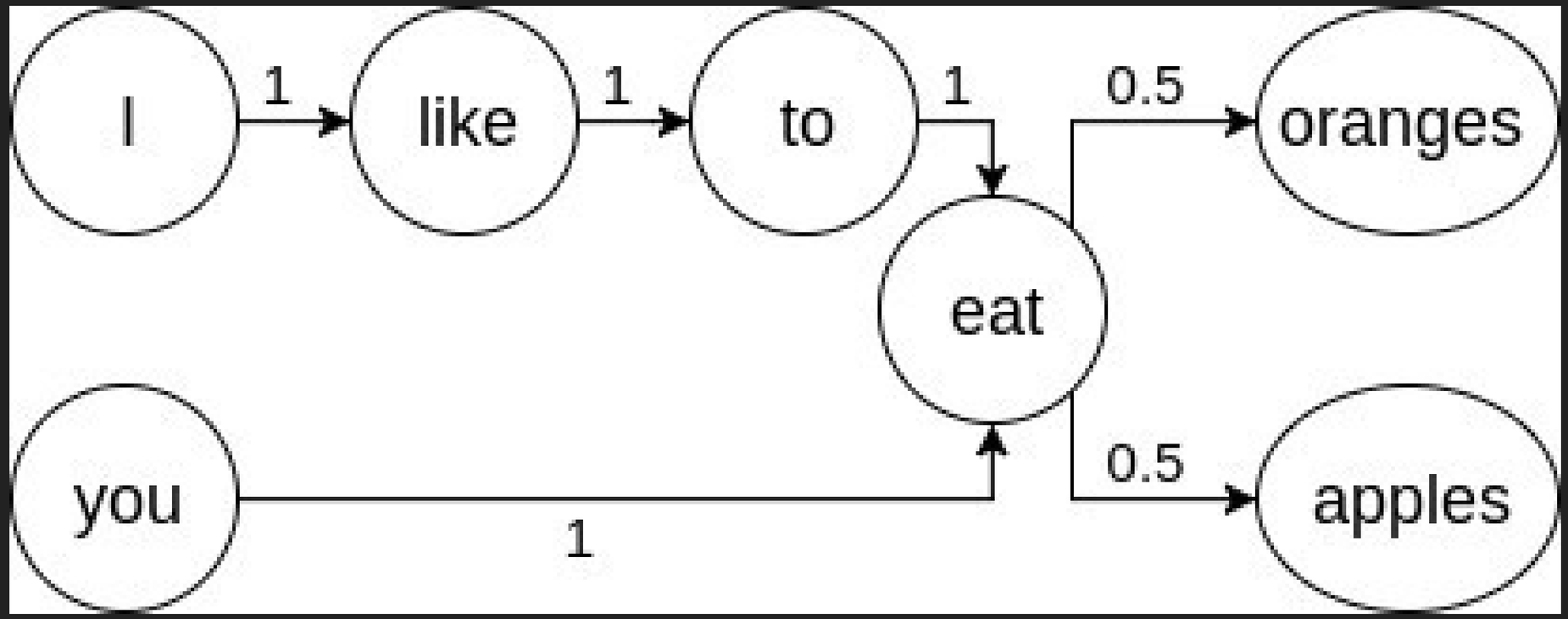
ESKENDER DZHEMADINOV
NATIONAL RESEARCH UNIVERSITY

MARKOV CHAINS APPLIED TO HSE TWITTER INFORMATION

13.11.2018

Part 1

Markov chain explained visually



Explanation

“Cat: Where are you going?
Alice: Which way should I go?
Cat: That depends on where you are going.
Alice: I don’t know.
Cat: Then it doesn’t matter which way you go.”
— Lewis Carroll, Alice in Wonderland

Definition

Марковский процесс — случайный процесс, эволюция которого после любого заданного значения временного параметра t не зависит от эволюции, предшествовавшей t, при условии, что значение процесса в этот момент фиксировано.

Марковская цепь — частный случай марковского процесса, когда пространство его состояний счётно.

ESKENDER DZHEMADINOV

INPUT

One fish two fish red fish blue fish

MARKING

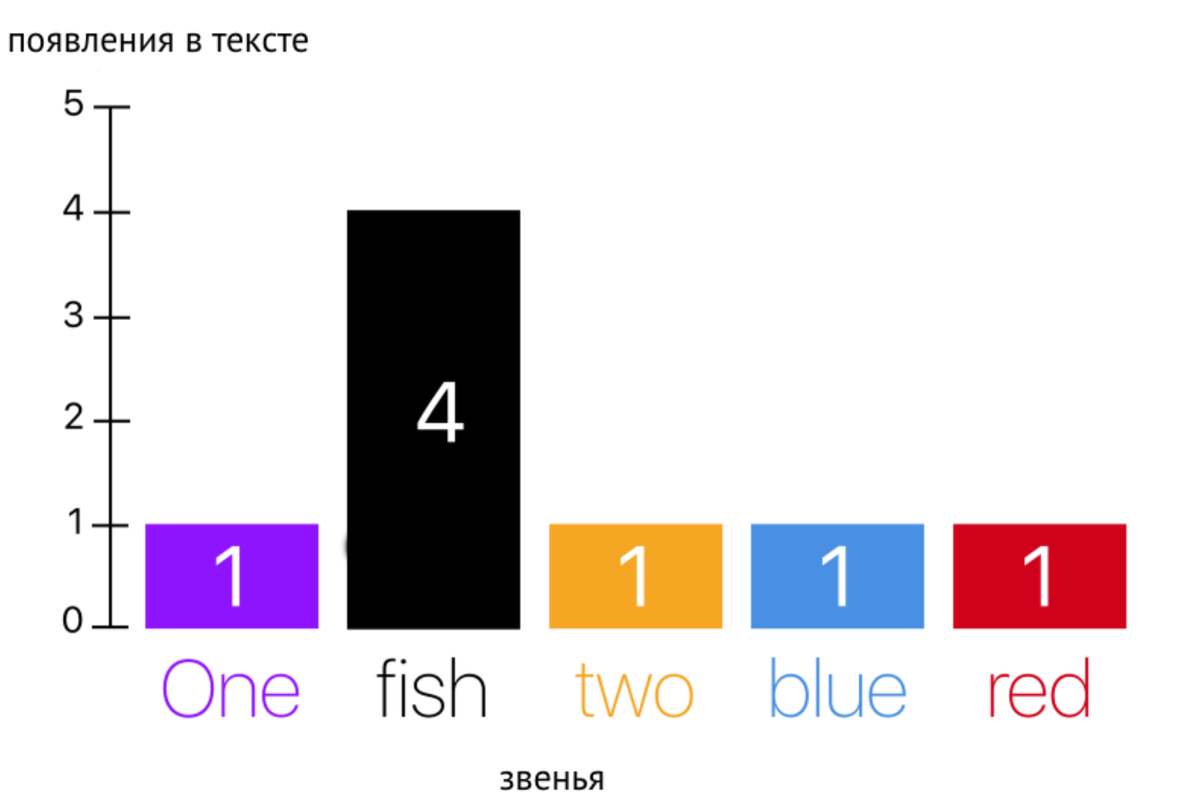
One fish two fish red fish blue fish

COUNTING

One : 1
fish : 4
two : 1
red : 1
blue : 1

Histogram

Here's a plot describing the density of the words in the sentence. The most frequent word is "fish" anothers have less frequency.



(*Start*, One)

(One, fish)

(fish, two) (fish, red) (fish, blue) (fish, *END*)

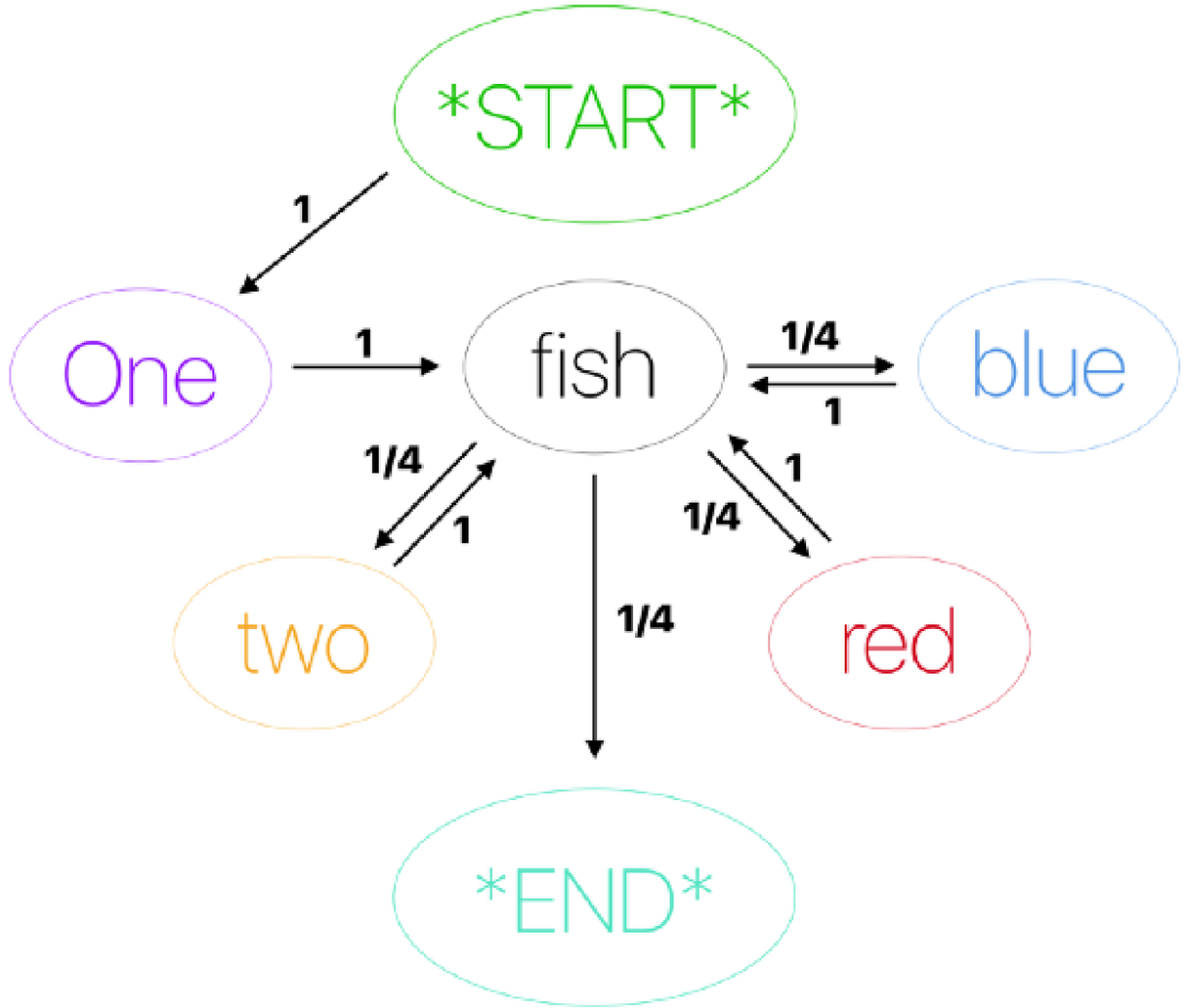
(two, fish)

(red, fish)

(blue, fish)

(*END*, none)

Graph



One fish two fish red fish blue fish

Part II

Example

THE FIRST STEP: CHOICE OF DATA

НИУ ВШЭ, HSE

https://twitter.com/su_hse

- 11,5 thousand tweets
- 544 k text corpora size
- russian language
- some links, hashtags, usersnames
- pictures

THE SECOND STEP: CLEANING DATA

We should delete links, tags and usernames. Alse we should delete some punctuation marks (except dots and question marks and exclamation marks) .

The result of cleaning

прошлые выходные мы узнали что некоторые близнецы
не могут отличить друг друга на фотографиях
научная карьера даже если вы считаете себя
гуманитарием знать точные науки на уровне
восьмиклассника пригодится в жизни каждому. не
только экономика в вышке присудили первую в
россии степень phd in computer science. звание
получил выпускник все неиспользованные вышкоины
сгорят уже 20 ноября но большинство проектов уже
успели собрать нужное вампиры и оборотни это
конечно страшно. но мы знаем чего студенты боятся
на самом деле свежие новости из мира рейтингов
вышка заняла 23 место в рейтинге qs ееса среди
вузов из стран европы с ребята из
благотворительной организации «открой глаза»

544 k
corpora size

including
numbers

including
dots etc.

THE THIRD STEP: CREATING AN ACCOUNT AND BIULDING A MODEL.

https://twitter.com/oa8k9wI4x1xE4wa

- Now we register a twitter account and build a Markov chain model.
- Also we create a script to post our tweets automatically

HERE'S SOME EXAMPLES OF GENERATED TWEETS

Hm.....

MORE EXAMPLES...

Wow.....

AND MORE...

Hm.....

Тень Декана @oa8k9wl4x1xE4wa · 14 мин.
это актуально повсюду даже в ней элементов массовой культуры.
Тень Декана @oa8k9wl4x1xE4wa · 15 мин.
подборка если кто и знает как сделать историю интересной и доступной то это выпускник вышки андрей ленвальский прочитает доклад об инвестиционном банкинге.
Тень Декана @oa8k9wl4x1xE4wa · 15 мин.
а может лучше обучение за счет индивидуальных достижений не получится.
Тень Декана @oa8k9wl4x1xE4wa · 15 мин.
преподаватели проекта data culture осилила проходные баллы снизятся скорее ух ты!

Thanks for your attention

Twitter account link

https://bit.ly/2RTVQJd | https://twitter.com/oa8k9wI4x1xE4wa

Email address

erdzhemadinov@gmail.com