Prediction by Partial Context Matching using Markov Chains

Rafael Natã

Prediction by Partial Matching is an *adaptive statistical data compression technique* based on *context modeling* and *prediction*

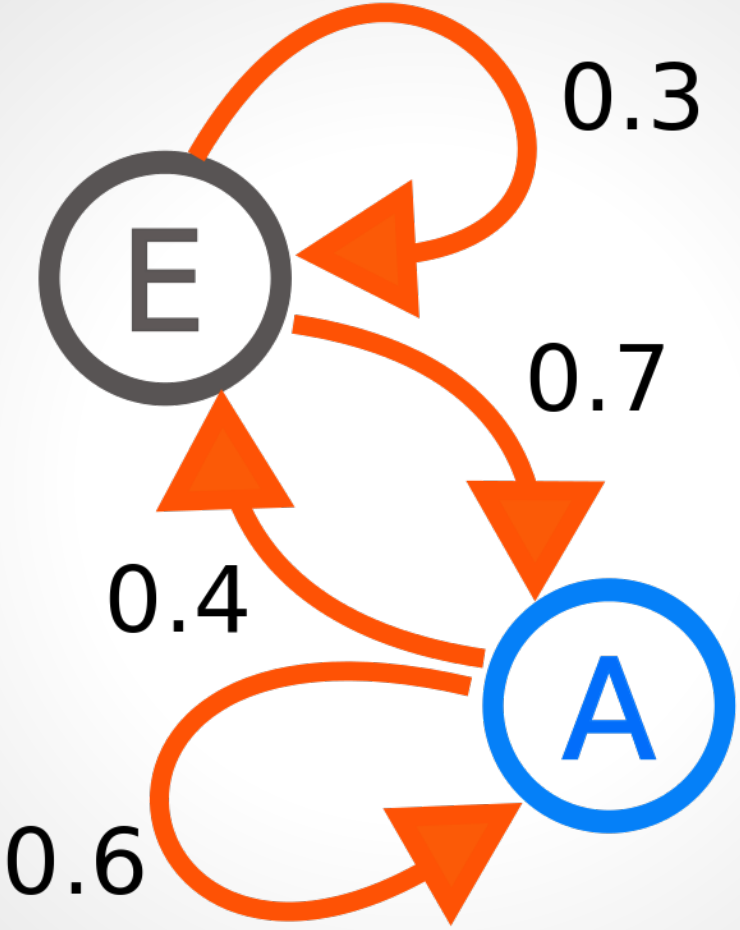
PPM models use a set of previous symbols in the uncompressed symbol stream to predict the next symbol in the stream

PPM compression implementations vary greatly in other details

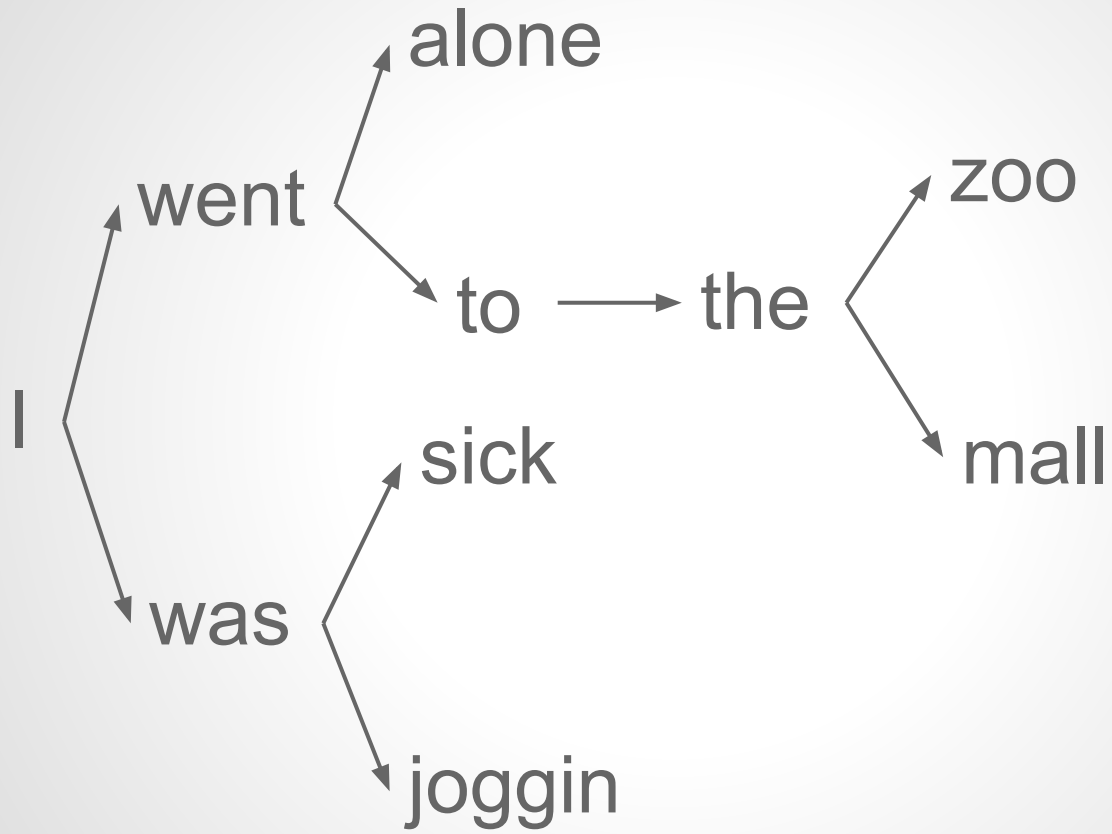
The actual symbol selection is usually recorded using *arithmetic coding*, though it is also possible to use *Huffman encoding* or even some type of *dictionary coding* technique

A Markov Chain is a mathematical system that undergoes transitions from one state to another on a state space

It is a *random process* usually characterized as *memoryless*: the next state depends only on the current state and not on the sequence of events that preceded it


0.3
E
0.7
0.4
A
0.6

I went alone
I went to the zoo
I went to the mall
I was sick
I was joggin



I went alone
I went to the zoo
I went to the mall
I was sick
I was joggin

Words

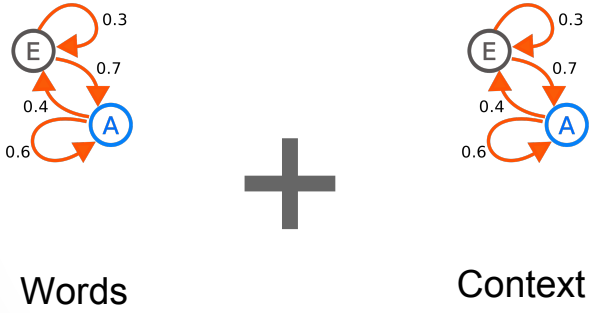
{I}
{w, went} {we, went} {wen, went}
{a, alone} {al, alone} {alo, alone} {alon, alone}
{t, to}
{t, the} {th, the}
{z, zoo} {zo, zoo}
{m, mall} {ma, mall} {mal, mall}
{w, was} {wa, was}
{s, sick} {si, sick} {sic, sick}
{j, joggin} {jo, joggin} {jog, joggin}, {jogg, joggin}
{joggi, joggin}

{I, went} {went, alone}
{I, went} {went, to} {to, the} {the, zoo}
{I, went} {went, to} {to, the} {the, mall}
{I, was} {was, sick}
{I, was} {was, joggin}

Context

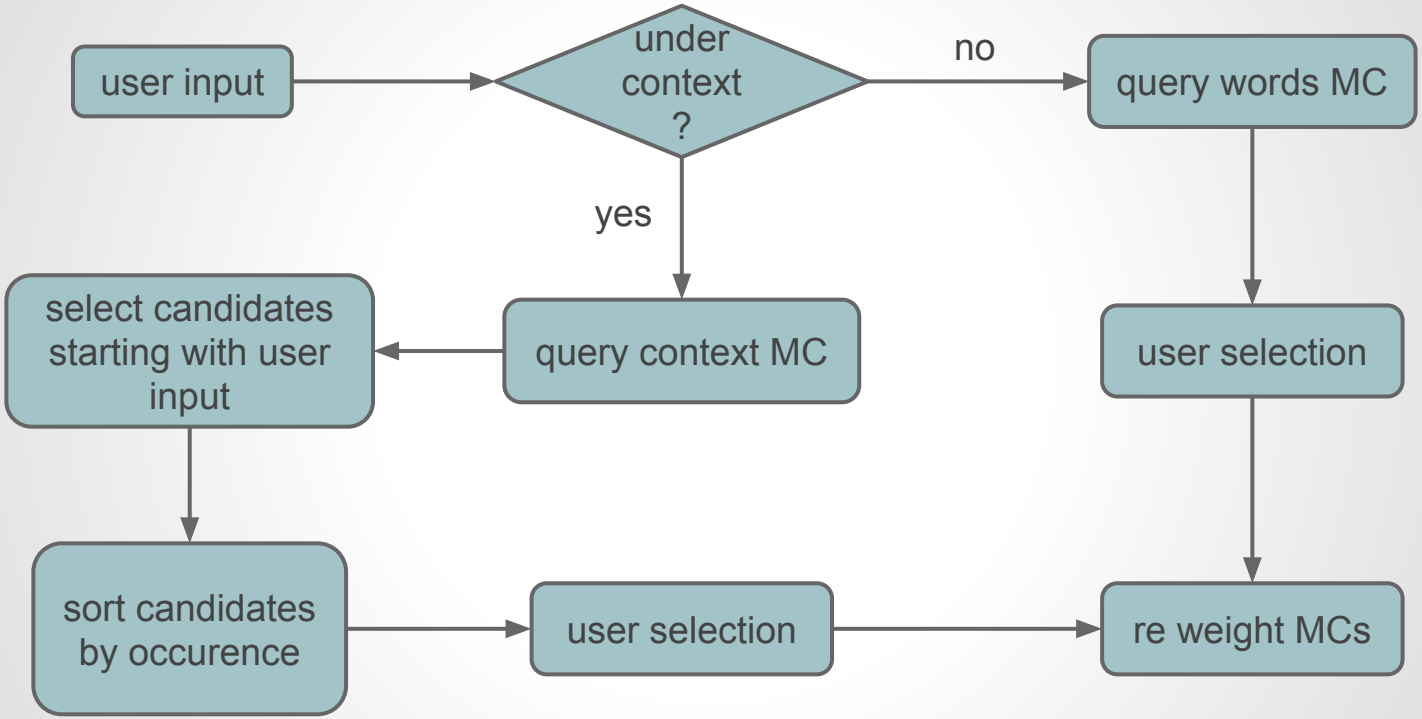
Train data

I went alone
I went to the zoo
I went to the mall
I was sick
I was joggin



Words

Context


user input
under context ?
no
yes
query words MC
query context MC
select candidates starting with user input
user selection
sort candidates by occurence
user selection
re weight MCs