

РАБОТА С РЕЧЕВОЙ БАЗОЙ ДАННЫХ НА КЛАСТЕРЕ IBM BLADE HS21 ИКИТ СФУ

М.С. Медведев, В.Н. Никитин, Г.С. Кирякова

На сегодняшний день задаче построение речевого интерфейса в вычислительных системах уделяется большое внимание. В связи с повышением производительности мобильных устройств и, как следствие, расширением их функциональных возможностей, возникает необходимость разработки пользовательских интерфейсов и приложений, использующих алгоритмы распознавания речи. Поэтому исследования в области создания систем речевого ввода являются наиболее перспективными на сегодняшний день.

В данной статье анализируются возможности сокращения временных затрат на предварительную обработку информации речевой базы данных за счет распараллеливания алгоритма нормализации, используя кластер IBM Blade HS21.

При обработке больших объемов данных возникает проблема скорости выполнения данной операции, особенно в системах распознавания речи, где этот параметр имеет критическое значение. Распараллеливание вычислительного процесса в алгоритмах разработанной системы «Преобразования речи в текст» [1] позволяет сократить время вычислений для целого комплекса задач:

1. Преобразование в цифровую форму и предварительная обработка речевого сигнала
2. Вычисление признаков речевых единиц на этапах формирования обучающей выборки и распознавания;
3. Обучение нейронной сети и ее использование для классификации фонем;
4. Построение грамматической формы распознаваемых образов слов.

Использование РБД.

Решение задач, использующих речевые технологии, помимо собственно речевых материалов, требует также наличия большого количества данных лингвистического характера, таких, как наборы текстов, их лингвистические описания, словари различного рода и т.п. Исследования в таких областях, как синтез и распознавание речи, требуют накопления все большего и большего количества речевых фрагментов, сопровождаемых описанием соответствующих деталей этих фрагментов. Поэтому при разработке системы преобразования речи в текст возникает необходимость создания речевой базы данных.

Речевая база данных (РБД) - это база данных, в которой хранятся речевые данные в виде массивов отсчетов оцифрованного речевого сигнала, а также необходимая справочная информация (данные о дикторе, параметры сигнала и т.д.). Каждый массив отсчетов обычно записывается в отдельном файле, а в самой базе данных хранится только имя этого файла. Современное развитие вычислительной техники позволяет собирать объемные БД и хранить их в достаточно компактном виде.

Предварительная обработка данных РБД.

Для анализа речи её необходимо преобразовать в цифровую форму. Параметрами, определяющими качество оцифровки сигнала, являются частота дискретизации и разрядность преобразования (сколько единиц информации кодируется один отсчет). Частота дискретизации определяет максимальную частоту сигнала, которую можно записать. Типичные значения – 11025, 22050, 44100 Гц. От разрядности зависит точность кодирования информации при аналого-цифровом преобразовании. Стандартные значения – 4 бит, 8 бит, 16 бит на отсчет. Параметры дискретизации также хранятся в РБД. Речевой сигнал необходимо подготовить для дальнейшей обработки. На данном этапе осуществляется очистка сигнала от шума, удаление пауз речи, его нормализация. Также проводится разбиение сигнала на фиксированные интервалы (сегменты) во временной области, на которых будут рассчитываться описательные признаки.

Нормализация сигнала.

Важным этапом предварительной обработки является нормализация амплитуды оцифрованного речевого сигнала. Это позволяет уменьшить ошибки распознавания, связанные с тем, что диктор может произносить слова с различным уровнем громкости. Нормализация амплитуды сигнала проводится по следующей формуле [2]:

$$pn_i = \frac{2 \cdot (p_i - \min p)}{(\max p - \min p)} - 1,$$

где p_i – значение амплитуды i -го отсчета вектора речевого сигнала; pn_i – нормализованное значение амплитуды i -го отсчета вектора речевого сигнала; $\min p$ – минимальное значение амплитуды вектора речевого сигнала; $\max p$ – максимальное значение амплитуды вектора речевого сигнала;

При этом максимальное и минимальное значения элементов вектора будут лежать в диапазоне [-1; 1].

Реализация алгоритма.

Для повышения эффективности предварительной обработки материала речевой базы данных, связанного с уменьшением временных затрат, анализировались возможности распараллеливания алгоритма нормализации, используя ресурс, имеющийся в вычислительном центре ИКИТ СФУ - кластер IBM Blade HS21 (224 узла) производительностью 9287 TFlops/s и пакет rMatLab. Кластер ИКИТ СФУ работает под управлением ОС SUSE Linux Enterprise Server 10. Параллельные вычисления обеспечивают пакеты Distributed Computing Toolbox (DCT) и Computing Engine (MDCE) (Рис.1).

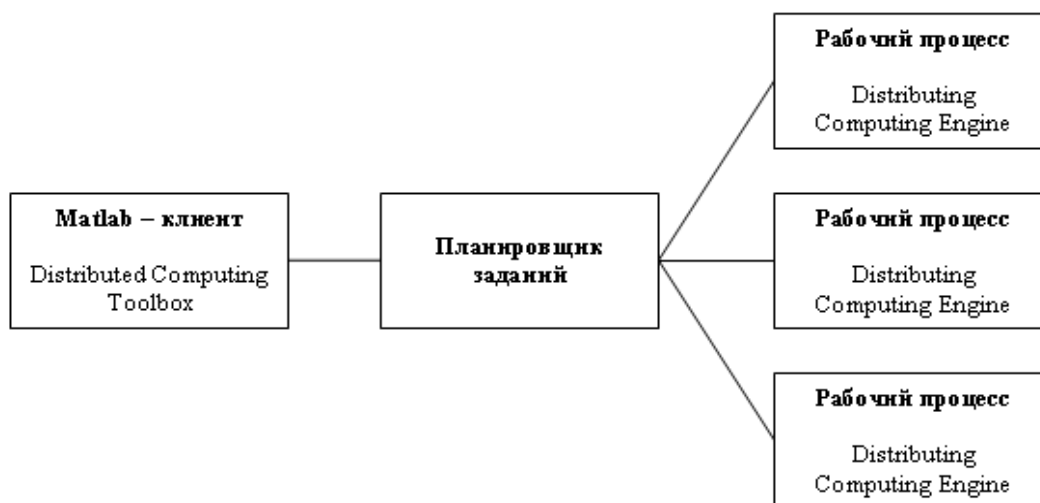


Рис. 1. Механизм распределенных вычислений Matlab

Инструменты Distributed Computing Toolbox и MATLAB Distributed Computing Engine предназначены для реализации кластерных вычислений в среде MATLAB. С помощью первого из них пользовательская программа, написанная на языке MATLAB, разделяется на несколько независимых задач. MATLAB Distributed Computing Engine занимается выполнением этих задач в автономных сессиях MATLAB, запущенных на кластере.

Для реализации интерфейса с базой данных рассматривались два подхода:

1. используется встроенный в пакет Matlab Database toolbox, который позволяет применять мощные инструменты анализа и визуализации Matlab для работы с информацией, хранящейся в базах данных. Предоставляется возможность работы с командами языка запросов SQL. Возможно взаимодействие с ODBC/JDBC-совместимыми базами данных: Oracle, Sybase SQL Server, Sybase SQL Anywhere, Microsoft SQL Server, Microsoft Access, Informix и Ingres. Из недостатков стоит отметить поддержку только локальных баз данных MySQL и низкую производительность.
2. разработка внешнего модуля расширения базовой функциональности командного интерфейса Matlab позволяет осуществлять работу с удаленными базами данных MySQL с сохранением высокой производительности.

Очевидно, что в случае удаленного расположения базы данных MySQL невозможно использование первого варианта реализации, поэтому рассматривается второй подход, а именно, сторонний модуль. Использование стороннего модуля позволяет решить поставленную задачу с минимальными потерями времени, так как затраты временных ресурсов на взаимодействие пакета Matlab с удаленной базой данных фактически нивелируются. При такой реализации решающим фактором является пропускная способность канала связи между удаленным сервером баз данных и кластерной вычислительной системой.

В рамках практической реализации стороннего модуля следует отметить следующие положения:

- язык высокого уровня (C++);
- кроссплатформенность;
- использование API MySQL, API Matlab;
- совместимость с любыми версиями Matlab и MySQL;
- прямое обращение к БД MySQL посредством Matlab. Это обеспечивается интеграцией новых функций для работы с базами данных в командный интерфейс Matlab.

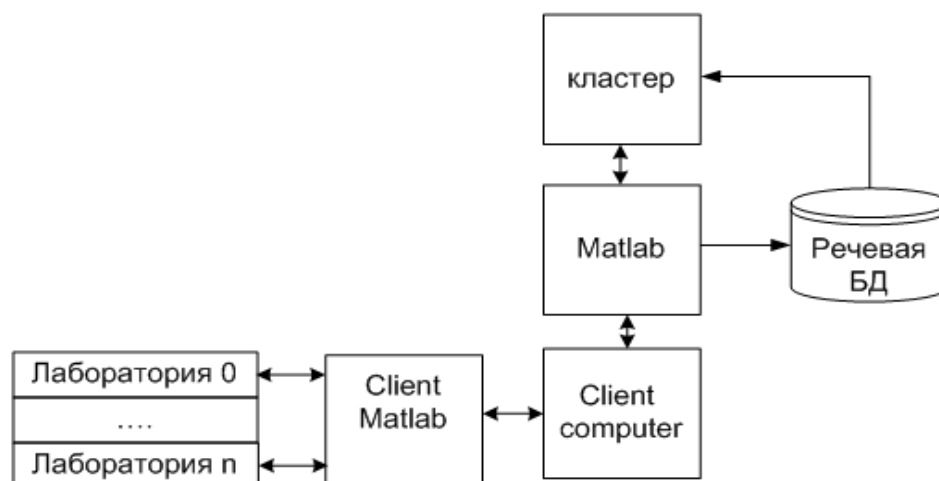


Рис. 2. Схема взаимодействия Matlab с РБД

Для хранения речевого материала использовалась база данных MySQL. Общее количество записей - 10 000, что составляет около трех часов оцифрованной речи диктора.

Речевые сигналы слов-эталонов в формате wav, оцифрованные с частотой дискретизации 22050 Гц, разрядностью 16 бит, длительностью не менее 1000 мс отделялись от пауз и записывались в базу данных с использованием типа поля BLOB.

Для исследования эффективности распараллеливания алгоритма нормализации на этапе предварительной обработки материала речевой базы данных, проводилось измерение общего времени обработки речевого материала, которое в свою очередь складывалось из времени соединения с базой данных, времени выполнения запроса к базе и времени обработки речевых сигналов с их последующим сохранением. В ходе эксперимента к базе данных формировался запрос на получение выборки из 5000 записей, после получения ответа, на сервере Matlab, по описанному алгоритму производилась нормализация каждого вектора речевого сигнала и оценивались временные затраты.

Время вычисления без использования параллельного режима составило 75 с, из них на соединение с базой отводилось 0,002 с и на получение данных – 3,2 с. Данная операция выполнялась с распараллеливанием вычислений на 2, 4 и 8 процессов, параллелизм реализовывался с использованием механизма parfor. Была проведена оценка сокращения времени вычислений в зависимости от числа используемых процессов.

Количество используемых процессов (ядер)	Сокращение времени вычисления, %
2	34%
4	64%
8	78%

Таблица 1 – Сокращение времени обработки данных в зависимости от числа процессов

Далее приведен листинг функции параллельной обработки речевых данных.

```

function res = speechnorm()
% Host
host = 'nodes-01-01';
% Database name
dbname = 'waves';
% Database user
user = 'user';
password = 'password';

% Connect to database
matlabpool open local 8;
mysql('open', host, user, password);
mysql('use waves');
vector = mysql('SELECT v FROM waves.vectors LIMIT 1, 5000');

tstart = tic;
  
```

```

parfor i=1:5000
    warning off;
    vec = str2double(vector(i));
    c = premnmx(vec); % нормализация
end
res = toc(tstart);
mysql('close');
matlabpool close;

```

Результаты экспериментов показали сокращение временных затрат на предварительную обработку речевого сигнала при увеличении количества задействованных параллельных процессов. При использовании алгоритма распознавания слитной речи, суммарное время вычислений, выполняемых для одного речевого сигнала слова, в соответствии со средним темпом речи, должно быть менее 500 мс. Минимальное время предварительной обработки вектора отсчетов одного слова составило 4 мс, что удовлетворяет указанным требованиям.

На основе полученных результатов можно сделать выводы о возможности применения разработанной схемы взаимодействия Matlab с базой данных MySQL для параллельной обработки речевого материала на кластере в задаче распознавания речи.

ЛИТЕРАТУРА:

1. Медведев М.С. Интеллектуальный комплекс для исследования речевого сигнала. / Кирякова Г.С., Ли А.С. // Информационные технологии в науке, образовании, телекоммуникации и бизнесе '09: материалы XXXVI Международн. конференц. молодых ученых. – Ялта-Гурзуф, 2009.
2. Медведев М.С. Распознавание речевых образов с использованием вейвлет-преобразования. / Модели и методы обработки изображений ММОИ-2007: Материалы Всерос. научн. конференц. - Красноярск: СФУ, 2007. – С. 53-59.
3. Дьяконов, В. MATLAB/В. Дьяконов.- СПб.: Питер, 2001. - 560 с.
4. Cohen, M.H. Voice user interface design/M.H. Cohen, J.P. Giagola, J. Balogh. – Redwood City: Addison Wesley Longman Publishing, 2004. – 336 с.
5. Minker, W. Speech and human-machine dialog/W. Minker. – London: Kluwer Acad. Publ., 2004. - 93 p.