

SVD-РАЗЛОЖЕНИЕ В ЗАДАЧЕ ИНФОРМАЦИОННОГО ПОИСКА

Охапкина Е.П., Охапкин В.П.

ФГБОУ ВО «Российский государственный гуманитарный университет»,
Институт информационных наук и технологий безопасности,
Факультет Информационных наук и технологий безопасности,
Кафедра информационных систем и моделирования,
Кафедра информационных технологий и ресурсов
Россия, 117534, г. Москва, ул. Кировоградская, д.25, корп.2
lenaokharkina@mail.ru, vpokharkin@yandex.ru

Использование косинусной меры сходства паттерна и сообщения обладает важными преимуществами, выражающимися в представлении текстовой информации в количественной, матричной форме и относительной простоте вычислений. Однако оценка релевантности паттерна и матрицы «термин-сообщение», насчитывающей тысячи строк и столбцов, подразумевает значительный объем вычислений, принимая во внимание, что нередко возникает ситуация, в которой термины паттерна не встречаются в сообщениях или же встречаются редко, тем самым, формируя в матрице «термин-сообщение» значительное количество нулевых элементов.

В задачах информационного поиска особое место занимает использование сингулярного разложения (далее SVD) матрицы «термин-документ», которое математически представляется следующей формулой:

$$A = U \Sigma V^T, \quad (1)$$

где $A_{m \times n}$ – матрица «паттерн-сообщения», $U_{m \times m}$ – матрица, столбцы которой образованы из ортогональных собственных векторов матрицы AA^T , а $V_{m \times n}$ – матрица, столбцы которой являются ортогональными векторами матрицы $A^T A$. Матрица $\Sigma_{m \times n}$ образована из нулей и чисел $\Sigma_{ii} = \sigma_i = \sqrt{\lambda_i}$, где $\lambda_i \geq \lambda_{i+1}$ являются собственными значениями матрицы AA^T и совпадают с собственными значениями матрицы $A^T A$.

Использование сингулярного разложения матрицы «паттерн-сообщения» A для решения задачи выявления деструктивного информационного воздействия в социальных сетях обладает рядом важных причин:

- В силу того, что в сингулярном разложении используются унитарные матрицы, то этот факт позволяет получить геометрическое представление преобразования матрицы «паттерн-сообщение» в n -мерном пространстве.
- Сингулярное разложение матрицы «паттерн-сообщение» A является устойчивым, другими словами, малым возмущениям матрицы «паттерн-сообщение» соответствуют малые возмущения диагональной матрицы Σ , использованной в разложении и наоборот.
- Исходя из матрицы Σ , можно определить является ли матрица A почти вырожденной. В том случае, если это подтверждается, SVD дает возможность понизить ранг матрицы A с наименьшей погрешностью.