Abstract

Without motivation, writing may be a cumbersome process. In this work, a methodology is proposed which will assist user by providing some reference information e.g. related words while composing an article or message. Smart systems with related word prediction have turned out to be extremely prevalent for English language but there is no such big efforts for Hindi language. The main goal of this dissertation work is to provide syntactically and semantically related words based on continuous feature vector representation. Continuous Bag of Words (CBOW) language model is used to get the feature vector representation of each word in training set. Cosine Distance and rule based strategy is used as measurement to find the most related word in context. In a comparative study we reasoned that our method excels in accuracy estimation than existing method. This approach will help Hindi writing in an effective and creative manner.

References
22. L. Qiu, Y. Cao, Z. Nie, Y. Yu, and Y. Rui, “Learning word representation considering
proximity and ambiguity,” AAAI Conference on Artificial Intelligence., June 2014.
25. Polyglot - Rami Al-Rfou - Google Sites
“https://sites.google.com/site/rmyeid/projects/polyglottocdownload-wikipedia-text-dumps.”
28. Hindi WordNet (A Lexical Database for Hindi):
http://www.cfilt.iitb.ac.in/wordnet/webhwn/.
29. Universal approximation theorem:
32. Polyglot - Rami Al-Rfou - Google Sites
“https://sites.google.com/site/rmyeid/projects/polyglottocdownload-wikipedia-text-dumps.”

**Index Terms**

Computer Science  Signal Processing

**Keywords**

Language Modelling, Curse of Dimensionality, Distributed Representation, CBOW, POS tagging.