Definition Learning
FutureAI Final Project

Colin McDonnell
how humans learn

definition

see a few usage examples
how machines learn

prediction

see a few hundred million usage examples
how machines learn

CBOC

Skip-gram
how machines learn

Billion Words Corpus

x2000
how machines learn

Billion Words Corpus

x23
teach machines to learn like us
approach

download Webster’s Dictionary

{ defn : ["extending", "far", "down", "from", "the", "top", "or", "surface"], word : “deep” }

pass definition words into network one at a time

final activation should be close to true word embedding
how to handle variable length inputs
how to handle variable length inputs
how to handle variable length inputs
architecture

\begin{align*}
\text{look up table} & \quad \text{(pretrained)} \\
h[t-1] & \\
\text{w}[t] & \\
\text{h}[T] & \\
\text{MSE} & \\
\text{BPTT} & \\
\end{align*}
results

days of continuous failure and misery
train on half of the dataset
see mean squared error of predicted embeddings for words in test set
calculate perplexity
test embeddings on analogical reasoning tasks
future work

curriculum learning: order matters

DAGify the dictionary

find axiomatic subset of language

learn that subset in an unsupervised way or with innate biases

pair with “definition detector” for high-impact updates
Turing Test system

working memory

  hierarchical SFA with increasing time constants

long-term memory

  definition learning + omni-scale skip-gram

recurrent attention mechanism

  over working memory
  over long term memory: associational vector space
  over internal output buffer (DRAW-style)
  over decision to output

text generation from vector representation

  train on SAT grammar questions, use as training signal
  self-play for AI chatbots?