Predicting Contradiction Intensity: Low, Strong or Very Strong?

Ismail Badache, Sébastien Fournier, Adrian Chifu

SIGIR 2018 Ann Arbor MICHIGAN

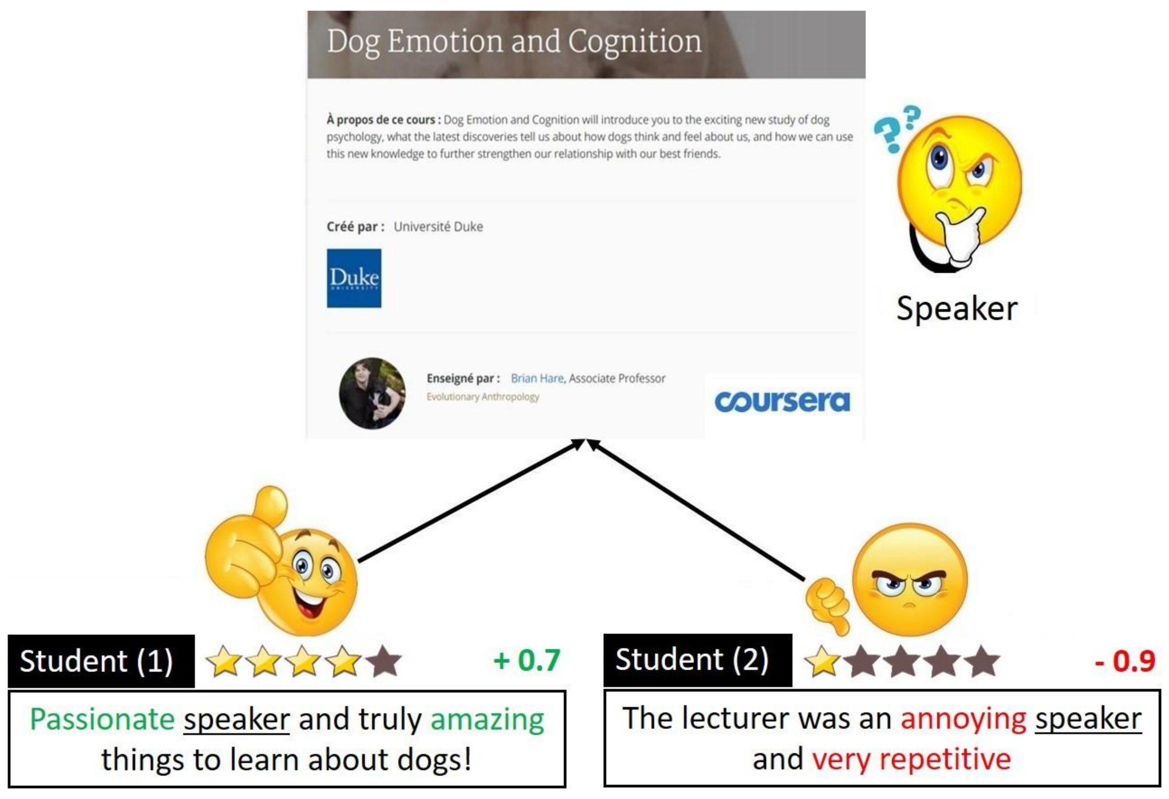
Predicting Contradiction Intensity: Low, Strong or Very Strong?

Ismail Badache – Sébastien Fournier – Adrian Chifu
Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France
Ismail.Badache@lis-lab.fr

1. Introduction

► **The goal:**

- Predicting intensity level of the contradiction using 2 dimensions (rating & polarity)



► **Research questions:**

- How to estimate the intensity of contradiction around a specific aspect?
- What is the impact of the joint consideration of polarity (Pol) and rating (Rat) on the measurement of contradiction intensity?

2. Contradiction Based-Aspect and Sentiment

► **Extraction of aspects:** The following treatments are applied:

(1) Term frequency calculation of the reviews corpus,
(2) Part-of-speech tagging of reviews using Stanford Parser,
(3) Selection of terms having nominal category (NN, NNS),
(4) Selection of nouns with emotional terms in their 5-neighborhoods (using SentiWordNet dictionary),
(5) Extraction of the most frequent (used) terms in the corpus among those selected in the previous step. These terms will be considered as aspects.

Step	Description
(1)	course : 44219, material : 3286, assignments : 3118, content : 2947, speaker : 2705,.......$term_i$
(2)	re = The/DT **lecturer**/NN was/VBD an/DT annoying/VBG **speaker**/NN and/CC very/RB repetitive/JJ ./. I/PRP found/VBD the/DT **formatting**/NN so/RB different/JJ from/IN other/JJ **courses**/NNS I/PRP 've/VBP taken/VBN ,/, that/IN it/PRP was/VBD hard/JJ to/TO get/VB started/VBN and/CC figure/VB **things**/NNS out/RP ./.
(3)	lecturer, speaker, formatting, things
(4)	lecturer, speaker
(5)	speaker

► **Sentiment analysis:** SentiNeuron (unsupervised)

- Trained over 82 million Amazon review dataset.
- LSTM with 4096 units, the 2389th neuron was found to be specifically focusing on the sentiment for a given sentence. We have normalized it between 0 and 1.
- Accuracy : 93% (error rate 7%).

3. Dataset and Judgments

► **Data:**

- Collected from *coursera.org* between October 10-14, 2016.

Table 1. Statistics on *coursera* data set

Field		Total Number
Courses		2244
Courses Rated		1115
Reviews		73873
Reviews	★☆☆☆☆	1705
Reviews	★★☆☆☆	1443
Reviews	★★★☆☆	3302
Reviews	★★★★☆	12202
Reviews	★★★★★	55221

Table 2. List of detected aspects

Assignment	Content	Exercise
Information	Instructor	Knowledge
Lecture	Lecturer	Lesson
Material	Method	Presentation
Professor	Quality	Question
Quiz	Slide	Speaker
Student	Teacher	Topic
Video		
22 aspects		

Table 3. Statistics on some aspects extracted from the reviews of *coursera.org*

Aspects	#Rat 1	#Rat 2	#Rat 3	#Rat 4	#Rat 5	#Negative	#Positive	#Review	#Course
Content	176	179	341	676	1641	505	1496	1883	207
Lecturer	32	41	48	85	461	55	193	236	39
Material	191	203	328	722	2234	784	1693	2254	237
Quiz	151	155	221	401	581	481	475	824	128

► **Judgments:** User study

- 3 users were asked to assess the sentiment class for each review-aspect.
- 3 other users assessed the degree of contradiction between reviews-aspect.
- In total, 66104 reviews-aspect of 1100 courses (instances) i.e. 50 courses for each aspect are judged manually for 22 aspects.

4. Identifying the Most Effective Features

Table 4. List of the exploited features

c_i	Feature	Description
c_1	#NegRev	Number of negative reviews on document
c_2	#PosRev	Number of positive reviews on document
c_3	#TotalRev	Total number of reviews on document
c_4	#Rat1	Number of reviews with rating ★☆☆☆☆
c_5	#Rat2	Number of reviews with rating ★★☆☆☆
c_6	#Rat3	Number of reviews with rating ★★★☆☆
c_7	#Rat4	Number of reviews with rating ★★★★☆
c_8	#Rat5	Number of reviews with rating ★★★★★
c_9	VarRat	Variation of ratings (using standard deviation)
c_{10}	VarPol	Variation of polarities (using standard deviation)

► **Training data:**

- The balanced collection for the 4-points scale as intensity class:
  - 230 *Very Low*
  - 230 *Low*
  - 230 *Strong*
  - 230 *Very Strong*

Table 5. Selected features by attribute selection algorithms

Algorithm	Metric	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}
CfsSubsetEval	[#Folds]	5	5	2	0	0	0	0	0	5	5
WrapperSubsetEval	[#Folds]	4	4	4	2	0	0	0	2	5	5
ConsistencySubsetEval	[#Folds]	5	5	4	2	1	1	2	2	5	5
FilteredSubsetEval	[#Folds]	5	5	4	3	2	2	3	3	5	5
	Average	4.75	4.75	3.5	1.75	0.75	0.75	1.25	1.75	5	5
ChiSquaredAttributeEval	[Rank]	3	4	5	7	9	10	8	6	2	1
FilteredAttributeEval	[Rank]	4	3	5	7	9	10	8	6	2	1
GainRatioAttributeEval	[Rank]	3	4	5	7	9	10	8	6	2	1
InfoGainAttributeEval	[Rank]	3	4	5	7	9	10	8	6	1	2
OneRAttributeEval	[Rank]	4	3	5	7	9	10	8	6	2	1
ReliefFAttributeEval	[Rank]	4	3	6	8	9	10	7	5	1	2
SVMAttributeEval	[Rank]	4	3	5	7	9	10	8	6	2	1
SymetricalUncertEval	[Rank]	3	4	5	7	9	10	8	6	2	1
	Average	3.5	3.5	5.12	7.12	9	10	7.87	5.87	1.75	1.25

5. Learning Features for Predicting Intensity

► **Learning process using the selection algorithms**

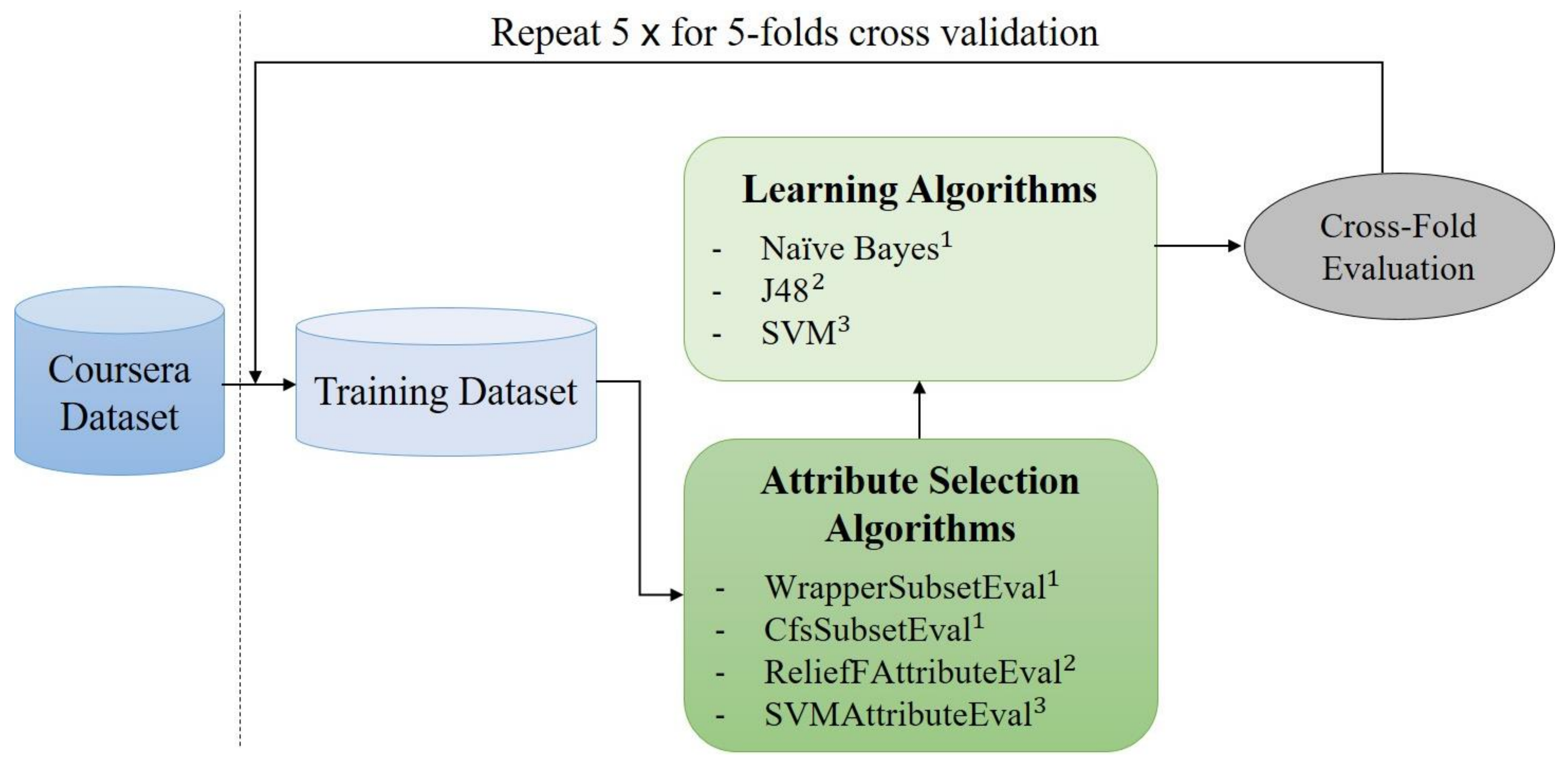


Table 6. Selected features sets

Algorithm	Features
CfsSubsetEval	$c_1, c_2, c_3, c_9, c_{10}$
WrapperSubsetEval	$c_1, c_2, c_3, c_4, c_8, c_9, c_{10}$
Other algorithms	$c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}$

- Features selected by CfsSubsetEval (CFS) and WrapperSubsetEval (WRP) are learned using Naive Bayes.
- Features selected by ReliefFAttributeEval (RLF) are learned using J48.
- Features selected by SVMAttributeEval (SVM) are learned using multi-class SVM (SMO function on Weka : Waikato Environment for Knowledge Analysis).

6. Experimental Results

► **Precision results for machine learning techniques:**

Classifiers	Contradiction intensity class	Features selection algorithms	All features
NaiveBayes	Very Low	0.81 (CFS)	0.71
	Low	0.38 (CFS)	0.34
	Strong	0.75 (CFS)	0.66
	Very Strong	0.78 (CFS)	0.69
	Average	0.68 (CFS)	0.60
	Very Low	0.86 (WRP)	0.72
	Low	0.46 (WRP)	0.38
	Strong	0.76 (WRP)	0.63
	Very Strong	0.80 (WRP)	0.67
	Average	0.72 (WRP)	0.60
SVM	Very Low	0.88^{*} (SVM)	0.88^{*}
	Low	0.72^{**} (SVM)	0.72^{**}
	Strong	0.78^{*} (SVM)	0.78^{*}
	Very Strong	0.90^{**} (SVM)	0.90^{**}
	Average	0.82^{**} (SVM)	0.82^{**}
J48	Very Low	0.97^{**} (RLF)	0.97^{**}
	Low	0.92^{**} (RLF)	0.92^{**}
	Strong	0.97^{**} (RLF)	0.97^{**}
	Very Strong	0.98^{**} (RLF)	0.98^{**}
	Average	0.96^{**} (RLF)	0.96^{**}

► **Findings**

- *#NegRev*, *#PosRev*, *VarRat* and *VarPol* are the most fruitful features to predict contradiction intensity.
- J48 algorithm brings the best improvement compared to Naïve Bayes and SVM.
- Approach weakness: dependence on the quality of sentiment and aspect models.