Learning Question Classifiers

Yuchen Zhong

Tongji University

March 27, 2018

Overview

Introduction

This paper introduces a two-layered classifiers to classifies questions. They used SNoW (which is a machine learnig system) to train and predict.

Paper Highlights

- hierarchical classifier
  - **coarse classes**
  - **fine classes**
- machine learning approach

Question Hierarchy

a two-layered taxonomy

- 6 coarse classes
- 50 fine classes

Class	#	Class	#
ABBREV.	9	description	7
abb	1	manner	2
exp	8	reason	6
ENTITY	94	**HUMAN**	65
animal	16	group	6
body	2	individual	55
color	10	title	1
creative	0	description	3
currency	6	**LOCATION**	81
dis.med.	2	city	18
event	2	country	3
food	4	mountain	3
instrument	1	other	50
lang	2	state	7
letter	0	**NUMERIC**	113
other	12	code	0
plant	5	count	9
product	4	date	47
religion	0	distance	16
sport	1	money	3
substance	15	order	0
symbol	0	other	12
technique	1	period	8
term	7	percent	3
vehicle	4	speed	6
word	0	temp	5
DESCRIPTION	138	size	0
definition	123	weight	4

The Ambiguity Problem

One Problem

There is no completely clear boundary between classes.

Example

- What is bipolar disorder? **definition** or **disease_medicine**
- What do bats eat? **animal** or **food**
- Waht is the PH scale? **numeric value** or **definition**

Solution

allow multiple class labels for a single question. So it becomes a multi-labels problem.

Process

1. C_0:The initial set of all the coarse classes
2. C_1:The selected coarse classes by Coarse Classifier $|C_1| \leq 5$
3. C_2:The corresponding fine classes determined by the class hierarchy
4. C_3:The selected fine classes by Fine Classifier $|C_3| \leq 5$

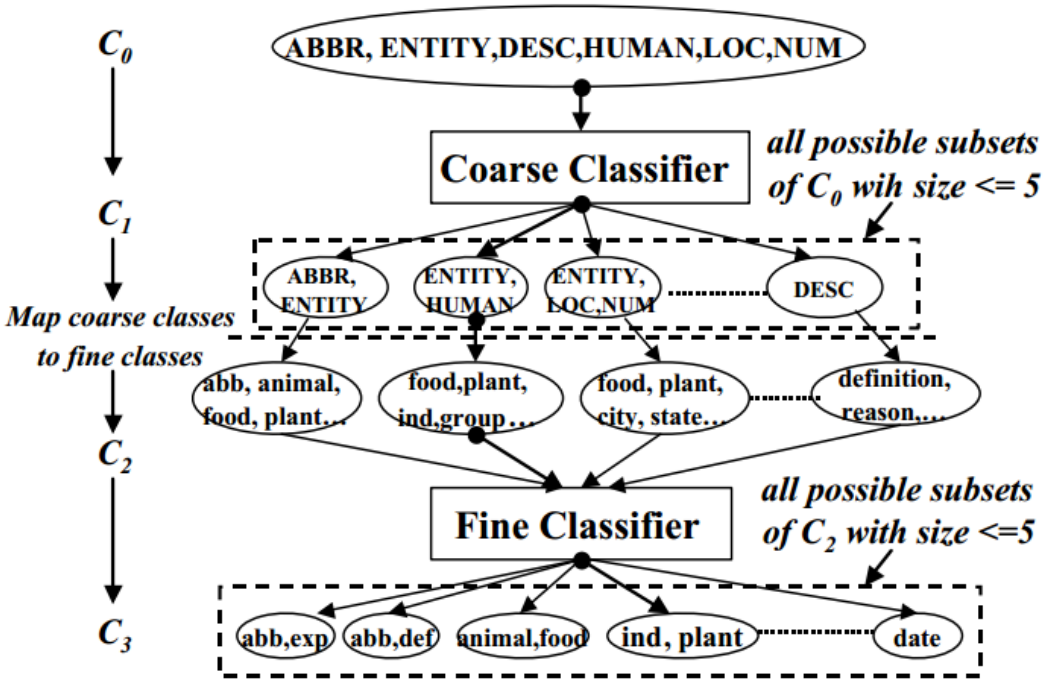


Figure 1: The hierarchical classifier

Features

Each question is analyzed and represented as a list of features

Primitive Features

- words
- pos tags
- chunks
- named entities
- head chunks
- semantically related words(RelWord)

Complex features

- conjunctive(n-grams)
- relational features

Decision Model

Model——SNoW

SNoW is a machine learning system developed in 1998. Its main algorithm is Winnow algorithm. SNoW deals with multi-class problems and will return possibilities of each class.

Details

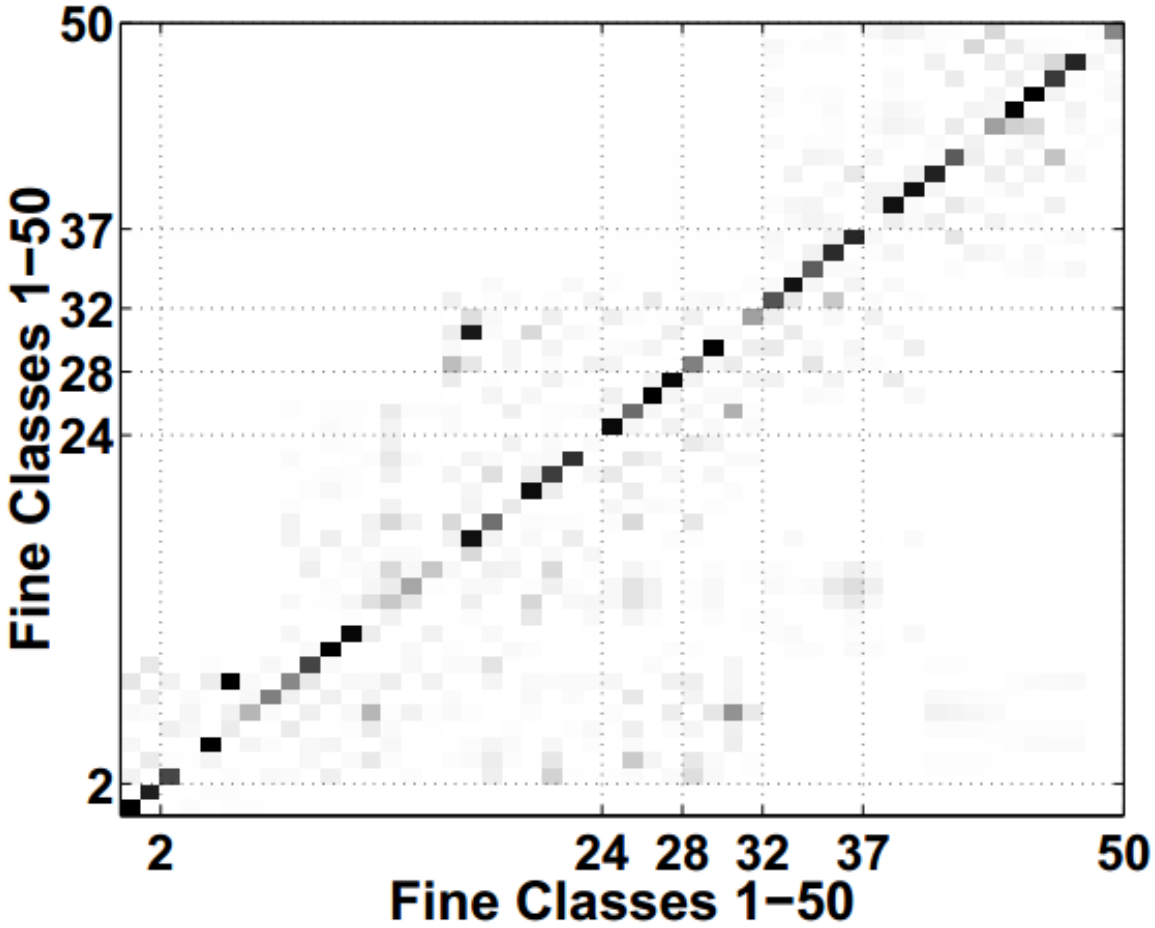
Given a confusion set and a question, SNoW outputs possibilities of each class ($P = \{p_1, p_2, ..., p_n\}$). Select the highest k classes($k \leq 5$) where k satisfies,

$$k = \min(\arg\min_t(\sum_{i=1}^{t} p_i \geq T), 5)$$

T is a threshold and $T = 0.95$ in the experiments.

Results

$P_{<=5}$	Word	Pos	Chunk	NE	Head	RelWord
Coarse	92.00	96.60	97.00	97.00	97.80	**98.80**
Fine	86.00	86.60	87.60	88.60	89.40	**95.00**



The End