

Modeling Annotators: A Generative Approach to Learning from Annotator Rationales

Omar F. Zaidan

Jason Eisner

The Center for Language
and Speech Processing
Johns Hopkins University



EMNLP 2008 – Honolulu, HI
Saturday October 25th, 2008

{ozaidan|jason}@cs.jhu.edu

Zaidan & Eisner – Modeling Annotators

Annotators are useful.

They provide us with data sets that we use everyday in the context of statistical learning.

Annotators are smart! 😊

We achieve good performance using such data in many NLP tasks.

Annotators are underutilized... 😞

A lot of thought goes into annotation, but little of that is captured.

Zaidan & Eisner – Modeling Annotators

Annotators are Underutilized?

“Hey annotator,
is this review
positive or
negative?”

Armageddon

This disaster flick is a disaster alright. Directed by Tony Scott (Top Gun), it's the story of an asteroid the size of Texas caught on a collision course with Earth. After a great opening, in which an American spaceship, plus NYC, are completely destroyed by a comet shower, NASA detects said asteroid and go into a frenzy. They hire the world's best oil driller (Bruce Willis), and send him and his crew up into space to fix our global problem.

Annotator says:



($y = -1$)

Useful!

The action scenes are over the top and too ludicrous for words. So much so, I had to sigh and hit my head with my notebook a couple of times. Also, to see a wonderful actor like Billy Bob Thornton in a film like this is a waste of his talents. The only real reason for making this film was to somehow out-perform Deep Impact. Bottom line is, Armageddon is a failure.

Zaidan & Eisner – Modeling Annotators

Annotators are Underutilized!

“Hey annotator,
is this review
positive or
negative?”

يوم الحساب

هذا الفيلم الكارثي هو بحق فيلم كارثي. الفيلم، من إخراج توني سكوت (مخرج فيلم توب جن)، يروي قصة كويكب بحجم ولاية تكساس في طريقه للاصطدام بالكرة الأرضية. بعد مشهد افتتاحي رائع، تدمر فيه سفينة فضاء أمريكية، إضافة إلى مدينة نيويورك، تكتشف ناسا الكويكب أنف الذكر ويجن جنونها. فيستعينون بأفضل منقب عن النفط في العالم (بروس ويليس)، ويرسلونه وطاقمه إلى الفضاء لإصلاح مشكلتنا العالمية هذه.

Annotator says:



($y = -1$)

Useful??

مشاهد الإثارة والأكشن مبالغ بها ويصعب وصف سخفها بأي كلمات، إلى حد وجدت نفسي فيه أتهد، بل وضربت جبهتي بمدونتي بضع مرات. و أيضا فإن رؤية ممثل مثل بيلي بوب ثورنتون في فيلم كهذا هو بحق مضيعة لمواهبه. السبب الوحيد لصنع هذا الفيلم هو محاولة التفوق على فيلم ديب امباكت بطريقة ما. خلاصة الأمر، يوم الحساب فيلم فاشل.

Annotators are Underutilized!

“Hey annotator,
is this review
positive or
negative?”

Annotator says:



($y = -1$)

يوم الحساب

هذا الفيلم الكارثي هو [redacted] الفيلم، من إخراج توني سكوت (مخرج فيلم توب جن)، يروي قصة كويكب بحجم ولاية تكساس في طريقه للاصطدام بالكرة الأرضية. بعد مشهد افتتاحي رائع، تدمر فيه سفينة فضاء أمريكية، إضافة إلى مدينة نيويورك، تكتشف ناسا الكويكب آنف الذكر ويجن جنونها. فيستعينون بأفضل منقب عن النفط في العالم (بروس ويليس)، ويرسلونه وطاقمه إلى الفضاء لإصلاح مشكلتنا العالمية هذه.

مشاهد الإثارة والأكشن

[redacted] إلى حد [redacted]

[redacted] بضع مرات. و أيضا فإن رؤية ممثل مثل بيلي بوب ثورنتون في فيلم كهذا هو بحق [redacted] السبب الوحيد لصنع هذا الفيلم هو محاولة التفوق على فيلم ديب امباكت بطريقة ما. خلاصة الأمر، يوم الحساب [redacted]

Annotators are Underutilized!

“Hey annotator,
is this review
positive or
negative?”

Annotator says:



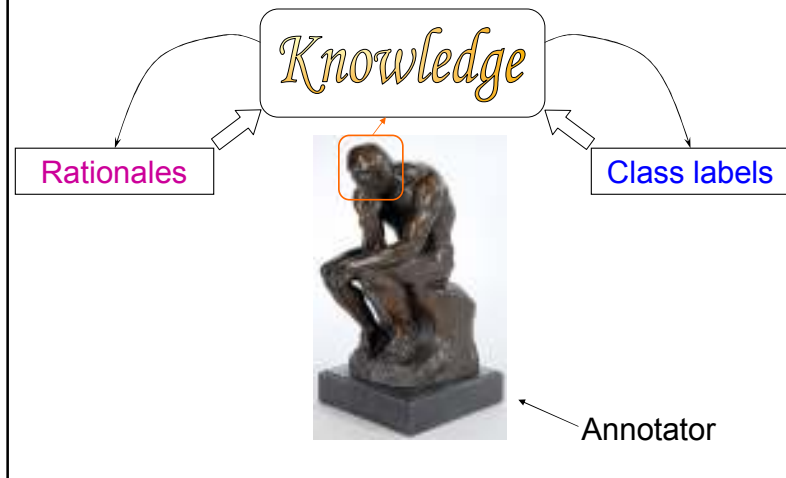
($y = -1$)

Armageddon

This disaster flick is [redacted] Directed by Tony Scott (Top Gun), it's the story of an asteroid the size of Texas caught on a collision course with Earth. After a great opening, in which an American spaceship, plus NYC, are completely destroyed by a comet shower, NASA detects said asteroid and go into a frenzy. They hire the world's best oil driller (Bruce Willis), and send him and his crew up into space to fix our global problem.

The action scenes are [redacted]. So much so, [redacted] a couple of times. Also, to see a wonderful actor like Billy Bob Thornton in a film like this is [redacted]. The only real reason for making this film was to somehow out-perform Deep Impact. Bottom line is, Armageddon is [redacted]

Why would Rationales be Useful?

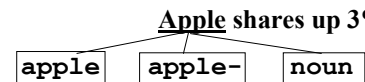


Related Work

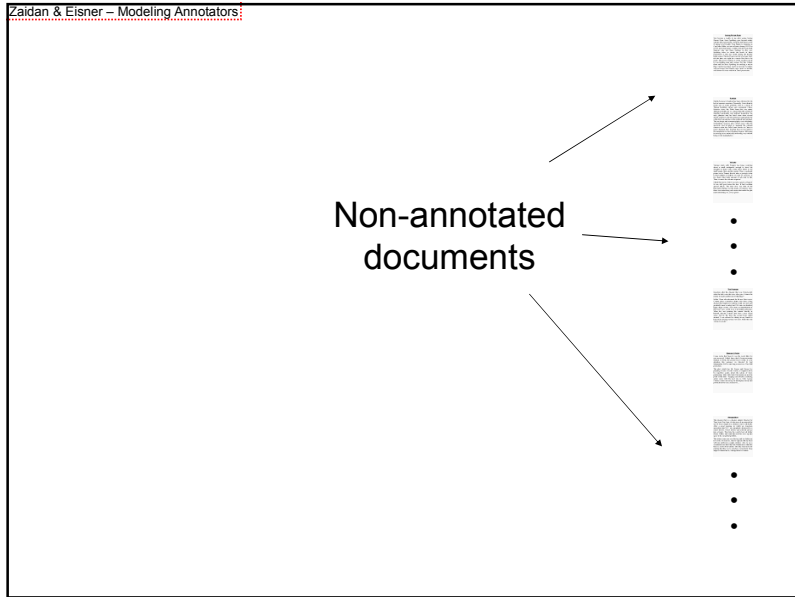
- “Annotators are underutilized. Let’s ask them to do more than just annotate class.”
- Let’s ask annotator to identify relevant **features**.
 - Haghighi & Klein (2006), Raghavan et al. (2006), Druck et al. (2008).

• But:

- Features could be hard to describe.



- Features could be hard for annotators to understand.
- We might want different features in the future.



Zaidan & Eisner – Modeling Annotators:

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.

The text is presented on a grid background. Several lines of text are obscured by black redaction boxes. To the right of the text is a vertical column of document thumbnails with vertical ellipses between them.

Zaidan & Eisner – Modeling Annotators:

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]. Mark this one [redacted].

The text is presented on a grid background. Several lines of text are obscured by black redaction boxes. To the left of the text are three thumbs-up icons, and to the right is a vertical column of document thumbnails with vertical ellipses between them.

Zaidan & Eisner – Modeling Annotators:

$\langle x, y, r \rangle$

Class and also "rationales"

Annotated documents

OK ... now what??

The diagram features a central text box 'Annotated documents' with an arrow pointing to a vertical column of document thumbnails on the left. Above this, a callout box contains the text 'Class and also "rationales"'. A large, tilted callout box at the bottom contains the text 'OK ... now what??. The text $\langle x, y, r \rangle$ is positioned at the top left.

Linear Classifier

- In our experiments, we use a linear classifier.
- Classifier is represented by a weight vector $\vec{\theta}$.
- Rationales will play a role when learning $\vec{\theta}$.
- Improvements **solely due to a better-learned $\vec{\theta}$** (vs. no rationales)
- At test time:
 - No change to decision rule.
 - No new features.
 - No need for rationales.

Linear Classifier

- At test time, when presented with document x :

$$\text{Choose } y^* = \begin{cases} +1 & , \vec{\theta} \cdot \vec{f}(x) > 0 \\ -1 & , \vec{\theta} \cdot \vec{f}(x) < 0 \end{cases}$$

- Where:
 - $\vec{f}(\cdot)$: feature vector (18k binary unigram features).
 - $\vec{\theta}$: corresponding weights (i.e. 18k weights).
- Positive weights favor $y = +1$
 Negative weights favor $y = -1$

x

Trees Lounge is the directoral debut from one of my favorite actors, Steve Buscemi. He gave memorable performances in in The Soup, Fargo, and Reservoir Dogs. Now he tries his hand at writing, directing and acting all in the same flick. The movie starts out awfully slow with Tommy (Buscemi) hanging around a local bar the "trees lounge" and him pestering his brother. It's obvious he a loser. But as he says "it's better I'm a loser and know I am, then being a loser and not thinking I am." Well put. The story starts to take off when his uncle dies, and Tommy, not having a job, decides to drive an ice cream truck. Well, the movie starts to pick up with him finding a love interest in a 17 year old girl named Debbie (Chloe Sevigny) and... I liked this movie alot even though it did not **reach** my expectation. After you've seen him in Fargo and Reservoir Dogs, you know he is capable of a better performance. I think his brother, Michael, did an excellent job for his debut performance. Mr. Buscemi is off to a good career as a director!

$\vec{f}(x)$: 0/1 vector; 18k elements | $f_{\text{and}} = 1, f_{\text{reach}} = 1, f_{\text{terrific}} = 0, \dots$

! " () , . a acting actors after all alot am an and around as at awfully bar being better brother buscemi but capable career chloe cream debbie debut decides did dies directing director dogs drive even excellent expectation fargo favorite finding flick for from gave girl good hand hanging having he him his i ice i'm in interest is it it's job know liked local loser lounge love memorable michael movie mr my named not now obvious of off old one out pick put reach reservoir same says seen sevigny slow soup starts steve story take the then think thinking this though to tommy trees tries truck uncle up well when with writing year you you've

$f(x)$

! " () , . a acting actors after all alot am an and around as at awfully bar being better brother buscemi but capable career chloe cream debbie debut decides did dies directing director dogs drive even excellent expectation fargo favorite finding flick for from gave girl good hand hanging having he him his i ice i'm in interest is it it's job know liked local loser lounge love memorable michael movie mr my named not now obvious of off old one out pick put reach reservoir same says seen sevigny slow soup starts steve story take the then think thinking this though to tommy trees tries truck uncle up well when with writing year you you've

$$\vec{\theta} \cdot \vec{f}(x) \quad \theta \begin{matrix} \xleftarrow{-0.2} & \xrightarrow{-0.1} & \xrightarrow{0.0} & \xrightarrow{+0.1} & \xrightarrow{+0.2} \end{matrix} \begin{matrix} \text{abc} & \text{abc} & \text{abc} & \text{abc} & \text{abc} \end{matrix} \quad \begin{matrix} \theta > 0 \text{ favor } y = +1 \\ \theta < 0 \text{ favor } y = -1 \end{matrix}$$

! " () , . a acting actors after all alot am an and around as at awfully bar being better brother buscemi but capable career chloe cream debbie debut decides did dies directing director dogs drive even excellent expectation fargo favorite finding flick for from gave girl good hand hanging having he him his i ice i'm in interest is it it's job know liked local loser lounge love memorable michael movie mr my named not now obvious of off old one out pick put reach reservoir same says seen sevigny slow soup starts steve story take the then think thinking this though to tommy trees tries truck uncle up well when with writing year you you've

Zaidan & Eisner – Modeling Annotators:

$\vec{\theta} \cdot \vec{f}(x) = -0.41 < 0 \implies y^* = -1$

! () ... acting actors after all abt an an and around as at awfully bar being better brother bussenti but capable career chloe cream debbie

debut decides did dies directing director dogs drive even excellent expectation fargo favorite finding flick for from gave girl

good hand hanging having he him his i ice i'm in interest is it's job know liked local loser lounge love memorable michael movie

mr my named not now obvious of off old one out pick put reach reservoir same says seen sevigny slow soup starts steve story take

the then think thinking this though to tommy trees tries truck uncle up well when with writing year you you've

This $\vec{\theta}$ was learned in a standard way:

Choose $\vec{\theta} = \arg \max_{\vec{\theta}} \prod_{i=1}^n p(y_i | x_i, \vec{\theta})$

choose $\vec{\theta}$ that models class labels well
(of training data)

PS: $p(y | x, \vec{\theta}) \propto \exp(y \cdot (\vec{\theta} \cdot \vec{f}(x)))$ i.e. log-linear model

Zaidan & Eisner – Modeling Annotators:

$\vec{\theta} \cdot \vec{f}(x) = +0.45 > 0 \implies y^* = +1$

! () ... a acting actors after all alot an an and around as at awfully bar being better brother bussenti but capable career chloe cream debbie

debut decides did dies directing director dogs drive even excellent expectation fargo favorite finding flick for from gave girl

good hand hanging having he him his i ice i'm in interest is it's job know liked local loser lounge love memorable michael movie

mr my named not now obvious of off old one out pick put reach reservoir same says seen sevigny slow soup starts steve story take

the then think thinking this though to tommy trees tries truck uncle up well when with writing year you you've

This $\vec{\theta}$ was learned in a novel way (our algorithm!):

Choose $\vec{\theta} = \arg \max_{\vec{\theta}} \prod_{i=1}^n p(y_i | x_i, \vec{\theta}) \cdot p(r_i | y_i, x_i, \vec{\theta})$

choose $\vec{\theta}$ that models class labels & rationales well
(of training data)

PS: $p(y | x, \vec{\theta}) \propto \exp(y \cdot (\vec{\theta} \cdot \vec{f}(x)))$ $p(r | y, x, \vec{\theta}) = ???$

Zaidan & Eisner – Modeling Annotators:

Choose $\vec{\theta} = \arg \max_{\vec{\theta}} \prod_{i=1}^n p(y_i | x_i, \vec{\theta}) \cdot p(r_i | y_i, x_i, \vec{\theta})$

From a review for “Prince of Egypt” ($y = +1$):

51 weeks into '98 , a champ has emerged .

the prince of egypt succeeds where other movies failed .

Zaidan & Eisner – Modeling Annotators:

Choose $\vec{\theta} = \arg \max_{\vec{\theta}} \prod_{i=1}^n p(y_i | x_i, \vec{\theta}) \cdot p(r_i | y_i, x_i, \vec{\theta})$

From a review for “Prince of Egypt” ($y = +1$):

o o o o o I I I I o r

51 weeks into '98 , a champ has emerged . x

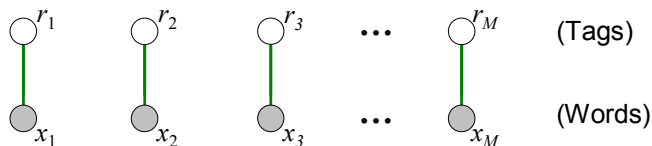
o o o o o I I I I I o r

the prince of egypt succeeds where other movies failed . x

- We encode rationales as a tag sequence.
- The model $p(r | y, x, \vec{\theta})$ should predict tag sequence with some help from $\vec{\theta}$. Let's describe it...

Designing the Model $p(r | y, x, \vec{\theta})$

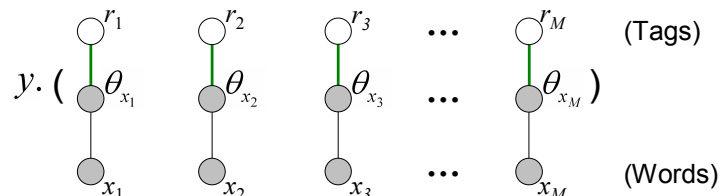
- Mission: model $\{\mathbf{I}, \mathbf{o}\}$ tag sequence with $\vec{\theta}$'s help.



Hmm... $\vec{\theta}$ has no role here.

Designing the Model $p(r | y, x, \vec{\theta})$

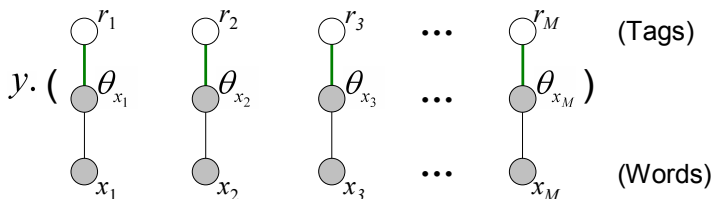
- Mission: model $\{\mathbf{I}, \mathbf{o}\}$ tag sequence with $\vec{\theta}$'s help.



- Makes sense: if for some word w , $(y \cdot \theta_w) \gg 0$, then annotator is likely to mark it \mathbf{I} .
- How likely? We learn a parameter ϕ_{rel} from that annotator's rationale data.
- ϕ_{rel} is learned **jointly** with $\vec{\theta}$.

Designing the Model $p(r | y, x, \vec{\theta})$

- Mission: model $\{\mathbf{I}, \mathbf{o}\}$ tag sequence with $\vec{\theta}$'s help.



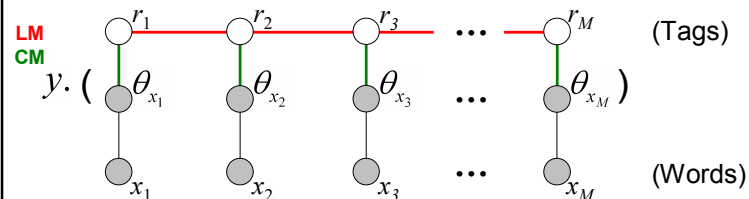
- Problem with this model: it is led to believe that any word w marked with \mathbf{I} has high $(y \cdot \theta_w)$.

- BUT:

... temptation to start shouting slogans, [REDACTED]

Designing the Model $p(r | y, x, \vec{\theta})$

- Mission: model $\{\mathbf{I}, \mathbf{o}\}$ tag sequence with $\vec{\theta}$'s help.



- So, learn 4 more weights: $\phi_{\mathbf{I}-\mathbf{I}}$ $\phi_{\mathbf{O}-\mathbf{O}}$ $\phi_{\mathbf{I}-\mathbf{O}}$ $\phi_{\mathbf{O}-\mathbf{I}}$
- Better! Now: " w is marked with \mathbf{I} either because of high $(y \cdot \theta_w)$ or being around others marked with \mathbf{I} ."
- CRF: Used in other labeling tasks too (POS, NER).

Recap

- Linear classifier: Choose $y^* = \begin{cases} +1 & , \vec{\theta} \cdot \vec{f}(x) > 0 \\ -1 & , \vec{\theta} \cdot \vec{f}(x) < 0 \end{cases}$
- Our work: choose $\vec{\theta}$ that models all of annotator's data well: **both class labels** and **rationales**.

$$\text{Choose } \vec{\theta} = \arg \max_{\vec{\theta}, \vec{\phi}} \prod_{i=1}^n p(y_i | x_i, \vec{\theta}) \cdot p_{\vec{\phi}}(r_i | y_i, x_i, \vec{\theta})$$

Learned jointly

$p(y | x, \vec{\theta})$: Standard log-linear model

$p_{\vec{\phi}}(r | y, x, \vec{\theta})$: CRF predicting \mathcal{I}/\mathcal{O} tag sequence

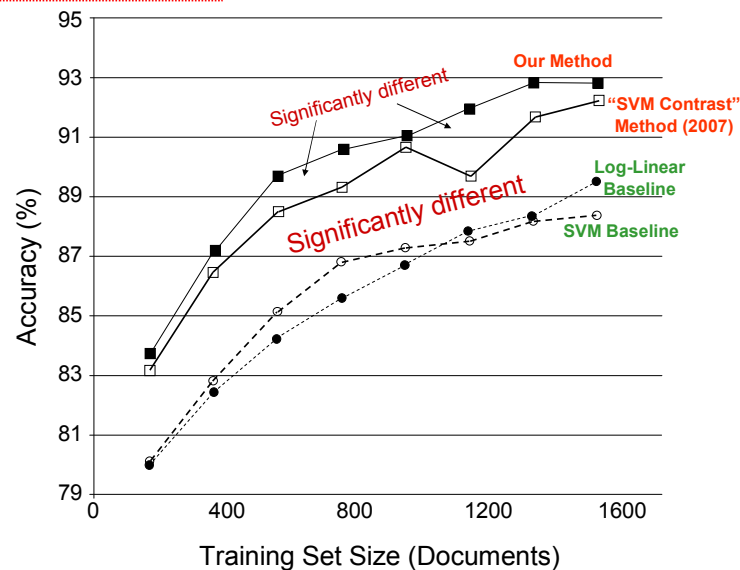
- Remember, at test time:

- No change to decision rule.
- No new features.
- No need for rationales.

Improvement due solely to better-learned $\vec{\theta}$.

Experiments

- Dataset: Pang & Lee's IMDB movie review dataset (Pang & Lee, 2004).
- 2000 document, each annotated with class label and enriched with rationales (Zaidan et al., 2007).
- We train a classifier jointly maximizing likelihood of class and rationale data.
- Compared it to 2 baseline models that account only for class data: log-linear model and SVM.
 - ...and to "SVM contrast" method of Zaidan et al. (2007).

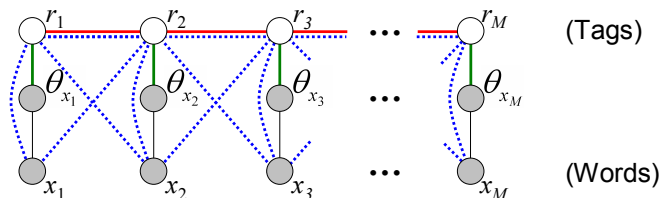


I like 'em curves, but...

- Q: "I see that including rationales helps with performance. But they do take extra time to collect. Is it worth the extra time?"
- A: "Yes!"
- Question addressed in Zaidan et al. (2007).
 - Collecting rationales doesn't take too long.
 - You don't even need so many to get much of benefit.
- *You should collect some in your next annotation project!*

Fancy ϕ -Features

- Remember this?



- Actually, we used an expanded set that includes about 100 “conditional” transition features, e.g.:
 - How often do you see I-O around punctuation?
 - How often do you see I-O around syntactic boundaries?
- Extra features model rationales better, but don’t help learn $\vec{\theta}$ any better than basic feature set.

ϕ Captures Annotator’s Style




- $p_\phi(r | y, x, \vec{\theta})$ is a noisy channel, parameterized by $\vec{\phi}$.
- Noisy? Annotators use r to tell us about $\vec{\theta}$, but they don’t do it “perfectly”:
 - They may mark too little (or too much).
 - They may prefer to start rationales at syntactic boundaries, and/or end around punctuation.
- To what degree? Remember that we train both models jointly:

$$\text{Choose } \vec{\theta} = \arg \max_{\vec{\theta}, \vec{\phi}} \prod_{i=1}^n p(y_i | x_i, \vec{\theta}) \cdot p_\phi(r_i | y_i, x_i, \vec{\theta})$$

- So, learned $\vec{\phi}$ captures style of a specific annotator.

ϕ Captures Annotator’s Style

- So, learned $\vec{\phi}$ captures style of a specific annotator.
- Empirical evidence?
 - An annotator’s own $\vec{\phi}$ predicts their rationales best.
- “So, perhaps you got lucky with A0 (whose data you used in experiments). Would this work with others?”

	A0	A3 	A4 	A5 
Log-linear baseline	71.0	73.0	71.0	70.0
Our method	76.0	76.0	77.0	74.0
SVM baseline	72.0	72.0	72.0	70.0
“SVM Contrast” method	75.0	73.0	74.0	72.0

Conclusions

- A model that accounts for class **and** rationale annotations outperforms two strong baselines.
- Generative approach generalizes to other classifiers (just incorporate $p_\phi(r | y, x, \vec{\theta})$ in objective).
- ...and to other domains:
 - E.g. vision: just model how relevant portions of image tend to trigger rationales (“channel model”) and what size/shape rationales tend to be (“language model”).
- Doing an annotation project? Collect rationales!
 - Even small number could help.
 - “You may get more bang for the buck!”



Jason Eisner
Salesman
1-800-for-RATS

And finally...

- Jason and I thank Christine Piatko (co-author on Zaidan et al., 2007) for helpful discussions and feedback on paper and presentation.
- Thank you!
- A woman in peril. A confrontation. An explosion. The end. ██████████. (*Metro*)
- As directed by Joel Schumacher, B&R is ██████████ ██████████. (*Batman & Robin*)
- ██████████ ██████████. (*The Postman*)

On The Internets

- Slides and more:

<http://cs.jhu.edu/~ozaidan/rationales>

FYI–BibTeX entry:

```
@InProceedings{zaidan-eisner:2008:EMNLP,
  author = {Zaidan, Omar and Eisner, Jason},
  title = {Modeling Annotators: {A} Generative Approach to Learning
    from Annotator Rationales},
  booktitle = {Proceedings of the 2008 Conference on Empirical Methods in
    Natural Language Processing (EMNLP)},
  month = {October},
  year = {2008},
  address = {Honolulu, Hawaii},
  publisher = {Association for Computational Linguistics},
  pages = {31--40},
  url = {http://www.aclweb.org/anthology/D08-1004}
}
```