

Рассмотрим пример использования слов в качестве функций в модели векторного пространства. Предположим, что у нас есть два текста: (1) игра, игра, (2) игрок, игрок. Он определяет 4-мерная модель векторного пространства(далее VSM) со следующими функциями: игрок, игра, игрок. У нас есть два вектора  $a$  и  $b$ :  $a = [1, 0, 1, 0]$  и  $b = [0, 1, 0, 1]$ . Традиционное косинусное сходство этих двух векторов равно 0. Но если принять во внимание сходство слов, то оказалось, что эти векторы очень похожи. В обработке естественного языка существует специальная процедура, называемая «выделение корней», направленная на учет подобного рода сходства между словами, но это особая специальная процедура. Более общий вопрос: как мы можем учесть сходство между функциями в модели векторного пространства? Традиционное сходство не рассматривает этот вопрос, т.е. Все признаки считаются разными. Мера косинуса широко применяется и обычно считается само собой разумеющимся

Наша идея более общая: мы предлагаем изменить способ вычисления подобия в модели векторного пространства с учетом подобия признаков. Если мы применим эту идею к косинусной мере, то будет введена «мягкая косинусная мера», в отличие от традиционного «жесткого косинуса», который игнорирует сходство характеристик. Обратите внимание, что когда мы рассматриваем сходство каждой пары функций, это эквивалентно введению новых функций в VSM. По сути, у нас есть матрица сходства между парами функций, и все эти функции представляют новые измерения в VSM. Обратите внимание, что если та же идея применяется к подобию при использовании алгоритмов машинного обучения в модели векторного пространства, то сходство преобразуется в «мягкое подобие».