

# Using “Annotator Rationales” to Improve Machine Learning for Text Categorization

Omar F. Zaidan

Jason Eisner

Christine D. Piatko

Johns Hopkins University

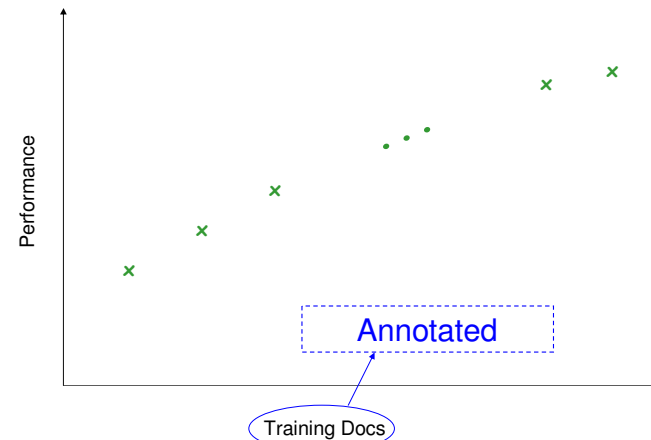
NAACL HLT 2007 – Rochester, NY

Tuesday April 24<sup>th</sup>, 2007

{[ozaidan](mailto:ozaidan@cs.jhu.edu) | [jason](mailto:jason@cs.jhu.edu) | [cpiatko](mailto:cpiatko@cs.jhu.edu)}@cs.jhu.edu

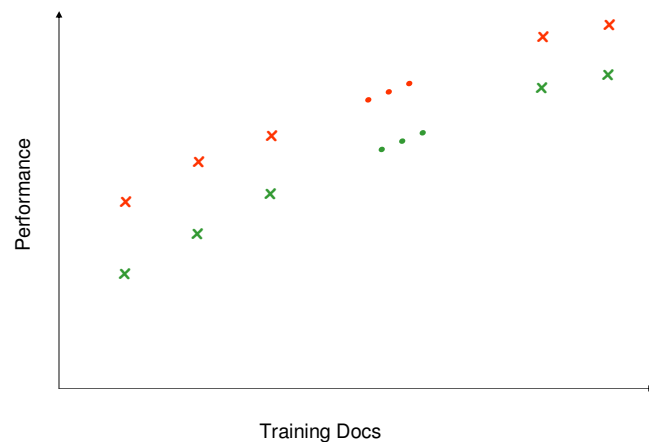
Zaidan et. al – Annotator Rationales

## Supervised Learning



Zaidan et. al – Annotator Rationales

## Richer Annotation



Zaidan et. al – Annotator Rationales

## Richer Annotation?

- Usually, an annotator indicates **what** the correct answer is.
  - We propose the annotator *also* indicate **why**.
- ⇒ Each training example provides data about **its class and why**.
- ⇒ Richer annotation provides more data.
- Idea #1: richer annotation can aid ML.
  - Idea #2: richer better use of our time than **more**.

# Rationales in Text Categorization

The following segments were taken from movie reviews. Did the reviewer have a positive or negative opinion of the movie?

- Trust me, you will enjoy the hell out of American Pie.
- He continues to be one of the most exciting artists on the big screen, performing his own stunts and dazzling audiences.
- ...and the romance was enchanting.
- The movie is so badly put together that even the most casual viewer may notice the miserable pacing and stray plot threads.
- ...and it even makes watching Eddie Murphy a tedious experience.
- A woman in peril. A confrontation. An explosion. The end. Yawn. Yawn. Yawn.

Non-annotated documents



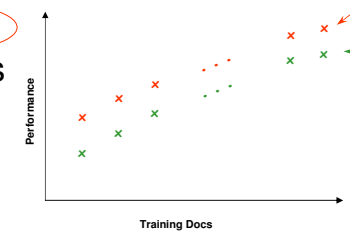
## Saving Private Ryan

War became a reality to me after seeing Saving Private Ryan. Steve Spielberg [redacted] with his latest production. Keep the kids home as the R rating is for Reality. Tom Hanks is [redacted] as Capt John Miller, set out in France during WW II to rescue and return home a soldier, Private Ryan (Matt Damon) who lost three brothers in the war. Spielberg takes us inside the heads of these individuals as they face death during the horrific battle scenes. Private Ryan is not for everyone, but [redacted] for a movie like this to be made. The movie reminds us of the sacrifices made by our fighting men and women. For this I thank them and for Steve Spielberg [redacted]. And I'm sure [redacted] come April, as [redacted] with be in Tom's possession.



Annotated documents

Class and also "rationales"



Only class; no "rationales"



Zaidan et. al – Annotator Rationales

### The Postman

Question: after the disaster that was Waterworld, [redacted] were the execs who gave Costner the money to make another movie thinking??

In this 3 hour advertisement for his new hair weave, Costner plays a nameless drifter who dons a long dead postal employee's uniform [redacted] and gradually turns a nuked-out USA into an idealized hippy-dippy society. (The main accomplishment of this brave new world is in re-inventing polyester.) When he's not pointing the camera directly at himself, director Costner does have a nice visual sense, but by the time the second hour rolled around, [redacted] [redacted] Mark this one [redacted].

[redacted]

[redacted]

[redacted]

[redacted]

Zaidan et. al – Annotator Rationales

Class and also "rationales"

Annotated documents

OK ... now what??

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

Zaidan et. al – Annotator Rationales

How sure was the annotator that this is a positive review?

### Saving Private Ryan

War became a reality to me after seeing Saving Private Ryan. Steve Spielberg [redacted] with his latest production. Keep the kids home as the R rating is for Reality. Tom Hanks is [redacted] as Capt John Miller, set out in France during WW II to rescue and return home a soldier, Private Ryan (Matt Damon) who lost three brothers in the war. Spielberg takes us inside the heads of these individuals as they face death during the horrific battle scenes. Private Ryan is not for everyone, but [redacted] for a movie like this to be made. The movie reminds us of the sacrifices made by our fighting men and women. For this I thank them and for Steve Spielberg [redacted] [redacted]. And I'm sure [redacted] [redacted] come April, as [redacted] [redacted] with be in Tom's possession.

Zaidan et. al – Annotator Rationales

How sure was the annotator that this is a positive review?

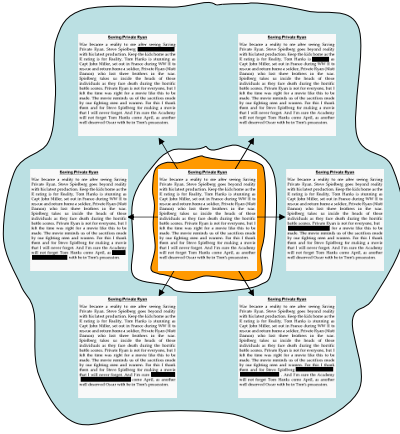
### Saving Private Ryan

War became a reality to me after seeing Saving Private Ryan. Steve Spielberg [redacted] with his latest production. Keep the kids home as the R rating is for Reality. Tom Hanks is [redacted] as Capt John Miller, set out in France during WW II to rescue and return home a soldier, Private Ryan (Matt Damon) who lost three brothers in the war. Spielberg takes us inside the heads of these individuals as they face death during the horrific battle scenes. Private Ryan is not for everyone, but [redacted] for a movie like this to be made. The movie reminds us of the sacrifices made by our fighting men and women. For this I thank them and for Steve Spielberg [redacted] [redacted]. And I'm sure [redacted] [redacted] come April, as [redacted] [redacted] with be in Tom's possession.

If a rationale is masked out, the annotator would not be as sure that this is a positive review.

Intuition: a good model should also be less sure.

# “Contrast” Examples

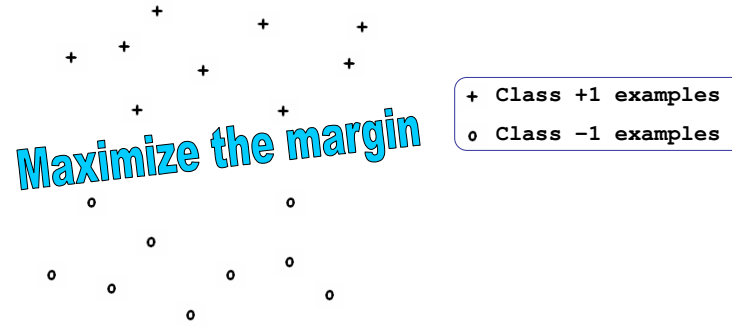


Intuition: a good model should be less sure of a positive classification on contrasts than on the original.

Our work: modified SVM that takes this intuition into account.

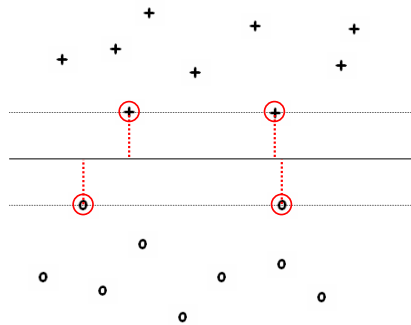
- Original Example
- Contrast Examples
- Obtain a contrast by masking out a rationale

# Standard SVM



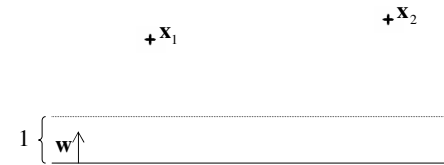
- + Class +1 examples
- o Class -1 examples

# Standard SVM



- + Class +1 examples
- o Class -1 examples
- Support vectors

# Standard SVM



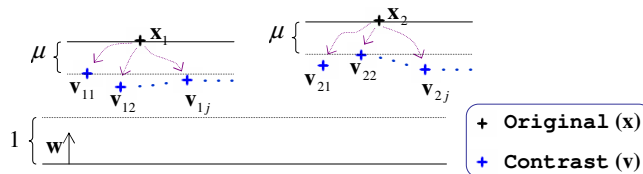
Minimize :

$$\frac{1}{2} \|\mathbf{w}\|^2$$

subject to :

$$\mathbf{w} \cdot \mathbf{x}_i \geq 1$$

## Incorporating Contrasts



Minimize:

$$\frac{1}{2} \|\mathbf{w}\|^2$$

subject to:

$$\mathbf{w} \cdot \mathbf{x}_i \geq 1$$

ALSO

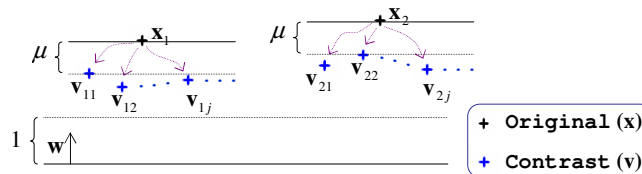
subject to:

$$\mathbf{w} \cdot \mathbf{x}_i - \mathbf{w} \cdot \mathbf{v}_{ij} \geq \mu$$

$$\mathbf{w} \cdot \frac{\mathbf{x}_i - \mathbf{v}_{ij}}{\mu} \geq 1$$

$$\mathbf{w} \cdot \mathbf{x}_{ij} \geq 1$$

## Slack Variables



Minimize:

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i + C_{contrast} \sum_{i,j} \xi_{ij}$$

subject to:

$$\mathbf{w} \cdot \mathbf{x}_i \geq 1 - \xi_i$$

ALSO

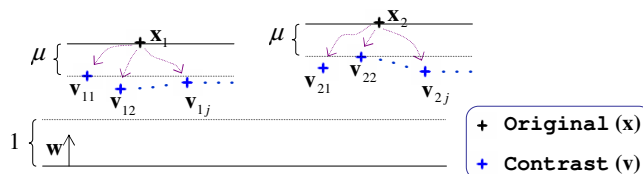
subject to:

$$\mathbf{w} \cdot \mathbf{x}_i - \mathbf{w} \cdot \mathbf{v}_{ij} \geq \mu(1 - \xi_{ij})$$

$$\mathbf{w} \cdot \frac{\mathbf{x}_i - \mathbf{v}_{ij}}{\mu} \geq 1 - \xi_{ij}$$

$$\mathbf{w} \cdot \mathbf{x}_{ij} \geq 1 - \xi_{ij}$$

## ((Include Negative Examples))



Minimize:

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i + C_{contrast} \sum_{i,j} \xi_{ij}$$

subject to:

$$y_i (\mathbf{w} \cdot \mathbf{x}_i) \geq 1 - \xi_i$$

ALSO

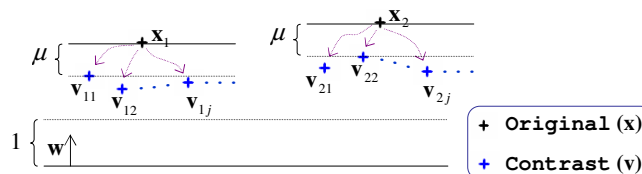
subject to:

$$y_i (\mathbf{w} \cdot \mathbf{x}_i - \mathbf{w} \cdot \mathbf{v}_{ij}) \geq \mu(1 - \xi_{ij})$$

$$y_i (\mathbf{w} \cdot \frac{\mathbf{x}_i - \mathbf{v}_{ij}}{\mu}) \geq 1 - \xi_{ij}$$

$$y_i (\mathbf{w} \cdot \mathbf{x}_{ij}) \geq 1 - \xi_{ij}$$

## The Modified SVM



Minimize:

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i + C_{contrast} \sum_{i,j} \xi_{ij}$$

subject to:

$$y_i (\mathbf{w} \cdot \mathbf{x}_i) \geq 1 - \xi_i$$

ALSO

subject to:

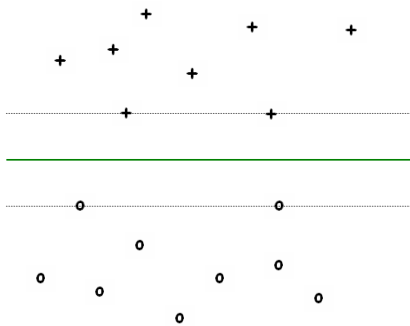
$$y_i (\mathbf{w} \cdot \mathbf{x}_i - \mathbf{w} \cdot \mathbf{v}_{ij}) \geq \mu(1 - \xi_{ij})$$

$$y_i (\mathbf{w} \cdot \frac{\mathbf{x}_i - \mathbf{v}_{ij}}{\mu}) \geq 1 - \xi_{ij}$$

$$y_i (\mathbf{w} \cdot \mathbf{x}_{ij}) \geq 1 - \xi_{ij}$$

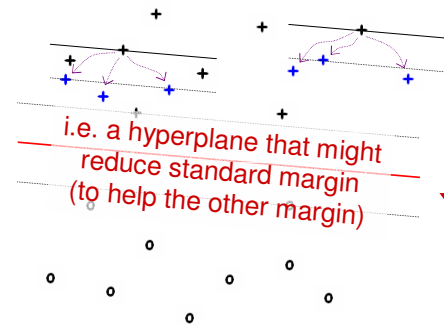
## What this Means in Practice

Standard SVM cares about this margin



## What this Means in Practice

Standard SVM cares about this margin



Modified SVM cares about both margins

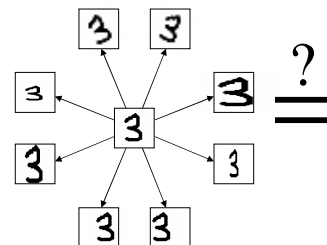
## Recap

- Training examples:  $(x_1, y_1), (x_2, y_2), \dots$
- $y_i$  has  $n_i$  rationales:  $r_{i1}, r_{i2}, \dots, r_{in}$
- $x_i$  gives  $n_i$  contrast examples:  $v_{i1}, v_{i2}, \dots, v_{in}$  (obtain  $j^{\text{th}}$  contrast by masking out  $j^{\text{th}}$  rationale.)
- We extend the SVM to determine best hyperplane subject to:
  - Constraints for standard margin,
  - and also
  - Constraints for original/contrast separating margin.

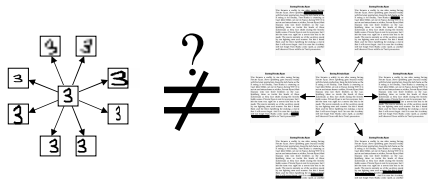
## What this is not

- In tasks like digit recognition, one can “generate” more training data from the existing examples

### Class-preserving transformations



## What this is *not*



Class-preserving

Information from new examples similar to that from real examples

Can get benefit by automatic preprocessing (rescale, deslant, etc)

Not necessarily (contrast)

Information from contrast examples is of a different kind

Actually provides new information via human insight

## The Dataset

- The movie review dataset (Pang & Lee)
  - 1000 positive reviews
  - 1000 negative reviews
- For each document, given the class annotation, we added the rationale annotation
  - Annotation process: in an HTML editor, rationale segments are boldfaced.

## Annotation Time

- How big is the overhead for annotating rationales?
- Ought to establish that richer annotation is a good use of an annotator's time.
  - vs. just annotating more documents
- One can imagine three annotation tasks:
  - **T1**: given document, annotate the **class**.
  - **T2**: given document *and* gold standard class, annotate the **rationales**.
  - **T3**: given document, annotate both the **class** and the **rationales**.
- 50 docs/task given to four annotators

## Annotation Time

- **T1**: given document, annotate the **class**.
- **T2**: given document *and* gold standard class, annotate the **rationales**.
- **T3**: given document, annotate both the **class** and the **rationales**.
- We found that  $\text{Time}(\mathbf{T3}) \approx 2 \times \text{Time}(\mathbf{T1})$
- Even though on average 8.3 rationales/doc + class!
- Annotator *already* needs to find rationales to determine class. Extra work is only to make them explicit:  
 $\text{Time}(\mathbf{T3}) < \text{Time}(\mathbf{T1}) + \text{Time}(\mathbf{T2})$  by about 20%

## Annotation Time

- **T1**: given document, annotate the **class**.
  - **T2**: given document *and* gold standard class, annotate the **rationales**.
  - **T3**: given document, annotate both the **class** and the **rationales**.
- **Synergy**:  $\text{Time}(\mathbf{T3}) < \text{Time}(\mathbf{T1}) + \text{Time}(\mathbf{T2})$
  - Extra time reduced with better annotation setup (e.g. automatic boldfacing of highlighting, stylus, etc) or smart use of eye tracking.
  - Note: the task of classifying full docs is a worst-case scenario for rationales.
    - Other tasks would have simpler/fewer rationales and more complex classes.

## Feature Vector

Word	Feature	val
!	1	1
"	2	0
.	.	.
.	.	.
.	.	.
a	333	1
a+	334	0
a-list	335	1
aaron	336	0
.	.	.
.	.	.
.	.	.
zoolander	17741	1
zorro	17742	0
zucker	17743	0
zwick	17744	0

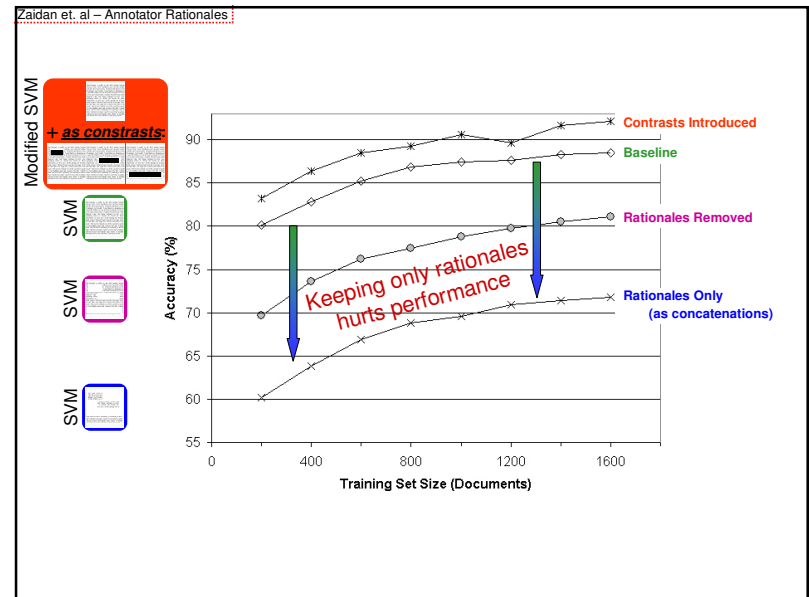
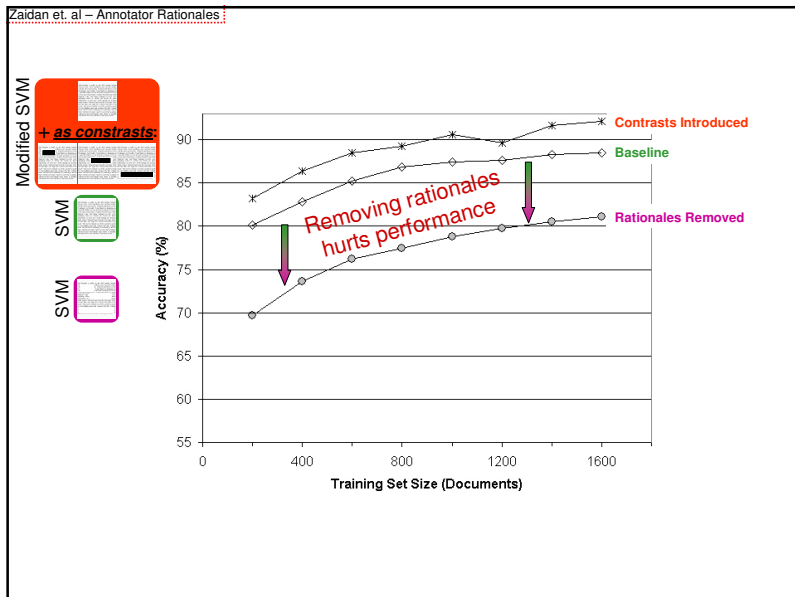
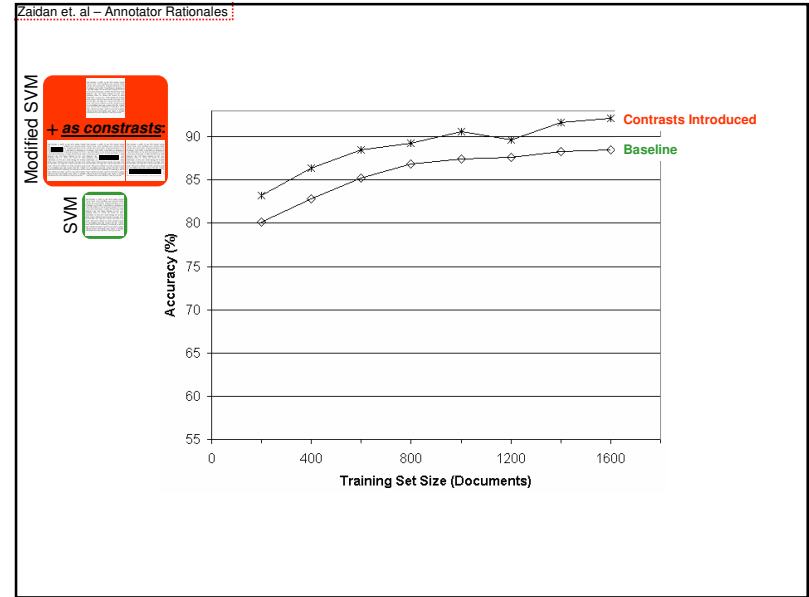
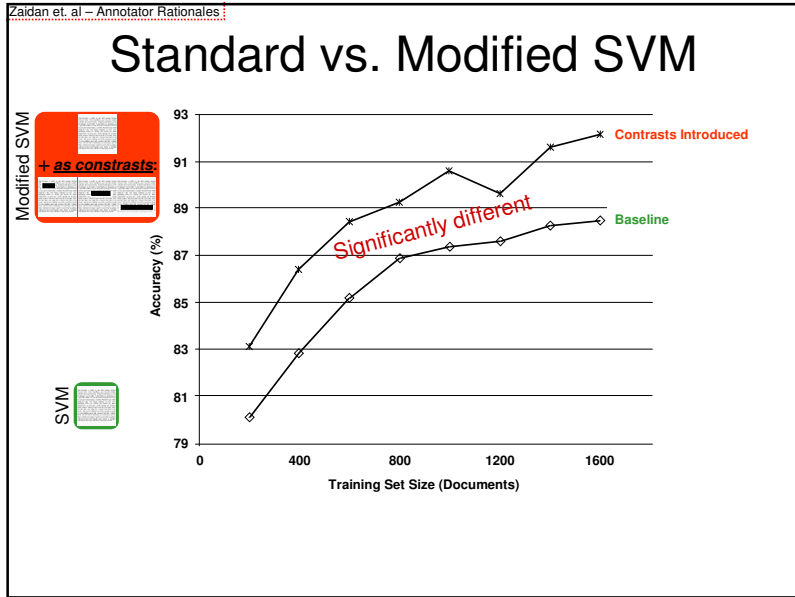
no a-list actor would  
star in @ movie like  
zoolander because that  
would be @ mistake !

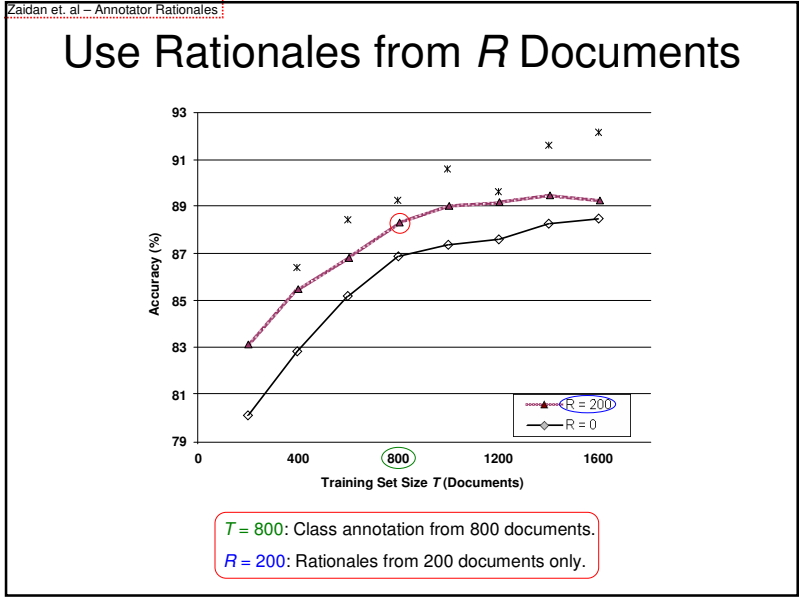
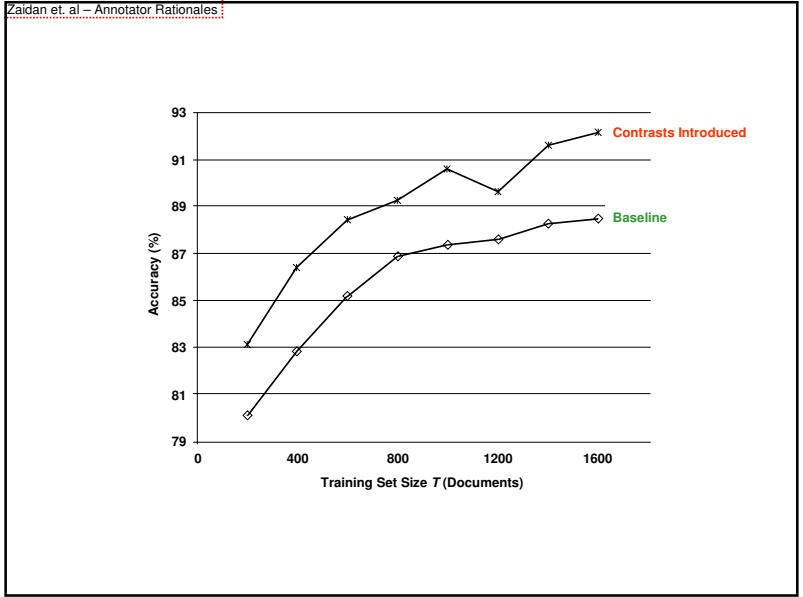
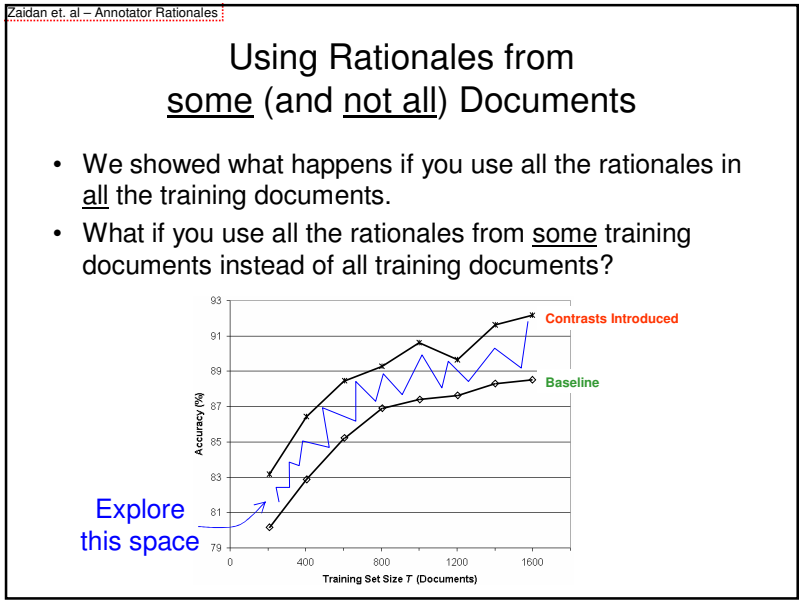
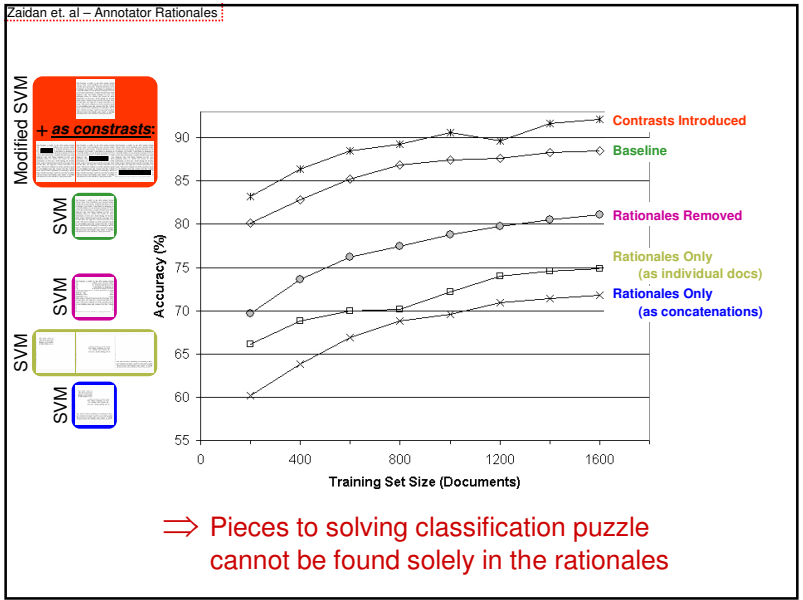
## Feature Set

- Binary unigram features
    - A document is reduced to a 0-1 vector with 17,744 dimensions.
  - Feature set too simple?
    - Goal is not to build the best classifier.
    - Goal is to improve an existing classifier regardless of its feature set.
    - We use this feature set to mirror previous work.
- (( Pang & Lee actually tried other features and found it did not matter much ))

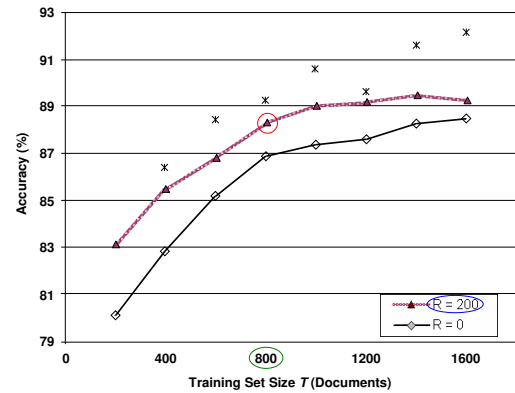
Let's see some experimental results...





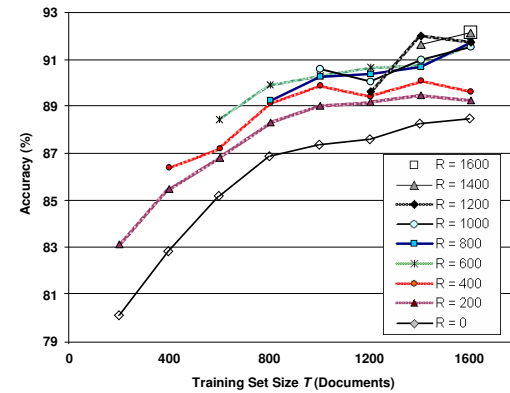


## Use Rationales from $R$ Documents



Class annotation from  $T$  documents.  
Rationales from  $R$  documents only.

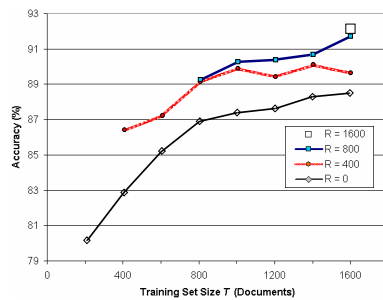
## Use Rationales from $R$ Documents



Class annotation from  $T$  documents.  
Rationales from  $R$  documents only.

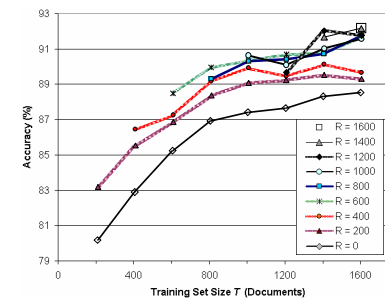
## Using Rationales from some (and not all) Documents

- Observation #1: much of the benefit can be obtained without annotating 100% of the documents
  - e.g. (0%, 50%, 100%) for  $T = 800$  and  $T = 1600$



## Using Rationales from some (and not all) Documents

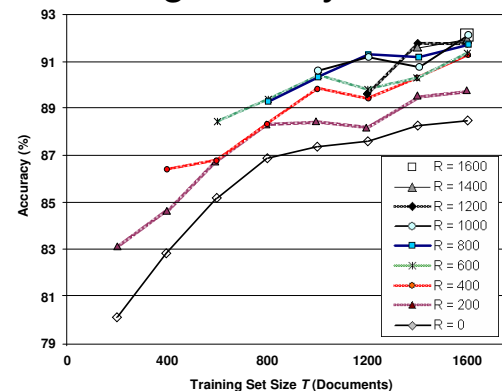
- Observation #2: if you have a lot of training documents, adding more may not help much (curves flatten out). BUT adding more rationales provides a fresh benefit.
  - ⇒ Benefit from  $R$  even if  $T$  “reaches its potential”



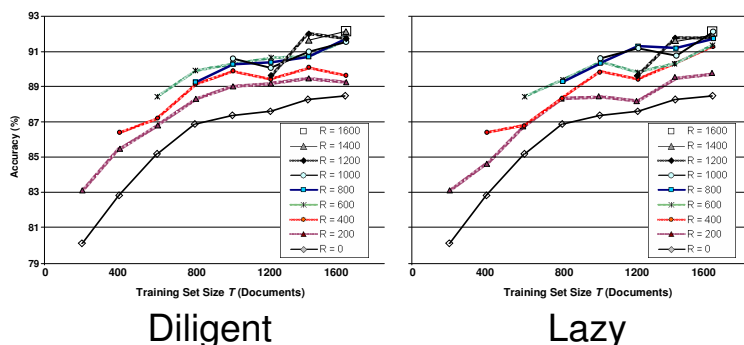
## Simulating a “Lazy Annotator”

- In last few experiments, we kept all rationales from some training documents.
  - $R=200$  and  $T=800$  means 600 documents contributed no contrast examples. Each of the 200  $R$  documents contributes *all* its rationales.
- What if we keep some rationales from all documents?
  - Instead of using all the rationales in 200 documents, use the same number of rationales *spread out* over all 800 documents.

## Simulating a “Lazy Annotator”



## Any Differences? Not Really.



The two ( $T=800, R=200$ ) points are comparable: same number of rationales. Difference is in distribution only.

## Simulating a “Lazy Annotator”

- Experiment simulates a not-so-diligent annotator
  - This might be more common in reality.
  - Likely to pick ‘obvious’ rationales, yielding faster rationale annotation.
  - Also, obvious rationales may prove to be better.  
(Though experiment doesn't test for that; rationales were picked at random)

## Big Picture

- Idea #1: richer annotation can aid ML.
- Idea #2: richer better use of our time than more.
- Example of richer annotation: rationales.
- Developed and tested one method to use rationales (our extended SVM).
- Simulated degree of annotator laziness.
- *Bonus*: annotator knows nothing about the ML method (or even feature set).

## Future Work

- More datasets:
  - Different task may require different kind of rationales.
  - Might also require different annotation tool.
- More experiments:
  - Examination of annotation process.
  - Real experiments to see effect of a lazy annotator.
- More models:
  - Generative models: model annotation of rationales as a noisy process (annotators are not perfect).
  - Potentially other discriminative methods.

## On The Internets

- The enriched dataset (and slides) here:  
<http://cs.jhu.edu/~ozaidan/rationales>

### FYI – BibTeX entry:

```
@InProceedings{zaidan-eisner-piatko:2007:MainConf,
  author = {Zaidan, Omar and Eisner, Jason and Piatko, Christine},
  title = {Using ``Annotator Rationales'' to Improve Machine Learning
  for Text Categorization},
  booktitle = {Human Language Technologies 2007: The Conference of the
  North American Chapter of the Association for Computational Linguistics;
  Proceedings of the Main Conference},
  month = {April},
  year = {2007},
  address = {Rochester, New York},
  publisher = {Association for Computational Linguistics},
  pages = {260--267},
  url = {http://www.aclweb.org/anthology/N/N07/N07-0133}
}
```