

باسمه تعالی

data structure

phase2

علی طاهری نستوه 9433597

احسان وخشوری 9433587

تابستان 96

الگوریتم

مقدمه

به طور کلی این الگوریتم وظیفه تبدیل یک سری زمانی را به یک رشته از string ها را دارد که این رشته میتواند طول های متفاوت و از حروف متفاوت استفاده کرد. و این رشته را میتوان نماینده ای از این سری زمانی دانست و کاربرد های متنوع میتواند داشته باشد.

نحوه عملکرد

در بطن این الگوریتم از الگوریتم PPA استفاده میشود ابتدا توضیح اجمالی در مورد این الگوریتم میدهیم و در ادامه الگوریتم SAX را تشریح میکنیم .

الگوریتم SAX

ابتدا ما از فاصله اقلیدسی برای تعیین فاصله دو نقطه استفاده میکنیم که با استفاده از فرمول زیر و جمع توان دو آنها (برای اینکه فاصله مثبت و منفی با هم خنثی نشوند) و با رادیکال گرفتن از اونها در واقع فاصله بین دو سری زمانی را محاسبه میکنیم.

Given two time series

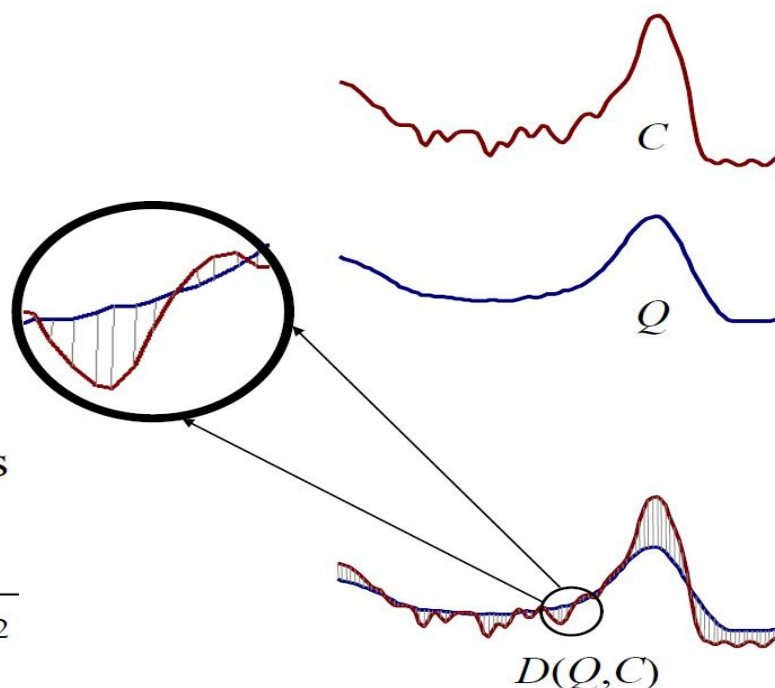
$$Q = q_1 \dots q_n$$

and

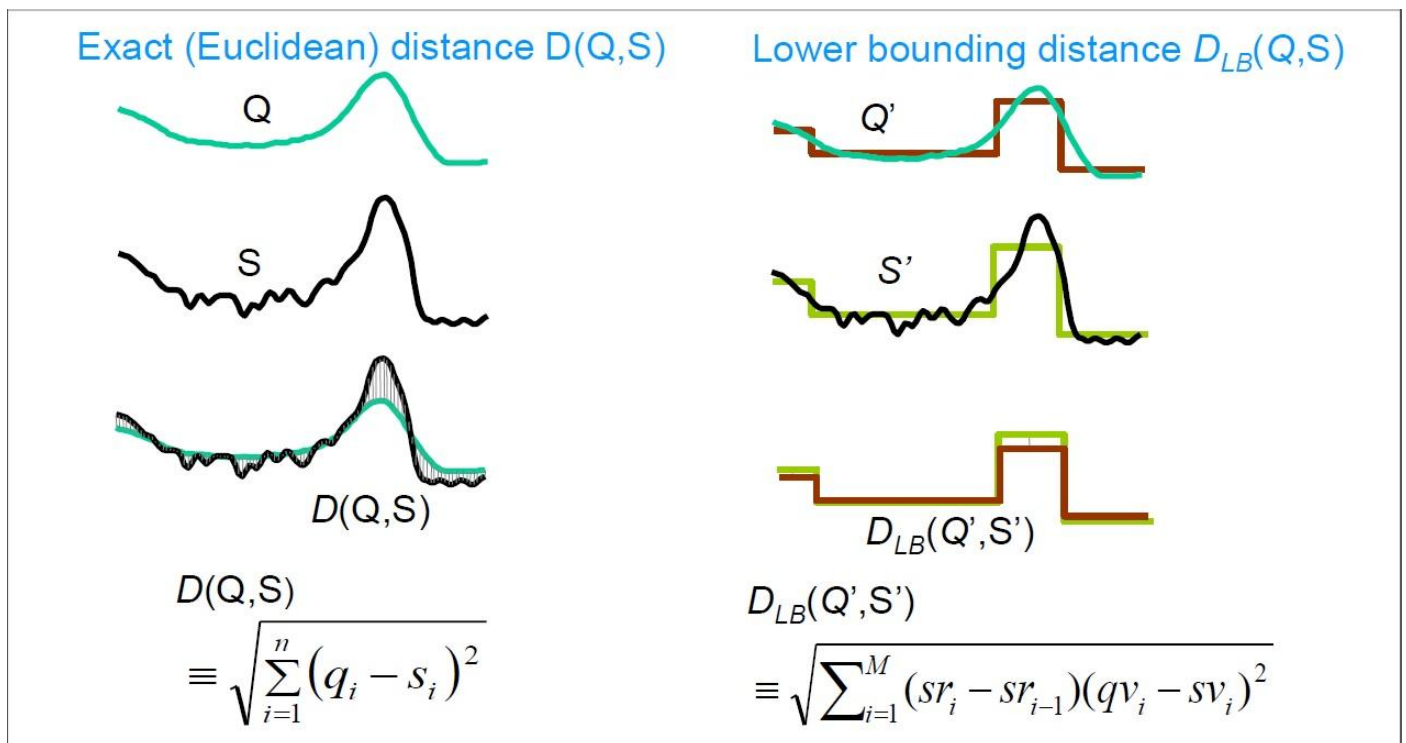
$$C = c_1 \dots c_n$$

their Euclidean distance is defined as:

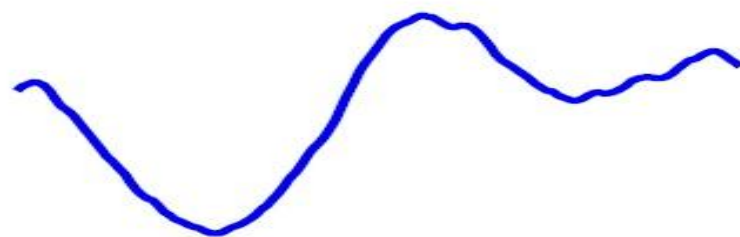
$$D(Q, C) \equiv \sqrt{\sum_{i=1}^n (q_i - c_i)^2}$$



و بعد از مفهومی به نام LOWER BOUND استفاده میکنیم که در واقع اعداد اعشاری را به اعداد صحیح کمتر از خود تبدیل میکند

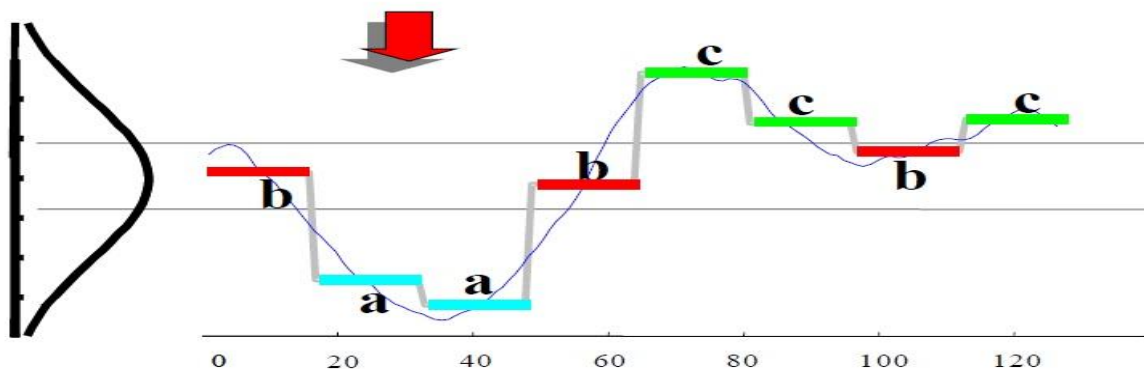
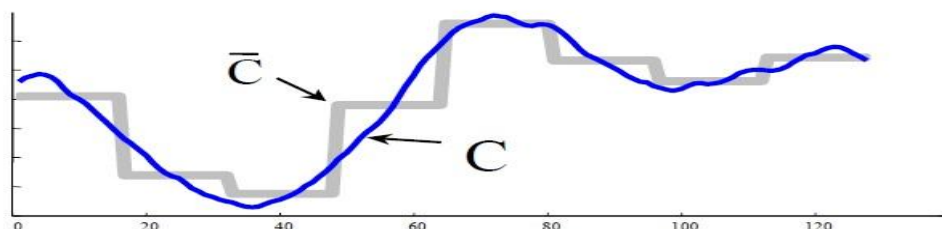
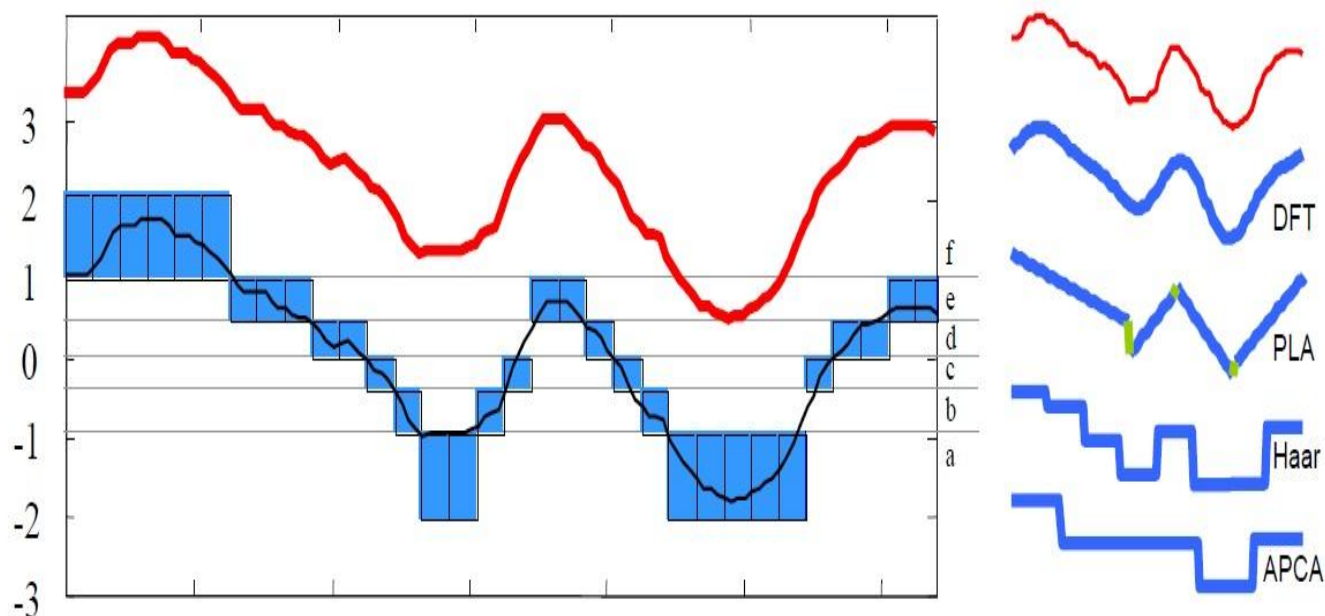


الگوریتم SAX



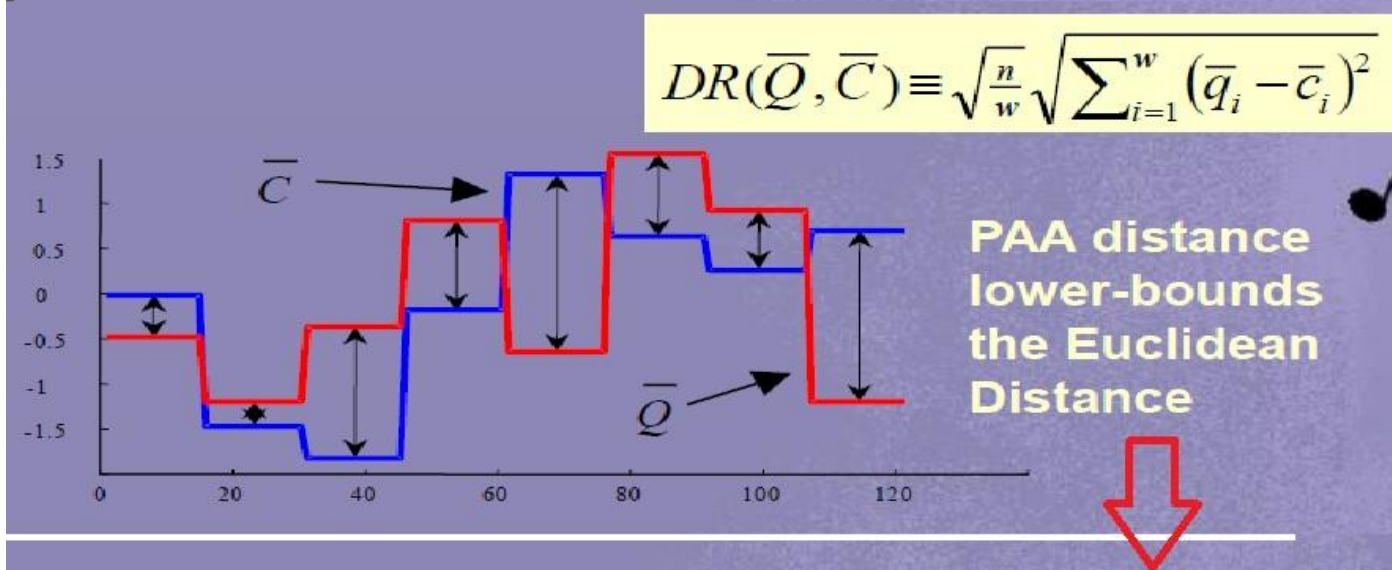
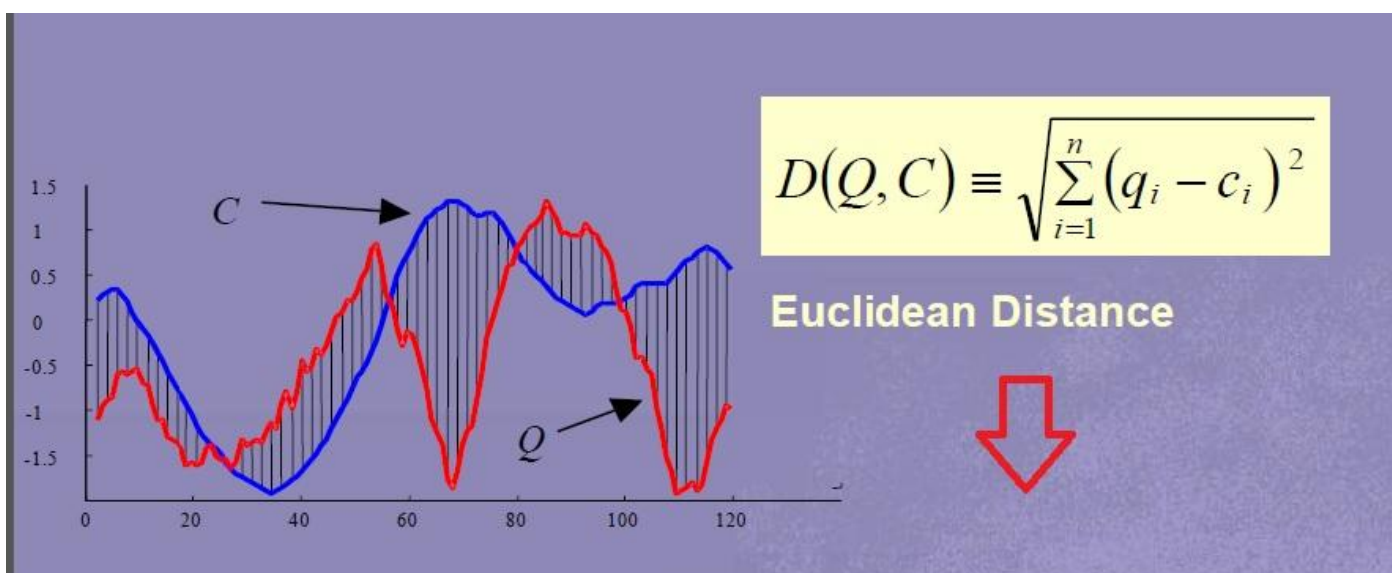
baabccbc

ابتدا سری زمانی را با استفاده از الگوریتم PPA تبدیل به LOWER BOUND میکنیم و بعد LOWER BOUND را با استفاده از SAX به رشته تبدیل میکنیم



baabccbc

زیر دنباله از سری زمانی در واقع تمایل زیادی به تابع نرمال خود دارد و وقتی سری را با تابع نرمال خود با استفاده از فاصله گوسی و PPA تقریب میزنیم و به اندازه های که میخواهیم طول string باشد ناحیه زیر تابع نرمال را به همان اندازه تقسیم میکنیم و قاعدتا اگر تعداد حروف زیاد بشوند ما برای تقسیم بندی نیاز به اعشار هم داریم



$\hat{C} = \text{baabccbc}$
 $\hat{Q} = \text{babacca}$

$MINDIST(\hat{Q}, \hat{C}) \equiv \sqrt{\frac{n}{w}} \sqrt{\sum_{i=1}^w (dist(\hat{q}_i, \hat{c}_i))^2}$

dist() can be implemented using a table lookup.

Output of our file

We search first string in data set

Data of: 6

Build time: 751 microSecond

Find time: 4 microSecond

Ans: 3

Data of: 7

Build time: 951 microSecond

Find time: 5 microSecond

Ans: 1

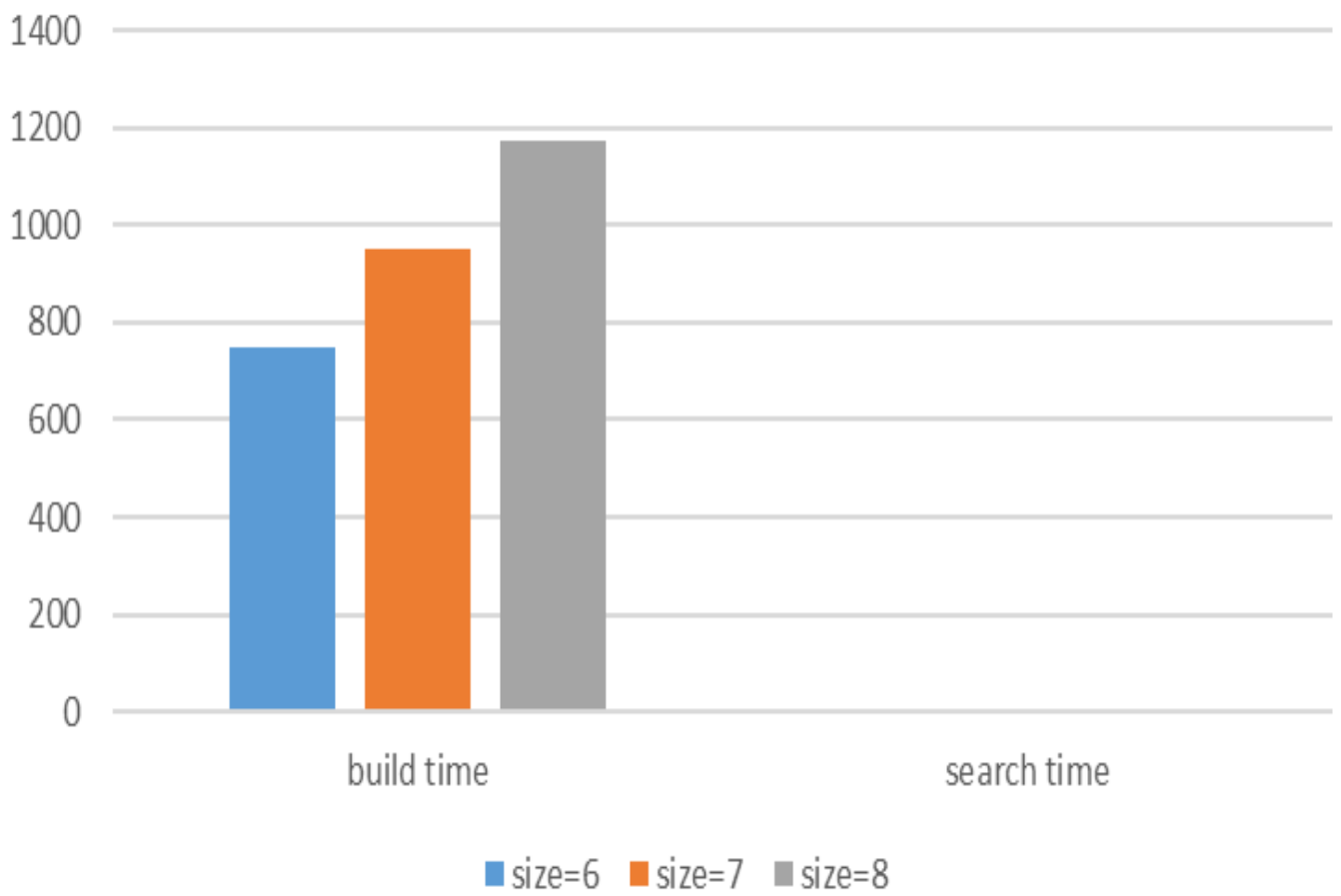
Data of: 8

Build time: 1176 microSecond

Find time: 5 microSecond

Ans: 1

Chart Title



Wikipedia

<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=9&cad=rja&uact=8&ved=0ahUKEwj0jfiyturUAhXMfRoKHeU-C48QFghUMAg&url=http%3A%2F%2Fwww.cs.utsa.edu%2F~qitian%2FCS7123%2Fslides%2F08-29-2011%2Ftalk2-08-29-2011.pptx&usg=AFQjCNG59CQyLZquJ69JpCMNwQ3ZzosDCA>