

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐH SƯ PHẠM KỸ THUẬT TP.HCM
KHOA ĐÀO TẠO CHẤT LƯỢNG CAO



BÁO CÁO ĐỒ ÁN 1

***ĐỀ TÀI : TÌM HIỂU VÀ ỨNG DỤNG THUẬT TOÁN K-NEAREST
NEIGHBORS***

Sinh viên thực hiện: Nguyễn Tiến Đạt - 16110048

Nguyễn Văn Đạt - 16110050

Giáo viên hướng dẫn: Thầy Lê Văn Vinh

TP.HCM, tháng 11 năm 2018

TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM

KHOA ĐÀO TẠO CHẤT LƯỢNG CAO

NHIỆM VỤ THỰC HIỆN ĐỒ ÁN 1

Họ tên SV : Nguyễn Tiến Đạt

MSSV : 16110048

Họ tên SV : Nguyễn Văn Đạt

MSSV : 16110050

Ngành : Công nghệ thông tin

Tên đề tài : Tìm hiểu và ứng dụng thuật toán k-nearest neighbors

Nội dung thực hiện :

- Tìm hiểu và cài đặt thuật toán k-nearest neighbors
- Demo ví dụ đơn giản về thuật toán k-nearest neighbors

Thời gian thực hiện từđến

TP.HCM, tháng 11 năm 2018

NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

TP.HCM, Ngày..... tháng 11 năm 2018

Trưởng khoa

(ký ghi rõ họ tên)

Giáo viên hướng dẫn

(ký ghi rõ họ tên)

LỜI CẢM ƠN

Chúng em xin cảm ơn khoa Đào tạo Chất lượng cao trường Đại học Sư Phạm Kỹ Thuật TP.HCM đã tạo điều kiện thuận lợi cho nhóm em thực hiện đề tài này.

Chúng em xin được gửi lời cảm ơn đến **thầy Lê Văn Vinh** đã hướng dẫn nhóm em thực hiện đề tài này.

Bên cạnh đó, nhóm em xin gửi lời cảm ơn đến các thầy cô, những người đã giảng dạy, tạo điều kiện cho em tích lũy được những kiến thức quý báu trong những năm học qua.

Do thời gian gấp rút nên đề án còn nhiều thiếu sót không thể tránh khỏi mong sự thông cảm từ các thầy cô.

TP.HCM, tháng 11 năm 2018

Nhóm sinh viên thực hiện

Nguyễn Tiến Đạt

Nguyễn Văn Đạt

MỤC LỤC

DANH MỤC HÌNH ẢNH	6
DANH MỤC BẢNG	7
PHẦN 1: TÌM HIỂU THUẬT TOÁN K-NEAREST NEIGHBOR	8
<i>I. GIỚI THIỆU THUẬT TOÁN K-NEAREST NEIGHBOR</i>	<i>8</i>
<i>II. MÔ TẢ THUẬT TOÁN K-NEAREST NEIGHBOR</i>	<i>8</i>
☉ Thuật toán K-NN được mô tả như sau:	8
<i>III. VÍ DỤ MINH HOẠ THUẬT TOÁN K-NEAREST NEIGHBOR</i>	<i>9</i>
PHẦN 2: CÀI ĐẶT THUẬT TOÁN K-NEAREST NEIGHBOR	11
<i>I. MÔI TRƯỜNG LẬP TRÌNH</i>	<i>11</i>
1. NGÔN NGỮ LẬP TRÌNH	11
2. CÔNG CỤ LẬP TRÌNH	11
3. MÔI TRƯỜNG LẬP TRÌNH	11
<i>II. CÀI ĐẶT THUẬT TOÁN K-NEAREST NEIGHBOR</i>	<i>12</i>
PHẦN 3: ĐÁNH GIÁ VÀ PHÂN CÔNG	15
<i>I. PHÂN CÔNG</i>	<i>15</i>
<i>II. ĐÁNH GIÁ</i>	<i>15</i>
1. Ưu điểm:	15
2. Nhược điểm:	15
PHẦN 4: TÀI LIỆU THAM KHẢO	16
<i>I. Tham khảo từ thầy.</i>	<i>16</i>
<i>II. Tham khảo từ Internet.</i>	<i>16</i>

DANH MỤC HÌNH ẢNH

[Hình 1.1 Ví dụ minh họa với \$k=5\$](#) 9

[Hình 1.2 Ví dụ minh họa với \$k=3\$](#) 10

DANH MỤC BẢNG

Bảng 1: Lớp	12
Bảng 2: Phương thức.....	13
Bảng 3 : Phân công.....	15

PHẦN 1: TÌM HIỂU THUẬT TOÁN K-NEAREST NEIGHBOR

I. GIỚI THIỆU THUẬT TOÁN K-NEAREST NEIGHBOR

- K-Nearest Neighbors algorithm (K-NN) được sử dụng rất phổ biến trong lĩnh vực Data Mining. K-NN là phương pháp để phân lớp các đối tượng dựa vào khoảng cách gần nhất giữa đối tượng cần xếp lớp (Query point) và tất cả các đối tượng trong Training Data. (1)(3)
- Một đối tượng được phân lớp dựa vào K láng giềng của nó. K là số nguyên dương được xác định trước khi thực hiện thuật toán. Người ta thường dùng khoảng cách Euclidean để tính khoảng cách giữa các đối tượng.(2)

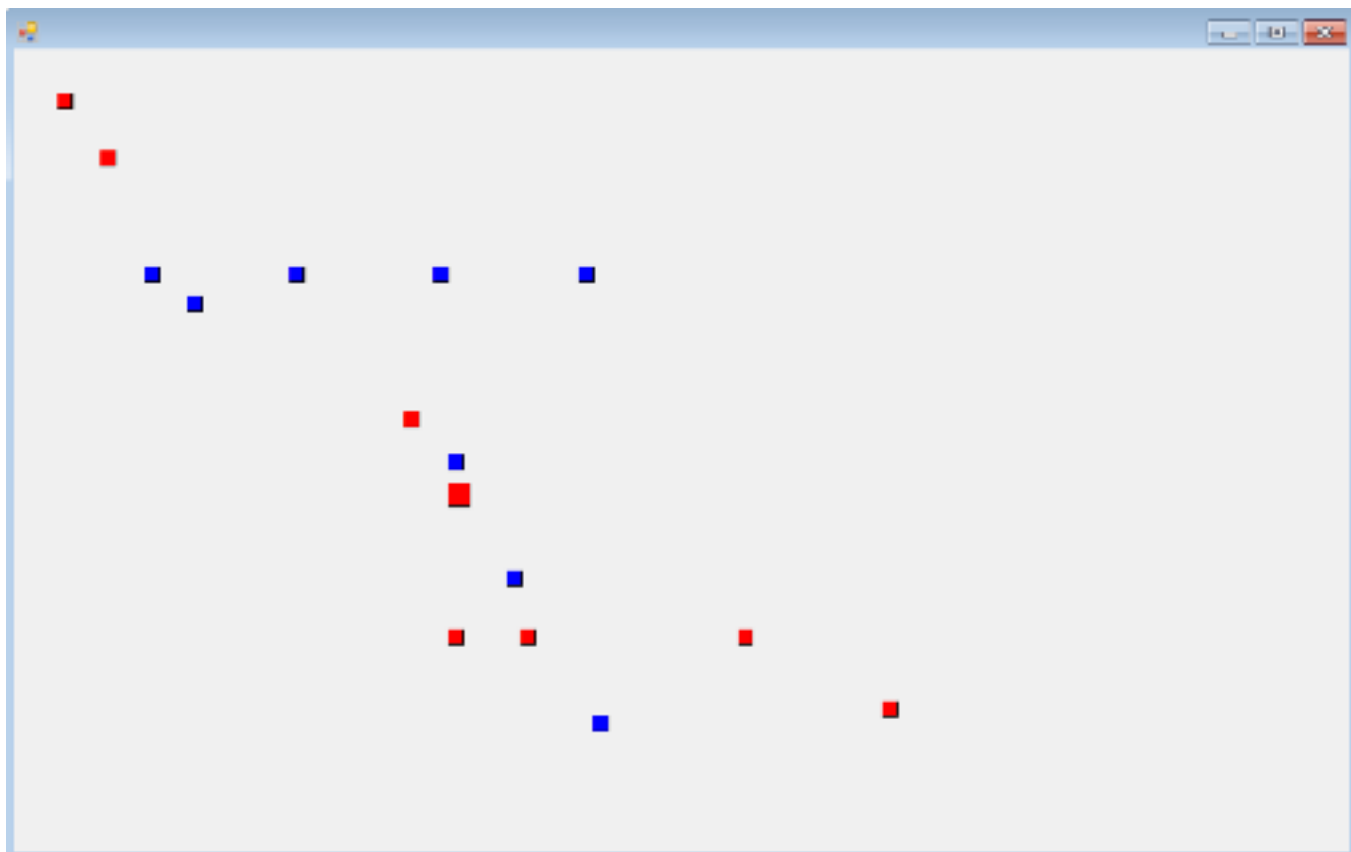
II. MÔ TẢ THUẬT TOÁN K-NEAREST NEIGHBOR

● Thuật toán K-NN được mô tả như sau:

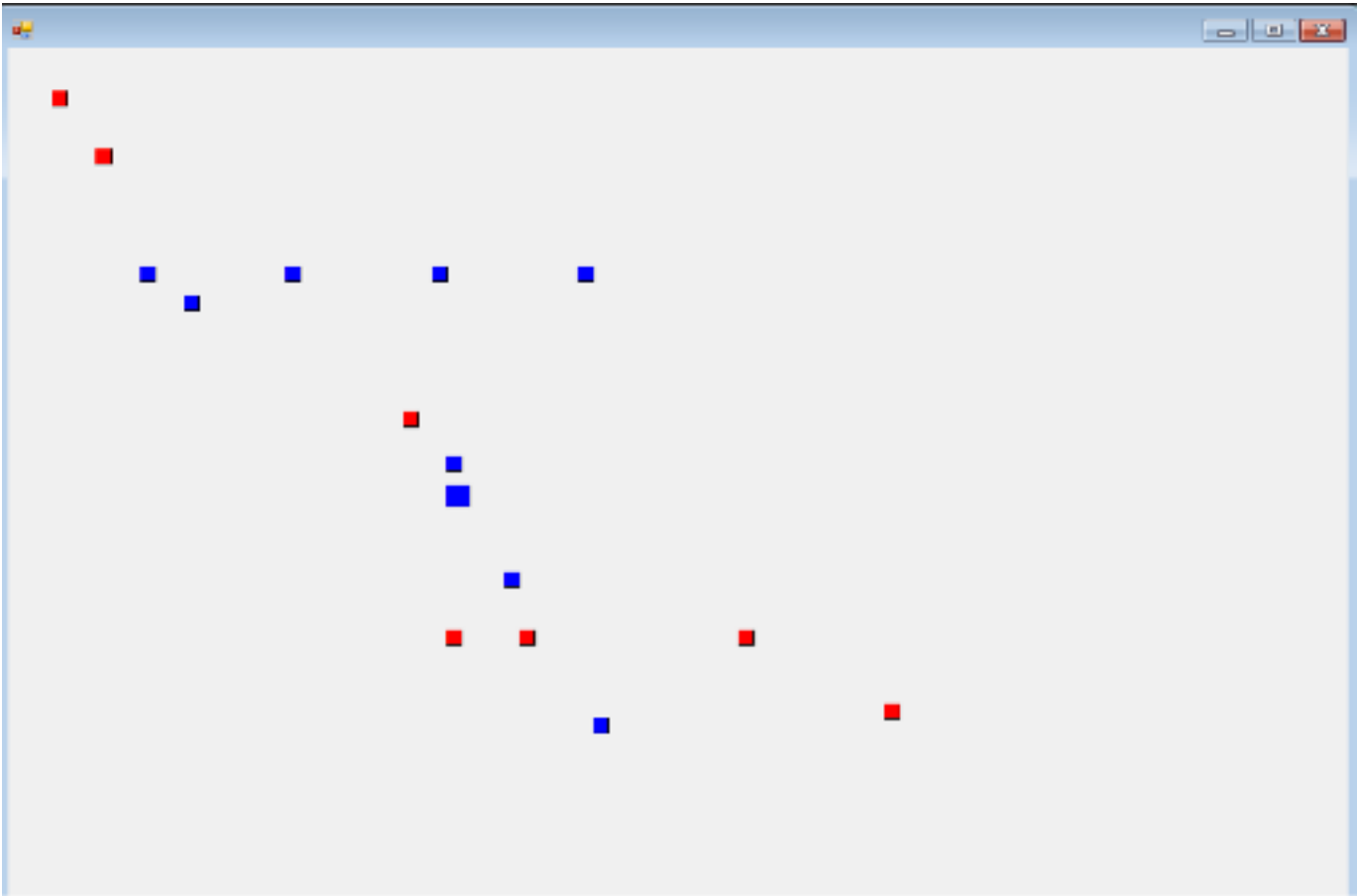
1. Xác định giá trị tham số k (số láng giềng gần nhất).
2. Tính khoảng cách giữa các đối tượng cần phân lớp (Query Point) với tất cả các đối tượng trong training data (thường sử dụng khoảng cách Euclidean).
3. Xếp xếp khoảng cách theo thứ tự tăng dần và xác định k láng giềng gần nhất với Query Point.
4. Lấy tất cả các lớp của k láng giềng gần nhất.
5. Dựa vào phân lớp của láng giềng gần nhất để xác định lớp cho Query Point.

III. VÍ DỤ MINH HOẠ THUẬT TOÁN K-NEAREST NEIGHBOR

Hình 1.1 Ví dụ minh hoạ với $k=5$



Hình 1.2 Ví dụ minh hoạ với $k=3$



PHẦN 2: CÀI ĐẶT THUẬT TOÁN K-NEAREST NEIGHBOR

I. MÔI TRƯỜNG LẬP TRÌNH

1. NGÔN NGỮ LẬP TRÌNH

- Ngôn ngữ được sử dụng để tìm hiểu và cài đặt thuật toán là C#.
- C# là một ngôn ngữ lập trình được phát triển bởi Microsoft và được phê chuẩn bởi European Computer Manufacturers Association (ECMA) và International Standards Organization (ISO).
- C# là một ngôn ngữ lập trình hướng đối tượng, là 1 phần của .NET Framework và có thể chạy được trên nhiều hệ thống khác nhau.

2. CÔNG CỤ LẬP TRÌNH

- Visual studio là một môi trường tích hợp (IDE) từ Microsoft. Nó được dùng để phát triển chương trình máy tính cho Windows, cũng như các website, ứng dụng web cũng như các dịch vụ web. Visual Studio được sử dụng nền tảng phát triển phần mềm của Microsoft như Windows API, Windows Forms, Windows Presentation Foundation, Windows Store và Microsoft Silverlight. Visual Studio có thể tạo ra cả hai ngôn ngữ máy và mã số quản lý. Visual studio là 1 trong những công cụ phát triển C#.
- GitHub là một website dùng để quản lý mã nguồn trực tuyến, dịch vụ này sử dụng Git làm nền tảng.

3. MÔI TRƯỜNG LẬP TRÌNH

- Được phát triển trên hệ điều hành Windows XP.
- Nền tảng .NET Framework 4.0.

II. CÀI ĐẶT THUẬT TOÁN K-NEAREST NEIGHBOR

Bảng 1: Lớp

TT	Tên lớp	Tên SV phụ trách viết	Mục đích chính của lớp trong chương trình
1	Diem	Nguyễn Tiến Đạt	<ul style="list-style-type: none">- Lớp Diem dùng tạo ra 1 đối tượng Diem gồm các thuộc tính như :<ul style="list-style-type: none">• toaDoX: là toạ độ x của đối tượng Diem• toaDoY: là toạ độ y của đối tượng Diem• chiSo: là giá trị đánh giá phân loại tính theo công thức Euclidean.• Loai : là phân loại các loại điểm khác nhau .

Bảng 2: Phương thức

TT	Phương thức	Mục đích	Tên file , số thứ tự dòng chứa khai báo	Tên SV phụ trách viết
1	<pre>public Diem(int toaDoX, int toaDoY, float chiSo, int Loai)</pre> <p>Input : toaDoX, toaDoY, chiSo, Loai</p> <p>Output: không</p>	<ul style="list-style-type: none"> • Là phương thức khởi tạo đối tượng Diem với đầy đủ các tham số truyền vào. 	Form1.cs,26	Nguyễn Tiến Đạt
2	<pre>public Diem()</pre> <p>Input: không</p> <p>Output: không</p>	<ul style="list-style-type: none"> • Là phương thức khởi tạo đối tượng Diem khi chưa có tham số truyền vào hay chưa phân loại. 	Form1.cs,27	Nguyễn Tiến Đạt
3	<pre>public khoiTao()</pre> <p>Input: không</p> <p>Output: không</p>	<ul style="list-style-type: none"> • Là phương thức khởi tạo dữ liệu đầu vào 	Form1.cs,32	Nguyễn Tiến Đạt
4	<pre>private int phanLoai(Diem diemMoi, int k)</pre> <p>Input: diemMoi , k</p> <p>Output: giá trị phân loại kiểu int</p>	<ul style="list-style-type: none"> • Là phương thức phân loại đối tượng dựa trên thuật toán Euclidean • Dữ liệu đầu vào là 1 đối tượng Diem chưa phân loại và với k đối tượng Diem gần nó nhất. 	Form1.cs,54	Nguyễn Tiến Đạt

TT	Phương thức	Mục đích	Tên file , số thứ tự dòng chứa khai báo	Tên SV phụ trách viết
5	public F1() Input: không. Output: không.	<ul style="list-style-type: none"> • Là bộ khởi tạo trong Windows Form, dùng tạo form hiển thị các đối tượng Diem 	Form1.cs,85	Nguyễn Tiến Đạt
6	private void F1_Paint(object sender, PaintEventArgs e) Input: không. Output: không.	<ul style="list-style-type: none"> • Là phương thức vẽ trong Windows Form dùng hiện giao diện trực quan cho thuật toán. 	Form1.cs,93	Nguyễn Tiến Đạt Nguyễn Văn Đạt

PHẦN 3: ĐÁNH GIÁ VÀ PHÂN CÔNG

I. PHÂN CÔNG

Bảng 3 : Phân công

TÊN SV	ĐÓNG GÓP	CÔNG VIỆC
Nguyễn Tiến Đạt	70%	<ul style="list-style-type: none">• Xây dựng ý tưởng.• Xây dựng giao diện.• Viết code.• Viết báo cáo.
Nguyễn Văn Đạt	30%	<ul style="list-style-type: none">• Tìm kiếm tài liệu.• Viết code.• Viết báo cáo.

II. ĐÁNH GIÁ

1. Ưu điểm:

- Ứng dụng chạy nhanh ổn định.

2. Nhược điểm:

- Dữ liệu đầu vào chưa lớn chưa thể hiện mức vĩ mô.

PHẦN 4:TÀI LIỆU THAM KHẢO

I. Tham khảo từ thầy.

II. Tham khảo từ Internet.

(1) https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

(2) <http://bis.net.vn/forums/t/370.aspx>

(3) <http://laptrinhaz.blogspot.com/2015/10/giai-thuat-k-lang-gieng-gan-knnk.html>