Genomics Paper

String Matching Algorithm (BM and KMP improved ⇒ CBM)

Abstract and Introduction :

- **what is string matching?**

  we can describe it simply in : it's the case of searching a string within another string.

- we will most be talking more about **exact matching which can be defined by :**

  Exact string matching algorithms is to find one, several, or all occurrences of a defined string (pattern) in a large string (text or sequences) such that each matching is perfect (so it's based on character comparison like BM and KMP).

- **what can we use it for?**

  String matching is widely used in data compression, text editing...etc.

  Therefore it's considered a basic component in many operating systems.

  but most important is that genes are made of group of bases which can be formulated as a large string, when we say large we mean it, to give you a glimpse of how large it could be, the human genome is more than 3 BILLION base pair so string matching is considered the base of bioinformatics, genetic engineering.

so designing an efficient and effective string matching algorithm will not only be useful for pattern matching but also benefits relevant application's development.

- in this paper, we give analysis to several classical algorithms such as : BM, KMP and their improvements.

- by composing the main method of the BM algorithm we propose a new algorithm : The Composite BM algorithm (CBM).

the advantages of CBM :

- BM algorithm only uses current matching information so that limits it's speed and efficiency.

- CBM tries to make use of the historical matching information.

- so CBM speed up the matching process effectively.

- we will explain how CBM is faster and more efficient.