

Android до версии 3.0 (версия API 11). Если в приложении требуется поддержка ранних версий Android, то есть возможно добавить код работы с FTS3 таблицами только для версий Android до одиннадцатой, так как программные интерфейсы FTS3 и FTS4 во многом идентичны.

Токенайзеры. FTS Токенайзер — алгоритм извлечения термов из текста. Токенайзер задействуется как при индексации текста документов во время добавления их в базу данных, так и при осуществлении запроса к базе данных (над текстом запроса). По умолчанию используется “simple” токенайзер. Основные положения алгоритма “simple” токенайзера:

Термом считается последовательность из UTF-символов, код которых больше или равен 128, из символов букв, цифр, символа “_”. Все остальные символы отбрасываются.

Все символы верхнего регистра набора ASCII преобразуются в свои эквиваленты нижнего регистра.

К примеру, применим алгоритм “simple” к тексту “This time we’re really breaking up”. Мы получим следующие термы: “this time we re really breaking up”. Термы будут добавлены в полнотекстовый индекс и данная фраза будет найдена по такому запросу, как “MATCH ‘Really’”, ведь благодаря “simple” токенайзеру полнотекстовые запросы являются регистронезависимыми.

Наряду с “simple” существуют и другие алгоритмы, например “porter” токенайзер на базе алгоритма стемминга Портера. Стеминг - это нахождение основы слова, которая не всегда совпадает с морфологическим корнем слова. К примеру “porter” включает в себя правила удаления окончаний слов: “ed”, “ing”, “ly”, а перед выделением основ термов проводится “simple” токенизация. Результат обработки того же документа, что и в предыдущем абзаце: “thi time we re realli break up”.

Таким образом, использование функциональности полнотекстового поиска чаще всего дает значительный выигрыш во времени, потраченном на нахождение результатов в больших объемах данных, позволяет довольно легко реализовать вывод результатов поиска в виде списка фрагментов документов. Данный подход должен быть применен во всех приложениях, в которых есть функционал поиска по относительно большим объемам данных, чтобы обеспечить положительный пользовательский опыт при работе с Android приложением.

Список использованных источников:

1. Википедия. Полнотекстовый поиск [Электронный ресурс] – Режим доступа. – URL https://ru.wikipedia.org/wiki/%D0%9F%D0%BE%D0%BB%D0%BD%D0%BE%D1%82%D0%B5%D0%BA%D1%81%D1%82%D0%BE%D0%B2%D1%8B%D0%B9_%D0%BF%D0%BE%D0%B8%D1%81%D0%BA (дата обращения 25.03.2017).
2. Adding Search Functionality . Storing and Searching for Data [Электронный ресурс] – Режим доступа. – URL <https://developer.android.com/training/search/search.html> (дата обращения 25.03.2017).
3. SQLite FTS3 and FTS4 Extensions [Электронный ресурс] – Режим доступа. – URL <http://www.sqlite.org/fts3.html> (дата обращения 25.03.2017).

ПРИМЕНЕНИЕ ФИЛЬТРА КАЛМАНА В ОБРАБОТКЕ ИЗОБРАЖЕНИЙ

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Галдунов Д.В.

Стройникова Е.Д. – ассистент кафедры информатики

В настоящее время проводятся множество исследований в области обработки изображений и компьютерного зрения, и результаты этих исследований активно используются в научной, производственной и других сферах деятельности. Методы обработки графической информации чаще всего применяются на начальных этапах работы алгоритмов и ставят своей целью очистку данных для дальнейших вычислений. В качестве одного из таких методов может выступать метод калмановской фильтрации, впервые опубликованный в 1960 г.

Фильтр Калмана – линейный рекурсивный фильтр, который используется для получения оценки состояния динамической системы по серии неточных измерений. Модель системы представлена уравнением

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1},$$

где вектор x_i – истинное состояние системы в момент времени i ; матрица A порядка $n \times n$ связывает состояние в предыдущий момент времени $k-1$ с состоянием в текущий момент k ; $n \times l$ матрица B связывает опциональный входной вектор управляющих воздействий u с вектором x ; w_i является случайным вектором. Измеренное состояние системы описывается уравнением

$$z_k = Hx_k + v_k,$$

где $m \times n$ матрица H связывает состояние x с измерением z ; случайный вектор v_i представляет собой ошибку измерений прибора. Предполагается, что векторы w и v являются взаимно независимыми и распределенными нормально с нулевым математическим ожиданием и ковариационными матрицами Q и R соответственно.

В области обработки графических данных фильтр Калмана может быть использован для решения

следующих задач:

- построение сверхразрешения;
- отслеживание перемещения объекта на серии изображений.

Применение методов сверхразрешения позволяет получить изображение высокого разрешения из последовательности изображений низкого разрешения, содержащих сходную информацию. Для упрощения расчетов исходные изображения предполагается переписать в вектор по столбцам. В этом случае x_k представляет собой изображение высокого разрешения, полученное в результате сдвига объекта, описываемого матрицей A ; $u_k = 0$; z_k – изображение низкого разрешения. Итерацию алгоритма фильтра Калмана можно разбить на 2 этапа: этап предсказания

$$\tilde{x}_{(k+1|k)} = A\tilde{x}_{(k|k)}, \quad P_{(k+1|k)} = A_k P_{(k|k)} A_k^T + Q_k$$

и этап корректировки

$$\tilde{x}_{(k|k)} = \tilde{x}_{(k|k-1)} + K_k (y_k - H\tilde{x}_{(k|k-1)}), \quad K_k = V_k U_k^{-1}, \quad U_k = H P_{(k|k-1)} H^T + R_k, \\ V_k = P_{(k|k-1)} H^T, \quad P_{(k|k)} = P_{(k|k-1)} - K_k V_k^T,$$

где оценка $\tilde{x}_{(k|k)}$ – изображение сверхразрешения; оценка $\tilde{x}_{(k|k-1)}$ – предсказание изображения сверхразрешения на следующем шаге; матрица ковариации $P_{(k|k-1)} = M[\varepsilon_k \varepsilon_k^T]$ и ошибка прогноза $\varepsilon_k = x_k - \tilde{x}_{(k|k-1)}$. Матрица K_k называется матрицей коэффициентов усиления. Под шагом k понимается поступление на вход очередного изображения y_k . Важно отметить тот факт, что вычисления производятся над большими матрицами, что отрицательно сказывается на производительности и может приводить к расходимости алгоритма. С этим борются путем разбиения исходного изображения на отдельные блоки.

Отслеживание перемещения объекта является одной из задач компьютерного зрения, и метод калмановской фильтрации может быть применен для предсказания положения объекта в следующий момент времени. Для двумерного случая вектор состояния $x_k = (x, y, \dot{x}, \dot{y})$ содержит текущие координаты и проекции скорости, а вектор измерений $z_k = (x, y)$ – зашумленные измеренные координаты положения объекта. Скорость предполагается постоянной величиной. Для инициализации процесса отслеживания объекта необходима оценка начального вектора состояния $\tilde{x}_{(0|0)}$ и ковариационная матрица $P_{(0|0)}$. Часто в качестве $P_{(0|0)}$ выбирают диагональную матрицу с дисперсиями на главной диагонали, предполагая, что компоненты независимы. Матрица перехода между состояниями

$$A = \begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

где Δt – интервал предсказания.

Список использованных источников:

1. Сирота, А. А. Блочные алгоритмы обработки изображений на основе фильтра Калмана в задаче построения сверхразрешения / А. А. Сирота, А. Ю. Иванков // Компьютерная оптика. – 2014. – т. 38, № 1. – С. 118–126.
2. Welch, G. An Introduction to the Kalman Filter : Technical Report / G. Welch, G. Bishop. – Chapel Hill : UNC, 1995. – 16 p.
3. KalmanFilterSimulation [Электронный ресурс]. – Режим доступа : https://cs.utexas.edu/~teammco/misc/kalman_filter/. – Дата доступа : 28.03.2017.
4. Фильтр Калмана – Введение [Электронный ресурс]. – Режим доступа : <https://habrahabr.ru/post/140274/>. – Дата доступа : 30.03.2017.

АНАЛИЗ ТОНАЛЬНОСТИ ТЕКСТА

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Гербик А.И., Дорошкевич П.Е., Свито А. И.

Стержанов М.В. – канд. техн. наук, доцент

В связи с накоплением больших объемов текстовых данных становится актуальной задача их автоматической обработки, свидетельством чего являются многочисленные исследования в области обработки естественного языка. Одна из наиболее исследуемых задач в данной области – задача анализа тональности текста.

Целью анализа тональности является нахождение мнений в тексте и определение позиции автора относительно упомянутой темы. Позиция автора может быть различной, и тональная оценка может принимать различные значения. Например: “положительная”, “отрицательная” и “нейтральная” либо “положительная” и