Identifying User Intents in Vietnamese Spoken Language Commands and Its Application in Smart Mobile Voice Interaction

Ph.D. student: Ngô Thị Lan

Supervisor: Assoc.Prof./Ph.D. Phạm Bảo Sơn - Co-Supervisor: Ph.D. Phan Xuân Hiếu

Abstract

This paper presents a lightweight machine learning model and a fast conjunction matching method to the problem of identifying user intents behind their spoken text commands. These model and method were integrated into a mobile virtual assistant for Vietnamese (VAV) to understand what mobile users mean to carry out on their smartphones via their commands. User intent, in the scope of our work, is an action associated with a particular mobile application. Given an input spoken command, its application will be identified by an accurate classifier while the action will be determined by a flexible conjunction matching algorithm.

Challenges

Ambiguity of natural language

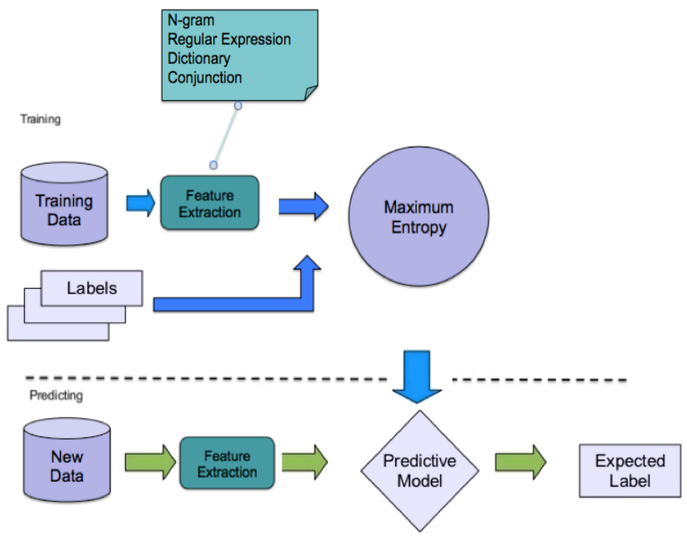
Output text of Automatic speech recognition services

- Short, less grammatical, no punctuation mark
- Numbers are interpreted as alphabetic text
- Continuous text chunks are recognized as discrete tokens

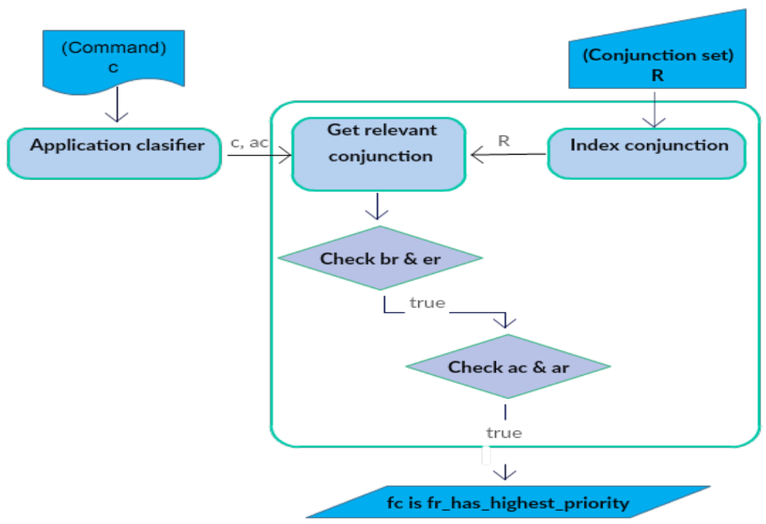
Lack of resources and the corpora for Vietnamese spoken text understanding.

Methods

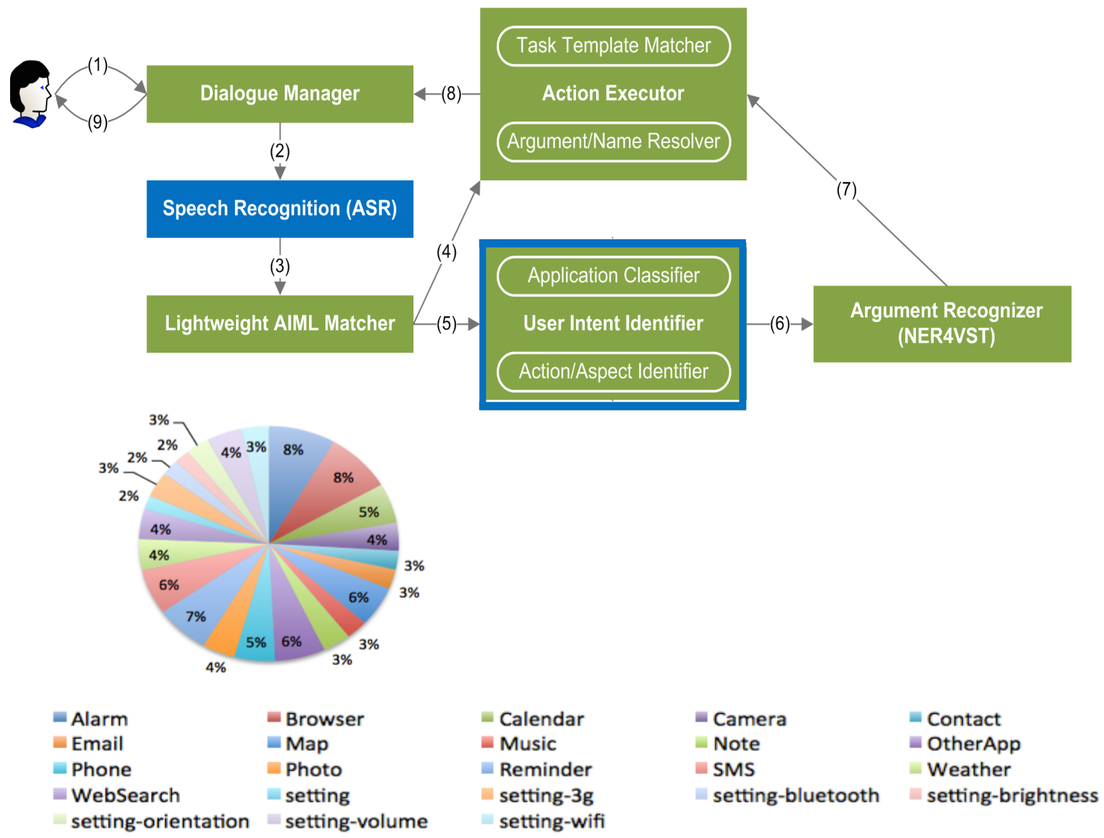
Application classiffication with maximum entropy model



Action identification with conjunction matching

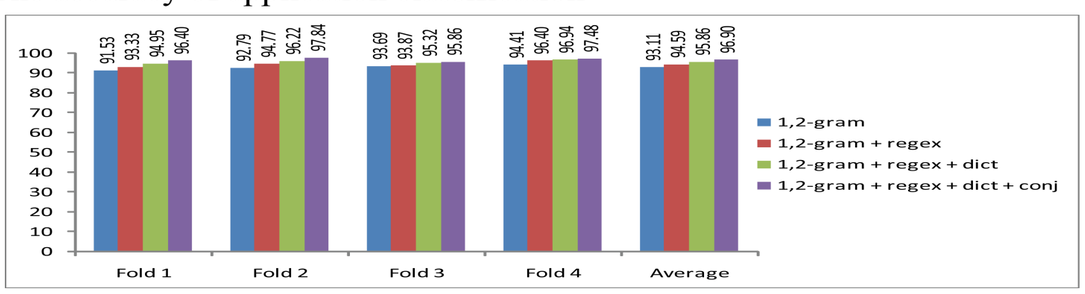


Experiment



Results

The accuracy of application classification



The accuracy of action identification

Application	Action	Accuracy	Application	Action	Accuracy
alarm	turn–off	100.0	phone	call	100.0
	set	99.06		query	100.0
	delete	100.0		open	90.0
	open	96.78	reminder	turn–off	100.0
browser	open	88.27		set	100.0
calendar	set	92.30		delete	97.96
	query	84.91		open	90.91
	delete	100.0	sms	send	96.15
	open	88.20		open	100.0
camera	take–photo	98.77	weather	query	100.0
	record–video	100.0	web–search	query	72.97
contact	add	100.0		open	50.00
	query	96.00	setting–wifi	turn–off	100.0
	share	71.43		turn–on	100.0
	open	100.0	setting	open	100.0
email	query	75.00	setting–3g	turn–off	93.94
	send	100.0		turn–on	95.92
	open	100.0	setting–volume	turn–down	95.24
map	find–direction	87.80		set	80.95
	locate	83.95		turn–up	95.56
	open	88.89	setting–orientation	turn–off	100.0
music	open	100.0		turn–on	100.0
note	add	97.56	setting–brightness	turn–down	93.33
	open	96.67		set	92.86
other–app	open	94.74		turn–up	93.75
photo	open	100.0	setting–bluetooth	turn–off	100.0
				turn–on	100.0

Conclusion

- Propose a definition of user intent for human–mobile voice interaction commands.
- Identify user intents in Vietnamese spoken texts.
- A lightweight approach to perform application & action identification.