

ПРОГНОЗИРОВАНИЕ ОЦЕНОК ПОЛЬЗОВАТЕЛЕЙ В ЭЛЕКТРОННОЙ КОММЕРЦИИ ПРИ ПОМОЩИ ФАКТОРНОГО АНАЛИЗА

ВВЕДЕНИЕ

При создании электронных систем часто возникает проблема отсутствия оценок и комментариев при добавлении в систему нового товара, фильма, аудиозаписи, статьи и т.д. Для решения этой проблемы можно составлять прогнозы, из которых, в зависимости от атрибутов и оценок старых товаров, можно узнать, как бы пользователи проголосовали или оценили новый товар. Для решения данной проблемы предлагается использовать факторный анализ. Основным плюсом факторного анализа - возможность работать с разреженными матрицами, когда пропуски в матрице могут составлять 90% от всех данных.

Факторный анализ находит множество применений в различных предметных областях. В электронной коммерции он может быть использован для прогноза цен, заполнения пропущенных данных, а также для сокращения большого количества записей в базе данных, что позволит улучшить рекомендательную систему.

I. ФАКТОРНЫЙ АНАЛИЗ

Факторный анализ - это многомерный анализ, который исследует внутреннюю структуру матриц ковариаций и корреляций [1]. При использовании факторного анализа следует соблюдать ряд требований:

1. Все признаки должны быть количественными;
2. Число признаков должно быть в два раза больше числа переменных;
3. Выборка должна быть однородна;
4. Исходные переменные должны быть распределены симметрично;
5. Факторный анализ осуществляется по коррелирующим переменным.

II. ПРИМЕНЕНИЕ ФАКТОРНОГО АНАЛИЗА В ЭЛЕКТРОННОЙ КОММЕРЦИИ

Чтобы решить проблему с пропущенными значениями и улучшить рекомендательную систему с помощью факторного анализа, вначале требуется высчитать матрицу корреляции из имеющихся данных. Например, в электронном магазине по продаже промышленного оборудования требуется узнать взаимосвязь атрибутов товаров и оценок пользователей, чтобы можно

было прогнозировать оценки новых товаров на основе уже имеющихся. Имея матрицу корреляции можно прогнозировать необходимые значения, но для повышения точности значений требуется выделить факторы, которые помогут сократить значения в базе данных. Для выделения факторов требуется проверить, имеет ли решение модель факторного анализа. Для этого необходимо вычислить количество степеней свободы и определить количество факторов. После того, как найдено количество факторов, при котором модель будет определенной, можно вычислить матрицу нагрузок. Матрица нагрузок представляет собой коэффициент корреляции между измеряемой переменной и латентным фактором. Далее необходимо определить матрицу специфичности. Специфичность характеризует остаточную дисперсию, не объясняемую общими факторами. Согласно фундаментальной теореме факторного анализа, редуцированную матрицу можно найти по формуле $R_h = A^*A^T$, где A - это матрица нагрузок при условии некоррелированности общих факторов. Для того, чтобы увидеть, как эта матрица отклоняется от исходной корреляционной матрицы, необходимо вычислить разность матриц остаточных корреляций. Эта матрица позволяет оценить, насколько точно удалось передать суммарную дисперсию, используя меньшее число общих факторов. Если остаточная корреляционная матрица не содержит высоких значений, то общих факторов достаточно для описания исходных данных. Зная редуцированную матрицу, можно прогнозировать оценки новых товаров и сократить малозначимые информацию о товаре в базе данных.

III. ВЫВОДЫ

Предлагаемый способ работы с продуктами в электронной коммерции позволит не только найти взаимосвязь между атрибутами и выделить факторы для уменьшения статистических переменных, но и заполнить пропущенные значения в базе данных с целью улучшения рекомендательной системы.

1. А. Максвелл, Д. Лоули Факторный анализ как статистический метод. - 1993г. - С. 9-20.

Пинчуков Артур Русланович, студент кафедры интеллектуальных информационных технологий БГУИР, pinchukovartur@outlook.com.

Научный руководитель: Шункевич Даниил Вячеславович, ассистент кафедры интеллектуальных информационных технологий БГУИР shu.dv@tut.by.