

<p align="center">Khoa Khoa học máy tính Trường Công nghệ Thông tin & Truyền thông Trường Đại học Bách khoa Hà Nội</p>	<p align="center">Đề thi cuối kỳ 20211 Học phần: Khai phá Web – IT4868 Thời gian: 90' (được sử dụng tài liệu giấy) Ngày thi: 18/02/2022</p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

ĐỀ BÀI

Câu 1 (2đ): Giả sử ta áp dụng một mô hình mạng nơ-ron tiến 3 tầng cho bài toán phân loại cảm xúc có giám sát với 3 lớp {tiêu cực, trung tính, tích cực}:

a/ Tầng đầu vào có bao nhiêu nơ-ron?

b/ Trọng số liên kết giữa hai tầng liên tiếp trong mạng gọi là gì của mô hình? Làm thế nào để xác định được giá trị tối ưu của các trọng số này?

c/ Tầng đầu ra có bao nhiêu nơ-ron?

d/ Gọi số nơ-ron của tầng ẩn là K. Làm thế nào để xác định giá trị tối ưu cho K?

e/ Mô hình sử dụng hàm mục tiêu cross-entropy. Hãy trình bày rõ công thức của hàm mục tiêu này dưới dạng một hàm của tín hiệu đầu vào và trọng số của mạng nơ-ron.

Câu 2 (1đ): Phương pháp phân rã ma trận cho lọc cộng tác được coi là áp dụng một mạng nơ-ron cơ bản

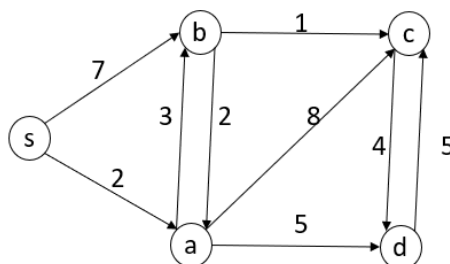
a/ Mạng nơ-ron này có kiến trúc như thế nào? (số tầng, số chiều mỗi tầng, liên kết giữa các tầng)

b/ Hàm mục tiêu của mô hình này là gì? Hãy trình bày rõ công thức của hàm mục tiêu này dưới dạng một hàm của tín hiệu đầu vào và trọng số của mạng nơ-ron.

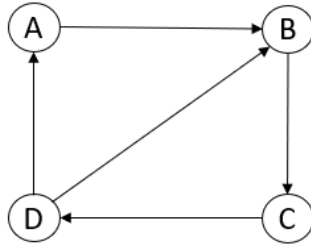
c/ Ưu điểm của phương pháp này so với knn-cf là gì?

d/ Nhược điểm của phương pháp này so với NCF là gì?

Câu 3 (2đ): Hãy trình bày các bước của thuật toán Dijkstra để tìm đường đi ngắn nhất từ b đến các đỉnh còn lại trong đồ thị dưới đây:



Câu 4 (2đ): Hãy dùng cách giải phương trình để tìm thứ hạng của các đỉnh trong đồ thị dưới đây sao cho thỏa mãn công thức của thuật toán Pagerank với damping factor $d = 1$ (không có yếu tố duyệt ngẫu nhiên) và tổng thứ hạng của các đỉnh = 1.



Câu 5 (2đ): Cho tập văn bản huấn luyện như dưới bao gồm các bình luận về điện thoại di động được phân vào hai lớp cảm xúc pos và neg. Hãy áp dụng phương pháp phân loại Multinomial Naive Bayes với kĩ thuật làm mịn thêm 1 để dự đoán cảm xúc của văn bản 5

Doc	Nội dung	Lớp
1	thích Bphone BKA V	pos
2	Bphone Bphone chất	pos
3	Bphone màn_hình	pos
4	lphone Bphone màn_hình	neg
5	màn_hình Bphone Bphone Iphone	?

Câu 6 (1đ): Cho ma trận tương tác người dùng – sản phẩm dưới đây. Giả sử độ tương đồng của hai sản phẩm i và j được tính theo công thức dưới đây:

$$sim(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_u)^2}}$$

trong đó U là tập người dùng cùng đánh giá sản phẩm i và j ; $R_{u,i}$ là đánh giá của người dùng u với sản phẩm i . \bar{R}_u là đánh giá trung bình của người dùng u .

a/ Hãy tính độ tương đồng của item 1 và item 2

b/ Hãy tính độ tương đồng của item 2 và item 3

	item 1	item 2	item 3
user_1	2		3
user_2	5	2	
user_3	3	3	1
user_4		2	2

-Hết-

Cán bộ ra đề

Nguyễn Kiêm Hiếu