

Различия между cosine similarity и soft similarity

В статье предлагается изменить способ вычисления подобия в модели векторного пространства с учетом подобия признаков. Если применить эту идею к косинусной мере, то будет введена «мягкая косинусная мера», в отличие от традиционного «жесткого косинуса», который игнорирует сходство характеристик. Традиционное косинусное сходство рассматривает функции модели векторного пространства (VSM) как независимые.

При работе с текстом мягкое косинусное сходство похоже на косинусное сходство, но, кроме того оно так же рассматривает семантические отношения между словами через их векторное представление. Такие функции, как слова, n-граммы или синтаксические n-граммы, могут быть похожими, хотя формально они рассматриваются как разные функции в VSM. Например, слова «бегать» и «бег» - это разные слова, и поэтому они соответствуют разным точкам в VSM; однако они семантически связаны.