

DỰ ĐOÁN HƯỚNG NGỮ NGHĨA CỦA CỤM TỪ TRONG KHAI PHÁ
QUAN ĐIỂM VỚI ĐỘ ĐO THÔNG TIN TƯƠNG HỖ

IDENTIFYING THE SEMANTIC ORIENTATION OF PHRASES IN OPINION
MINING WITH MUTUAL INFORMATION MEASURES

Tác giả: Tác giả: Trần Uyên Trang, Hoàng Thị Thanh Hà, Huỳnh Xuân Hiệp

Trường Đại học Sư phạm - Đại học Đà Nẵng ; trang.tranuyen@gmail.com
Trường Đại học Kinh tế - Đại học Đà Nẵng; ha.htt@due.edu.vn
Trường Đại học Cần Thơ; hxhiep@ctu.edu.vn

Tóm tắt:

Khai phá quan điểm được sử dụng để tìm hiểu, nắm bắt được xu thế, quan điểm của cộng đồng về một đối tượng cụ thể. Khai phá quan điểm liên quan đến việc xây dựng hệ thống thu thập và phân loại quan điểm. Yếu tố quan trọng trong các tác vụ chính của khai phá qua điểm là khả năng dự đoán được hướng ngữ nghĩa của từ hoặc cụm từ được trích rút trong văn bản. Khả năng này có thể đạt được bằng cách sử dụng các độ đo và mô hình khác nhau. Bài báo này đề xuất áp dụng các độ đo thông tin tương hỗ vào mô hình dự đoán hướng ngữ nghĩa của cụm từ được trích rút từ tài liệu để đo độ phụ thuộc giữa các cụm từ với các từ hạt nhân, từ đó xác định hướng quan điểm của cụm từ trong bài toán phân loại quan điểm mức tài liệu.

Từ khóa : Khai phá quan điểm; Hướng ngữ nghĩa; Cụm từ quan điểm; Độ đo PMI; Độ đo thông tin tương hỗ.

Abstract:

Opinion mining or sentiment analysis is used to find and understand the trends and attitudes of the community about a particular product. Opinion mining involves in building a system for collecting and categorizing opinions. The important task in opinion mining is the ability to predict the semantic orientation of words or phrases extracted in the text. This can be achieved by using different measures and models. This paper proposes applying mutual information measures to the semantic estimation model of extracted phrases in the document to measure the dependence between phrases and seed words to determine the opinion orientation of phrases in document-level sentiment classification.

Key words: Opinion mining; Semantic orientation; Opinion phrases; PMI measure; Mutual information measure.