asked 6 years, 10 months ago

viewed 2,043 times

active 6 years, 10 months ago

1

1

i am trying to build a procedure to obtain k nearest neighbor points to a point with a selected ID. I need to do this without using any spatial locator features like sdo_geometry or nn.

basically i have a table in oracle with ID, Data_X, Data_Y. let's say i have 10 entries in my table, and i need the 3 nearest points to a fictional point target_x, target_y.

we would need to calculate the euclidean distance of each point in the table with my given fictional point. I just dont know an algorithm in pl/sql that would return me the nearest neighbor ids.

algorithm oracle plsql nearest-neighbor

share improve this question

asked Apr 4 '11 at 12:18
nackif99
15 ● 5

add a comment

2 Answers

active oldest votes

3

✓

Calculate the distance (Pythagoras) between each point and the selected point and order by the distance. Pseudo sql:

```
select id from points
order by sqrt(sqr(Data_x - target_x) + sqr(Data_y - target_y))
```

The first 3 rows are the the nearest 3 points.

share improve this answer

answered Apr 4 '11 at 12:38
nang
286 ● 4 ● 14

thanks! ill give this a go – nackif99 Apr 4 '11 at 13:13

add a comment

2

nang's answer is a great starting point, and if it does the job, I'd use it. Unfortunately it will most probably require a full table scan (or perhaps a full index scan, if you have a covering index).

If performance becomes a problem, you could probably have a look at making a poor-man's spatial index over the data. It won't be as simple as "create index" but it just might work.

The proper method would be to create a custom index, but that would just be reinventing the sdo_geometry wheel, a path you said you wish to avoid.

A simple-but-rough method (disclaimer: this is just an idea off the top of my head, not tested) might be to create a function-based index that groups all the points in 2D space into square blocks. You basically create an index to map each (x,y) pair onto a list of blocks. Each block would have a defined width and height, and to do a search you would first work out which grid of blocks need to be searched, then query across just the list of points in that grid.

An example index would be something like:

Related

118 Nearest neighbors in high-dimensional data?

7 how to find k-th nearest neighbor of a point in a set of point

11 K-Nearest Neighbor Query in PostGIS

2 Transforming a table of Nearest Neighbor Distances into a Matrix

0 K Nearest Neighbors classification Special Case with Identical Points

3 Travel from a point A to all remaining points via nearest neighbor point first

8 Remove points to maximize shortest nearest neighbor distance

-1 K-Nearest Neighbor Implementation for Strings (Unstructured data) in Java

0

```
      where TRUNC(Data_X/100) BETWEEN TRUNC(:target_x/100)) - :threshold
                                  AND TRUNC(:target_x/100)) + :threshold
      and   TRUNC(Data_Y/100) BETWEEN TRUNC(:target_y/100)) - :threshold
                                  AND TRUNC(:target_y/100)) + :threshold
     )
order by sqrt(sqr(Data_x - :target_x) + sqr(Data_y - :target_y))
```

You can then set :threshold to basically eliminate a large set of blocks of points from the query. I reckon if the values for the functional index (i.e. 100) and the threshold are set correctly, you'll see the query use the function-based index to get a small set of candidates, instead of calculating the distance for every single point in the table.

The downside is that if :threshold is too low, the query might return no rows. On the other hand, that might be a useful feature, depending on your needs.

share improve this answer

answered Apr 4 '11 at 13:55

Jeffrey Kemp
43.8k ● 10 ● 75 ● 117

1

It's still probably more efficient to do a binary back-off of the threshold until you get the k nearest neighbors than to calculate the distance to every other point every time. In other words, if you don't get enough points with a threshold of 1, try 2, 4, 8, 16, etc. until you get k rows returned. If you do this, however, be aware that you may need to go one step beyond the point where you get k rows to ensure that you get the k closest points. If the distance to the k-th closest point in a particular iteration is > `threshold` , there could be closer points outside the bounding box. – Justin Cave Apr 4 '11 at 15:24

add a comment

Your Answer

Sign up or log in

Sign up using Google

Sign up using Facebook

Sign up using Email and Password

Post as a guest

Name

Email

required, but never shown

Post Your Answer

By posting your answer, you agree to the privacy policy and terms of service.

Not the answer you're looking for? Browse other questions tagged algorithm oracle plsql nearest-neighbor or ask your own question.

How to describe an everyday routine?

Terminology for 24 hour wake-sleep cycle on world with longer dark-light cycle?

Why was Dolores Umbridge not arrested after the end of Harry's fifth year?

Mathematical ideas that took long to define rigorously

Is there a reason for heavy doors in hotels?

more hot questions

question feed

STACK OVERFLOW

Questions

Jobs

Developer Jobs Directory

Salary Calculator

- Help
- Mobile

STACK OVERFLOW BUSINESS

- Talent
- Ads
- Enterprise

COMPANY

- About
- Press
- Work Here
- Legal
- Privacy Policy
- Contact Us

STACK EXCHANGE NETWORK

- Technology
- Life / Arts
- Culture / Recreation
- Science
- Other

Blog • Facebook • Twitter • LinkedIn