

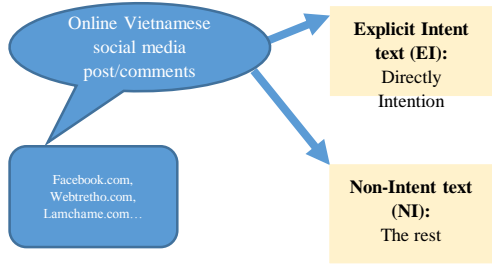


User Intent Identification

Ph.d student: Lương Thái Lê
 Advisor: Phan Xuân Hiếu

Introduction

"Intent or intention is a mental states that represents a commitment to carrying out an action or actions in the future"
 -Michael Bratman-

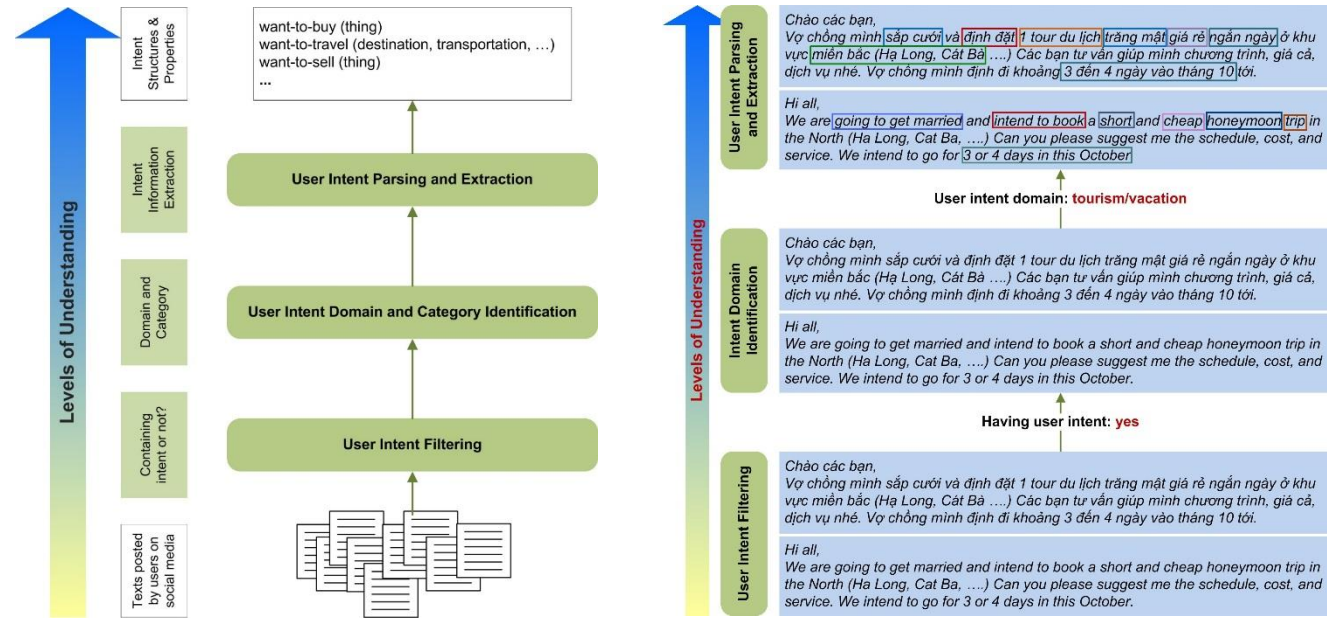


EI / NI	Post/Comment in Vietnamese
EI	"Tình hình là mình đang cần thuê nhà quanh khu vực Phương Mai, Bách Khoa hoặc Tôn Thất Tùng cho 3 người lớn và một cháu nhỏ. Tầm tiền đến khoảng 3 triệu. Bạn nào có thông tin gì liên hệ với mình theo số 0905231880. Cảm ơn nhiều" <i>(I'm looking for a house to rent near Phuong Mai, Bach Khoa or Ton That Tung street for three adults and one child. The price is about 3 million vnd. Please contact me at 0905231880 if you have any information. Thanks a lot)</i>
NI	"Với số tiền bạn có thì khó có thể mua được một căn hộ tại khu vực Cầu Giấy hoặc Thanh Xuân" <i>(It's impossible to buy an apartment in Cau Giay or Thanh Xuan areas with your amount of money)</i>
EI	"Mình đang định vay ngân hàng một khoản bằng bằng lương của mình. Không biết có mẹ nào ở đây có kinh nghiệm về việc này có thể tư vấn cho mình được không ạ?..." <i>(I intend to borrow an amount of money from any bank using my payroll. If any mom here has experience about this, please give me a tip?..)</i>

Motivations

- Help enterprises and businesses
 - Understand the potential customer's need
 - Give the better online advertisement and marketing plan to the customers
 - Penetrate the market faster and more efficiently
- Improve the users satisfaction

Problem of User Intent Identification



Explicit Intent Formulation

$$I_u^e = (u, c, d, w, p)$$

where:

- u : user identifier (nick name, id, user name...)
- c : context or condition (honey moon, pregnant, ...)
- d : domain of intent (vacation, education, finance,...)
- w : key word of intent (book-trip, rent-house,...)
- p : list of properties associated with an intent
 $p = \{price="giá rẻ (cheap)", duration="3 đến 4 ngày (3 or 4 days)", destination="Hà Long, Cát Bà" \dots\}$

Challenges

- Natural language: Post/Comments usually have informal and lack grammatical structures
- There are a lot of noisy data around the main intention in online post/comment texts
- Ambiguity between EI and NI:
 - ✓ Intention in the past:
"Cách đây vài năm mình đã định mua xe Camry nhưng sau đó..."
(I intended to buy a Camry couple of years ago but after that...)
 - ✓ Non-Intent texts may contain the **key word**:
"Chị em nào định mua loại sữa này cho em bé thì suy nghĩ kỹ nhé..."
(Think thoroughly if you want to buy this milk product...)

Filtering User Intent Module

Problem statement

- Input: Texts (posts/comments on Social media such as facebook, webtretho...)
- Output: Classify Texts into EI or NI
- Solutions: Maximum Entropy classifier

• Classify function: $P_\theta(y|x) = \frac{1}{Z_\theta(x)} \exp \sum_{i=1}^n \lambda_i f_i(x, y)$

where:

- $\theta = (\lambda_1, \lambda_2, \dots, \lambda_n)$: vector of weights (identified after training by L-BFGS)
- $F = (f_1, f_2, \dots, f_n)$: vector of features
- $Z_\theta(x) = \sum_{y \in \mathcal{Y}} \exp \sum_{i=1}^n \lambda_i f_i(x, y)$: normalizing factor
- $f_{cp, \lambda} (x, y) = [cp(x)][y = I]$: feature function

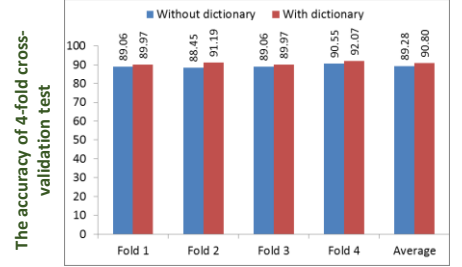
• Given x then y is determined by: $y^* = \arg \max_{y \in \mathcal{Y}} P_\theta(y|x)$

Feature Template

N-grams	Context predicate templates
1-grams	$[w_{-2}], [w_{-1}], [w_0], [w_1], [w_2]$
2-grams	$[w_{-2}, w_{-1}], [w_{-1}, w_0], [w_0, w_1], [w_1, w_2]$
3-grams	$[w_{-2}, w_{-1}, w_0], [w_{-1}, w_0, w_1], [w_0, w_1, w_2]$
Dictionaries	Text template for matching dictionaries
2-words	$[w_{-2}, w_{-1}], [w_{-1}, w_0], [w_0, w_1], [w_1, w_2]$ in dictionary
Examples	<i>muốn mua (want-buy), cần tìm (looking-for),...</i>

Experiment & Result

Class	Human	Model	Match	Precision	Recall	F1-score
Non-intent	181	185	170	91.89	93.92	92.90
Explicit Intent	147	143	132	92.31	89.80	91.03
Average _{macro}				92.10	91.86	91.98
Average _{micro}	328	328	302	92.07	92.07	92.07



Conclusions

- What have been done:
 - Proposed a three-stage process for full understanding of user explicit intents.
 - Also proposed the formulation of explicit intent.
 - Built a classification model based on maximum entropy method with the average accuracy of more than 90%.
- Future works:
 - Improve the User intent filtering module
 - Build the User intent domain identification module
 - Build the User intent parsing and extraction module