CS 3750: Advanced Topics in Machine Learning

Final Project

Exploring Applications of Topic Model to SWoRD

Huy Viet Nguyen

SWoRD Project[*]

‣ Web-based reciprocal peer review system
  ‣ Students turn their class papers into SWoRD
  ‣ The paper is assigned to 3-6 peers in class
  ‣ The peers grade the paper and give advices (do reviewing)
  ‣ Students revise the paper and turn it back in to SWoRD

‣ Large corpus of review, grade from peers and instructors

‣ SWoRD is designed so that review are on different dimensions of paper, e.g. flow, logic, insight…

How NLP benefits SWoRD

‣ From instructor perspective…

  ‣ Also large number of papers and corresponding reviews from students,

  ‣ Of course, grading always is not easy,

  ‣ But more essential, how utilize this rich source to understand student overall performance.

‣ Currently, SWoRD provides little support for instructors discover insights from student writing.

‣ How to model such information and extract them

  ‣ Can topic models give interesting partitions on students' reviews

Topics models on review

- ## Training data
  - Annotated reviews in History class
  - Type-review = {praise, problem, summary}
  - Identifies$_{\text{problem review}}$ = {solution, problem, both}
  - pLocalization$_{\text{problem review}}$ = {yes, no}
  - and many other annotations...
- ## Topic model[GibbsLDA++]
  - What inside the review?
  - What similar feedbacks say?
  - Any correlation between topics and annotations above?

Annotated students' reviews

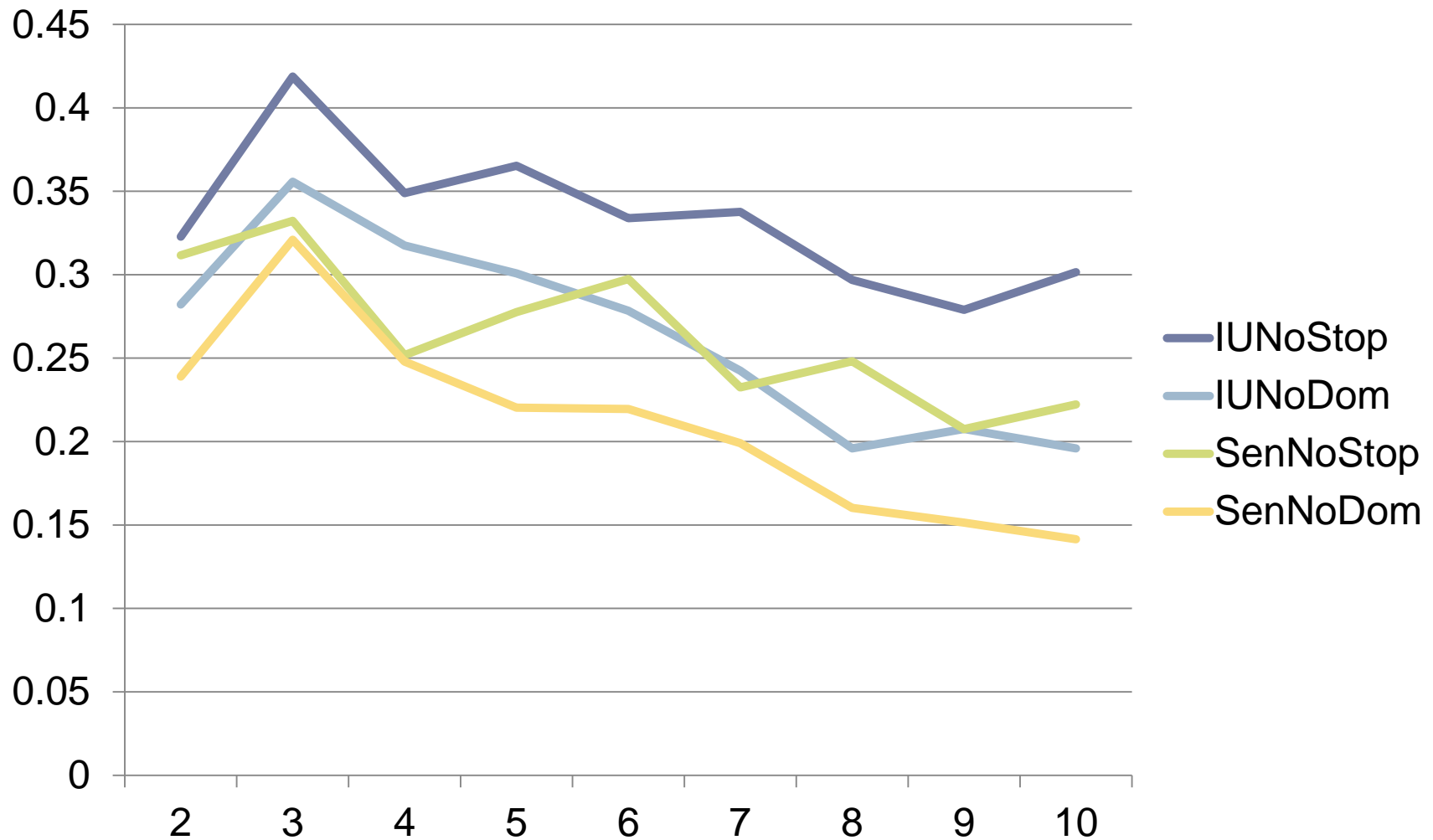| writerNo	reviewerNo	reviewer	dim	rating	review
394	1	Writing	Flow	5	Your theses/arguments are stated fairly clearly and ear
394	1	Writing	Flow	5	For revision, be careful to keep restatements/reinforce
394	311	Content	Flow	6	'This essay argues that American democracy did not ac
394	311	Content	Flow	6	The first two ideas are clearer in the paper than the thi
394	311	Content	Flow	6	The writer needs to clarify the following sentence on p
394	1	Writing	Logic	5	Your attention to important details of each situation co
394	1	Writing	Logic	5	As noted in Dimension 1, greater attention to ongoing s
394	311	Content	Logic	5	The paragraph about Abraham Lincoln's actions towar
394	311	Content	Logic	5	The section of the essay on African Americans needs
394	1	Writing	Insight	5	'You are, in many ways, moving very promisingly into
394	1	Writing	Insight	5	As noted in the argument and logic dimensions, careful
394	311	Content	Insight	4	'This essay only uses one quote from an historical docu
394	446	High Peer	Flow	7	The essay is well written and clearly flows from point
394	446	High Peer	Flow	7	It was very easy to see the main idea and easy to follo
394	446	High Peer	Flow	7	There were several minor grammatical errors (wrong
394	453	High Peer	Flow	5	'arguments were good but some poor word choice and
394	460	Low Peer	Flow	5	'Your arguement was that the United States democracy

Annotated students' reviews

ID	dim	type-review	Identifies	pLocalization
1	Flow	praise	N/A	N/A
2	Flow	problem	solution	not localized
3	Flow	summary	N/A	N/A
4	Flow	problem	problem	not localized
5	Flow	problem	both	localized
6	Logic	praise	N/A	N/A
7	Logic	problem	solution	not localized
8	Logic	problem	problem	localized
9	Logic	problem	solution	localized
10	Insight	praise	N/A	N/A
11	Insight	problem	solution	not localized
12	Insight	problem	problem	not localized
13	Flow	praise	N/A	N/A
14	Flow	praise	N/A	N/A
15	Flow	problem	problem	localized

Choose the best number of topics

▶ Cluster validation procedure[Levine and Domany 2001]

  ▶ Resample at random of the original data set

  ▶ Compare cluster connectivity matrices of resample subset VS. original set

  ▶ Return consistency score for each number of topics

▶ 4 variants of data set

  ▶ A document can be the a review or a sentence of review

  ▶ Only stop-words or both stop-words and domain-words are removed

▶ The best number of topics is 3

▶ Levin and Domany 2001, Resampling Method for Unsupervised Estimation of Cluster Validity.

Consistency score with different numbers of topics

Top words of topic

Domain	Insight	Structure
African	argument	paper
women	evidence	paragraph
Americans	insight	good
rights	examples	point
democratic	claims	page
immigrants	support	argument
vote	specific	essay
United	clarity	reader
blacks	logic	sentence
government	consequences	flow
laws	revision	facts
political	careful	topic
southern	clearly	thesis
writer	reader	clear
Amendments	significant	introduction

Correlation to annotation

- Cluster validation shows that 3 is the best number of topics
- Assign topic for each review
- For reviews
  - 3 types: praise, problem, summary
  - 3 dimension: flow, insight, logic
- For problem review
  - problem, solution or both
  - problem localized or not

Topic VS. review's type and dimension

- ▸ Very low kappa
- ▸ No agreement between topics and labels

Type	Topic		
Labels	408	268	198
	206	86	93
1401 reviews	6	130	6
Kappa	0.13		

Dimension	Topic		
Labels	251	289	110
	77	162	74
1401 reviews	156	169	113
Kappa	0.07		

Topic VS. problem review

	Topic	
pLocalization	203	258
873 reviews	107	305
Kappa	0.18	

	Topic		
Identifies	101	45	179
	53	97	135
873 reviews	71	39	153
Kappa	0.11		

Discussion

- Very first exploring application of topic models to students' reviews
- Extracted topics have no correlation with annotation
- However, topics represent the information at essay level, consistent results with
  - Sentence level
  - Dimension level
- So how we can use topic models to group reviews?
  - There is no previous work of this kind on SWoRD

Future directions

‣ Use topic models as a feature to learn labels
  ‣ Not necessary topic ID, can use topic words vector
  ‣ Compute the best topic words, making topics more readily human interpretable
‣ Use topic models to learn students' writing performance
  ‣ Students have writing graded by their peers
  ‣ Grades usually correlate to criticism of reviews
  ‣ Topic models for reviewers' sentiment
‣ Summarize students' reviews for instructor
  ‣ Instructors want snippet on reviews
  ‣ Topic models for multi-document summarization