**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ GIAO THÔNG VẬN TẢI**

**KHOA CÔNG NGHỆ THÔNG TIN**  
A logo with orange letters and text

AI-generated content may be incorrect.

**BÁO CÁO BÀI TẬP LỚN**

**MÔN HỌC:**

**HỆ TRỢ GIÚP QUYẾT ĐỊNH**

**ĐỀ TÀI:**

**TÌM HIỂU LÝ THUYẾT TẬP THÔ VÀ ỨNG DỤNG TRONG TRÍCH CHỌN DỮ LIỆU**

|  |  |
| --- | --- |
| **GV HƯỚNG DẪN:**  **LỚP:** | **ĐOÀN THỊ THANH HẰNG**  **73DCHT23** |
| **NHÓM THỰC HIỆN:** | **NHÓM 7** |
|  | 1. **ĐINH QUANG ĐỨC** 2. **DƯƠNG THỊ KIỀU DIỄM** 3. **NGUYỄN VĂN HỒNG ANH** 4. **NGUYỄN TẤT THẮNG** 5. **PHẠM THỊ VÂN** |
|  |

**HÀ NỘI – 2025**

**CHƯƠNG 1: TỔNG QUAN VỀ MÔN HỌC**

1. Tổng quan môn học Hệ trợ giúp quyết định (Decision Support System – DSS)

[1.1.Gi](http://1.1.gi)ới thiệu chung

1.1.1.Bối cảnh ra đời

Trước thập niên 1970, các tổ chức/doanh nghiệp chủ yếu sử dụng **Hệ thống xử lý giao dịch (TPS)** và **Hệ thống thông tin quản lý (MIS)** để lưu trữ, xử lý dữ liệu cơ bản.

Tuy nhiên, khi môi trường kinh doanh ngày càng phức tạp, dữ liệu ngày càng nhiều, các hệ thống này **chỉ cung cấp báo cáo tĩnh** mà không hỗ trợ phân tích sâu cho việc ra quyết định.

Năm 1971, khái niệm **Decision Support System (DSS)** lần đầu tiên được đề cập bởi **Gorry và Scott Morton** trong nghiên cứu về hệ thống hỗ trợ quyết định dựa trên máy tính.

Đến thập niên 1980, với sự phát triển của **cơ sở dữ liệu, hệ thống mô hình và trí tuệ nhân tạo**, DSS bắt đầu được triển khai rộng rãi trong doanh nghiệp và quản lý nhà nước.

1.1.2.Sự cần thiết của DSS

**Môi trường quyết định phức tạp**: Dữ liệu nhiều chiều, có yếu tố không chắc chắn, không đầy đủ.

**Ra quyết định kịp thời**: Trong cạnh tranh toàn cầu, tốc độ ra quyết định ảnh hưởng trực tiếp đến hiệu quả kinh doanh.

**Hạn chế của con người**: Nhà quản trị khó phân tích thủ công khối lượng dữ liệu khổng lồ.

**Nhu cầu dự báo và tối ưu hóa**: Không chỉ xem xét dữ liệu quá khứ mà cần dự báo tương lai và so sánh nhiều phương án.

**Xu thế công nghệ**: Sự phát triển của **Big Data, AI, Data Mining, IoT, Cloud** càng làm DSS trở thành công cụ tất yếu trong quản trị hiện đại.

1.1.3. Khái niệm DSS

**DSS (Decision Support System – Hệ thống hỗ trợ ra quyết định)** là một hệ thống thông tin dựa trên máy tính, được thiết kế nhằm hỗ trợ nhà quản trị hoặc người ra quyết định trong quá trình phân tích, lựa chọn phương án, giải quyết vấn đề phức tạp.

DSS **không thay thế con người** trong việc ra quyết định, mà cung cấp **các công cụ, dữ liệu, mô hình và tri thức** giúp người dùng đưa ra quyết định **nhanh chóng, chính xác và hiệu quả hơn**.

Theo định nghĩa của **Turban & Aronson (1998)**:  
 “DSS là hệ thống thông tin dựa trên máy tính, cung cấp sự hỗ trợ trực tiếp cho người ra quyết định trong các tình huống bán cấu trúc và phi cấu trúc, nơi mà việc sử dụng trực giác con người kết hợp với phân tích máy tính là cần thiết.”

Đặc điểm nổi bật của DSS:

**Hỗ trợ (Support)**, không thay thế người ra quyết định.  
 Xử lý được các tình huống **cấu trúc, bán cấu trúc và phi cấu trúc**.  
 Kết hợp **cơ sở dữ liệu + mô hình + giao diện người dùng**.  
 Cho phép người dùng **tương tác, mô phỏng, thử nghiệm nhiều phương án** trước khi đưa ra quyết định.

1.1.4. Vai trò tổng quát của DSS trong  môn học Hệ trợ giúp quyết định

**Cầu nối giữa dữ liệu và quyết định** DSS giúp biến dữ liệu thô thành thông tin có ý nghĩa, từ đó hỗ trợ người học hiểu cách thông tin được chuyển hóa thành cơ sở cho quyết định.

**Môi trường thực hành ra quyết định** DSS cung cấp các công cụ, mô hình, kịch bản để sinh viên mô phỏng và thử nghiệm việc ra quyết định trong nhiều tình huống khác nhau.

**Hình thành tư duy phân tích** Thông qua DSS, sinh viên được rèn luyện khả năng phân tích dữ liệu, so sánh phương án, đánh giá rủi ro và dự báo kết quả.

**Gắn kết lý thuyết với thực tiễn** DSS không chỉ dừng ở khái niệm, mà còn cho thấy ứng dụng thực tế trong kinh doanh, y tế, tài chính, sản xuất... từ đó giúp môn học mang tính ứng dụng cao

**Hỗ trợ kỹ năng công nghệ** Sinh viên làm quen với các công cụ phân tích dữ liệu, mô hình toán học, hệ thống thông tin và công nghệ trí tuệ nhân tạo – những kỹ năng quan trọng trong kỷ nguyên số.

1.1.5. Phạm vi nghiên cứu của môn học

**Về đối tượng nghiên cứu** Nghiên cứu các **khái niệm, mô hình và công cụ** hỗ trợ quá trình ra quyết định trong tổ chức/doanh nghiệp.  
 Tập trung vào **bài toán bán cấu trúc và phi cấu trúc**, nơi các quyết định khó có thể lập trình hoặc tự động hóa hoàn toàn.

**Về nội dung kiến thức** Cơ sở lý thuyết về DSS: khái niệm, đặc điểm, thành phần, vai trò.  
 Các mô hình ra quyết định: định lượng, định tính, mô phỏng, tối ưu hóa.  
 Công cụ và công nghệ: cơ sở dữ liệu, kho dữ liệu (Data Warehouse), OLAP, trí tuệ nhân tạo, khai phá dữ liệu (Data Mining).  
 Các phương pháp phân tích và hỗ trợ: phân tích kịch bản, dự báo, mô hình hóa, heuristic.

**Về phạm vi ứng dụng** Ứng dụng DSS trong nhiều lĩnh vực: kinh doanh, tài chính, y tế, giao thông, giáo dục, hành chính công.  
 Tập trung vào việc **nâng cao chất lượng và hiệu quả ra quyết định** trong bối cảnh dữ liệu lớn và môi trường nhiều bất định.

**Về giới hạn môn học** DSS chỉ hỗ trợ chứ **không thay thế hoàn toàn con người** trong ra quyết định.  
 Không đi sâu vào kỹ thuật hạ tầng phần cứng, mà tập trung vào **mô hình, công cụ và ứng dụng DSS trong quản trị và ra quyết định**.

1.1.6.Ý nghĩa khi học DSS

**Nâng cao năng lực phân tích và ra quyết định** Người học được rèn luyện kỹ năng khai thác và phân tích dữ liệu để hỗ trợ các quyết định quan trọng.  
 Giúp hiểu rõ mối quan hệ giữa dữ liệu – thông tin – tri thức – quyết định.

**Tiếp cận công cụ và công nghệ hiện đại** Làm quen với các mô hình, phương pháp phân tích, và công nghệ hỗ trợ (CSDL, OLAP, Data Mining, AI/ML).  
 Trang bị kiến thức cần thiết để vận dụng vào các hệ thống quản trị trong thực tế.

**Tư duy hệ thống và tư duy quản trị** DSS giúp sinh viên phát triển tư duy nhìn nhận vấn đề một cách toàn diện.  
 Rèn luyện khả năng đánh giá các phương án khác nhau, phân tích rủi ro, dự báo xu hướng.

**Ứng dụng thực tiễn cao**

Kiến thức từ môn học có thể áp dụng trong nhiều lĩnh vực: kinh doanh, tài chính, y tế, sản xuất, giao thông, giáo dục…

Góp phần nâng cao năng lực cạnh tranh, hiệu quả quản trị và chuyển đổi số trong tổ chức.

**Phát triển kỹ năng mềm** Học DSS không chỉ là học công cụ, mà còn rèn luyện **kỹ năng ra quyết định, làm việc nhóm, tư duy phản biện, trình bày kết quả** – rất quan trọng trong công việc sau này.

1.2.  Đặc điểm của hệ trợ giúp quyết định

**Hỗ trợ, không thay thế con người** DSS được thiết kế để **hỗ trợ nhà quản trị ra quyết định**, chứ không thay thế hoàn toàn vai trò của con người.  
 Người dùng vẫn là trung tâm trong quá trình đưa ra quyết định.

**Xử lý các bài toán bán cấu trúc và phi cấu trúc** DSS đặc biệt hữu ích trong những tình huống mà dữ liệu và thông tin chưa đầy đủ, có sự mơ hồ, bất định.  
 Khác với các hệ thống thông tin quản lý (MIS) vốn chỉ xử lý bài toán có cấu trúc rõ ràng.

**Khả năng tích hợp và linh hoạt** DSS có thể tích hợp nhiều nguồn dữ liệu (cơ sở dữ liệu, kho dữ liệu, dữ liệu lớn).  
 Có khả năng kết hợp nhiều mô hình phân tích, công cụ mô phỏng, thuật toán tối ưu.

**Khả năng tương tác cao** Người dùng có thể dễ dàng giao tiếp với DSS thông qua giao diện thân thiện, truy vấn dữ liệu, thử nghiệm các kịch bản.  
 DSS cung cấp phản hồi nhanh chóng và trực quan (bảng biểu, đồ thị, báo cáo).

**Khả năng phân tích và dự báo** DSS không chỉ báo cáo dữ liệu quá khứ mà còn có khả năng **dự báo, mô phỏng và so sánh các phương án quyết định**.

**Tính thích ứng và mở rộng** DSS có thể được điều chỉnh để phù hợp với từng tổ chức, lĩnh vực và quy mô khác nhau.  
 Dễ mở rộng, nâng cấp với công nghệ mới như trí tuệ nhân tạo (AI), học máy (ML), điện toán đám mây.

1.3. Mục tiêu của DSS

**Hỗ trợ quá trình ra quyết định** Cung cấp công cụ phân tích dữ liệu, xử lý thông tin để nhà quản trị đưa ra lựa chọn tốt hơn.  
 Không thay thế mà **tăng cường khả năng ra quyết định của con người**.

**Nâng cao chất lượng quyết định** Giúp quyết định **chính xác hơn, kịp thời hơn, hợp lý hơn**.  
 Hạn chế sai lầm do thiếu thông tin hoặc do chủ quan.

**Giảm chi phí và rủi ro** Hỗ trợ mô phỏng, thử nghiệm nhiều phương án trước khi thực hiện.  
 Từ đó tiết kiệm nguồn lực và hạn chế rủi ro trong môi trường bất định.

**Tăng cường khả năng phân tích và dự báo** DSS không chỉ nhìn vào dữ liệu quá khứ mà còn dự báo xu hướng tương lai.  
 Hỗ trợ lập kế hoạch chiến lược, chiến thuật và tác nghiệp.

**Cải thiện hiệu quả quản lý và cạnh tranh** Giúp doanh nghiệp/ tổ chức phản ứng nhanh hơn với thay đổi.  
 Nâng cao năng lực cạnh tranh trong bối cảnh dữ liệu lớn và thị trường biến động.

1.4.Đối tương phục vụ của DSS

Hệ trợ giúp quyết định (DSS) được xây dựng nhằm **phục vụ những người tham gia vào quá trình ra quyết định trong tổ chức**. Cụ thể:

**Nhà quản trị các cấp  
 Cấp chiến lược (Top Managers):** DSS hỗ trợ phân tích môi trường vĩ mô, dự báo xu hướng dài hạn, lựa chọn chiến lược phát triển.  
 **Cấp trung gian (Middle Managers):** Hỗ trợ trong việc lập kế hoạch, phân bổ nguồn lực, theo dõi tiến độ, đánh giá hiệu quả.  
 **Cấp tác nghiệp (Operational Managers):** DSS hỗ trợ xử lý công việc hằng ngày, tối ưu hoá quy trình và giải quyết vấn đề phát sinh nhanh chóng.

**Chuyên gia phân tích và ra quyết định**

Sử dụng DSS để khai thác dữ liệu, mô hình hóa tình huống, kiểm tra giả thuyết, mô phỏng các kịch bản khác nhau.

**Doanh nghiệp / tổ chức nói chung** DSS cung cấp thông tin giúp doanh nghiệp **nâng cao năng lực cạnh tranh**, **giảm thiểu rủi ro** và **tối ưu hóa nguồn lực**.

**Người dùng không chuyên về CNTT** DSS thường được thiết kế giao diện thân thiện, trực quan, giúp những người không chuyên về công nghệ (như giám đốc, trưởng phòng, nhân viên nghiệp vụ) vẫn có thể sử dụng hiệu quả.

1.5. Thành phần của DSS

**Hệ thống quản lý dữ liệu (Data Management Subsystem)** Bao gồm cơ sở dữ liệu (Database) chứa dữ liệu nội bộ và dữ liệu bên ngoài.  
 Có hệ thống quản lý cơ sở dữ liệu (DBMS) để lưu trữ, truy vấn và cập nhật dữ liệu.  
 Có thể mở rộng bằng **Data Warehouse, OLAP, Big Data** để xử lý dữ liệu lớn.

**Hệ thống quản lý mô hình (Model Management Subsystem)** Chứa các mô hình toán học, mô hình mô phỏng, tối ưu hóa, thống kê, dự báo.  
 Có trình quản lý mô hình (Model Base Management System – MBMS) để gọi, chạy và kết hợp mô hình.  
 Giúp người dùng thử nghiệm nhiều kịch bản, so sánh phương án khác nhau.

**Hệ thống quản lý đối thoại (User Interface / Dialog Management Subsystem)** Giao diện người dùng để tương tác với hệ thống.  
 Cho phép nhập yêu cầu, thực hiện truy vấn, hiển thị kết quả (báo cáo, biểu đồ, dashboard).  
 Yếu tố quan trọng để DSS có tính thân thiện và dễ sử dụng.

**Thành phần tri thức (Knowledge Management Subsystem – mở rộng)** Xuất hiện trong DSS hiện đại.  
 Lưu trữ tri thức, luật, kinh nghiệm chuyên gia.  
 Ứng dụng **AI, Machine Learning, hệ chuyên gia (Expert System)** để DSS trở nên thông minh hơn.

1.6. Các kỹ thuật và phương pháp trong DSS

**Kỹ thuật xử lý dữ liệu (Data-oriented techniques)  
 Kho dữ liệu (Data Warehouse):** Tập hợp dữ liệu từ nhiều nguồn để phục vụ phân tích.  
 **OLAP (Online Analytical Processing):** Phân tích dữ liệu đa chiều, giúp so sánh, tổng hợp nhanh chóng.  
 **Khai phá dữ liệu (Data Mining):** Phát hiện mẫu, xu hướng ẩn trong dữ liệu lớn.

**Kỹ thuật mô hình hóa và tối ưu hóa (Model-oriented techniques)  
 Mô hình toán học:** Tối ưu hóa (linear programming, integer programming…), mô hình thống kê, mô hình dự báo.  
 **Mô phỏng (Simulation):** Tái hiện quá trình hoặc hệ thống phức tạp để dự báo kết quả.  
 **Mô hình ra quyết định đa tiêu chí (MCDM – Multi-Criteria Decision Making):** Hỗ trợ lựa chọn phương án khi có nhiều tiêu chí.

**Kỹ thuật trí tuệ nhân tạo (AI-oriented techniques)  
 Hệ chuyên gia (Expert Systems):** Mô phỏng tri thức chuyên gia trong lĩnh vực cụ thể.  
 **Máy học (Machine Learning):** Dùng dữ liệu huấn luyện để dự đoán và gợi ý quyết định.  
 **Xử lý ngôn ngữ tự nhiên (NLP):** Hỗ trợ người dùng giao tiếp với hệ thống bằng ngôn ngữ tự nhiên.

**Phương pháp trực quan hóa và giao diện (Visualization & UI techniques)  
 Dashboard, báo cáo động (Interactive Reports):** Hiển thị dữ liệu trực quan, dễ hiểu.  
 **Biểu đồ, bản đồ, mô hình 3D:** Hỗ trợ phân tích không gian, địa lý (GIS – Geographic Information Systems).

**Phương pháp dựa trên tri thức và kinh nghiệm** Kết hợp dữ liệu, mô hình và kinh nghiệm chuyên gia để đưa ra gợi ý tối ưu.  
 Cho phép người dùng điều chỉnh mô hình theo bối cảnh thực tế.

1.7. Ứng dung của DSS

**Kinh doanh và quản trị doanh nghiệp** Hỗ trợ phân tích thị trường, dự báo nhu cầu khách hàng.  
 Xây dựng chiến lược marketing, quản lý sản phẩm và dịch vụ.  
 Quản trị chuỗi cung ứng (Supply Chain Management), quản lý tồn kho.

**Tài chính – Ngân hàng** Phân tích rủi ro tín dụng, hỗ trợ quyết định cho vay.  
 Quản lý danh mục đầu tư, dự báo biến động thị trường.  
 Phát hiện gian lận trong giao dịch.

**Y tế và chăm sóc sức khỏe** Hỗ trợ bác sĩ chẩn đoán dựa trên dữ liệu bệnh án điện tử.  
 Đề xuất phác đồ điều trị cá nhân hóa cho bệnh nhân.  
 Quản lý phân bổ nguồn lực (trang thiết bị, thuốc men, nhân sự).

**Sản xuất và công nghiệp** Tối ưu hóa quy trình sản xuất, giảm chi phí.  
 Quản lý chất lượng sản phẩm.

Lập kế hoạch sản xuất và phân phối.

**Giao thông và logistics** Hỗ trợ điều phối giao thông đô thị (ví dụ: hệ thống quản lý đèn tín hiệu thông minh).  
 Quản lý đội xe, tối ưu hóa tuyến đường vận tải.  
 Hệ thống đặt vé, quản lý chuyến bay, tàu hỏa.

**Hành chính công và chính phủ điện tử** Hỗ trợ ra quyết định trong quy hoạch đô thị, quản lý tài nguyên.  
 Phân tích dữ liệu dân cư, dịch vụ công.  
 Dự báo và quản lý khủng hoảng (thiên tai, dịch bệnh)

**Giáo dục và đào tạo** Phân tích dữ liệu học tập để đánh giá chất lượng đào tạo.

Đề xuất lộ trình học tập cá nhân hóa.  
 Quản lý nguồn lực trường học (giảng viên, phòng học, thiết bị).

1.8. Thách thức và xu hướng phát triển

1.8.1. Thách thức

**Khối lượng và tốc độ dữ liệu lớn (Big Data):** Khó khăn trong việc xử lý, lưu trữ và phân tích dữ liệu ngày càng phức tạp.

**Chất lượng dữ liệu:** Dữ liệu có thể thiếu, sai lệch hoặc không đồng bộ, ảnh hưởng đến độ chính xác của quyết định.

**Tính bảo mật và riêng tư:** DSS thường xử lý dữ liệu nhạy cảm (tài chính, y tế, cá nhân), đòi hỏi an toàn và tuân thủ pháp luật.

**Độ phức tạp của mô hình:** Một số mô hình phân tích khó hiểu với người dùng, gây khó khăn trong việc giải thích kết quả.

**Phụ thuộc vào công nghệ:** Chi phí triển khai, bảo trì và cập nhật hệ thống cao, dễ gây khó khăn cho tổ chức nhỏ.

**Khả năng thích ứng:** DSS cần liên tục cập nhật để phù hợp với môi trường kinh doanh thay đổi nhanh chóng.

1.8.2. Xu hướng phát

**DSS thông minh dựa trên AI và Machine Learning:** Học hỏi từ dữ liệu để đưa ra gợi ý, dự báo chính xác hơn.

**DSS dựa trên dữ liệu lớn (Big Data DSS):** Khai thác dữ liệu đa dạng, khối lượng lớn theo thời gian thực.

**DSS trên nền tảng điện toán đám mây (Cloud-based DSS):** Giúp tiết kiệm chi phí, tăng khả năng mở rộng và truy cập từ mọi nơi.

**DSS di động (Mobile DSS):** Cho phép nhà quản trị ra quyết định mọi lúc, mọi nơi qua smartphone, tablet.

**DSS tích hợp IoT (Internet of Things):** Thu thập dữ liệu trực tiếp từ cảm biến, thiết bị kết nối để hỗ trợ quyết định trong thời gian thực.

**DSS hướng người dùng (User-centric):** Giao diện trực quan, dễ sử dụng, tăng khả năng giải thích và tương tác.

**Kết hợp DSS với hệ thống hỗ trợ nhóm (GDSS – Group DSS):** Hỗ trợ ra quyết định theo nhóm, tăng tính hợp tác trong tổ chức.

1.9.Kết luận

Hệ trợ giúp quyết định (Decision Support System – DSS) là một công cụ quan trọng trong quản lý và ra quyết định hiện đại. DSS không thay thế con người mà hỗ trợ nhà quản trị phân tích dữ liệu, so sánh phương án, dự báo kết quả và lựa chọn quyết định tối ưu. Với sự phát triển mạnh mẽ của công nghệ dữ liệu lớn, trí tuệ nhân tạo, điện toán đám mây và IoT, DSS ngày càng trở nên thông minh, linh hoạt và gần gũi với người dùng. Tuy còn nhiều thách thức về dữ liệu, bảo mật và chi phí triển khai, nhưng DSS đã và sẽ tiếp tục đóng vai trò then chốt trong việc nâng cao hiệu quả quản lý, giảm rủi ro, gia tăng năng lực cạnh tranh và thúc đẩy quá trình chuyển đổi số trong mọi lĩnh vực của đời sống – từ kinh doanh, tài chính, y tế đến giáo dục, giao thông và hà

**CHƯƠNG 2: CƠ SỞ LÝ THUYẾT**

2. Cơ sở lý thuyết

2.1. Khái niệm chung về tập thô

Lý thuyết tập thô (Rough Set Theory – RST) được Zdzisław Pawlak đề xuất vào năm 1982 nhằm xử lý các vấn đề liên quan đến dữ liệu không đầy đủ, không chắc chắn hoặc mơ hồ.

Ý tưởng cốt lõi là: thay vì biểu diễn một tập hợp chính xác, ta dùng hai xấp xỉ (dưới và trên) để mô tả tập hợp đó.

RST trở thành một công cụ mạnh mẽ trong khai phá dữ liệu, trí tuệ nhân tạo và học máy, đặc biệt trong trích chọn thuộc tính và sinh luật quyết định.

2.2. Hệ thông tin và hệ quyết định

Một hệ thông tin được biểu diễn bởi bộ bốn:

IS = (U, A, V, f)

Trong đó:

* U: tập hữu hạn các đối tượng (universe).
* A: tập hữu hạn các thuộc tính (attributes).
* V = ⋃a∈A​Va​: miền giá trị của các thuộc tính.
* f: U x A→V: hàm gán giá trị cho mỗi cặp (đối tượng, thuộc tính).

Ví dụ:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Đối tượng | Sốt | Ho | Đau đầu | Bệnh (d) |
| u1 | Cao | Có | Có | Cúm |
| u2 | Cao | Có | Không | Cúm |
| u3 | Thấp | Có | Có | Cảm lạnh |
| u4 | Thấp | Không | Có | Khỏe |

Nếu trong bảng dữ liệu có thêm thuộc tính quyết định d, ta có hệ quyết định (Decision System):

DS=(U,C∪{d},V,f)

Trong đó C là tập các thuộc tính điều kiện (condition attributes).

2.3. Quan hệ tương đương và phân hoạch dữ liệu

Với tập thuộc tính B⊆A, định nghĩa quan hệ tương đương:

x≡B​y⟺∀a∈B,f(x,a)=f(y,a)

Quan hệ này chia tập U thành các lớp tương đương [x]B​. Các lớp tương đương chính là những cụm đối tượng không thể phân biệt khi chỉ xét các thuộc tính trong .

Ví dụ (dựa vào bảng trên, chọn B={Sốt, Ho}:

* Lớp 1: {u1,u2} (đều có Sốt = Cao, Ho = Có).
* Lớp 2: {u3} (Sốt = Thấp, Ho = Có).
* Lớp 3: {u4} (Sốt = Thấp, Ho = Không).

2.4. Xấp xỉ dưới và xấp xỉ trên

Trong nhiều trường hợp, do dữ liệu chưa đủ, ta không thể xác định chính xác một tập con X⊆U. Lý thuyết tập thô sử dụng khái niệm xấp xỉ để mô tả:

* Xấp xỉ dưới theo B:

B​(X)={x∈U∣[x]B​⊆X}

→ Tập các đối tượng chắc chắn thuộc X.

* Xấp xỉ trên theo B:

B(X)={x∈U∣[x]B​∩X=∅}

→ Tập các đối tượng có thể thuộc X.

Nếu:

B​(X)=B(X)

thì X là một tập chính xác (crisp set). Ngược lại, X được gọi là tập thô (rough set).

* Biên mờ (Boundary region):

 BNB​(X)=B(X)−B​(X)

→ Vùng không chắc chắn, thể hiện sự mơ hồ của dữ liệu.

Nếu BNB(X)=∅BN\_B(X) = \varnothingBNB​(X)=∅ thì XXX là một tập chính xác (crisp set). Ngược lại, XXX là một tập thô (rough set).

2.5. Giảm thuộc tính: Reduct và Core

Trong hệ quyết định, nhiều thuộc tính có thể dư thừa. RST đưa ra khái niệm:

* Reduct: Reduct: tập con tối thiểu của C vẫn giữ được khả năng phân biệt các đối tượng giống như tập C ban đầu.
* Core: giao của tất cả các reduct, gồm các thuộc tính cốt lõi, không thể loại bỏ.

Ý nghĩa:

* Reduct giúp giảm chiều dữ liệu mà không mất thông tin quan trọng.
* ore chỉ ra những thuộc tính bắt buộc trong mô hình.

Ví dụ: Trong bảng trên, nếu thuộc tính “Đau đầu” không ảnh hưởng đến phân loại bệnh, thì reduct có thể chỉ cần {Sốt, Ho}. Khi đó “Đau đầu” là dư thừa.

2.6. Luật quyết định (Decision Rules)

Từ hệ quyết định, ta có thể sinh ra các luật ở dạng:

(a1​=v1​∧a2​=v2​∧…)⇒(d=vd​)

Ví dụ:

* (Sốt = Cao ∧ Ho = Có) ⇒ (Bệnh = Cúm).
* (Sốt = Thấp ∧ Ho = Có) ⇒ (Bệnh = Cảm lạnh).

Các luật này hỗ trợ phân loại, dự đoán và ra quyết định.

2.7. Vai trò của lý thuyết tập thô trong trích chọn dữ liệu

* Cho phép giảm số chiều của dữ liệu, loại bỏ thuộc tính dư thừa.
* Xây dựng được tập luật quyết định đơn giản, dễ hiểu.
* Giúp cải thiện hiệu suất và độ chính xác của các thuật toán học máy.
* Không cần thêm thông tin ngoài dữ liệu gốc (không yêu cầu xác suất hoặc lý thuyết mờ).
* Có thể kết hợp với các kỹ thuật khác như mạng nơ-ron, SVM, hoặc thuật toán di truyền để tăng hiệu quả.

**CHƯƠNG 3: PHÂN TÍCH VÀ TRÌNH BÀY ỨNG DỤNG VỀ LÝ THUYẾT TẬP THÔ**

3. Trình bày ví dụ

Ví dụ minh họa cho lý thuyết tập thô (Rough Set)

Dữ liệu (bảng quan sát)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Đối tượng | Sốt | Ho | Đau đầu | Bệnh (d) |
| u1 | Cao | Có | Có | Cúm |
| u2 | Cao | Có | Không | Cúm |
| u3 | Thấp | Có | Có | Cảm lạnh |
| u4 | Thấp | Không | Có | Khỏe |

* Tập đối tượng: U = {u1, u2, u3, u4}
* Tập thuộc tính điều kiện: A = {Sốt, Ho, Đau đầu}
* Thuộc tính quyết định: d = Bệnh
* Ví dụ hệ quyết định: DS = (U, A ∪ {d})

Bước 1 — Chọn tập thuộc tính điều kiện để xét

Ví dụ ta xét B = {Sốt, Ho} (bỏ tạm Đau đầu).

Dựa trên B, ta nhóm các đối tượng thành các lớp tương đương theo giá trị (Sốt, Ho):

* [u1]\_B = {u1, u2} vì cả hai đều có (Cao, Có).
* [u3]\_B = {u3} vì (Thấp, Có) chỉ có u3.
* [u4]\_B = {u4} vì (Thấp, Không) chỉ có u4.

Phân hoạch U theo B: {{u1,u2}, {u3}, {u4}}.

Bước 2 — Xác định tập quan tâm X (ví dụ)

Chọn tập đối tượng có Bệnh = Cúm:

X = { u | d(u) = "Cúm" } = {u1, u2}

Ta tính xấp xỉ dưới và xấp xỉ trên theo B:

Xấp xỉ dưới (lower approximation):  
  
 B\_\*(X) = { x ∈ U | [x]\_B ⊆ X }

* Ở đây [u1]\_B = {u1,u2} ⊆ X và [u2]\_B = {u1,u2} ⊆ X ⇒  
   B\_\*(X) = {u1, u2}.

Xấp xỉ trên (upper approximation):  
  
 B^\*(X) = { x ∈ U | [x]\_B ∩ X ≠ ∅ }

* Các lớp có giao với X là {u1,u2} ⇒  
   B^\*(X) = {u1, u2}.
* Vùng biên (boundary) = B^\*(X) \ B\_\*(X) = ∅ → tập X là chính xác (crisp) theo B.

Bước 3 — Quan sát về thuộc tính

So sánh sử dụng cả ba thuộc tính A = {Sốt, Ho, Đau đầu} và chỉ B = {Sốt, Ho}:

* Với B = {Sốt, Ho} ta đã phân biệt được đầy đủ các giá trị quyết định trong ví dụ (Cúm, Cảm lạnh, Khỏe).
* Thuộc tính Đau đầu trong bộ A không cần thiết để phân biệt các lớp bệnh trong ví dụ này — tức là nó có thể dư thừa.

Bước 4 — Luật quyết định rút ra (từ B)

Từ phân hoạch theo {Sốt, Ho} ta có thể viết các luật quyết định dạng if-then:

1. Nếu Sốt = Cao và Ho = Có ⇒ Bệnh = Cúm.
2. Nếu Sốt = Thấp và Ho = Có ⇒ Bệnh = Cảm lạnh.
3. Nếu Sốt = Thấp và Ho = Không ⇒ Bệnh = Khỏe.

4. Phân tích về phần tìm hiểu

* Phân tích – tìm hiểu để làm rõ ý nghĩa và cách áp dụng lý thuyết tập thô (Rough Set Theory – RST).

1. Mục tiêu của việc tìm hiểu

* Xử lý dữ liệu không chắc chắn, dư thừa: Trong thực tế, dữ liệu thường có thông tin trùng lặp, không đầy đủ hoặc khó phân biệt. RST được dùng để làm rõ đâu là thuộc tính quan trọng (cốt lõi), đâu là thuộc tính dư thừa.
* Sinh luật quyết định: Từ dữ liệu ban đầu có thể rút ra những luật dạng “If … then …” để hỗ trợ phân loại hoặc dự đoán.

2. Qua ví dụ

1. Phân hoạch dữ liệu:  
     
   * Tập {Sốt, Ho} đủ để phân biệt bệnh → tức là có thể giảm thuộc tính.
   * Tập {Sốt, Ho, Đau đầu} vẫn đúng nhưng rườm rà hơn.
2. Khái niệm xấp xỉ (Approximation):  
     
   * Với tập bệnh = Cúm, ta thấy xấp xỉ dưới = xấp xỉ trên → đó là tập chính xác (crisp).
   * Nếu dữ liệu có chồng chéo (ví dụ, cùng triệu chứng dẫn đến nhiều loại bệnh khác nhau), khi đó vùng biên ≠ ∅ → xuất hiện tập thô (rough set).
3. Giảm thuộc tính (Reduct và Core):  
     
   * Đau đầu không ảnh hưởng đến quyết định → bị loại bỏ.
   * Sốt và Ho là thuộc tính cốt lõi (core).
4. Sinh luật quyết định:  
     
   * Các luật rút ra từ reduct {Sốt, Ho} ngắn gọn, chính xác và dễ áp dụng.

3. Ý nghĩa của phân tích

* Trong lý thuyết: RST giúp phát hiện thuộc tính quan trọng, làm dữ liệu gọn nhẹ mà không mất thông tin quyết định.
* Trong ứng dụng: Ví dụ với bài toán y tế, ta có thể xây dựng hệ chuyên gia hỗ trợ chẩn đoán nhanh bệnh dựa trên một vài triệu chứng cốt lõi thay vì hỏi quá nhiều câu.

5. Trình bày kỹ thuật

Trong phần này, ta trình bày kỹ thuật (quy trình, công thức) của Lý thuyết Tập thô (Rough Set Theory – RST) dựa trên ví dụ đã đưa.

1. Hệ thông tin (Information System)

Một hệ thông tin được biểu diễn:

IS=(U,A,V,f)IS = (U, A, V, f)IS=(U,A,V,f)

* UUU: tập hữu hạn các đối tượng (universe).
* AAA: tập hữu hạn các thuộc tính.
* V=⋃a∈AVaV = \bigcup\_{a \in A} V\_aV=⋃a∈A​Va​: tập giá trị thuộc tính.
* f:U×A→Vf: U \times A \to Vf:U×A→V: hàm thông tin, ánh xạ mỗi cặp (đối tượng, thuộc tính) → giá trị.

Khi có thêm thuộc tính quyết định d, ta có hệ quyết định:

DS=(U,A∪{d},V,f)DS = (U, A \cup \{d\}, V, f)DS=(U,A∪{d},V,f)

2. Quan hệ tương đương & Phân hoạch

Cho tập thuộc tính B⊆A

Hai đối tượng x,y∈U gọi là tương đương theo B nếu:  
 ∀a∈B:f(x,a)=f(y,a)

Quan hệ tương đương IND(B)  sinh ra phân hoạch U/IND(B), chia dữ liệu thành các lớp đối tượng có cùng giá trị trên B.

3. Xấp xỉ tập hợp

Cho X⊆U, với tập thuộc tính B:

* Xấp xỉ dưới (lower approximation):  
   B∗(X)={x∈U∣[x]B⊆X} → tập đối tượng chắc chắn thuộc X.
* Xấp xỉ trên (upper approximation):  
   B∗(X)={x∈U∣[x]B∩X≠∅} → tập đối tượng có thể thuộc X.
* Vùng biên (boundary region):  
   B∗(X)∖B∗(X)

 Nếu vùng biên ≠ ∅ → X là tập thô (rough set).

4. Giảm thuộc tính (Attribute Reduction)

* Một reduct là tập con tối thiểu R⊆A

 sao cho:  
 IND(R)=IND(A)

 tức là phân hoạch theo R giống phân hoạch theo toàn bộ A.

* Core = giao của tất cả reducts. Đây là tập thuộc tính cốt lõi, không thể bỏ.

5. Sinh luật quyết định (Decision Rules)

* Từ các lớp tương đương trong reduct, ta sinh luật dạng:

(Điều kiện trên R)⇒(Quyết định d)

* Các luật được chia thành:  
    
  + Luật chắc chắn (certain rule): khi lớp tương đương hoàn toàn thuộc về một quyết định.
  + Luật có thể (possible rule): khi lớp tương đương chứa nhiều quyết định khác nhau.

6. Quy trình áp dụng RST (tóm tắt kỹ thuật)

1. Xây dựng hệ thông tin từ dữ liệu.
2. Xác định quan hệ tương đương từ tập thuộc tính.
3. Tính xấp xỉ dưới/trên cho các tập quyết định.
4. Xác định reduct & core để giảm thuộc tính dư thừa.
5. Sinh luật quyết định từ reduct.
6. Ứng dụng các luật này cho phân loại, dự đoán hoặc phân tích dữ liệu.