PHÂN LOẠI HÀNH ĐỘNG NGÔN TỪ TRONG TƯƠNG TÁC CỦA NGƯỜI DÙNG VỚI ỨNG DỤNG TRÊN ĐIỆN THOẠI DI ĐỘNG

Ngô Thị Lan[1*], Nguyễn Thị Dung, Nguyễn Lan Oanh
Trường Đại học Công nghệ Thông tin và Truyền thông – ĐH Thái Nguyên

TÓM TẮT

Trong các hệ thống tương tác với con người thông qua giọng nói như trợ lý cá nhân ảo (Apple Siri, Microsoft Cortana, Amaron Alexa, Facebook, Google Assistant, ...) hay chat-bot, việc máy có thể nhận biết được ý định của người dùng trong tương tác là vấn đề quan trọng khi xây dựng hệ thống. Xác định hành động ngôn từ tự động giúp cho hệ thống phát hiện được ý định của người dùng. Phát hiện được hành động ngôn từ trong phát ngôn của người dùng sẽ xác định được ý định của người dùng là hỏi, yêu cầu hành động, chào, thông báo…Từ đó cung cấp các chỉ dẫn hữu ích để hệ thống cải thiện hiệu quả tương tác với con người. Trong bài báo này, chúng tôi trình bày một nghiên cứu về phân loại hành động ngôn từ trong phát ngôn của người dùng và ứng dụng trong hệ thống GoldHealth360 – ứng dụng theo dõi quá trình rèn luyện sức khoẻ. Kết quả thử nghiệm cho thấy khả năng phân loại ngôn từ của mô hình có thể hoạt động tốt trên điện thoại di động.

Từ khóa: *Hành động ngôn từ, hiểu ngôn ngữ nói, hiểu ý định người dùng, nhận biết ý định người dùng, xử lý văn bản nói tiếng Việt*

GIỚI THIỆU

Trong thời đại ngày nay, điện thoại di động thông minh đã đóng vai trò quan trọng trong cuộc sống của nhiều người. Việc tương tác với điện thoại di động sử dụng giọng nói đang mang lại nhiều tiện lợi cho người dùng. Xong việc thiết kế giao diện đàm thoại sử dụng ngôn ngữ tự nhiên là một việc phức tạp, cần có nhiều nghiên cứu sâu hơn để hiểu được ý định của người dùng trong các phát ngôn của họ. Bước đầu tiên hiểu được ý định của người dùng là hiểu được hành động ngôn từ trong phát ngôn ấy. Bởi vì hành động ngôn từ (speech act – SA), theo Austin [1] người đã đề xuất lý thuyết hành động lời nói, là lực ngôn trung của phát ngôn và đặc trưng cho ý định của người nói. Phân loại SA có một vai trò quan trọng trong nhiều ứng dụng khác nhau như giải quyết sự mơ hồ trong nhận dạng ngôn ngữ nói, hạn chế không gian tìm kiếm thông tin trong hệ thống tìm kiếm sử dụng giọng nói, giúp chọn giải pháp tốt nhất khi một số bản dịch có sẵn, giúp tác tử phản hồi phù hợp với yêu cầu của người dùng và gợi ý hành động tiếp theo cho người dùng. Phân tích SA đã được nghiên cứu không chỉ trong tiếng Anh mà còn trong nhiều ngôn ngữ khác như tiếng Trung Quốc [2], Hàn Quốc [3, 4], Ả Rập [5, 6], ... Nhưng trong tiếng Việt, mới chỉ các nghiên cứu về SA trong ngôn ngữ học mà chưa có nghiên cứu về phân loại hành động ngôn từ tự động. Trong bài báo này, chúng tôi nghiên cứu về việc phân loại hành động ngôn từ trong tiếng Việt và ứng dụng tronghệ thống GoldHealth360 – một hệ thống theo dõi quá trình rèn luyện và giữ gìn sức khoẻ của người dùng.

Trên thực tế, phân loại hành động ngôn từ là một nhiệm vụ phức tạp, và một hành động ngôn từ không thể được suy ra trực tiếp từ sự diễn giải theo nghĩa đen của một lời nói mà nó còn phụ thuộc ngữ cảnh. Có ba thách thức của việc tự động nhận dạng hành động ngôn từ trong phát ngôn. Thứ nhất, mối quan hệ giữa các tính năng và nhãn hành động ngôn từ của một phát ngôn là phức tạp. Ý định của người sử dụng không phải lúc nào cũng tường minh và thể hiện rõ ràng trong lời nói. Con người có nhiều cách diễn đạt bằng lời khác nhau để thể hiện cùng một ý định. Thứ hai, nhiều biến thể trong lớp và sự phân phối của các loại hành động ngôn từ là không cân bằng. Thách thức cuối cùng là tiếng Việt thiếu các nguồn lực và kỹ thuật xử lý với

* Tel: 0943 870272, Email: ntlan@ictu.edu.vn

ngôn ngữ nói tiếng Việt. Để khắc phục các thách thức trên, trong bài báo này chúng tôi sử dụng mô hình học máy maximum entropy cho việc phân loại hành động ngôn từ.

Đóng góp chính của chúng tôi trong nghiên cứu này là:

- Thứ nhất, chúng tôi đã xây dựng một bộ ngữ liệu có gán nhãn hành động ngôn từ.

- Thứ hai, chúng tôi đã khảo sát hiệu quả của mô hình máy học maximum entropy để phân loại một phát ngôn vào các loại hành động ngôn từ.

Cấu trúc của bài báo được tổ chức như sau. Trong phần 2, chúng tôi giới thiệu tóm tắt các công trình nghiên cứu liên quan. Phần 3 trình bày về ứng dụng GoldHeath360 và ứng dụng mô hình phân loại hành động ngôn từ. Phần tiếp theo, chúng tôi mô tả về thực nghiệm phân loại hành động ngôn từ sử dụng maximum entropy. Cuối cùng, chúng tôi rút ra một số kết luận và thảo luận về công việc tương lai.

CÁC NGHIÊN CỨU LIÊN QUAN

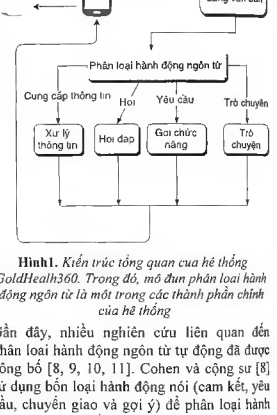
Lý thuyết về hành động ngôn từ được phát biểu đầu tiên trong ngôn ngữ học bởi Austin [1]. Sau đó học trò của ông, Searle, đã trình bày một nghiên cứu sâu hơn hành động ngôn từ [7]. Hành động ngôn từ trong ngôn ngữ học còn được dịch là hành vi ngôn ngữ hay hành động ngữ vi. Trong các nghiên cứu về hệ thống hội thoại, hành động ngôn từ còn được hiểu là hành động hội thoại (dialog act). Theo Austin và Searle, một phát ngôn có hành động ngôn từ được thể hiện trong ba loại hành động sau:

1) Hành động tạo lời: hành động sử dụng các đơn vị, các quan hệ ngôn ngữ để tạo nên phát ngôn có nghĩa.

2) Hành động tại lời (lực ngôn trung): là hành động mà đích của nó nằm ngay trong việc tạo nên phát ngôn, hiệu quả của nó đạt được ngay tại thời điểm nói. Ví dụ: ra lệnh, yêu cầu, chào, hứa hẹn …

3) Hành động mươn lời: hiên ngay tức thì thông qua phương tiện ngôn ngữ để tạo ra hiệu quả tức thì, gây ra một tác động nào đó làm biến đổi ngữ cảnh

Hành động ngôn từ được đề cập trong nghiên cứu này là hành động tại lời hay lực ngôn trung của phát ngôn.



Hình 1. *Kiến trúc tổng quan của hệ thống GoldHeath360. Trong đó, mô đun phân loại hành động ngôn từ là một trong các thành phần chính của hệ thống*

Gần đây, nhiều nghiên cứu liên quan đến phân loại hành động ngôn từ tự động đã được công bố [8, 9, 10, 11]. Cohen và cộng sự [8] sử dụng bốn loại hành động nói (cam kết, yêu cầu, chuyển giao và gợi ý) để phân loại hành động ngôn từ, kết hợp tính năng từ vựng và thông tin thời gian. Bhatia và cộng sự [9] đã tiến hành dự đoán các hành động hội thoại trong một bảng thảo luận trực tuyến về lập trình cơ sở dữ liệu và đề xuất một mô hình kết hợp các tính năng từ vựng, số bài viết trước và vị trí của bài đăng trong chủ đề. Arguello và cộng sự [10] tập trung vào dự đoán hành động hội thoại trong các bài viết trên diễn đàn trực tuyến. Họ điều tra các tính năng khác nhau (bắt nguồn từ nội dung bài viết, tác giả, và bối cảnh xung quanh) để dự đoán phát biểu của họ thuộc loại hành động ngôn từ nào. Vosoughi và cộng sự [11] tập trung vào phân loại các twice. Họ đã tiến hành phân loại hành động ngôn từ với ba mức độ chi tiết khác nhau.

Tuy nhiên, chưa có sự đồng thuận đầy đủ về các danh mục các loại hành động ngôn từ của phát ngôn được đưa ra. Các loại hành động ngôn từ được đưa ra khác nhau trong các hệ thống hội thoại hoặc lĩnh vực khác nhau [12]. Trong bài báo này chúng tôi định nghĩa một tập các loại hành động ngôn từ phù hợp với ứng dụng GoldHealth360 và áp dụng cho tiếng Việt.

DANH MỤC HÀNH ĐỘNG NGÔN TỪ

Phân loại hành động ngôn từ là phần cốt lõi của hệ thống tương tác với người dùng sử dụng ngôn ngữ nói. Kiến trúc tổng quan của hệ thống GoldHeal360 được thể hiện trong Hình 1.

Hệ thống cho phép người dùng nhập chỉ số cân nặng, chiều cao, lên kế hoạch tập luyện và ăn uống để đạt được mục tiêu giảm cân, tăng cân hoặc giữ dáng. Hàng ngày người dùng có thể nhập thông tin tập thể dục, thể thao và ăn uống của mình vào hệ thống để theo dõi quá trình luyện tập của mình. Ngoài việc tương tác với hệ thống bằng bàn phím thì người dùng có thể giao tiếp bằng giọng nói. Phát ngôn của người dùng được chuyển đổi sang dạng văn bản bằng cách sử dụng một dịch vụ chuyển đổi văn bản tự động từ giọng nói. Ở đây chúng tôi sử dụng dịch vụ Google để nhận dạng giọng nói. Phát ngôn ở dạng văn bản này được chuyển sang mô hình phân loại để xác định hành động ngôn từ. Sau đó, tuỳ theo loại hành động ngôn từ mà hệ thống có những hành động thích hợp. Ví dụ: nếu phát ngôn đầu vào là một câu hỏi, hệ thống sẽ gọi đến thành phần hỏi đáp, và hiển thị thông tin thích hợp. Nếu nó là hành động yêu cầu, hệ thống sẽ gọi mô đun phân tích yêu cầu để thực hiện chức năng mà người dùng yêu cầu. Nếu câu là thuộc loại cung cấp thông tin thì hệ thống trích xuất thông tin mà hệ thống cần tuỳ theo trạng thái hiện tại của hệ thống. Nếu câu đầu vào thuộc loại trò chuyện, hệ thống sẽ gọi mô đun chat để trả lời người dùng.

Theo yêu mục đích của ứng dụng, chúng tôi định nghĩa các loại hành động ngôn từ thành 4 loại như trong Bảng 1.

Bảng 1. *Định nghĩa các loại hành động ngôn từ*

Loại	Giải thích
Cung cấp thông tin	Phát ngôn nhằm mục đích cung cấp thông tin cho hệ thống, thường gồm các câu thông báo (ví dụ: "đi bộ từ 5 giờ đến 6 giờ"; "nặng 40 cân cao một mét sáu hai")
Hỏi	Phát ngôn nhằm mục đích yêu cầu hệ thống cung cấp thông tin (ví dụ: "Tôi nặng 50 cân cao mét rưỡi là béo hay gầy"?)
Yêu cầu	Phát ngôn nhằm yêu cầu hệ thống thực hiện một chức năng nào đó (ví dụ: "hiển thị thông tin ăn uống của tôi ngày hôm nay")
Trò chuyện	Phát ngôn nhằm mục đích nói chuyện vui (ví dụ: "chào bạn"; "hôm nay mình tập được nhiều quá"; "cảm ơn nhà")

MÔ HÌNH MAXIMUM ENTROPY

Bài toán phân lớp hành động ngôn từ được mô hình hoá như bài toán phân lớp. Cho tập các phát ngôn $S = \{x_1, x_2, \ldots x_n\}$ và tập các nhãn $L = \{$cung cấp thông tin, hỏi, yêu cầu, trò chuyện$\}$, xác định các phát ngôn x_i có nhãn tương ứng l_i là gì, $l_i \in L$. Để phân lớp chúng tôi sử dụng mô hình Maximum entropy (MaxEnt)[13]. Cho mẫu đầu vào (x, y), $x \in S$, y là nhãn $\in L$, mô hình học cần xác định $p(y|x)$ cao nhất gọi là $p^*(y|x)$ thoả mãn công thức sau:

$$p^*(y|x) = Z(x)\exp(\textstyle\sum_i \lambda_i f_i(x, y)) \quad (1)$$

trong đó, $Z(x)$ gọi là nhân tố bình thường hoá, được tính theo công thức:

$$Z(x) = \textstyle\sum_y \exp(\sum_i \lambda_i f_i(x, y)) \quad (2)$$

và f_i là các đặc trưng của mô hình. Việc tìm $p^*(y|x)$ được đưa về bài toán tìm $\{\lambda_1, \lambda_2 \ldots \lambda_n\}$. Chúng tôi sử dụng phương pháp L_BFGS [14] để xác định các λ_i thoả mãn.

Đặc trưng được sử dụng trong mô hình học MaxEnt là n-gram (1, 2, 3- gram). Để nâng cao hiệu quả của mô hình, chúng tôi xây dựng các từ điển gồm các từ điển hình để nhận biết cho các lớp. Một số ví dụ về các từ dấu hiệu nhận biết các hành động ngôn từ trong từ điển được trình bày trong Bảng 2. Các đặc trưng trong mô hình được biểu diễn trong Bảng 3.

Bảng 2. *Ví dụ một số từ trong từ điển*

Loại	Từ dấu hiệu
Hỏi	Bao nhiêu, bao lâu, như thế nào, ra sao, sao vậy, cho mình hỏi, cho em hỏi, làm thế nào, ra làm sao, tôi muốn hỏi, gì, muốn hỏi, muốn biết …
Yêu cầu	Hãy, mở chức năng, vào menu, nhập thông tin…
Trò chuyện	Chào, xin chào, cảm ơn, ơi, ê, ôi, quá, quá cơ, bye, cảm ơn, thank, …

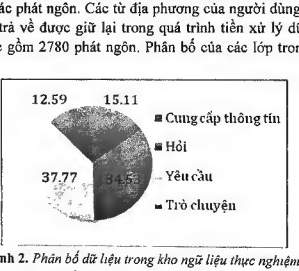
Bảng 3. *Các đặc trưng của mô hình MaXent phân lớp hành động ngôn từ*

n-gram	*Mẫu context predicate*
1-grams	$[\omega_{-2}]$, $[\omega_{-1}]$, $[\omega_0]$, $[\omega_1]$, $[\omega_2]$
2-grams	$[\omega_{-2}\omega_{-1}]$, $[\omega_{-1}\omega_0]$, $[\omega_0\omega_1]$, $[\omega_1\omega_2]$
3-grams	$[\omega_{-2}\omega_{-1}\omega_0]$, $[\omega_{-1}\omega_0\omega_1]$, $[\omega_0\omega_1\omega_2]$
Từ điển	***Mẫu đoạn văn bản trong câu để so khớp với từ trong từ điển***
2-words	$[\omega_{-2}\omega_{-1}]$, $[\omega_{-1}\omega_0]$, $[\omega_0\omega_1]$, $[\omega_1\omega_2]$ trong từ điển
3-words	$[\omega_{-2}\omega_{-1}\omega_0]$, $[\omega_{-1}\omega_0\omega_1]$, $[\omega_0\omega_1\omega_2]$ trong từ điển

THỰC NGHIỆM

Xây dựng dữ liệu

Để xây dựng dữ liệu, một nhóm sinh viên tình nguyện được yêu cầu sử dụng thử ứng dụng phiên bản đầu tiên của ứng dụng GoldHealth360 – một phiên bản chưa có mô đun nhận biết hành động ngôn từ. Chúng tôi tập hợp các phát ngôn của người dùng tương tác với hệ thống và gán nhãn hành động ngôn từ cho các phát ngôn. Các từ địa phương của người dùng và lỗi do dịch vụ nhận dạng giọng nói tự động trả về được giữ lại trong quá trình tiền xử lý dữ liệu. Bộ ngữ liệu của chúng tôi xây dựng được gồm 2780 phát ngôn. Phân bố của các lớp trong ngữ liệu được chỉ ra trong Hình 2.



Hình 2. *Phân bố dữ liệu trong kho ngữ liệu thực nghiệm.*

Bảng 4. *Kết quả trên một lần thực hiện tốt nhất*

Nhãn	Thực tế	Mô hình	Mô hình khớp với thực tế	Độ chính xác	Độ hồi tưởng	F1
Cung cấp thông tin	84	92	56	60.87	66.67	63.64
Hỏi	192	190	172	90 53	89.58	90.05
Yêu cầu	210	210	188	89.52	89.52	89.52
Trò chuyện	70	70	66	87.14	87.14	87.14
Trung bình (Average – macro)				82.02	83.23	82.62
Trung bình (Average – micro)	280	280	241	84.88	84.88	84.88

Kết quả

Chúng tôi chạy thực nghiệm theo phương pháp đánh giá chéo, chạy trong 5 lần. Dữ liệu được chia thành 2 tập: tập huấn luyện 80% dữ liệu, 20% còn lại cho tập kiểm tra. Kết quả thực nghiệm cao nhất trong 1 lượt thử nghiệm được trình bày tại Bảng 4.

Trong đó, số lượng các phát ngôn được gán nhãn thủ công, thể hiện tại cột "thực tế". Số lượng phát ngôn do mô hình dự đoán ra được trình bày trong cột "mô hình". Số lượng các phát ngôn do mô hình dự đoán khớp với nhãn thực tế thể hiện trong cột "khớp". Tiếp theo là độ chính xác (precision), độ hồi tưởng (recall) và độ đo F1 của mô hình. Độ chính xác F1 trung bình trên 5 lần là 80.26% theo Macro (độ chính xác trung bình theo từng lớp) và 82.14% theo Micro (độ chính xác theo tổng các phát ngôn). Kết quả này có thể áp dụng vào ứng dụng thực tế. Việc xác định chính xác hành động ngôn từ của câu là không đơn giản. Do ngôn ngữ tự nhiên của con người đa dạng, có nhiều nhập nhằng. Có nhiều phát ngôn ngắn, thậm chí chỉ một hoặc hai từ. Điều này dẫn tới mô hình không đủ thông tin ngữ cảnh cho việc phân lớp. Như chúng ta có thể thấy, hiệu suất của lớp cung cấp thông tin thấp hơn các lớp khác vì nó bao gồm nhiều loại từ và cú pháp khác nhau, không có từ dấu hiệu đặc trưng để xác định. Điều này cần những đặc trưng phức tạp và cao cấp hơn để mô hình có thể phân biệt.

KẾT LUẬN

Việc xác định được hành động ngôn từ của phát ngôn đóng vai trò quan trọng trong việc hiểu ý định của người dùng khi tương tác với hệ thống bằng giọng nói. Trong bài báo này, chúng tôi đã trình bày một phương pháp phân loại hành động ngôn từ để xác định ý định của người nói ở cấp độ diễn ngôn. Chúng tôi đã tích hợp phân loại hành động ngôn từ vào ứng dụng thực tế - hệ thống theo dõi việc rèn luyện sức khoẻ hàng ngày cho người dùng. Thông tin theo ngữ cảnh có thể là một đầu mối quan trọng cho việc phân loại hành động ngôn từ. Vì vậy, trong tương lai, chúng tôi sẽ tập trung vào các thông tin theo ngữ cảnh kết hợp với các đặc trưng về cú pháp để cải thiện hiệu quả của mô hình phân lớp.

LỜI CẢM ƠN

Nghiên cứu này được hỗ trợ bởi đề tài NCKH cấp cơ sở T2017-07-02, ĐH CNTT & TT, ĐH Thái Nguyên.

TÀI LIỆU THAM KHẢO

1. Austin, J. *How to Do Things with Words*. In OxfordUniversity Press, 1962.
2. Xu,H., Huang, C.-R.: Annotate and Identify Modalities, *Speech Acts and Finer-Grained Event Types in Chinese Text*. In Workshop on Lexical and GrammaticalResources for Language Processing, 2014.
3. Seon, Choong-Nyoung, Harksoo Kim, and JungyunSeo. *A statistical prediction model of speakers' intentions using multi-level features in a goal-oriented dialog system*.In PatternRecognition Letters 33.10, trang 1397-1404, 2012.
4. Kim, J., and Kang, J.-H. *Towards identifying unre-solved discussions in student online forums*. InApplied Intelligence 40(4), 2014.
5. Dbabis, S. B., Mallek, F., Ghorbel, H., Belguith, L :*Dialogue Acts Annotation Scheme within Arabic discussions*. In SemDial, 2012.
6. Zaghouani, W.: *Critical Survey of the Freely Available Arabic Corpora*. In Workshop LREC, 2014.
7. Searle, J. R.: *A classication of illocutionary acts. In: Language in society*, 5(01), trang 1-23, 1976.
8. Cohen, W.; R. Carvalho, V.; and M. Mitchell, *T.Learning to classify email into speech acts*. In ACL, 2004.
9. Bhatia, S., Biyani, P , Mitra, P. *Classifying usermessages for managing web forum data*. In International Workshop in Web andDatabases, 2012.
10. Arguello, J., Sha_er, K., *Predicting Speech Acts inMOOC Forum Posts*.In international conference on web and social media, 2015.
11. Vosoughi, S., Roy, D. *Tweet acts. A speech act classier for twitter* In arXiv preprint arXiv:1605.05156, 2016.
12. Kr_al, P., Cerisara, C.: *Dialogue act recognitionapproaches*. In Computing and Informatics, 2012.
13. Berger, A., Pietra, S.A.D., Pietra, V.J.D.: *A maximumentropy approach to natural language processing*. InComputational Linguistics, 22(1), 1996.
14. Liu, D., Nocedal, J.: *On the limited memory BFGS method for large scale optimization*. Mathematical Programming, 45, pp.503_528, 1989.

SUMMARY

IMPROVINGTHE FOCL TO LEARN RECURSIVE THEORIES

Ngo Thi Lan [*], Nguyen Thi Dung, Nguyen Lan Oanh
University of Information and Communication Technology - TNU

In voice-based human-machine interaction systems, user's intents understanding of devices in human-machine interactions is one of the most important modules when building a dialog system. Automate speech act identificationplays an important role in user intents understanding. Speech act identification in their utterances can define the users intents such as asking, requesting action, greeting, informing and so on. It helps the system to improve the efficiency of human-machine interactions. In this paper, we present a study on speech act identification in the user's utterances and its application in GoldHealth360 - an application to tracking the user's training and keeping process the health. The experiment results show that the model can work well on mobile phones.

Key words: *Suggestion mining, Suggestion analysis, Vietnamese suggestion, maximum entropy, Vietnamese suggestion detection, advice extraction, online forum*

Ngày nhận bài: 08/9/2017; Ngày phản biện: 13/10/2017; Ngày duyệt đăng: 30/11/2017

[*] *Tel: 0943 870272, Email: ntlan@ictu.edu.vn*