



UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Documentation – Workshop 1

Juan Sebastian Bedoya Serna

2023102029

Faculty of engineering; Universidad Distrital Francisco Jose De Caldas

System's Analysis

Carlos Andrés Sierra Virgüez

September 14 – 2024

MOTIFFINDER CLASS

The MotifFinder class is a Java program that generates random DNA sequences, finds motifs (recurring patterns) in these sequences, and calculates the entropy of each sequence.

MAIN METHOD

The main method is the entry point of the program. It generates random DNA sequences, saves them to a file, finds motifs in these sequences, and calculates the entropy of each sequence.

Parameters

- **n**: The number of sequences to generate (randomly chosen between 1000 and 2000000).
- **m**: The length of each sequence (randomly chosen between 5 and 100).
- **probabilities**: An array of probabilities for each base (A, C, G, and T) in the sequence.
- **s**: The length of the motif to find (randomly chosen between 4 and 10).

Steps

1. Generate **n** random DNA sequences of length **m** using the **generateSequences** method.
2. Save the generated sequences to a file named "dataBaseSequences.txt" using the **saveSequencesInTxt** method.
3. Find motifs of length **s** in the generated sequences using the **findMotif** method.
4. Calculate the entropy of each sequence using the **calculateEntropy** method.
5. Print the motifs found and the sequence with the highest entropy.

GENERATESEQUENCES METHOD

Generates **n** random DNA sequences of length **m** using the provided probabilities.

Parameters

- **n**: The number of sequences to generate.
- **m**: The length of each sequence.

- **probabilities:** An array of probabilities for each base (A, C, G, and T) in the sequence.

Return

- An ArrayList of generated DNA sequences.

GETINDEXBASE METHOD

Returns a random index based on the provided probabilities.

Parameters

- **random:** A Random object used to generate a random number.
- **probabilities:** An array of probabilities for each base (A, C, G, and T) in the sequence.

Return

- A random index (0, 1, 2, or 3) corresponding to the base (A, C, G, or T).

GETBASE METHOD

Returns the base corresponding to the provided index.

Parameters

- **baseIndex:** The index of the base (0, 1, 2, or 3).

Return

- The base (A, C, G, or T) corresponding to the provided index.

SAVESEQUENCESINTXT METHOD

Saves the provided sequences to a file.

Parameters

- **database:** An ArrayList of DNA sequences to save.
- **fileName:** The name of the file to save the sequences to.

Throws

- **IOException:** If an error occurs while writing to the file.

LOADSEQUENCESFROMTXT METHOD

Loads DNA sequences from a file.

Parameters

- **fileName:** The name of the file to load the sequences from.

Return

- An ArrayList of loaded DNA sequences.

Throws

- **IOException:** If an error occurs while reading from the file.

FINDMOTIF METHOD

Finds motifs of length *s* in the provided sequences.

Parameters

- **database:** An ArrayList of DNA sequences to find motifs in.
- **s:** The length of the motif to find.

Return

- An ArrayList of found motifs.

GETMAXCOUNT METHOD

Returns the maximum count of a motif in the provided HashMap.

Parameters

- **motifCounts:** A HashMap of motif counts.

Return

- The maximum count of a motif.

CALCULATEENTROPYMETHOD

Calculates the entropy of a DNA sequence.

Parameters

- **sequence:** The DNA sequence to calculate the entropy of.

Return

- The entropy of the sequence.

PERFORMEXPERIMENTS CLASS

The PerformExperiments class is a Java program that performs experiments to measure the performance of the MotifFinder class.

MAIN METHOD

The main method is the entry point of the program. It performs experiments with different parameters and saves the results to a CSV file.

Parameters

- **n:** The number of sequences in the database (varies between 1000, 5000, and 10000).
- **m:** The length of each sequence (varies between 50 and 100).
- **probabilities:** The probabilities of each base (A, C, G, and T) in the sequence (varies between {0.25, 0.25, 0.25, 0.25} and {0.3, 0.2, 0.3, 0.2}).
- **s:** The length of the motif to find (varies between 4, 5, and 6).

Steps

1. Generate a database of random DNA sequences using the **MotifFinder.generateSequences** method.
2. Measure the time it takes to find motifs in the database using the **MotifFinder.findMotif** method.
3. Count the occurrences of each motif in the database using the **getMotifCounts** method.
4. Save the results to a CSV file using the **saveResultsToCSV** method.

EXPERIMENTRESULT CLASS

Represents the result of an experiment, including the database size, sequence length, probabilities, motif size, motifs, motif counts, and elapsed time.

Fields

- **dataBaseSize:** The number of sequences in the database.
- **sequenceLength:** The length of each sequence.
- **probabilities:** The probabilities of each base (A, C, G, and T) in the sequence.
- **motifSize:** The length of the motif to find.
- **motifs:** The found motifs.
- **motifCounts:** The counts of each motif.
- **elapsedTime:** The time it took to find the motifs.

GETMOTIFCOUNTS METHOD

Returns an array of motif counts for the given database and motifs.

Parameters

- **database:** The database of DNA sequences.
- **motifs:** The found motifs.

Return

- An array of motif counts.

COUNTMOTIFOCCURRENCES METHOD

Returns the count of a motif in a database.

Parameters

- **database:** The database of DNA sequences.
- **motif:** The motif to count.

Return

- The count of the motif in the database.

COUNTMOTIFOCURRENCESINSEQUENCE METHOD

Returns the count of a motif in a sequence.

Parameters

- **sequence:** The DNA sequence.
- **motif:** The motif to count.

Return

- The count of the motif in the sequence.

SAVERESULTSTOCSV METHOD

Saves the experiment results to a CSV file.

Parameters

- **results:** The list of experiment results.
- **filename:** The name of the file to save the results to.

Throws

- **IOException:** If an error occurs while writing to the file.

ARRAYTOSTRING METHOD

Converts an array to a string.

Parameters

- **array:** The array to convert.

Return

- The string representation of the array.

LISTTOSTRING METHOD

Converts a list to a string.

Parameters

- **list:** The list to convert.

Return

- The string representation of the list.