

Xây dựng mô hình phân tích cảm xúc dựa trên mối quan hệ phụ thuộc ngữ nghĩa

Trần Khải Thiện, Phan Thị Tươi

Tóm tắt — Trong bài báo này, các tác giả đề xuất một mô hình hiệu quả cho bài toán phân tích cảm xúc mức khía cạnh. Thứ nhất, từ điển cảm xúc và các luật văn phạm phụ thuộc được kết hợp để trích xuất các cặp từ (cảm xúc – khía cạnh). Sau đó, nhờ vào ontology, các khía cạnh được gom nhóm và xác định độ đo cảm xúc cho từng khía cạnh. Những điểm đặc biệt của công trình được tóm tắt như sau: 1) để “làm mịn” thang độ cảm xúc, các tác giả đã tiến hành trích xuất theo cụm từ (thay vì chỉ xét từ đơn lập) và 2) bên cạnh việc trích xuất các tính từ như nhiều nghiên cứu khác, các danh từ và động từ có trong từ điển cảm xúc cũng được xét đến. Các thực nghiệm cho thấy hệ thống cho những kết quả khả quan.

Từ khóa — phân tích cảm xúc, khai phá quan điểm, luật văn phạm phụ thuộc, biểu thức chính quy, mô hình phân tích cảm xúc.

1 GIỚI THIỆU

Bài toán khai phá quan điểm (hay còn gọi là “phân tích cảm xúc” [1]) có nhiệm vụ xác định các yếu tố tình cảm, các đánh giá đến các đối tượng sản phẩm, dịch vụ, nhân vật hay các chủ đề qua ba mức: văn bản, câu, và khía cạnh. Hiện nay đã có nhiều công trình làm về phân tích cảm xúc cho các ngôn ngữ phổ biến như tiếng

Anh, tiếng Hoa và tiếng Nhật nhưng với tiếng Việt, đây vẫn là một chủ đề “nóng” cần giải quyết. Trong bài báo này, các tác giả đề xuất mô hình cho bài toán phân tích cảm xúc mức khía cạnh tiếng Việt bằng cách kết hợp giữa luật của văn phạm phụ thuộc (gọi là luật phụ thuộc), biểu thức chính quy, ontology và từ điển cảm xúc. Để “làm mịn” thang độ cảm xúc, các tác giả đã tiến hành rút trích cụm từ mang cảm xúc (thay vì chỉ từ) bằng cách xét chọn cả các từ loại là phó từ. Các công trình [2,3] đã cho thấy phó từ đóng vai trò quan trọng trong sự dịch chuyển thang độ cảm xúc của cả cụm từ. Đối với các từ mang cảm xúc, ngoài việc trích chọn tính từ như những cách tiếp cận khác, các danh từ, động từ có điểm số cảm xúc cũng được xét chọn. Theo hiểu biết của các tác giả, đây là công trình đầu tiên áp dụng văn phạm phụ thuộc cho phân tích cảm xúc của ý kiến trong tiếng Việt.

Phần còn lại của bài báo được tổ chức như sau: Trong Phần 2, các công trình liên quan được giới thiệu. Phần 3, mô hình rút trích đặc trưng và từ mang cảm xúc được trình bày. Phần 4, các tác giả tập trung vào đánh giá thực nghiệm. Cuối cùng là kết luận và kế hoạch nghiên cứu tiếp theo.

2 CÔNG TRÌNH LIÊN QUAN

Phân tích cảm xúc mức khía cạnh đi vào phân tích các biểu hiện cảm xúc của người dùng nhằm đến các đối tượng là các khía cạnh, đặc trưng hay thuộc tính của một hay nhiều thực thể trong một văn bản cho trước. Phân tích ở mức này có ba tác vụ: 1) xác định và trích xuất các đối tượng được nhận xét, 2) xác định quan điểm và 3) gom nhóm các đối tượng đồng nghĩa. G. Qiu và cộng sự [4,5] tiếp cận phương pháp xác định đối tượng dựa vào quan sát sự phụ thuộc giữa các từ thay vì chỉ quan tâm đến vị trí của từ trong câu như nhiều nghiên cứu khác. Các tác giả này đề xuất giải thuật “truyền kép” để trích xuất đối tượng và từ cảm xúc

Bản thảo nhận ngày 10 tháng 4 năm 2017, hoàn chỉnh sửa chữa ngày 05 tháng 10 năm 2017.

Nghiên cứu được tài trợ bởi Đại học Quốc gia Thành phố Hồ Chí Minh (ĐHQG-HCM) trong khuôn khổ Đề tài mã số C2016-20-32.

Trần Khải Thiện. Khoa Khoa học và Kỹ thuật Máy tính, Trường Đại học Bách Khoa - ĐHQG-HCM. Số 268 Lý Thường Kiệt, Phường 14, Quận 10, Hồ Chí Minh
(E-mail: thientk@cse.hcmut.edu.vn)

Phan Thị Tươi. Khoa Khoa học và Kỹ thuật Máy tính, Trường Đại học Bách Khoa - ĐHQG-HCM. Số 268 Lý Thường Kiệt, Phường 14, Quận 10, Hồ Chí Minh
(E-mail: tuoi@cse.hcmut.edu.vn)

từ việc quan sát mối quan hệ giữa chúng. Các mối quan hệ này được xác định bởi bộ phân tích cú pháp của văn phạm phụ thuộc. Ban đầu, các từ “mồi” được sử dụng để trích xuất các từ mang cảm xúc cùng các đối tượng. Các từ mang quan điểm và các đối tượng mới này lại được dùng để trích xuất các từ mang quan điểm và các đối tượng tiếp theo. Quá trình được tiếp diễn cho đến khi không tìm được từ mang quan điểm nào khác nữa. Giải thuật “truyền kép” đã giúp rút trích được nhiều đối tượng. Tuy nhiên, phương pháp này không thích hợp cho việc phát hiện một số đối tượng quan trọng mà không được đề cập trực tiếp bởi các từ mang quan điểm. Công trình [6] sử dụng mẫu “part-whole” và “no” để xác định các đối tượng bị bỏ qua bởi giải thuật này. Ngoài ra còn một số công trình đáng lưu ý sử dụng văn phạm phụ thuộc để rút trích đối tượng và từ mang cảm xúc như [7 - 9].

3 MÔ HÌNH RÚT TRÍCH ĐỐI TƯỢNG – TỪ MANG CẢM XÚC.

Chúng tôi đề xuất mô hình rút trích đối tượng – từ mang cảm xúc cho bài toán phân tích cảm xúc mức khía cạnh. Điểm đặc biệt của mô hình này là chúng tôi đã kết hợp giữa ontology, từ điển cảm xúc và tập các luật rút trích được xây dựng dựa trên mối quan hệ phụ thuộc ngữ nghĩa giữa các từ trong câu của văn bản tiếng Việt cho trước.

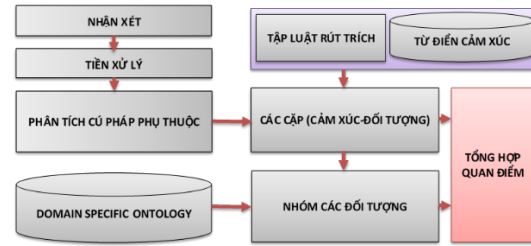
3.1 Định nghĩa

Theo B. Liu [10], một quan điểm là một bộ 5, gồm: $(E_i, A_{ij}, S_{ijkl}, H_k, T_i)$ với E_i : tên của thực thể, A_{ij} : khía cạnh E_i , S_{ijkl} là ý kiến cảm xúc về khía cạnh A_{ij} của thực thể E_i cho bởi chủ thể H_k tại thời điểm T_i , H_k là chủ thể thể hiện ý kiến, T_i thời gian thể hiện ý kiến của chủ thể H_k . Ở bài báo này, chúng tôi sử dụng thuật ngữ “đối tượng” để nói về cho khía cạnh A_{ij} và thực thể E_i .

3.2 Mô hình

Mô hình rút trích đối tượng – từ mang cảm xúc được mô tả ở Hình 1. Trong mô hình, các nhận xét được tiền xử lý qua việc loại bỏ các ký tự đặc biệt, các “stop word”. Dữ liệu sau đó được cho vào bộ phân tích cú pháp phụ thuộc. Trong công trình này, bộ phân tích cú pháp VNDP (Vietnamese dependency parsing toolkit) [11] được lựa chọn sử dụng. Đầu ra của giai đoạn này là các cây cú pháp của văn phạm phụ thuộc chứa các quan hệ phụ thuộc và các nhãn từ loại của các từ trong câu. Các

luật rút trích được áp dụng và kết hợp với từ điển cảm xúc để trích chọn ra các cặp từ (cảm xúc– đối tượng). Thao tác gom nhóm các đối tượng dựa vào ontology được tiến hành sau đó. Kết quả của cả quá trình là tổng hợp các quan điểm của từng đối tượng. Các công đoạn trong mô hình được trình bày chi tiết như sau: Tập luật rút trích được mô tả tại mục 3.3. Mục 3.4 giới thiệu về từ điển cảm xúc tiếng Việt. Mục 3.5 trình bày cách các tác giả xây dựng ontology.



Hình 1. Mô hình rút trích đối tượng – từ cảm xúc.

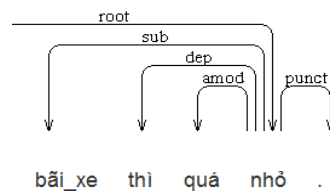
3.3 Tập luật rút trích

Chúng tôi xây dựng tập luật rút trích dựa trên sự phụ thuộc ngữ nghĩa giữa các từ trong câu. Các cặp từ loại được xét đến gồm danh từ - tính từ, danh từ - động từ, danh từ - danh từ, tính từ - tính từ, động từ - động từ, phó từ - tính từ. Có 12 luật được đề xuất và cài đặt như sau:

Luật 1: Nếu có cụm danh từ là chủ ngữ (sub) tác động trực tiếp đến tính từ mang cảm xúc thì rút trích cặp từ này.

$$T \rightarrow T\text{-Dep} \rightarrow S, \\ S \in \text{SentiDict}, T\text{-Dep} \in \{\text{sub}\}, \\ \text{POS}(T) \in \{N\}, \text{POS}(S) \in \{A\}$$

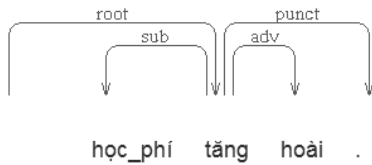
Ví dụ 1: bãi xe thì quá nhỏ. Output: (nhỏ - bãi xe)



Luật 2: Nếu có cụm danh từ là chủ ngữ (sub) tác động trực tiếp đến động từ mang cảm xúc thì rút trích cặp từ này.

$$T \rightarrow T\text{-Dep} \rightarrow S, S \in \text{SentiDict}, T\text{-Dep} \in \{\text{sub}\}, \\ \text{POS}(T) \in \{N\}, \text{POS}(S) \in \{V\}$$

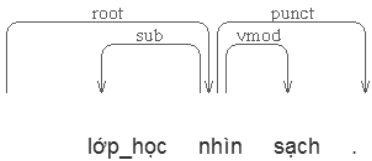
Ví dụ 2: học phí tăng hoài. Output: (tăng – học phí)



Luật 3: Nếu có cụm danh từ T là subject của một từ H và H có mối quan hệ verb modifier (vmod) với từ mang cảm xúc S thì rút trích cặp từ (S-T).

$T \rightarrow T\text{-Dep} \rightarrow H \leftarrow S\text{-Dep} \leftarrow S,$
 $S \in \text{SentiDict}, T\text{-Dep} \in \{\text{sub}\},$
 $S\text{-Dep} \in \{\text{vmod}\}, \text{POS}(T) \in \{\text{N}\}, \text{POS}(S) \in \{\text{A}\}$

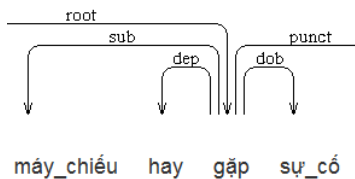
Ví dụ 3: lớp học nhìn sạch. Output: (sạch – lớp học)



Luật 4: Nếu có cụm danh từ T là subject của một từ H và H có mối quan hệ direct object (dob) với từ mang cảm xúc S thì rút trích cặp từ (S-T).

$T \rightarrow T\text{-Dep} \rightarrow H \leftarrow S\text{-Dep} \leftarrow S$
 $S \in \text{SentiDict}, T\text{-Dep} \in \{\text{sub}\},$
 $S\text{-Dep} \in \{\text{dob}\}, \text{POS}(T) \in \{\text{N}\}, \text{POS}(S) \in \{\text{N}\}$

Ví dụ 4: máy chiếu hay gặp sự cố. Output: (sự cố - máy chiếu).

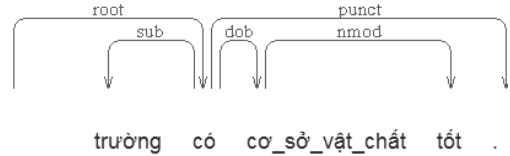


Luật 5: Nếu một tính từ mang cảm xúc có mối liên hệ noun modifier với cụm danh từ thì rút trích cặp từ này.

$S \rightarrow S\text{-Dep} \rightarrow T, S \in \text{SentiDict},$

$S\text{-Dep} \in \{\text{nmod}\}, \text{POS}(T) \in \{\text{N}\}, \text{POS}(S) \in \{\text{A}\}$

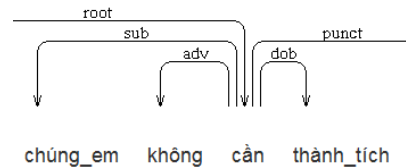
Ví dụ 5: trường có cơ sở vật chất tốt. Output: (tốt – cơ sở vật chất)



Luật 6: Nếu một cụm danh từ có mối liên hệ direct object với động từ mang cảm xúc thì rút trích cặp từ này.

$T \rightarrow T\text{-Dep} \rightarrow S, S \in \text{SentiDict},$
 $T\text{-Dep} \in \{\text{dob}\}, \text{POS}(T) \in \{\text{N}\}, \text{POS}(S) \in \{\text{V}\}$

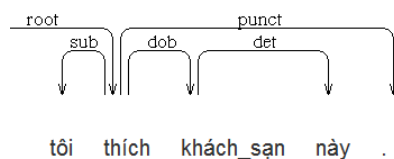
Ví dụ 6: chúng em không cần thành tích. Output: (không cần – thành tích).



Luật 7: Nếu một cụm danh từ có mối liên hệ direct object (dob) với tính từ mang cảm xúc thì rút trích cặp từ này.

$T \rightarrow T\text{-Dep} \rightarrow S, S \in \text{SentiDict},$
 $T\text{-Dep} \in \{\text{dob}\}, \text{POS}(T) \in \{\text{N}\}, \text{POS}(S) \in \{\text{A}\}$

Ví dụ 7: tôi thích khách sạn này. Output: (thích – khách_sạn).



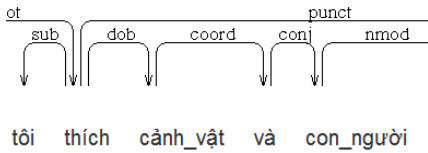
Luật 8: Nếu nhiều đối tượng được nối nhau bởi các liên từ “và”, “hoặc”, “hay” thì trích chọn các đối tượng này cùng từ cảm xúc đề cập đến chúng.

$T_i \rightarrow T_i\text{-Dep} \rightarrow H \rightarrow T_j\text{-Dep} \rightarrow T_j$

$T_i\text{-Dep} \in \{\text{coord}\},$

$T_j\text{-Dep} \in \{\text{conj}\}$, $\text{POS}(T_{i(j)}) \in \{N\}$

Ví dụ 8: tôi thích cảnh vật và con người nơi này. Output: (thích – cảnh_vật), (thích – con_người).



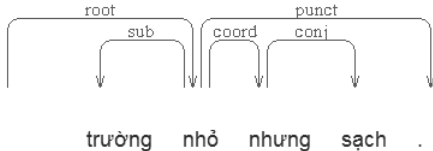
Luật 9: Nếu nhiều từ mang cảm xúc được nối nhau bởi các từ “nhưng”, “tuy”, “tuy nhiên” ... thì trích chọn các từ này cùng target mà chúng đề cập đến.

$S_i \rightarrow S_i\text{-Dep} \rightarrow H \rightarrow S_j\text{-Dep} \rightarrow S_j$

$S_{i(j)} \in \text{SentiDict}$, $S_i\text{-Dep} \in \{\text{conj}\}$,

$S_j\text{-Dep} \in \{\text{coord}\}$, $\text{POS}(S_{i(j)}) \in \{A\}$

Ví dụ 9: trường nhỏ nhưng đẹp. Output: (nhỏ - trường), (đẹp - trường).



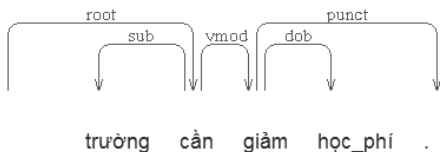
Luật 10: Nếu trong câu có động từ mang quan điểm có mối liên hệ verb modifier với động từ khác thì rút trích hai động từ này cùng target mà chúng đề cập đến.

$S_i \rightarrow S\text{-Dep} \rightarrow S_j \leftarrow T\text{-Dep} \leftarrow T$

$S_{i(j)} \in \text{SentiDict}$, $\text{POS}(S_{i(j)}) \in \{V\}$,

$S\text{-Dep} \in \{\text{vmod}\}$, $T\text{-Dep} \in \{\text{dob}\}$

Ví dụ 10: trường cần giảm học phí. Output: (cần giảm - học phí).

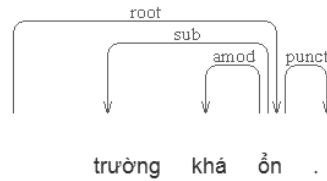


Luật 11: Nếu trong câu xuất hiện phó từ (phụ từ, adv) thì trích từ này cùng từ cảm xúc mà nó bổ nghĩa.

$RC \rightarrow RC\text{-Dep} \rightarrow S$, $\text{POS}(RC) \in \{R\}$,

$\text{POS}(S) \in \{A\}$, $RC\text{-Dep} \in \{\text{amod}\}$

Ví dụ 11: trường khá ổn. Output: (khá ổn - trường).



Luật 12: Nếu trong câu xuất hiện mẫu “không” thì rút trích cặp “không” và cụm danh từ.

Ví dụ 12: không máy lạnh, không quạt. Output: (không - máy lạnh), (không - quạt)

Trong đó: SentiDict là từ điển cảm xúc; S là từ mang cảm xúc; T là đối tượng mà cảm xúc chỉ đến; S-Dep (hoặc T-Dep) là quan hệ phụ thuộc của S (hoặc T); POS(S) hoặc POS(T) là nhãn từ loại của S hoặc T; N, A, V, R tương ứng là từ loại động từ, tính từ, danh từ, phó từ; và sub, amod, dob, nmod, vmod, sub, conj, coord là các nhãn quan hệ phụ thuộc theo VNDP [11].

3.4 Từ điển cảm xúc (SentiDict)

Chúng tôi tiến hành xây dựng thủ công từ điển cảm xúc khoảng 1.500 từ (term), gồm các động từ và tính từ. Số lượng từ là tương đối đủ để chúng tôi tiến hành các thử nghiệm. Các từ này được thu thập từ tập ngữ liệu ý kiến của 420 sinh viên HUFLIT và từ công bố [12]. Để xác định trọng số các từ mang quan điểm này, chúng tôi chuyển ngữ chúng sang tiếng Anh và gán trọng số của chúng bằng với trọng số cao nhất của từ tiếng Anh tương ứng có trong bộ từ điển cảm xúc tiếng Anh, SentiWordNet [13]. Từ điển cảm xúc được lưu trữ trong cơ sở dữ liệu với các trường chính như sau: term, điểm số tích cực, điểm số tiêu cực và từ loại. Bảng 1 trình bày một số từ cảm xúc có trong từ điển.

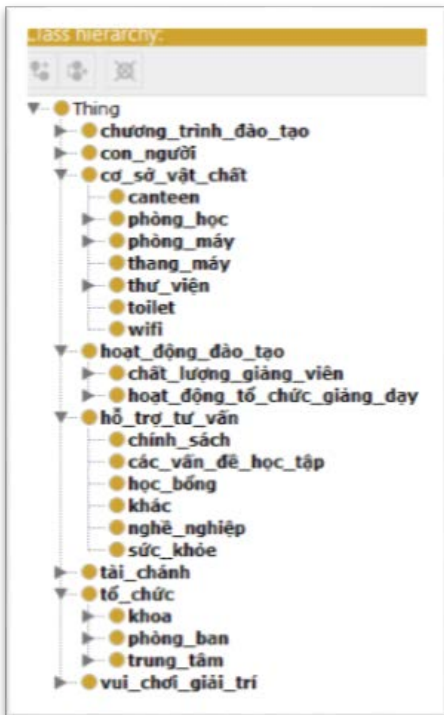
Bảng 1. Cấu trúc từ điển cảm xúc tiếng Việt.

Term	Điểm số tích cực	Điểm số tiêu cực	Từ loại
đẹp	0,875	0	tính từ
xấu	0	0,875	tính từ

3.5 Ontology

Cho mục đích thử nghiệm, chúng tôi xây dựng một ontology cho trường đại học để quản lý các

mối quan hệ ngữ nghĩa giữa các thực thể của trường đại học và các thuộc tính của chúng. Bằng cách xác định các mối quan hệ ngữ nghĩa, chúng ta có thể dễ dàng gom nhóm các đối tượng đã thu thập được tại mục 3.3. Trong bài báo này, các tác giả xây dựng một ontology cung cấp các biểu diễn tri thức về trường đại học (cụ thể là Hufalit). Hầu hết các thông tin về trường đại học được mô tả thông qua các khái niệm về thực thể, các thuộc tính, và các mối quan hệ giữa chúng. Các khái niệm được tổ chức thành một cấu trúc cây phân cấp có mỗi nút đại diện cho một khái niệm. Việc định nghĩa các khái niệm được dựa trên Bộ Tiêu chuẩn Đánh giá Chất lượng trường Đại học của Bộ GD-ĐT (QĐ 65/2007/QĐ-BGDĐT). Có tám nhóm đối tượng trong phạm vi đánh giá một trường Đại học: Chương trình đào tạo; Con người; Cơ sở vật chất; Hoạt động đào tạo; Hỗ trợ tư vấn; Tài chính; Tổ chức; Vui chơi giải trí. Phiên bản hiện tại gồm 85 khái niệm được xây dựng bằng phần mềm Protégé [14]. Một đoạn của ontology được mô tả bởi Hình 2.



Hình 2. Ontology về trường đại học.

4 ĐÁNH GIÁ.

Để xác định tính hiệu quả của hệ thống, các thực nghiệm đã được tiến hành trên tập dữ liệu

kiểm thử là 320 nhận xét thu thập thực tế từ sinh viên của trường đại học Hufalit.

Chúng tôi sử dụng ba độ đo Precision (P), Recall (R), và F-measure (F1) được định nghĩa theo công thức (1), (2), và (3). Trong đó GT là tập các từ được gán nhãn bởi chuyên gia. T là tập các từ được rút trích bởi luật.

Kết quả thực nghiệm trên mô hình đề xuất được thể hiện ở Bảng 2. Hệ thống đạt trên 81% về độ đo Precision, độ đo Recall là trên 70% và độ đo F-measure là trên 75%.

$$P = \frac{GT \cap T}{T} \tag{1}$$

$$R = \frac{GT \cap T}{GT} \tag{2}$$

$$F_1 = \frac{2 * P * R}{P + R} \tag{3}$$

Bảng 2. Kết quả thực nghiệm của hệ thống rút trích đối tượng và từ mang cảm xúc.

GT	T	GT∩T	P	R	F ₁
453	395	320	81%	70%	75%

So sánh với phương pháp biểu thức chính quy

Chúng tôi cũng tiến hành so sánh với cách tiếp cận sử dụng biểu thức chính quy ở [12] với các mẫu được cài đặt bởi GATE/JAPE như sau:

$$\langle \text{adv} \rangle \langle \text{word in dict.} \rangle \rightarrow \langle \text{sentiment word} \rangle \tag{4}$$

$$\langle \text{sentiment word} \rangle (\langle \text{conj.} \rangle \langle \text{sentiment word} \rangle)^* \rightarrow \langle \text{sentiment words} \rangle \tag{5}$$

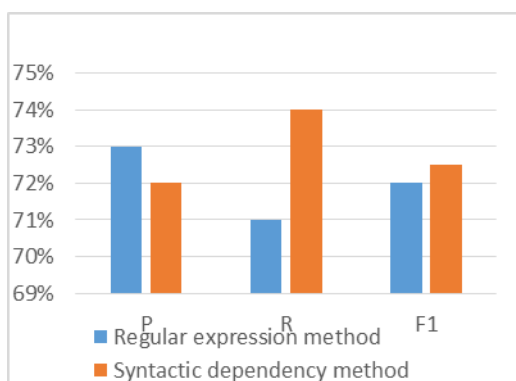
$$\langle \text{reverse word} \rangle \langle \text{positive word (negative word)} \rangle \rightarrow \langle \text{negative word (positive word)} \rangle \tag{6}$$

Trong đó: <adv> có thể là “rất”, “khá”, “siêu”... hoặc “bị”, “hơi”, “gây”; <reverse word> có thể là “không”, “chả”, “chẳng”.

Chúng tôi đánh giá hai hệ thống trong việc xác định được các từ cảm xúc và các khía cạnh mà chúng hướng đến. Chúng tôi cũng sử dụng 437 nhận xét trong [12] làm dữ liệu kiểm thử. Bảng 3 và Hình 3 biểu diễn các các kết quả so sánh.

Bảng 3. So sánh hai hệ thống.

Phương pháp	P	R	F ₁
Biểu thức chính quy	73%	71%	72%
Hệ thống	72%	74%	73%



Hình 3. So sánh hai hệ thống.

Kết quả ở Hình 3 cho thấy hệ thống sử dụng biểu thức chính quy cho kết quả tốt hơn với độ đo Precision nhưng tệ hơn Recall và F-measure so với hệ thống đề xuất trong bài báo này.

5 KẾT LUẬN.

Trong bài báo này, mô hình và cơ chế rút trích đối tượng và từ mang cảm xúc trong bài toán phân tích cảm xúc tiếng Việt đã được trình bày và cài đặt. Với sự kết hợp của ontology, từ điển cảm xúc và các luật phụ thuộc cũng như biểu thức chính quy, kết quả của giai đoạn thực nghiệm đã cho thấy tính hiệu quả của đề xuất này. Xác định đối tượng và từ mang quan điểm là nhiệm vụ trọng tâm của bài toán phân tích cảm xúc nói chung và bài toán phân tích cảm xúc mức khía cạnh nói riêng. Theo hiểu biết của chúng tôi, đây cũng là công trình đầu tiên áp dụng văn phạm phụ thuộc cho phân tích cảm xúc thực hiện cho tiếng Việt. Nghiên cứu tiếp theo, chúng tôi dự định sẽ mở rộng tập từ vựng cảm xúc cũng như mở rộng các luật rút trích. Chúng tôi cũng xem xét việc tích hợp học máy nhằm nâng cao hiệu năng của hệ thống.

TÀI LIỆU THAM KHẢO

[1] Nasukawa, T., and Yi, J., "Sentiment Analysis: Capturing Favorability Using Natural Language Processing", in *Proceedings of the 2nd International Conference on Knowledge Capture (K-CAP 2003)*, Sanibel Island, FL, USA, ACM, 2003, pp. 70-77.

[2] T.K. Tran, T.T. Phan, "Computing Sentiment Score of Adjective Phrases for Vietnam." *International Workshop on Multi-disciplinary Trends in Artificial Intelligence*. Springer, Cham, 2016, pp. 288-296.

[3] T.K. Tran; T.T. Phan, "Computing Sentiment Scores of Verb Phrases for Vietnamese." In *Proceedings of the 28th Conference on Computational Linguistics and Speech Processing (ROCLING 2016)*, 2016, pp. 204-213.

[4] G. Qiu, B. Liu, J. Bu and C. Chen, "Expanding Domain Sentiment Lexicon through Double Propagation", in *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI-09)*, California, USA, July 11-17, 2009, pp. 1199-1204.

[5] G. Qiu, B. Liu, J. Bu and C. Chen, "Opinion Word Expansion and Target Extraction through Double Propagation", *Computational Linguistics*, vol. 37, no. 1, pp. 9-27, 2011.

[6] L. Zhang, et al, "Extracting and Ranking Product Features in Opinion Documents", in *Proceedings of the 23rd International Conference on Computational Linguistics*, August 23-27, Beijing, China, 2010, pp. 1462-1470.

[7] A.Mukherjee and B. Liu, "Aspect Extraction through Semi-Supervised Modeling," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers, Vol. 1*, the Association for Computer Linguistics, Jeju Island, Korea, July 08 – 14, 2012, pp. 339-348.

[8] S. Mukherjee and S. Joshi, "Sentiment aggregation using ConceptNet ontology," in *Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP '13)*, 2013, pp. 570-578.

[9] B. Agarwal, S. Poria, N. Mittal, A.F. Gelbukh, A. Hussain, "Concept-Level Sentiment Analysis with Dependency-Based Semantic Parsing: A Novel Approach." *Cognitive Computation* vol. 7, no. 4, pp. 487-499, 2015.

[10] B. Liu, *Sentiment Analysis and Subjectivity*, Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkha and F. J. Damerau), 2010, pp.627-666.

[11] D.Q. Nguyen, S.B. Pham, P-T. Nguyen and M.L. Nguyen, "From Treebank Conversion to Automatic Dependency Parsing for Vietnamese", in *Proceedings of 19th International Conference on Application of Natural Language to Information Systems, NLDB'14*, Springer LNCS, 2014, pp. 196-207.

[12] T.K. Tran, T.T. Phan (2015), "An upgrading SentiVoice-a system for querying hotel service reviews via phone," *International Conference on Asian Language Processing (IALP)*, Suzhou, China, 2015, pp. 115-118.

[13] S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet," 2010.

[14] Stanford Center for Biomedical Informatics Research (BMIR) <http://protege.stanford.edu/> Internet: Truy cập ngày 1.4.2017.

Trần Khải Thiện

Trần Khải Thiện tốt nghiệp thủ khoa Thạc sỹ về Khoa học Máy tính vào năm 2013 tại Trường Đại học Công nghệ Thông tin (ĐHQG-HCM). Hiện nay, Trần Khải Thiện là giảng viên trường Đại học Ngoại ngữ - Tin học TP. Hồ Chí Minh và đang là nghiên cứu sinh ngành Khoa học Máy tính tại Đại học Bách khoa TP Hồ Chí Minh (ĐHQG-HCM). Hướng nghiên cứu chính là Xử lý ngôn ngữ tự nhiên và ứng dụng.

Phan Thị Tươi

Phan Thị Tươi là Giáo sư Tiến sĩ công tác tại khoa Khoa học và Kỹ thuật Máy tính, Đại học Bách

Khoa, ĐHQG-HCM. Phan Thị Tươi nhận bằng Tiến sĩ Khoa học Máy tính từ Đại học Charles, Cộng hòa Séc năm 1985. Các hướng nghiên cứu bao gồm Trình biên dịch, Truy vấn thông tin và Xử lý ngôn ngữ tự nhiên. Phan Thị Tươi là nghiên cứu viên chính của các dự án trọng điểm cấp Quốc gia và đã xuất bản nhiều bài báo trên các tạp chí và hội nghị uy tín Quốc gia và Quốc tế.

Towards a sentiment analysis model based on semantic relation analysis

Thien Khai Tran, Tui Thi Phan

Abstract — In this paper, we propose an effective model for aspect-based sentiment analysis. First, we combined a sentiment dictionary and syntactic dependency rules to extract reliable word pairs (sentiment — aspect). Then, thanks to ontology, we grouped those aspects and determined the sentiment polarity of each. When we conducted experiments on real reviews, the system showed positive results.

Index Terms — opinion mining, sentiment analysis, semantic relations, sentiment analysis model.