

Использование суффиксных деревьев для оптимизации поиска элемента звуковой базы по технологии Unit Selection

Аннотация

Синтез речи широко используется для различных задач, таких как, разработка речевого пользовательского интерфейса, чтение произвольного текста без предварительной разметки. Данные проблемы весьма актуальны для людей с ограниченными возможностями.

Для получения качественных результатов, компилятивный (компиляционный, конкатенативный) синтез требует существенных вычислительных мощностей для обработки большой звуковой базы. В данной работе рассматривается использование суффиксных деревьев для увеличения скорости работы метода Unit Selection.

Введение

Синтез речи широко используется для различных задач, таких как, разработка речевого пользовательского интерфейса, чтение произвольного текста без предварительной разметки. Данные проблемы весьма актуальны для людей с ограниченными возможностями.

Предоставляя лучшее качество синтезированной речи (при наличии качественной речевой базы), компилятивный синтез требует существенных вычислительных мощностей.

Предполагается, что использование суффиксных деревьев позволит снизить вычислительную стоимость синтеза. Таким образом, основной проблемой является вычислительная сложность алгоритмов компилятивного синтеза речи, объектом исследования является суффиксное дерево в задаче поиска элементов звуковой базы по технологии Unit Selection

Сравнение аналогов

Поиск полным перебором

Использует алгоритм Ханта-Блэка для вычисления стоимости последовательности элементов звуковой базы, как сумму стоимости замены (англ. targets cost) и стоимости связи (англ. concatenation cost) пар элементов следующих друг за другом. В качестве признака, используемого для вычисления стоимости связи, могут использоваться различные метрики, такие как энергия, спектральные параметры (Mel Frequency Cepstral Coefficients, Multiple Centroid Analysis, ...)

Алгоритм Витерби

Алгоритм Витерби (в некоторых источниках поиск Витерби, решетчатая функция декодирования Витерби) - динамического программирования, реализующий поиск

наиболее подходящего списка состояний (называемого путём Витерби), который в контексте цепей Маркова получает наиболее вероятную последовательность произошедших событий. Для построения матрицы переходов используются стоимости замены и связи[3].

Алгоритм Витерби с отсечениями

Представляет собой алгоритм Витебри, с добавлением методики отсечения (англ. pruning), направленной на уменьшение числа рассматриваемых последовательностей.

Используются две основные методики отсечения: предварительный отбор (англ. pre-selection) и отсечение лучей (англ. beam pruning). Предварительный отбор оставляет фиксированное количество лучших кандидатов для каждого элемента, отсечение лучей рассматривает фиксированное количество локально оптимальных путей. Использование любой методики отсечения создаёт риск отсечь оптимальный путь.

Критерии сравнения аналогов

Скорость

Является ключевым критерием, поскольку качество синтезируемой речи напрямую зависит от размера звуковой базы, что требует возможность быстрой обработки большой базы.

Сложность реализации

Влияет на экономическую стоимость реализации системы распознавания.

Качество синтеза

Мера схожести синтезированной речи с человеческой.

Таблица сравнения по критериям

Аналог	Скорость	Сложность реализации	Качество синтеза
Поиск полным перебором	экспоненциальная	-	зависит от базы
Алгоритм Витерби	квадратичная	+	зависит от базы
Алгоритм Витерби с отсечениями	квадратичная	+	зависит от базы, и от того, было ли отсечено лучшее решение

Выводы по итогам сравнения

В результате сравнения аналогов, можно сделать вывод, что увеличение скорости вычислений является главной задачей. Идея использовать суффиксных деревьев для поиска оптимальной последовательности преследует именно эту цель.

Выбор метода решения

В результате проведенного анализа существующих алгоритмов синтеза речи по технологии Unit Selection, определена главная проблема - скорость поиска подходящих элементов, для больших (100 000 и более элементов) речевых баз. Следует отметить, что в работе не рассматриваются вопросы разметки звуковой базы, предобработки, эффективности использования различных метрик (MFCC, PCT, MSE, ...) и их комбинаций для определения стоимости замены и стоимости связи, постобработки выбранных юнитов, способов программной реализации суффиксного дерева. Данный подход позволит использовать разработанный алгоритм не только для технологии Unit Selection, но и для метода гибридного синтеза. Поскольку в гибридном методе, к примеру, с использованием скрытых марковских моделей, статистический блок определяет акустические признаки (частоту основного тона и длительность) элементов, после чего, на основе функции стоимости, методом Unit Selection определяется оптимальная последовательность. Таким образом, определены два возможных вектора работы: разработка непосредственно алгоритма поиска элемента в базе, и разработка алгоритма вычисления весовых коэффициентов используемых в построении оценки стоимости связи. Любой из разработанных алгоритмов, должен соответствовать следующим качествам: * скорость лучше чем у поиска Витерби; * гарантированное нахождение лучшего пути.

Описание метода решения

Как уже упоминалось ранее, в данной работе рассматривается исключительно алгоритм поиска элемента звуковой базы. Пусть даны требуемые параметры звуковых элементов, необходимых для синтеза некоторого предложения или слова. Задача Unit Selection - выбрать наилучшую последовательность реализаций звуковых элементов по полученным параметрам.[3]

В качестве метрики характеризующей насколько элемент *u* подходит для синтеза текущей фонетической единицы, используются понятия стоимости замены (англ. target cost) и стоимости связи (англ. concatenation cost). Данные оценки представляют собой взвешенные суммы различий в признаках между двумя элементами. Используемые признаки не регламентируются Unit Selection и выбираются разработчиком. Как правило используются следующая информация: * частота основного тона * длительность * контекст * позиция элемента в слоге/слове * количество ударных слогов * энергия * спектральные параметры: * MFCC (Mel Frequency Cepstral Coefficients); * LPC (Linear Predictive Coding Coefficients); * LSF (Line Spectral Frequencies); * MCA (Multiple Centroid Analysis).

Множество исследований, имеющих своей целью определить какая комбинация спектральное представление/метрика даст лучшую корреляцию с человеческим восприятием речи, не привели к однозначному ответу. Однако, можно отметить, что расстояние Махаланобиса на коэффициентах MFCC даёт хорошие результаты в большинстве тестов.[2]

Таким образом, стоимость замены, характеризующая близость выбранного элемента к искомому, вычисляется по формуле:

equation

где C - расстояние между k -ми характеристиками элементов, w - вес для k -ой характеристики,
 u - элемент из звуковой базы.

А стоимость связи, отображающая насколько хорошо сочетаются два последовательно выбранных элемента:

equation

где C - расстояние между k -ми характеристиками элементов,
 w - вес для k -ой характеристики.

Общая стоимость последовательности из n элементов, вычисляется как сумма описанных выше стоимостей:

equation

Поскольку различия одних характеристик оказывают больше влияния, чем различия других характеристик, возникает необходимость использования весовых коэффициентов w . Существуют следующие методики определения w : * автоматический подбор на основе объективной меры; Идея - найти набор тестов, который минимизирует акустическое расстояние между синтезированной и эталонной речью. Задача определения данной метрики является отдельной проблемой. * перцепционный; Проводится эксперимент, в ходе которого, испытуемые оценивают качество синтезированной речи. Полученные оценки используют для тренировки модели. * ручная настройка. Коэффициенты полностью определяются разработчиком системы.

Задача Unit Selection - выбрать множество элементов из звуковой базы таким образом, чтобы минимизировать общую стоимость. Классическим решением является использование поиска Витерби с отсечениями, поскольку полный поиск является слишком долгой операцией. Используются две основные методики отсечения: предварительный отбор (англ. pre-selection) и отсечение лучей (англ. beam pruning). Предварительный отбор оставляет фиксированное количество лучших кандидатов для каждого элемента, отсечение лучей рассматривает фиксированное количество локально оптимальных путей.

В данной работе рассматривается возможность использования суффиксных деревьев для оптимизации поиска.

Определение[5]

Суффиксное дерево для m -символьной строки S - это ориентированное дерево с корнем, имеющее ровно m листьев, занумерованных от 1 до m . Каждая внутренняя вершина, отличная от корня, имеет не меньше двух детей, а каждая дуга помечена непустой подстрокой строки S . Никакие две дуги, выходящие из одной и той же вершины, не могут иметь пометок, начинающихся с одного и того же символа.

Для того, чтобы можно было построить суффиксное дерево для любой строки, вводится терминальный символ конца строки, не входящий в основной алфавит.

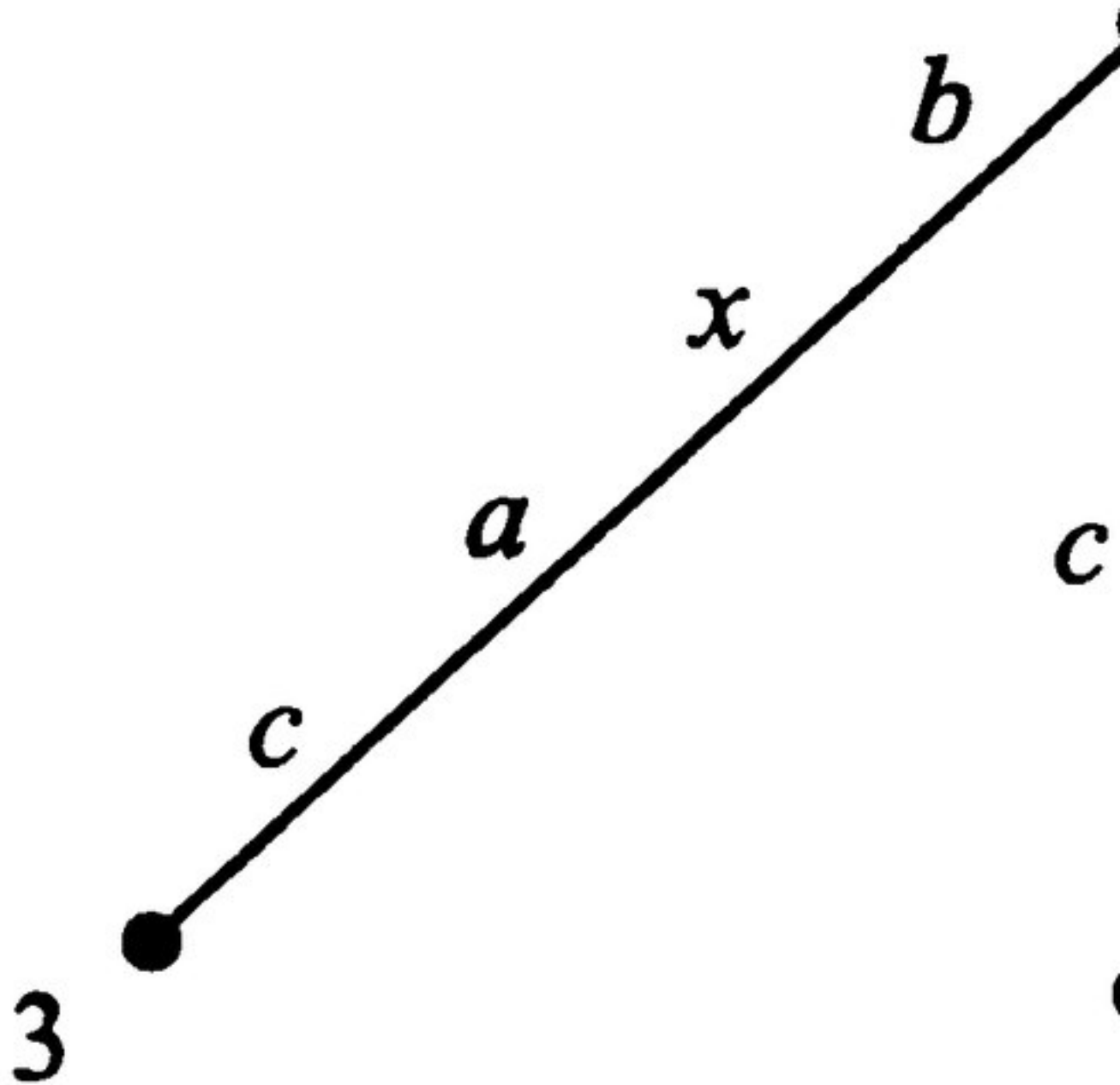


Рис. 1 - Суффиксное дерево $S=хавхас$

Рис. 1 - Суффиксное дерево $S=хавхас$

Суффиксные деревья позволяют решать множество задач поиска на неструктурированных данных. Для строки T длиной m , суффиксное дерево занимает $O(m)$ памяти, построение дерева требует время $O(m)$. Поиск вхождения строки S длиной n в строку T , занимает $O(n)$. Не смотря на то, что обычно, суффиксные деревья используются для задач точного совпадения, они позволяют так же решать задачи неточного совпадения (нечеткого поиска), необходимые для Unit Selection.

Заключение

В данной работе изучена проблема поиска элемента звуковой базы при синтезе речи по технологии Unit Selection для больших речевых баз. Изучение механизмов обработки (сглаживания) сформированной звуковой последовательности, получения акустических данных искомой последовательности не входят в рамки данной работы. Изучены существующие реализации, проанализированы их проблемы, а именно низкая скорость работы. Рассмотрено использование суффиксных деревьев, с целью оптимизации поиска элемента звуковой базы. Выявлены и описаны проблемы и ограничения использования суффиксных деревьев. Дальнейшим направлением исследований может быть реализация суффиксного дерева, для решения задач поиска в методе Unit Selection, разработка алгоритма направленного на повышение производительности гибридной модели речевого синтеза.

Источники

1. Motirz M., Suffix Trees and their Applications // Munchen Technical University, Munchen, 1999
2. Рыбин С. В. СИНТЕЗ РЕЧИ Учебное пособие по дисциплине "Синтез речи". – СПб: Университет ИТМО, 2014. – 92 с.
3. Black A.W., Hunt A.J. Unit Selection in a Concatenative Speech Synthesis Using a Large Speech Database // In Proceedings of ICASSP 96. Atlanta, Georgia, 1996. Vol. 1, pp. 373-376.
4. Balint M. Suffix Trees and Their Applications // Lorand University, Budapest, 2013
5. Гасфилд Дэн Строки, деревья и последовательности в алгоритмах: Информатика и вычислительная биология / Пер. с англ. И.В.Романовского. - СПб.: Невский Диалект; БХВ-Петербург, 2003. - 654 с.: ил.
6. Корольков В.А., Главатских И.А., Таланов А.О. Синтез естественной русской речи при помощи метода Unit selection // Тр. XXXVI межд. филолог. конф. «Формальные методы анализа русской речи». Россия, 2008