – Character Grams

“the bird is flying on the blue sky”

N-Gram Modelling – Character Grams

“the bird is flying on the blue sky”

$N = 2$

Bigrams = ‘th’, ‘he’, ‘e ’, ‘ b’, ‘bi’, ‘ir’, ‘rd’, ‘d ’, ‘ i’ etc.

$N = 3$

Trigrams = ‘the’, ‘he ’, ‘e b’, ‘ bi’, ‘bir’, ‘ird’, ‘rd ’, ‘d i’ etc

N-Gram Modelling – Character Grams

“the bird is flying on the blue sky”

Consider only trigrams,

N-Gram Modelling – Character Grams

“the bird is flying on the blue sky”

Consider only **trigrams**,

Window size = $N = 3$

N-Gram Modelling – Character Grams

Trigrams

“the bird is flying on the blue sky”

N-Gram Modelling – Character Grams

Trigrams

the bird is flying on the blue sky

N-Gram Modelling – Character Grams

Trigrams
the

the bird is flying on the blue sky

N-Gram Modelling – Character Grams

Trigrams

the

he

the bird is flying on the blue sky

N-Gram Modelling – Character Grams

Trigrams

the

he

e b

the  bird is flying on the blue sky

N-Gram Modelling – Character Grams

Trigrams

the

he

e b

bi

the  bird is flying on the blue sky

N-Gram Modelling – Character Grams

the  bird is flying on the blue sky

Trigrams

the

he

e b

bi

bir

N-Gram Modelling – Character Grams

the bird is flying on the blue sky

Trigrams

the

he

e b

bi

bir

ird

N-Gram Modelling – Character Grams

the bird is flying on the blue sky

Trigrams	Next
the	→
he	→ b
e b	→ i
bi	→ r
bir	→ d
ird	→

N-Gram Modelling – Word Grams

Words are the states of Markov Chains

N-Gram Modelling – Word Grams

“the bird is flying on the blue sky”

N-Gram Modelling – Word Grams

“the bird is flying on the blue sky”

Consider only trigrams,

Window size = $N = 3$ (Words)

N-Gram Modelling – Word Grams

Trigrams

“the bird is flying on the blue sky”

N-Gram Modelling – Word Grams

Trigrams
the bird is

the bird is flying on the blue sky

N-Gram Modelling – Word Grams

Trigrams
the bird is
bird is flying

the bird is flying on the blue sky

N-Gram Modelling – Word Grams

Trigrams
the bird is
bird is flying
is flying on

the bird is flying on the blue sky

N-Gram Modelling – Word Grams

Trigrams
the bird is
bird is flying
is flying on
flying on the

the bird is flying on the blue sky

N-Gram Modelling – Word Grams

the bird is flying on the blue sky

Trigrams
the bird is
bird is flying
is flying on
flying on the
on the blue

N-Gram Modelling – Word Grams

the bird is flying on the blue sky

Trigrams
the bird is
bird is flying
is flying on
flying on the
on the blue
the blue sky

N-Gram Modelling – Word Grams

the bird is flying on the blue sky

Trigrams		Next
the bird is	→	flying
bird is flying	→	on
is flying on	→	the
flying on the	→	blue
on the blue	→	sky
the blue sky	→	

N-Gram Modelling – Usage



N-Gram Modelling – Usage

Trigrams		Next
the bird is	→	[flying, eating, sleeping]
bird is flying	→	[on, through, on]
is flying on	→	the
flying on the	→	[blue, orange]
on the blue	→	sky
the blue sky	→	[.]

N-Gram Modelling – Usage

“the bird is”

N-Gram Modelling – Usage

Trigrams		Next
the bird is	→	[flying, eating, sleeping]
bird is flying	→	[on, through, on]
is flying on	→	the
flying on the	→	[blue, orange]
on the blue	→	sky
the blue sky	→	[.]

N-Gram Modelling – Usage

“the bird is flying”

N-Gram Modelling – Usage

“the bird is flying”

N-Gram Modelling – Usage

Trigrams		Next
the bird is	→	[flying, eating, sleeping]
bird is flying	→	[on, through, on]
is flying on	→	the
flying on the	→	[blue, orange]
on the blue	→	sky
the blue sky	→	[.]

N-Gram Modelling – Usage

“the bird is flying on”

N-Gram Modelling – Usage

“the bird is flying on”

N-Gram Modelling – Usage

Trigrams		Next
the bird is	→	[flying, eating, sleeping]
bird is flying	→	[on, through, on]
is flying on	→	the
flying on the	→	[blue, orange]
on the blue	→	sky
the blue sky	→	[.]

N-Gram Modelling – Usage

“the bird is flying on the”

N-Gram Modelling – Usage

“the bird is flying on the”

N-Gram Modelling – Usage

Trigrams		Next
the bird is	→	[flying, eating, sleeping]
bird is flying	→	[on, through, on]
is flying on	→	the
flying on the	→	[blue, orange]
on the blue	→	sky
the blue sky	→	[.]

N-Gram Modelling – Usage

“the bird is flying on the blue”

N-Gram Modelling – Usage

“the bird is flying on the blue”

N-Gram Modelling – Usage

Trigrams		Next
the bird is	→	[flying, eating, sleeping]
bird is flying	→	[on, through, on]
is flying on	→	the
flying on the	→	[blue, orange]
on the blue	→	sky
the blue sky	→	[.]

N-Gram Modelling – Usage

“the bird is flying on the blue sky”

N-Gram Modelling – Usage

“the bird is flying on the blue sky”

N-Gram Modelling – Usage

Trigrams		Next
the bird is	→	[flying, eating, sleeping]
bird is flying	→	[on, through, on]
is flying on	→	the
flying on the	→	[blue, orange]
on the blue	→	sky
the blue sky	→	[.]

N-Gram Modelling – Usage

“the bird is flying on the blue sky.”

N-Gram Modelling – Usage

“the bird is flying on the blue sky.”

N-Gram Modelling – Usage

Trigrams		Next
the bird is	→	[flying, eating, sleeping]
bird is flying	→	[on, through, on]
is flying on	→	the
flying on the	→	[blue, orange]
on the blue	→	sky
the blue sky	→	[.]

N-Gram Modelling – Usage

“the bird is flying on the blue sky.”

N-Gram Modelling – Usage

“the bird is sleeping”

“the bird is flying on the orange”