IT-309
EXPERIMENT-10 REPORT

Indian Institute of Information Technology, Vadodara

Group-1

Contents

0.1 Word2Vec

- Word2Vec approach uses deep learning and neural networks-based techniques to convert words into corresponding vectors in such a way that the semantically similar vectors are close to each other in N-dimensional space, where N refers to the dimensions of the vector.
- Word2Vec model comes in two flavors: Skip Gram Model and Continuous Bag of Words Model (CBOW).
- In this experiment, we have used Python Genism library for Word2Vec implementation.
- For training data, we have used 100 MB of text from http://mattmahoney.net/

0.2 Results

Results for some of the queries are shown below:

- 10 Words similar to 'dog'

  [('cat', 0.7946374416351318),
  ('hound', 0.7834134697914124),
  ('cow', 0.7432703971862793),
  ('pig', 0.735741138458252),
  ('goat', 0.7230333089828491),
  ('dogs', 0.7201663255691528),
  ('haired', 0.711754560470581),
  ('pie', 0.7058099508285522),
  ('deer', 0.7057844400405884),
  ('rat', 0.703100323677063)]

- 10 Words similar to 'computer' [('computers', 0.6866356134414673),
  ('computing', 0.6243542432785034),
  ('console', 0.6194486618041992),
  ('hardware', 0.5985126495361328),

('mainframe', 0.584159255027771),
('programmer', 0.5807563066482544),
('handheld', 0.5782939195632935),
('pc', 0.5771670341491699),
('digital', 0.5723547339439392),
('cpu', 0.5669999122619629)]

- Word similar to 'king' - 'man' + 'woman'

  [('queen', 0.6278063058853149)]