Generera rubriker med deep learning

DATX02-16-27

Rickard Lantz **Jacob Genander**
Alex Evert Nicklas Lallo Filip Nilsson

27 februari 2016

Chalmers Tekniska Högskola

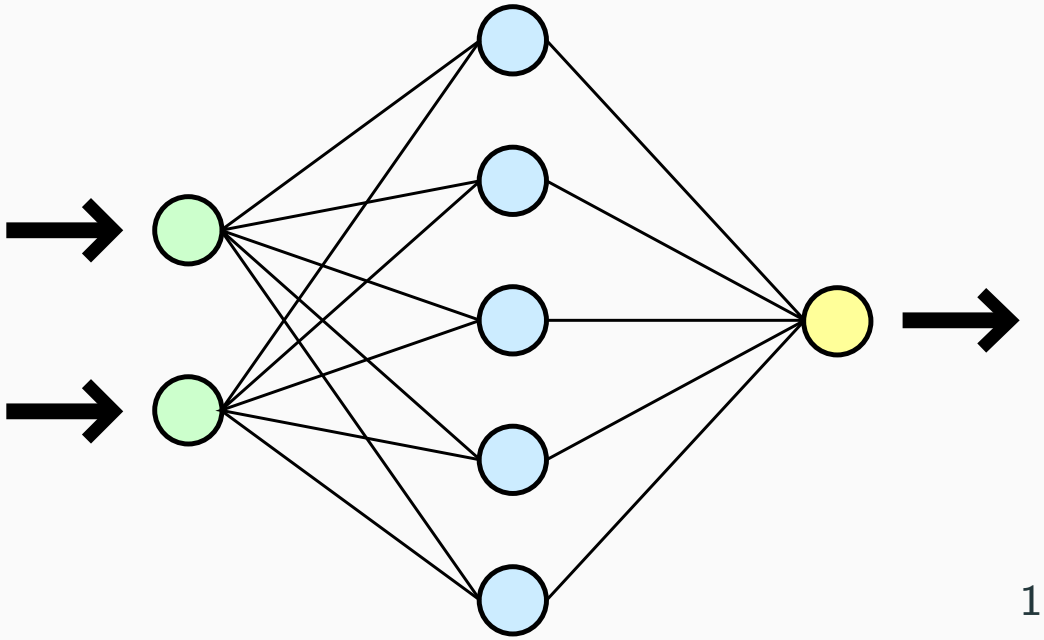
Introduktion

Projektets mål

Generering av:

1. Slumpmässiga men ”vettiga” rubriker
2. Rubriker baserade på artiklar

Artificiella Neuronnät



[1]

- Metod med inspiration från den mänskliga hjärnan
- Kan hitta mönster i mycket komplexa data
- Nätverket identifierar mönster som experter tidigare behövt representera i kod.

[1] https://commons.wikimedia.org/wiki/File:Neural_network.svg

Arbetsmetodik

Arbetsmetodik

Separata modeller för varje delmål:

Slumpmässiga rubriker: Sekvensmodellering

Rubriker till artiklar: Sekvens till sekvens-modellering

Liknande arbetssätt

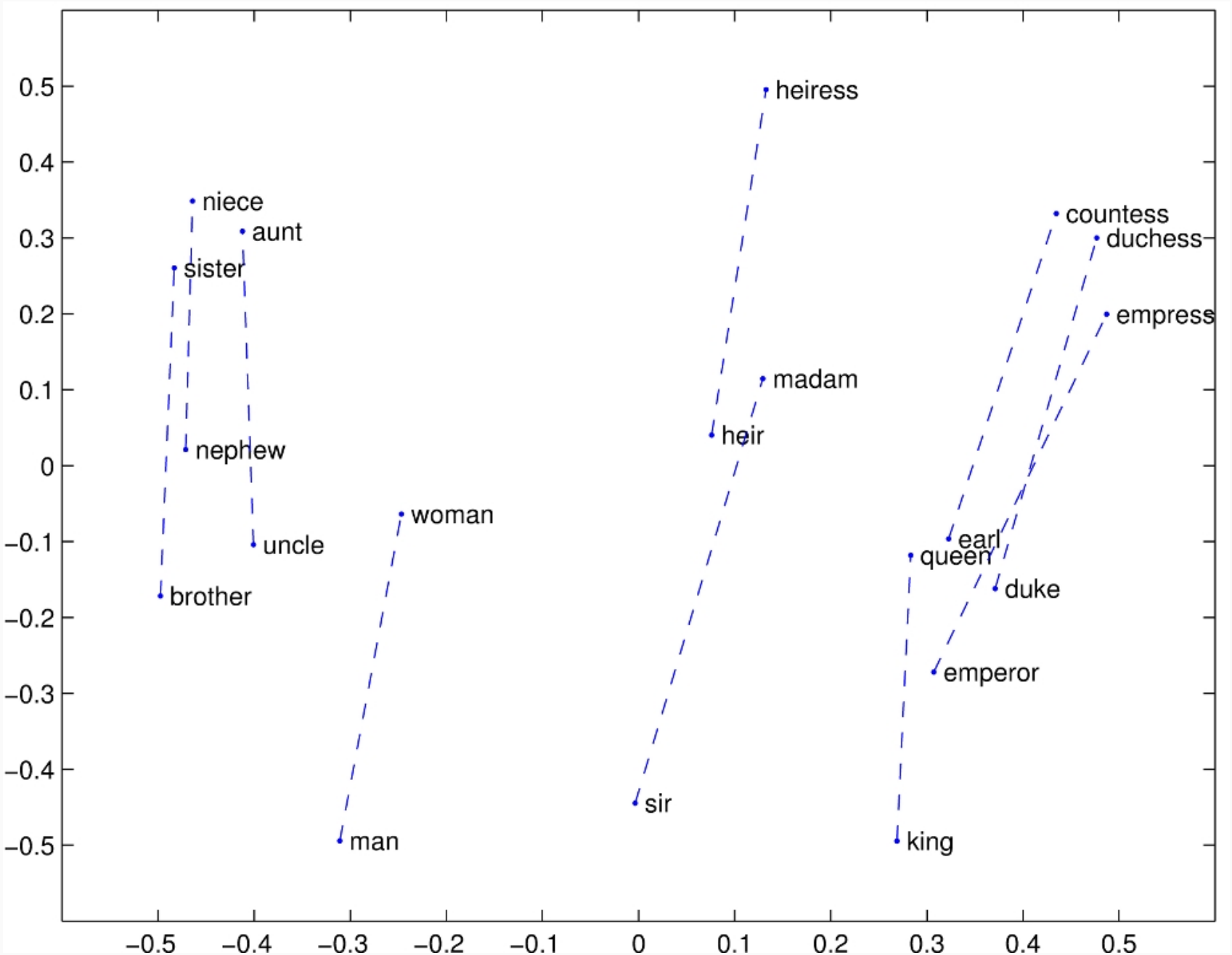
1. Datainsamling och bearbetning
   - Artiklar med titlar/rubriker
   - Ordinbäddningar
2. Modellkonstruktion
3. Modelljustering

Datainsamling

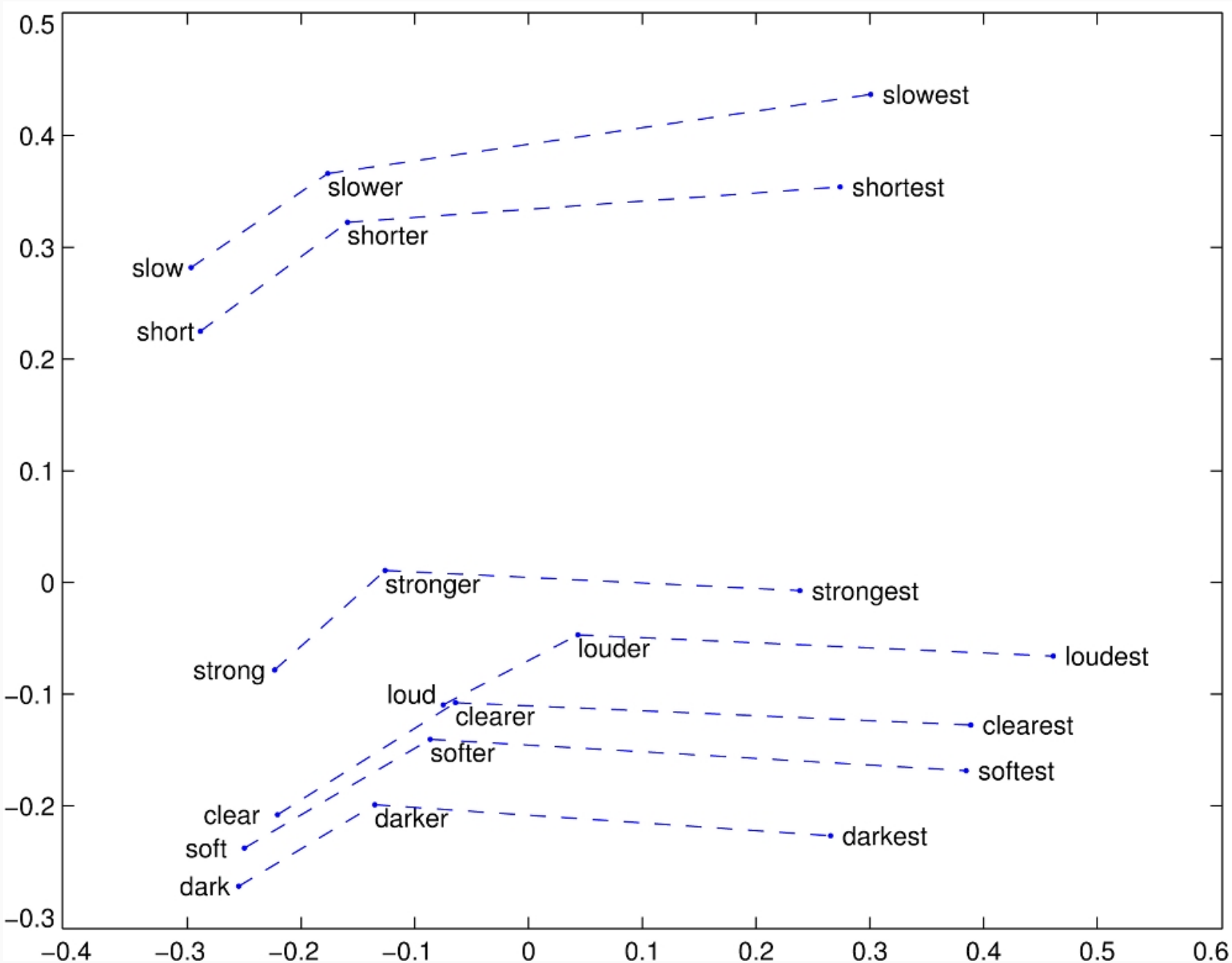
Två sorters data:

- Artiklar och titlar
  - Större dataset ger bättre generaliseringar
  - Homogent dataset
- Ordinbäddningar
  - Neuronnät kan endast hantera numeriska data
  - Ger vektorrepresentation av ord

GloVe: Global Vectors for Word Representation



GloVe: Global Vectors for Word Representation



Modellkonstruktion

- LSTM-celler, ”Long Short-Term Memory”
  - Viktar in ny data
  - Bra på sekvenser
- Ramverket TensorFlow™ från Google
  - Färdiga implementationer av ex. LSTM-celler
  - Parallell exekvering på CPU:er och GPU:er

Modelljustering

Finns inget ”facit”, endast riktlinjer

- Iterativ, testbaserad process
    1. Träning av nätverket
    2. Utvärdering av nätverkets prestanda
    3. Justerande av hyperparametrar, ex.
        - Nätverkets storlek
        - Inlärningstakt

Sammanfattning

Nuvarande status

Projektets fas:

- ✓ Datainsamling och bearbetning
- ✓ Modellkonstruktion
- Modelljustering

Uppnådda mål:

- ✓ Slumpmässiga men ”vettiga” rubriker
- Rubriker baserade på artiklar

Erfarenheter och lärdomar

- Tar tid att sätta sig in i ett komplext ramverk inom ett nytt fält
- Ingen exakt vetenskap
- Verkar som magi på ytan, men går ändå att förstå på ett intuitivt plan