

Dự báo tính khách quan của bài viết thể thao bằng mô hình Perceptron

Giảng Viên Hướng Dẫn:
TS. Trần Nguyễn Minh Thư

Nhóm 17:
Lê Quang Sang B1606927
Nguyễn Phước Thành B1610669

Dự báo tính khách quan của bài viết thể thao bằng giải thuật Perceptron

Giới Thiệu

Giải thuật học

Đánh giá kết quả

Thảo Luận

Giới thiệu tập dữ liệu

Giới Thiệu

Giải thuật học

Đánh giá kết quả

Thảo Luận

1. Tên tập dữ liệu: Sport articles for objectivity analysis.
2. Thông tin tập dữ liệu: Trích xuất từ Phòng thí nghiệm trí tuệ nhân tạo, American University of Beirut
3. Nhiệm vụ: Dự đoán tính chủ quan khách quan của bài viết thể thao
4. Dữ liệu gồm: 1000 phần tử với 59 thuộc tính

Giới thiệu tập dữ liệu

Giới Thiệu

Giải thuật học

Đánh giá kết quả

Thảo Luận

5. 59 thuộc tính và 1 nhãn

- CC: Tần suất sử dụng liên từ (*Frequency of coordinating conjunctions*)
- CD: Tần suất sử dụng chữ số liệt kê (*Frequency of numerals and cardinals*)
- DT: Tần suất của các yếu tố quyết định (*Frequency of determiners*)
- EX: Tần suất xuất hiện từ chỉ số nhiều (*Frequency of existential there*)
- FW: Tần suất sử dụng từ ngữ nước ngoài (*Frequency of foreign words*)
- Ins: Tần suất kết hợp liên từ (*Frequency of subordinating preposition or conjunction*)
- ...
- Nhãn: khách quan (objective) / chủ quan (subjective)

Giới thiệu tập dữ liệu

TextID	URL	Label	totalWords	semanticc	semantics	CC	CD	DT	EX	FW	INs	J.
Text0001	http://msn.foxsports.com/	objective	109	0	1	7	9	0	5	8	6	
Text0002	http://msn.foxsports.com/	objective	309	21	4	1	19	1	4	35	23	
Text0003	http://uk.eurosport.yahoo	objective	149	6	1	8	14	0	5	15	11	
Text0004	http://uk.eurosport.yahoo	objective	305	18	5	7	26	0	10	37	21	
Text0005	http://uk.eurosport.yahoo	objective	491	23	8	33	47	0	12	61	36	
Text0006	http://msn.foxsports.com/	objective	314	14	1	17	17	0	0	36	15	
Text0007	http://uk.eurosport.yahoo	objective	314	10	6	1	37	1	10	33	20	
Text0008	http://msn.foxsports.com/	objective	462	19	6	5	47	0	0	40	39	
Text0009	http://msn.foxsports.com/	objective	808	40	11	49	71	1	6	94	44	
Text0011	http://msn.foxsports.com/	objective	860	44	22	13	111	0	4	101	56	
Text0012	http://msn.foxsports.com/	objective	497	25	9	15	44	0	19	64	22	
Text0013	http://msn.foxsports.com/	objective	106	4	0	1	4	0	9	7	7	
Text0014	http://msn.foxsports.com/	objective	792	57	22	28	93	0	26	121	62	
Text0016	http://msn.foxsports.com/	objective	246	11	2	6	27	0	12	30	23	
Text0017	http://msn.foxsports.com/	objective	614	30	10	21	68	1	13	90	38	
Text0018	http://msn.foxsports.com/	objective	143	16	1	2	20	0	0	13	12	
Text0019	http://msn.foxsports.com/	objective	246	12	2	3	20	0	4	21	11	

Mô hình

Giới Thiệu

Giải thuật học

Đánh giá kết quả

Thảo Luận

Chọn Tập dữ liệu và Mô hình:

- ✓ Mô hình: Perceptron đơn tầng
- ✓ Chọn ra 3 phần tử với 3 thuộc tính và 1 nhãn.

Nhãn	CC	CD	DT
Objective	7	9	0
Objective	1	19	1
Subjective	8	14	0

Mô hình

Chọn Tập dữ liệu và Mô hình:

✓ 1 cho objective 0 cho subjective

Giới Thiệu

Giải thuật học

Đánh giá kết quả

Thảo Luận

Nhãn	CC	CD	DT
1	7	9	0
1	1	19	1
0	8	14	0

Mô hình

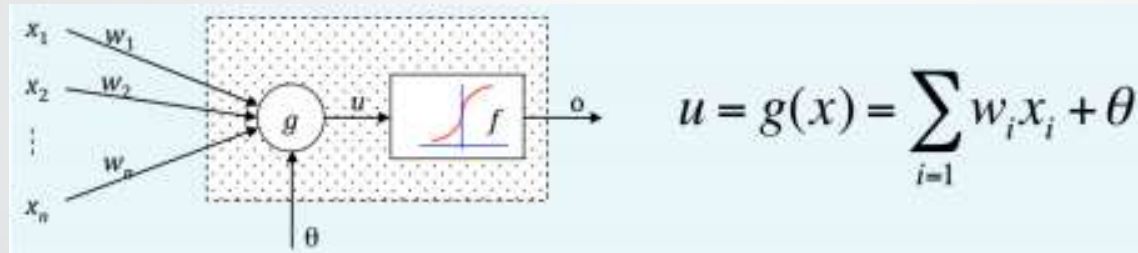
Giới Thiệu

Giải thuật học

Đánh giá kết quả

Thảo Luận

Sử dụng hàm kết hợp tuyến tính



Đầu ra

$$o = f(u) = f(g(x))$$

$$f(u) = \frac{1}{1 + e^{-u/T}} \quad f(u)[1 - f(u)] / T$$

GIẢI THUẬT HỌC

Giới Thiệu

Giải thuật học

Đánh giá kết quả

Thảo Luận

Khởi tạo ngẫu nhiên các trọng số w với : $w_0 = 0.25$; $w_1 = 0.5$; $w_2 = -0.2$; $w_3 = 0$; $\eta = 0.2$

$$o = f(u) = f(g(x)) = \begin{cases} 1 & g(x) \geq 0 \\ 0 & g(x) < 0 \end{cases}$$

Nhãn (y)	CC (x1)	CD (x2)	DT (x3)
1	7	9	0
1	1	19	1
0	8	14	0

Phần tử 1: $x_1^1 = 7$, $x_2^1 = 9$, $x_3^1 = 0$

$$g(x) = 0.25*1 + 0.5*7 + (-0.2)*9 + 0*1 = 1.95$$

do $g(x) > 0 \Rightarrow \text{output} = 1 = y$

GIẢI THUẬT HỌC

Phần tử 2: $x_1^2 = 1$, $x_2^2 = 19$, $x_3^2 = 1$

$$g(x) = 0.25*1 + 0.5*1 + (-0.2)*19 + 0*1 = -3.05$$

do $g(x) < 0 \Rightarrow \text{output} = 0 \neq y$

Cập nhật:

$$\begin{aligned} W0 &= W0 + \text{eta}(y^2 - o^2)*x^0 \\ &= 0.25 + 0.2(1 - 0)*1 = 0.45 \end{aligned}$$

$$\begin{aligned} W1 &= W1 + \text{eta}(y^2 - o^2)*x_1^2 \\ &= 0.5 + 0.2(1 - 0)*1 = 0.7 \end{aligned}$$

$$\begin{aligned} W2 &= W2 + \text{eta}(y^2 - o^2)*x_2^2 \\ &= (-0.2) + 0.2(1 - 0)*19 = 3.6 \end{aligned}$$

$$\begin{aligned} W3 &= W3 + \text{eta}(y^2 - o^2)*x_3^2 \\ &= 0 + 0.2(1 - 0)*1 = 0.2 \end{aligned}$$

Giới Thiệu

Giải thuật học

Đánh giá kết quả

Thảo Luận

Nhãn (y)	CC (x1)	CD (x2)	DT (x3)
1	7	9	0
1	1	19	1
0	8	14	0

GIẢI THUẬT HỌC

Phần tử 3: $x_1^3 = 1, x_2^3 = 19, x_3^3 = 1$

$$g(x) = 0.45*1 + 0.7*1 + 3.6*14 + 0.2*0 = 51.55$$

do $g(x) > 0 \Rightarrow \text{output} = 1 \neq y$

Cập nhật:

$$\begin{aligned} W0 &= W0 + \text{eta}(y^3 - o^3)*x_0 \\ &= 0.45 + 0.2(0 - 1)*1 = 0.25 \end{aligned}$$

$$\begin{aligned} W1 &= W1 + \text{eta}(y^3 - o^3)*x_1^3 \\ &= 0.7 + 0.2(0 - 1)*8 = -0.9 \end{aligned}$$

$$\begin{aligned} W2 &= W2 + \text{eta}(y^3 - o^3)*x_2^3 \\ &= 3.6 + 0.2(0 - 1)*14 = 0.8 \end{aligned}$$

$$\begin{aligned} W3 &= W3 + \text{eta}(y^3 - o^3)*x_3^3 \\ &= 0.2 + 0.2(0 - 1)*0 = 0.2 \end{aligned}$$

Giới Thiệu

Giải thuật học

Đánh giá kết quả

Thảo Luận

Nhãn (y)	CC (x1)	CD (x2)	DT (x3)
1	7	9	0
1	1	19	1
0	8	14	0

ĐÁNH GIÁ KẾT QUẢ

Giới Thiệu

Giải thuật học

Đánh giá kết quả

Thảo Luận

- ✓ Nghi thức Hold-out:
 - Lấy ngẫu nhiên $3/4$ tập dữ liệu để học và $1/4$ tập dữ liệu còn lại dùng cho kiểm tra.
- ✓ Các mô hình được xây dựng dựa trên:
 - Số lượng tập huấn luyện: 750
 - Số lượng tập kiểm tra: 250

ĐÁNH GIÁ KẾT QUẢ

- ✓ So sánh hiệu quả của các phương pháp theo RMSE

	Perceptron	Cây quyết định
Độ chính xác	0.82934131	0.832335329

[Giới Thiệu](#)[Giải thuật học](#)[Đánh giá kết quả](#)[Thảo Luận](#)

ĐÁNH GIÁ KẾT QUẢ

- ✓ Ta có thể thấy các giá trị của nghi thức đánh giá :
Giải thuật Perceptron không có khác nhiều về mặt kết quả so với Cây Quyết Định



Cảm ơn Cô và các bạn đã lắng nghe !!!