

Задачи оценивания значимости выравнивания при помощи скрытых марковских моделей

Власенко Даниил Владимирович, гр.19.Б04-мм

Научный руководитель: к.ф.-м.н. Коробейников А.И.

Санкт-Петербургский государственный университет
Прикладная математика и информатика
Вычислительная стохастика и статистические модели

Отчет по производственной практике

Санкт-Петербург, 2022

Оценивание значимости выравнивания

Задачи оценивания значимости выравнивания
при помощи скрытых марковских моделей

Власенко Даниил Владимирович, гр.19.Б04-мм
Научный руководитель: к.ф.-м.н. Коробейников А.И.

Санкт-Петербургский государственный университет
Прикладная математика и информатика
Вычислительная стохастика и статистические модели

Отчет по производственной практике
Санкт-Петербург, 2022

Научный руководитель к.ф.-м.н., Коробейников А.И.,
кафедра статистического моделирования

Пусть X и Y последовательности, состоящие из символов алфавита A .

Определение

Выравнивание Q последовательностей X и Y — это установление соответствия между символами этих строк с помощью добавлений символов пропуска и замен символов, так что порядок символов внутри строк сохраняется.

$$\begin{array}{cccccccc} X = & A & C & E & A & A & F & A & E \\ Y = & C & E & A & F & D & C & E \end{array}$$
$$\begin{array}{cccccccccc} X = & A & C & E & A & A & F & A & - & E \\ Y = & - & C & E & A & - & F & D & C & E \end{array}$$

Рис. 1: Последовательности до и после выравнивания.

Оценивание значимости выравнивания

Введение

Последовательность длины L — строка D состоящая из L символов алфавита Σ . Выравнивание последовательностей — это установление соответствия между символами строк с помощью добавлений символов пропуска и замен символов, так что порядок символов внутри строк сохраняется. Например, даны последовательности ACEAAFAE и CEAFDCE, если расположить их друг под другом, то не будет ни одного совпадения соответствующих символов, но если вставить пропуск восьмого символа в первой последовательности и пропуски первого и пятого символов во второй последовательности, то мы получим 5 совпадений.

Оценка выравнивания $s(Q)$ — функция, отражающая выравнивание последовательностей в действительные числа. Способом вычисления оценки выравнивания s может быть, например, увеличение оценки на 1 при совпадении символов, стоящих друг под другом, и уменьшение на $\frac{1}{2}$ при несовпадении. Тогда оценка s приведенного выше выравнивания будет равна 3. Способ вычисления оценки выравнивания выбирается исходя из целей и вида выравнивания.

Введение

Пусть X и Y последовательности, состоящие из символов алфавита A .

Определение

Выравнивание Q последовательностей X и Y — это установление соответствия между символами этих строк с помощью добавлений символов пропуска и замен символов, так что порядок символов внутри строк сохраняется.

$$\begin{array}{cccccccc} X = & A & C & E & A & A & F & A & E \\ Y = & C & E & A & F & D & C & E \end{array}$$
$$\begin{array}{cccccccccc} X = & A & C & E & A & A & F & A & - & E \\ Y = & - & C & E & A & - & F & D & C & E \end{array}$$

Рис. 1: Последовательности до и после выравнивания.

A	C	E	A	A	F	A	E
C	A	A	F	D	C	E	
A	E	A	F	A	C	E	

A	C	E	A	A	F	A	—	E
—	C	A	A	—	F	D	C	E
A	—	E	A	—	F	A	C	E

Рис. 2: Последовательности до и после выравнивания.

Определение

Оценка выравнивания $s(Q)$ — функция, отражающая выравнивание последовательностей в действительные числа.

Оценивание значимости выравнивания

— Введение

Введение

A	C	E	A	A	F	A	E
C	A	A	F	D	C	E	
A	E	A	F	A	C	E	

A	C	E	A	A	F	A	—	E
—	C	A	A	—	F	D	C	E
A	—	E	A	—	F	A	C	E

Рис. 2: Последовательности до и после выравнивания.

Определение

Оценка выравнивания $s(Q)$ — функция, отражающая выравнивание последовательностей в действительные числа.

На предыдущем слайде приведен пример попарного выравнивания двух строк, но если сходство последовательностей слабое, то через такое выравнивание может не выйти идентифицировать взаимосвязь описываемых последовательностями объектов. Однако сравнение сразу трех и более последовательностей может позволить выявить эту взаимосвязь, такое выравнивание называется множественным. Проводить множественное выравнивание стандартными методами динамического программирования для попарного выравнивания вычислительно неэффективно, но оказывается, что аппарат скрытых марковских моделей (СММ) позволяет эффективно решать эту задачу.

Сходство последовательностей может отражать функциональные, структурные или эволюционные взаимосвязи объектов, которые описывают эти последовательности. Таким образом вычисление оценки выравнивания последовательностей может быть полезно в задаче определения степени родства биологических организмов путем сравнения их ДНК или РНК, нуклеотидных последовательностей, задаче анализа свойств белков, аминокислотных последовательностей, задаче распознавания речи человека или письменного языка и многих других приложениях.

Оценивание значимости выравнивания

Введение

└ Введение

Рассмотрим последовательности над алфавитом

- достаточно ли высокая оценка, чтобы считать последовательность не шумом, или шум мог добиться такой оценки.
- достаточно ли низкая оценка, чтобы считать последовательность шумом, или не шум мог получить такую оценку.

Определение

Ложноположительная вероятность оценки s — это вероятность того, что шум получит оценку равную или выше s .

Оценивание значимости выравнивания

— Введение

Введение

- достаточно ли высокая оценка, чтобы считать последовательность не шумом, или шум мог добиться такой оценки.
- достаточно ли низкая оценка, чтобы считать последовательность шумом, или не шум мог получить такую оценку.

Определение

Ложноположительная вероятность оценки s — это вероятность того, что шум получит оценку равную или выше s .

СММ будут описаны далее, пока что зададимся следующим вопросом. Если есть множество последовательностей, описывающих взаимосвязанные объекты, имеется еще одна последовательность и была посчитана оценка выравнивания этой последовательности ко всему множеству каким-либо способом, то

- достаточно ли высокая эта оценка, чтобы считать объект, описываемый последовательностью, родственным к объектам, описываемым множеством, или шум, т.е. случайная последовательность, мог добиться такой оценки.
- достаточно ли низкая эта оценка, чтобы считать объект описываемый последовательностью, не родственным к объектам, описываемым множеством, или сигнал, т.е. последовательность, описывающая взаимосвязанный с множеством объект, мог получить такую оценку.

Ложноположительная вероятность оценки s — это вероятность того, что шум получит оценку равную или выше s .

Далее будут описаны метод, который позволяет эффективно вычислять введенный термин.

Определение

Пусть X_n и Y_n дискретные стохастические процессы, $n \geq 1$. Пара (X_n, Y_n) называется скрытой марковской моделью, если

- X_n — марковский процесс, поведение которого напрямую не наблюдается ("скрытый");
- $P(Y_n = y_n | X_1 = x_1, \dots, X_n = x_n) = P(Y_n | X_n = x_n)$ для любого $n \geq 1$, где x_1, \dots, x_n — значения, принимаемые процессом X_n (**состояния модели**), y_n — значение, принимаемое процессом Y_n (**наблюдаемый символ модели**).

Оценивание значимости выравнивания

— Обозначения и известные результаты

Обозначения и известные результаты

Определение

Пусть X_n и Y_n дискретные стохастические процессы, $n \geq 1$. Пара (X_n, Y_n) называется скрытой марковской моделью, если

- X_n — марковский процесс, поведение которого напрямую не наблюдается ("скрытый");
- $P(Y_n = y_n | X_1 = x_1, \dots, X_n = x_n) = P(Y_n | X_n = x_n)$ для любого $n \geq 1$, где x_1, \dots, x_n — значения, принимаемые процессом X_n (**состояния модели**), y_n — значение, принимаемое процессом Y_n (**наблюдаемый символ модели**).

Сначала опишем модели, затем алгоритмы, которые используются для манипуляции ими.

Метод предполагает, что даны профильная СММ, с помощью которой будут оцениваться последовательности, и фоновая модель B , которая будет описывать шум.

Определение

Пусть X_n и Y_n дискретные стохастические процессы, $n \geq 1$. Пара (X_n, Y_n) называется скрытой марковской моделью, если

- X_n — марковский процесс, поведение которого напрямую не наблюдается ("скрытый");
- $P(Y_n = y_n | X_1 = x_1, \dots, X_n = x_n) = P(Y_n | X_n = x_n)$ для любого $n \geq 1$, где x_1, \dots, x_n — значения, принимаемые процессом X_n (**состояния модели**), y_n — значение, принимаемое процессом Y_n (**наблюдаемый символ модели**).

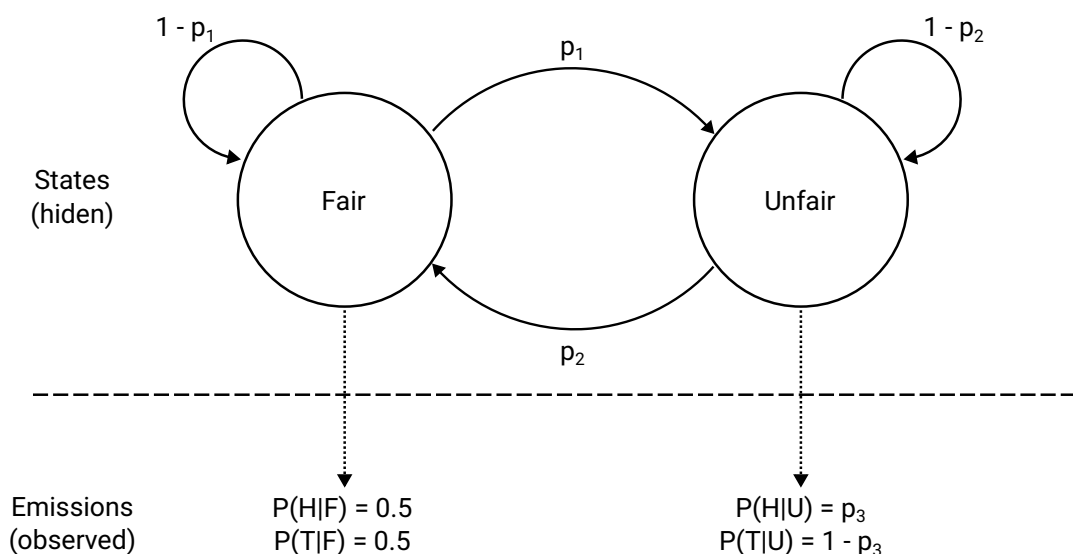
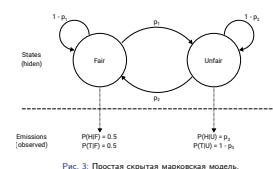


Рис. 3: Простая скрытая марковская модель.

Оценивание значимости выравнивания

Обозначения и известные результаты

Обозначения и известные результаты



Примером простой СММ может быть модель, изображенная на слайде и описывающая подбрасывание двух монет. Пусть между наблюдателем и человеком с монетами стоит ширма, которая позволяет наблюдателю видеть только пол, куда падают монеты. Пусть есть две монеты: одна — честная монета, вторая — нечестная монета с перевесом в одну из сторон. Пусть человек с монетами с некоторой вероятностью либо подбрасывает монету, которую он бросил в прошлый раз, либо меняет монеты и бросает новую. При этом наблюдатель не знает, какая монета используется в конкретный момент времени, так как он не видит рук бросающего монеты и не может отличить одну монету от другой по их внешнему виду, он видит только последовательность результатов бросков.

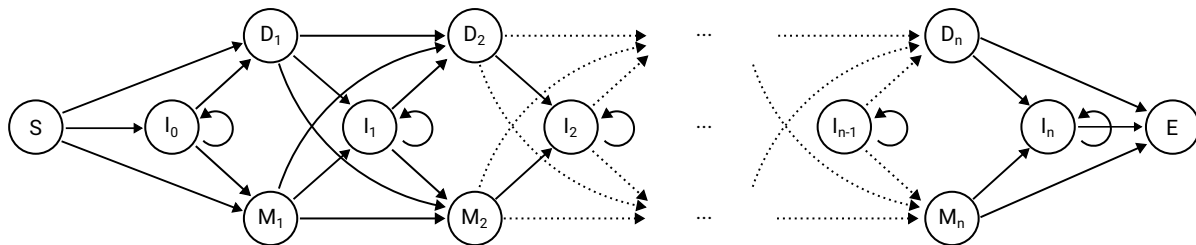


Рис. 4: Профильная скрытая марковская модель.

Оценивание значимости выравнивания

Обозначения и известные результаты

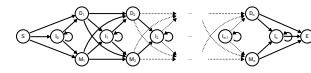


Рис. 4: Профильная скрытая марковская модель.

Профильная СММ — это СММ со специальной линейной архитектурой состояний, которая позволяет выравнивать последовательность к множеству последовательностей.

Если для удобства реализации алгоритмов добавить специальное *начальное* и специальное *конечное* состояния, в которых профильная СММ начинает и заканчивает работу и не испускает наблюдаемых символов, как показано на слайде, тогда *путь* π в профильной СММ начинается в начальном состоянии, заканчивается в конечном состоянии и проходит от состояния к состоянию, испуская в каждом состоянии наблюдаемый символ, то есть мы считаем, что путь π включает в себя и состояния, и наблюдаемые символы.

Последовательность D , связанная с путем π — последовательность наблюдаемых символов, которая была получена в результате прохода профильной СММ пути π .

Определение

Вероятность последовательности D может интерпретироваться и считаться по-разному — алгоритмом Витерби или Форвард алгоритмом.

$$s_{max}(D) = \max_{\pi \in \pi_D} (s(\pi));$$

$$s_{fw}(D) = \sum_{\pi \in \pi_D} s(\pi);$$

$$Z(D, T) = \sum_{\pi \in \pi_D} s(\pi)^{\frac{1}{T}}.$$

Оценивание значимости выравнивания

— Обозначения и известные результаты

Обозначения и известные результаты

Определение

Вероятность последовательности D может интерпретироваться и считаться по-разному — алгоритмом Витерби или Форвард алгоритмом.

$$s_{max}(D) = \max_{\pi \in \pi_D} (s(\pi));$$

$$s_{fw}(D) = \sum_{\pi \in \pi_D} s(\pi);$$

$$Z(D, T) = \sum_{\pi \in \pi_D} s(\pi)^{\frac{1}{T}}.$$

Вероятность пути $s(\pi)$ — произведение всех переходных вероятностей от состояний к состоянию и вероятностей наблюдаемых символов, которые излучаются в каждом состоянии, кроме начального и конечного, на протяжении всего пути π .

Вероятность последовательности D может интерпретироваться и считаться по-разному — алгоритмом Витерби или Форвард алгоритмом.

Вероятность Витерби $s_{max}(D)$ последовательности D — это максимальная вероятность последовательности D среди всех путей π , которые могли бы ее испустить:

$$s_{max}(D) = \max_{\pi \in \pi_D} (s(\pi)),$$

Несмотря на большое количество возможных путей, которые могли бы испустить последовательность D , алгоритм Витерби позволяет эффективно решать эту задачу.

Форвард вероятность $s_{fw}(D)$ последовательности D — это общая вероятность того, что в результате работы СММ будет получена последовательность D :

$$s_{fw}(D) = \sum_{\pi \in \pi_D} s(\pi).$$

Определение

Вероятность последовательности D может интерпретироваться и считаться по-разному — алгоритмом Витерби или Форвард алгоритмом.

$$s_{max}(D) = \max_{\pi \in \pi_D} (s(\pi));$$

$$s_{fw}(D) = \sum_{\pi \in \pi_D} s(\pi);$$

$$Z(D, T) = \sum_{\pi \in \pi_D} s(\pi)^{\frac{1}{T}}.$$

Оценивание значимости выравнивания

— Обозначения и известные результаты

Обозначения и известные результаты

Определение

Вероятность последовательности D может интерпретироваться и считаться по-разному — алгоритмом Витерби или Форвард алгоритмом.

$$s_{max}(D) = \max_{\pi \in \pi_D} (s(\pi));$$

$$s_{fw}(D) = \sum_{\pi \in \pi_D} s(\pi);$$

$$Z(D, T) = \sum_{\pi \in \pi_D} s(\pi)^{\frac{1}{T}}.$$

Форвард алгоритм работает за то же время, что и алгоритм Витерби. Третий способ оценивать последовательности, позволяющий уменьшить дисперсию дальнейших вычислений оценки ложноположительной вероятности оценки, заключается в том, что каждая вероятность перехода из одного состояния в другое и вероятность излучения символа состоянием будут возводиться в степень $\frac{1}{T}$, где $T \in (0; +\infty)$. При этом логика вычислений остается та же, то есть $s(\pi)^{\frac{1}{T}}$ и $s(D)^{\frac{1}{T}}$ будут вычисляться как вероятность произведения независимых событий и как сумма непересекающихся событий соответственно, хотя они уже могут не являться вероятностями (Например, сумма всех $s(\pi)^{\frac{1}{T}}$ не обязательно равна единице):

$$Z(D, T) = \sum_{\pi \in \pi_D} s(\pi)^{\frac{1}{T}}.$$

Функция $Z(D, T)$ называется статистической суммой и вычисляется через модификацию Форвард алгоритма. Параметра T подбирается экспериментально под конкретную интересующую оценку выравнивания.

Мы предполагаем наличие простой фоновой модели B для последовательностей длины L такой, что все L символьных позиций независимы и одинаково распределены в соответствии с некоторым распределением $P(d|B)$, где d отражает возможный наблюдаемый символ:

$$P(D|B) = \prod_{i=1}^L P(d_i|B),$$

где d_i — это i -ый наблюдаемый символ последовательности D .

Оценивание значимости выравнивания

└ Обозначения и известные результаты

Обозначения и известные результаты

Мы предполагаем наличие простой фоновой модели B для последовательностей длины L такой, что все L символьных позиций независимы и одинаково распределены в соответствии с некоторым распределением $P(d|B)$, где d отражает возможный наблюдаемый символ:

$$P(D|B) = \prod_{i=1}^L P(d_i|B),$$

где d_i — это i -ый наблюдаемый символ последовательности D .

Мы предполагаем наличие простой фоновой модели B для последовательностей длины L такой, что все L символьных позиций независимы и одинаково распределены в соответствии с некоторым распределением $P(d|B)$, где d отражает возможный наблюдаемый символ:

$$P(D|B) = \prod_{i=1}^L P(d_i|B),$$

где d_i — это i -ый наблюдаемый символ последовательности D .

Определение

Ложноположительная вероятность оценки s_0 для строк длины L :

$$fpr(s_0) = \sum_{D \in D_L} P(D|B) \Theta(s(D) \geq s_0), \quad (1)$$

где $P(D|B)$ — условная вероятность последовательности D , описываемая фоновой моделью, $s(D)$ — вероятность последовательности D , считаемая профильной СММ, и

$$\Theta(s(D) \geq s_0) = \begin{cases} 1, & s(D) \geq s_0 \\ 0, & s(D) < s_0 \end{cases}.$$

Оценивание значимости выравнивания

— Обозначения и известные результаты

Обозначения и известные результаты

Определение

Ложноположительная вероятность оценки s_0 для строк длины L :

$$fpr(s_0) = \sum_{D \in D_L} P(D|B) \Theta(s(D) \geq s_0), \quad (1)$$

где $P(D|B)$ — условная вероятность последовательности D , описываемая фоновой моделью, $s(D)$ — вероятность последовательности D , считаемая профильной СММ, и

$$\Theta(s(D) \geq s_0) = \begin{cases} 1, & s(D) \geq s_0 \\ 0, & s(D) < s_0 \end{cases}.$$

Вероятность последовательности D длины L сравнивается с остальными последовательностями той же длины. Определим ложноположительную вероятность оценки:

$$fpr(s_0) = \sum_{D \in D_L} Pr(D|B) \Theta(s(D) \geq s_0),$$

где $Pr(D|B)$ — условная вероятность последовательности D , описываемая фоновой моделью, $s(D)$ — вероятность последовательности D , считаемая профильной СММ, и

$$\Theta(s(D) \geq s_0) = \begin{cases} 1, & s(D) \geq s_0 \\ 0, & s(D) < s_0 \end{cases}.$$

То есть $fpr(s_0)$ — это вероятность того, что шум достигнет или превзойдет оценку s_0 . В определении $fpr(s_0)$ вероятность D отмечена как $s(D)$, потому что способ оценки последовательности может выбираться относительно интересующего приложения, подходит $s(D) = s_{max}(D)$ и $s(D) = s_{fw}(D)$.

Вычисление $fpr(s_0)$ через формулу (1) обычно неосуществимо, значение $fpr(s_0)$ может быть оценено через выборку по значимости.

Пусть $P(D|T)$ — это условная вероятность строки D относительно некоторой модели строк длины L параметризованной значением T . Тогда можно переписать $fpr(s_0)$:

$$fpr(s_0) = \sum_{D \in D_L} P(D|T) f(D, s_0),$$

где

$$f(D, s_0) = \frac{P(D|B) \Theta(s(D) \geq s_0)}{P(D|T)}. \quad (2)$$

Оценивание значимости выравнивания

Обозначения и известные результаты

Обозначения и известные результаты

Вычисление $fpr(s_0)$ через формулу (1) обычно неосуществимо, значение $fpr(s_0)$ может быть оценено через выборку по значимости.

Пусть $P(D|T)$ — это условная вероятность строки D относительно некоторой модели строк длины L параметризованной значением T . Тогда можно переписать $fpr(s_0)$:

$$fpr(s_0) = \sum_{D \in D_L} P(D|T) f(D, s_0),$$

где

$$f(D, s_0) = \frac{P(D|B) \Theta(s(D) \geq s_0)}{P(D|T)}. \quad (2)$$

Так как вычисление $fpr(s_0)$ через формулу 5 обычно неосуществимо, значение $fpr(s_0)$ может быть оценено через выборку по значимости, то есть через моделирование строк в соответствии с фоновой моделью B и оценивание значения $fpr(s_0)$ долей тех из них, что достигают оценки s_0 .

Построим распределение, относительно которого будем моделировать строки. Пусть $P(D|T)$ — это условная вероятность строки D относительно некоторой модели строк длины L параметризованной значением T . Тогда можно переписать $fpr(s_0)$:

$$fpr(s_0) = \sum_{D \in D_L} Pr(D|T) f(D, s_0),$$

где

$$f(D, s_0) = \frac{Pr(D|B) \Theta(s(D) \geq s_0)}{Pr(D|T)}.$$

Мы можем оценить значение $fpr(s_0)$ через моделирование последовательностей в соответствии с этой альтернативной моделью и подсчет среднего значения $f(D, s_0)$. Этот подход и называется *выборкой по значимости*, он полезен, потому что если правильно подобрать альтернативную модель, то удастся уменьшить дисперсию оценки $fpr(s_0)$.

Определим модель, используемую для выборки по важности параметризованную T :

$$P(D|T) = \frac{P(D|B)Z(D, T)}{Z(T)},$$

где

$$Z(T) = \sum_{D \in D_L} P(D|B)Z(D, T).$$

Подставив определение $P(D, T)$ в уравнение (2) получим

$$f(D, s_0) = \frac{Z(T)\Theta(s(D) \geq s_0)}{Z(D, T)}.$$

Оценивание значимости выравнивания

— Обозначения и известные результаты

Обозначения и известные результаты

Определим модель, используемую для выборки по важности параметризованную T :

$$P(D|T) = \frac{P(D|B)Z(D, T)}{Z(T)},$$

где

$$Z(T) = \sum_{D \in D_L} P(D|B)Z(D, T).$$

Подставив определение $P(D, T)$ в уравнение (2) получим

$$f(D, s_0) = \frac{Z(T)\Theta(s(D) \geq s_0)}{Z(D, T)}.$$

Определим модель, используемую для выборки по важности параметризованную T следующим образом:

$$Pr(D|T) = \frac{P(D|B)Z(D, T)}{Z(T)},$$

где

$$Z(T) = \sum_{D \in D_L} Pr(D|B)Z(D, T).$$

Подставив определение $Pr(D, T)$ в уравнение 7 получим

$$f(D, s_0) = \frac{Z(T)\Theta(s(D) \geq s_0)}{Z(D, T)}.$$

В итоге мы хотим смоделировать последовательности в соответствии с распределением $Pr(D|T)$, вычислить $f(D, s_0)$ для каждой последовательности и использовать среднее этих значений как оценку $fpr(s_0)$.

Опуская подробности того как устроены моделирование, построенное на модификациях классических алгоритмов, связанных с НММ, и метод подбора параметра T , перейдем к полученным результатам.

Вычислим оценку $\widehat{fpr}(s_0)$ для строк длины $L = 100$, состоящих из 5 символов, и доверительные интервалы уровня $\gamma = 0.99$:

Таблица 1: Результаты.

s_0	T	$\widehat{fpr}(s_0)$	$[c_1(\gamma); c_2(\gamma)]$
10^{-85}	7	0.0000000183	[0.0; 0.00066349]
10^{-90}	7	0.003175	[0.001884; 0.004779]
10^{-100}	7	0.615709	[0.597540 0.622677]

Оценивание значимости выравнивания

Полученные результаты

Полученные результаты

Вычислим оценку $\widehat{fpr}(s_0)$ для строк длины $L = 100$, состоящих из 5 символов, и доверительные интервалы уровня $\gamma = 0.99$:

Таблица 1: Результаты.

s_0	T	$\widehat{fpr}(s_0)$	$[c_1(\gamma); c_2(\gamma)]$
10^{-85}	7	0.0000000183	[0.0; 0.00066349]
10^{-90}	7	0.003175	[0.001884; 0.004779]
10^{-100}	7	0.615709	[0.597540 0.622677]

Вычислим оценку $\widehat{fpr}(s_0)$ для строк длины $L = 100$ и доверительные интервалы уровня $\gamma = 0.99$:

s_0	T	$\widehat{fpr}(s_0)$	$[c_1(\gamma); c_2(\gamma)]$
10^{-85}	7	0.0000000183	[0.0; 0.00066349]
10^{-90}	7	0.003175	[0.001884; 0.004779]
10^{-100}	7	0.615709	[0.597540 0.622677]

Вычисление $\widehat{fpr}(s_0)$ перебором привело бы к перебору 5^{100} строк, что неосуществимо, если бы мы вычисляли перебором всех строк длины L .

- Была изучена тема алгоритмов парного и множественного выравнивания последовательностей и тема СММ и алгоритмов взаимодействия с ними.
- Был реализован алгоритм, позволяющий эффективно вычислять оценку $fpr(s_0)$.
- Предстоит подробно верифицировать реализованный алгоритм и сравнить его с имеющимися методами вычисления оценки $fpr(s_0)$.





Оценивание значимости выравнивания

└ Заключение

Заключение

- Была изучена тема алгоритмов парного и множественного выравнивания последовательностей и тема СММ и алгоритмов взаимодействия с ними.
- Был реализован алгоритм, позволяющий эффективно вычислять оценку $fpr(s_0)$.
- Предстоит подробно верифицировать реализованный алгоритм и сравнить его с имеющимися методами вычисления оценки $fpr(s_0)$.





- Была изучена тема алгоритмов парного и множественного выравнивания последовательностей и тема СММ и алгоритмов взаимодействия с ними.
- Был реализован алгоритм, позволяющий эффективно вычислять оценку $fpr(s_0)$.
- Предстоит подробно верифицировать реализованный алгоритм и сравнить его с имеющимися методами вычисления оценки $fpr(s_0)$.

-  Newberg Lee A. Error statistics of hidden Markov model and hidden Boltzmann model results // MC Bioinformatics. — 2009.
-  Compeau Phillip, Pevzner Pavel. How do we compare DNA sequences // Bioinformatics Algorithms: An Active Learning Approach, 2nd Ed. Vol. 1. — Active Learning Publishers, 2015.
-  Compeau Phillip, Pevzner Pavel. Why have biologists still not developed an HIV vaccine // Bioinformatics Algorithms: An Active Learning Approach, 2nd Ed. Vol. 2. — Active Learning Publishers, 2015.
-  A tutorial on Hidden Markov Models : Rep. / Signal Processing and Artificial Neural Networks Laboratory Department of Electrical Engineering Indian Institute of Technology ; Executor: Rakesh Dugad, U. B. Desai : 1996.

Оценивание значимости выравнивания

— Список литературы

Список литературы

-  Newberg Lee A. Error statistics of hidden Markov model and hidden Boltzmann model results // MC Bioinformatics. — 2009.
-  Compeau Phillip, Pevzner Pavel. How do we compare DNA sequences // Bioinformatics Algorithms: An Active Learning Approach, 2nd Ed. Vol. 1. — Active Learning Publishers, 2015.
-  Compeau Phillip, Pevzner Pavel. Why have biologists still not developed an HIV vaccine // Bioinformatics Algorithms: An Active Learning Approach, 2nd Ed. Vol. 2. — Active Learning Publishers, 2015.
-  A tutorial on Hidden Markov Models : Rep. / Signal Processing and Artificial Neural Networks Laboratory Department of Electrical Engineering Indian Institute of Technology ; Executor: Rakesh Dugad, U. B. Desai : 1996.

На данном слайде представлен список основных источников, используемых в моей работе.