So với các nghiên cứu khác có liên quan đến đề tài. Đề tài của chúng tôi có tập đặc trưng được chọn lọc, thử nghiệm nhiều lần. Đồng thời, khác với việc sử dụng kết hợp 5 bộ từ điển tiếng Anh khác nhau để xây dựng 1 bộ từ điển cảm xúc tiếng Việt. Chúng tôi xây dựng bộ từ điển cảm xúc tiếng Việt dựa trên duy nhất 1 bộ từ điển cảm xúc tiếng Anh và bổ sung khá nhiều cho bộ từ điển tiếng Việt này những từ đồng nghĩa tương đương và những từ phổ biến thường được sử dụng. Ngoài ra, việc đánh trọng số cảm xúc cho bộ từ điển cảm xúc tiếng Việt từ -5 đến +5 sẽ giúp đánh giá giá trị cảm xúc của câu chi tiết hơn là những điều cơ bản giúp nâng cao độ chính xác trong việc phân loại cảm xúc tiếng Việt.

**DỮ LIỆU TEXT TIẾNG VIỆT**

**RÚT TRÍCH ĐẶC TRƯNG**

**GÁN NHÃN TỪ LOẠI**

**BỘ TỪ ĐIỂN CẢM XÚC TIẾNG VIỆT**

**SUPPORT VECTOR MACHINE**

**TIỀN XỬ LÝ DỮ LIỆU**

**BỘ DỮ LIỆU HUẤN LUYỆN**

**DỮ LIỆU ĐÃ ĐƯỢC PHÂN LỚP**

Rút trích đặc trưng được chia làm 2 phần lần lượt thực hiện theo thứ tự phân loại chủ quan rồi đến phân loại cảm xúc:

* Modul rút trích đặc trưng phân loại chủ quan

***Input:*** tập tin tagger và bộ từ điển SO-CAL tiếng Việt.

***Output:*** tập tin có cấu trúc vector, với mỗi dòng là 01 vector đặc trưng.

**Các thao tác áp dụng:**

Với mỗi câu trong bộ dữ liệu, rút trích các giá trị

*1) Tổng số từ.*

*2) Tổng giá trị cảm xúc của các tính từ.*

*3) Tổng giá trị cảm xúc của các trạng từ.*

*4) Tổng giá trị cảm xúc của các danh từ.*

*5) Tổng giá trị cảm xúc của các động từ.*

*6) Giá trị cảm xúc của cả câu:*

*Nếu câu thuộc câu nghi vấn hoặc câu điều hiện thì trả về 0*

*Ngược lại, trả về tổng của các đặc trưng số 2, 3, 4 và 5.*

Trả về vector đặc trưng

* Modul rút trích đặc trưng phân loại cảm xúc

***Input:*** tập tin tagger và bộ từ điển SO-CAL tiếng Việt.

***Output:*** tập tin có cấu trúc vector, với mỗi dòng là 01 vector đặc trưng.

**Các thao tác áp dụng:**

Với mỗi câu trong bộ dữ liệu, rút trích các giá trị

*1) Tổng giá trị cảm xúc của các tính từ.*

*2) Tổng giá trị cảm xúc của các trạng từ.*

*3) Tổng giá trị cảm xúc của các danh từ.*

*4) Tổng giá trị cảm xúc của các động từ.*

*5) Giá trị cảm xúc của cả câu: tổng của các đặc trưng số 2, 3, 4 và 5.*

*6) Giá trị cảm xúc phụ thuộc vào từ tăng cường.*

*7) Giá trị cảm xúc phụ thuộc vào từ liên kết mang nghĩa trái ngược.*

*8) Giá trị cảm xúc phụ thuộc vào từ khiếm khuyết.*

*9) Giá trị cảm xúc của câu có xu hướng tích cực.*

*10) Giá trị cảm xúc phụ thuộc vào từ phủ định thay đổi.*

Trả về vector đặc trưng