

赤倉応用情報工学演習第2回課 4620028 貝津俊一

・Q2 演習1で扱った cos 類似度を用いた意味の類似度計算について、精度を上げるためのアイデアを理由とともに述べなさい

演習1で扱った cos 類似度を用いた意味の類似度計算は、単語の意味の近いもの・遠いものを考慮せずに、各単語をそのまま軸にとっていた。なので、単語の意味によって、重みをつけたり、意味が近いものほど大きく寄与するようにすれば、精度を上げることが出来ると考えられる。

Q3 本日の講義まとめと感想

文書、単語を人間が理解している形に近づけるために、文章の類似度を求める必要がある。目的に応じた類似度を設定して、文書・単語を分類していく。

前回から進展して、かなり身近なレベルの言語処理にまつわる内容であったなと感じた。BOW や Word2Vec、潜在的意味解析などは他の授業の「データマイニング」等と被る部分も多く、言語処理の大事な部分であるのだなと再認識した。