A model for information summarization, and some basic results

Eric Graves
ARL

Qiang Ning
UIUC

Prithwish Basu
Raytheon

July 8, 2019

Presentation Goals

- Derive model for summarizer.
- Present basic results for summarizer.
- Discuss how results are achieved.

Previous work

Past Work

- Y. Lin, G. Cao, J. Gao, and J.-Y. Nie, "*An information-theoretic approach to automatic evaluation of summaries*," in 2006 N.A.A.C.L.-H.L.T.
- H. Lin and J. Bilmes, "*Multi-document summarization via budgeted maximization of submodular functions*," in 2010 N.A.A.C.L-H.L.T.
- H. Lin and J. Bilmes, "*Learning mixtures of submodular shells with application to document summarization*," in 2012 U.A.I.
- A. See, P. J. Liu, and C. D. Manning, "*Get to the point: Summarization with pointer-generator networks*," in 2017 A.M.A.C.L.
- S. Gehrmann, Y. Deng, and A. Rush, "*Bottom-up abstractive summa-rization*," in 2018 E.M.N.L.P.
- . . . and many more

Motivating Example

Phenomena	
High winds	✓
High UV index	✓
Heavy Rain	✓
Snow	
Low visibility	✓
Smog	
Hurricane	✓

Weather Report

Motivating Example

Phenomena	
High UV index	✓
Hurricane	✓

Weather Summary

Phenomena	
High winds	✓
High UV index	✓
Heavy Rain	✓
Snow	
Low visibility	✓
Smog	
Hurricane	✓

Motivating Example

Phenomena	
High UV index	✓
Hurricane	✓

or, if LA

Phenomena	
High UV index	✓
Smog	
Hurricane	✓

Weather Summary

Phenomena	
High winds	✓
High UV index	✓
Heavy Rain	✓
Snow	
Low visibility	✓
Smog	
Hurricane	✓

Motivating Example

Phenomena	
High UV index	✓
Hurricane	✓

or, if LA

Phenomena	
High UV index	✓
Smog	

Is this just compression/rate distortion?

High UV index	✓
Heavy Rain	✓
Snow	
Low visibility	✓
Smog	
Hurricane	✓

Motivating Example

Phenomena	
High UV index	✓
Hurricane	✓

or, if LA

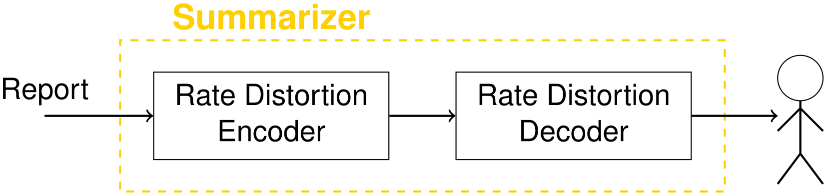
Phenomena	
High UV index	✓
Smog	

Is this just compression/rate distortion?

Yes

High UV index	✓
Heavy Rain	✓
Snow	
Low visibility	✓
Smog	
Hurricane	✓

Model formulation



Goal

Minimize $\mathbb{E}[d(T, S)]$*, for some distortion function** d*, between the report,* T*, and summary,* S*, subject to some restrictions on how the summary is generated.*

Summarizer Restrictions

Phenomena	
High UV index	✓
Smog	
Hurricane	✓

⇒ _1_ _ _01

Phenomena	
High winds	✓
High UV index	✓
Heavy Rain	✓
Snow	
Low visibility	✓
Smog	
Hurricane	✓

⇒ 1110101

Summarizer introduces erasures

Bad Summary

Phenomena	
High UV index	✓
Smog	
Hurricane	

⇒ _1_ _ _00

Phenomena	
High winds	✓
High UV index	✓
Heavy Rain	✓
Snow	
Low visibility	✓
Smog	
Hurricane	✓

⇒ 1110101

Summarizer *only* introduces erasures

Summarizer Restrictions

Phenomena	
High UV index	✓
Smog	
Hurricane	✓

$\Rightarrow$ _1_ _ _01

Summarizer *only* introduces erasures

Phenomena	
High winds	✓
High UV index	✓
Heavy Rain	✓
Snow	
Low visibility	✓
Smog	
Hurricane	✓

$\Rightarrow$ 1110101

Summary $\subset$ Report

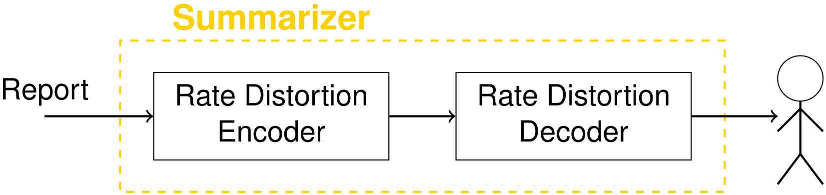
$S \subset T$

$$\mathcal{T}(s) = \{t : s \subset t\}$$

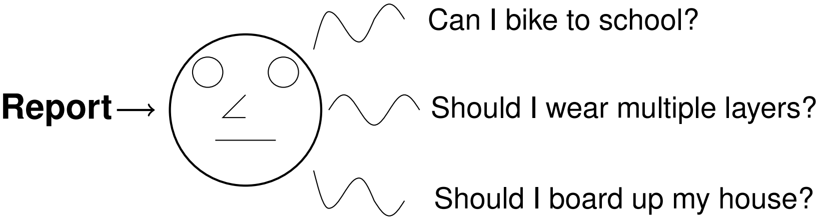
Model formulation

Goal

Minimize $\mathbb{E}[d(T, S)]$, *for some distortion function** d, *between the report,* T, *and summary,* $S \subset T$.



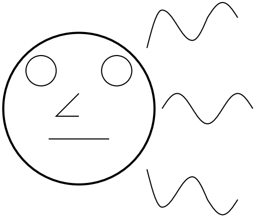
Distortion function, *d*



Distortion function, *d*

Summaries should answer these questions.

Report⟶



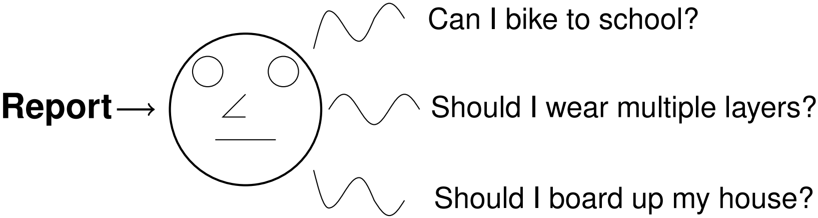
Can I bike to school?

Should I wear multiple layers?

Should I board up my house?

Assumptions

Answers are deterministic function of report.



Assumptions (Example)

Answers are deterministic function of report.

Can I bike to school?

$$\text{If } \{\text{Heavy Rain}\} \cup \left\{ \begin{array}{c} \text{High Winds} \\ \text{and} \\ \text{Low visibility} \end{array} \right\} \cup \{\text{Snow}\} \cup \{\text{Hurricane}\}$$

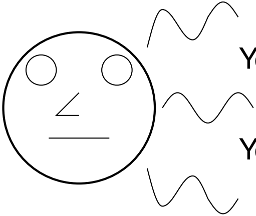
then **NO.**

Otherwise **YES.**

Assumptions

Answer depends on to which sets the report belongs.

Report ⟶



Can I bike to school?

Yes if $T \in \mathcal{W}_1$, No if $T \in \mathcal{W}_1^c$

Should I wear multiple layers?

Yes if $T \in \mathcal{W}_2$, No if $T \in \mathcal{W}_2^c$

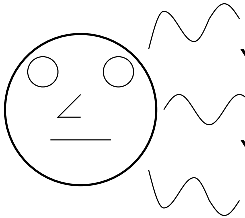
Should I board up my house?

Yes if $T \in \mathcal{W}_3$, No if $T \in \mathcal{W}_3^c$

Model construction

Cost $c(t, \mathcal{W})$ represents importance summary convey $t \in \mathcal{W}$.

Report⟶



Can I bike to school?

Yes if $T \in \mathcal{W}_1$, No if $T \in \mathcal{W}_1^c$

Should I wear multiple layers?

Yes if $T \in \mathcal{W}_2$, No if $T \in \mathcal{W}_2^c$

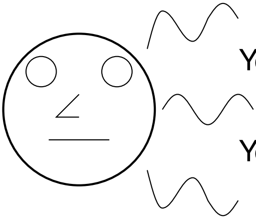
Should I board up my house?

Yes if $T \in \mathcal{W}_3$, No if $T \in \mathcal{W}_3^c$

Model construction

$$d(t,s) = \sum_{\mathcal{W}:t\in\mathcal{W}} c(t,\mathcal{W}) \cdot \textbf{ambiguity penalty}(\mathcal{W}|s)$$

Report⟶



Can I bike to school?

Yes if $T \in \mathcal{W}_1$, No if $T \in \mathcal{W}_1^c$

Should I wear multiple layers?

Yes if $T \in \mathcal{W}_2$, No if $T \in \mathcal{W}_2^c$

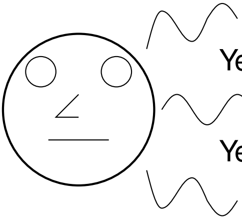
Should I board up my house?

Yes if $T \in \mathcal{W}_3$, No if $T \in \mathcal{W}_3^c$

Ambiguity penalty

$i_S(\cdot)$ models user's estimated probability of report $\cdot$ given summary s.

Summary $\longrightarrow$



Can I bike to school?

Yes $\sim i_S(\mathcal{W}_1)$, No $\sim i_S(\mathcal{W}_1^c)$

Should I wear multiple layers?

Yes $\sim i_S(\mathcal{W}_2)$, No $\sim i_S(\mathcal{W}_2^c)$

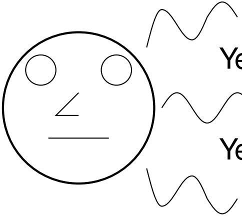
Should I board up my house?

Yes $\sim i_S(\mathcal{W}_3)$, No $\sim i_S(\mathcal{W}_3^c)$

Ambiguity penalty

$$i_s(\mathcal{W}) = \Pr\left(T \in \mathcal{W} | s \subset T\right) = \frac{p_T(\mathcal{W} \cap \mathcal{T}(s))}{p_T(\mathcal{T}(s))}$$

Summary $\longrightarrow$



Can I bike to school?

Yes $\sim i_S(\mathcal{W}_1)$, No $\sim i_S(\mathcal{W}_1^c)$

Should I wear multiple layers?

Yes $\sim i_S(\mathcal{W}_2)$, No $\sim i_S(\mathcal{W}_2^c)$

Should I board up my house?

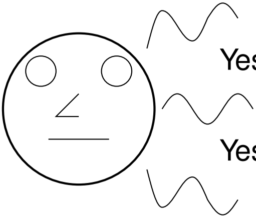
Yes $\sim i_S(\mathcal{W}_3)$, No $\sim i_S(\mathcal{W}_3^c)$

Ambiguity penalty

Use *f*-divergence to measure distance, variational distance gives

$$\textbf{ambiguity penalty}(\mathcal{W}|s) = 1 - i_s(\mathcal{W})$$

Summary $\longrightarrow$



Can I bike to school?

Yes $\sim i_S(\mathcal{W}_1)$, No $\sim i_S(\mathcal{W}_1^c)$

Should I wear multiple layers?

Yes $\sim i_S(\mathcal{W}_2)$, No $\sim i_S(\mathcal{W}_2^c)$

Should I board up my house?

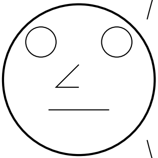
Yes $\sim i_S(\mathcal{W}_3)$, No $\sim i_S(\mathcal{W}_3^c)$

Model construction

$$d(t,s) = \sum_{\mathcal{W}:t\in\mathcal{W}} c(t,\mathcal{W})(1 - i_s(\mathcal{W}))$$

$$= \sum_{\mathcal{W}:t\in\mathcal{W}} c(t,\mathcal{W}) \left(1 - \frac{p_T(\mathcal{W}\cap\mathcal{T}(s))}{p_T(\mathcal{T}(s))}\right)$$

Summary $\longrightarrow$



Can I bike to school?

Yes $\sim i_S(\mathcal{W}_1)$, No $\sim i_S(\mathcal{W}_1^c)$

Should I wear multiple layers?

Yes $\sim i_S(\mathcal{W}_2)$, No $\sim i_S(\mathcal{W}_2^c)$

Should I board up my house?

Yes $\sim i_S(\mathcal{W}_3)$, No $\sim i_S(\mathcal{W}_3^c)$

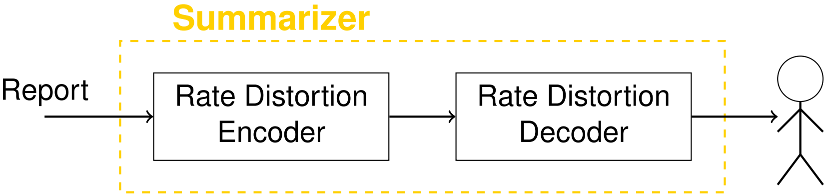
Model

Goal

Minimize $\mathbb{E}[d(T,S)]$, *where*

$$d(t,s) = \sum_{\mathcal{W}:t\in\mathcal{W}} c(t,\mathcal{W}) \left(1 - \frac{p_T(\mathcal{W}\cap\mathcal{T}(s))}{p_T(\mathcal{T}(s))}\right),$$

between the report, T*, and summary,* $S \subset T$.



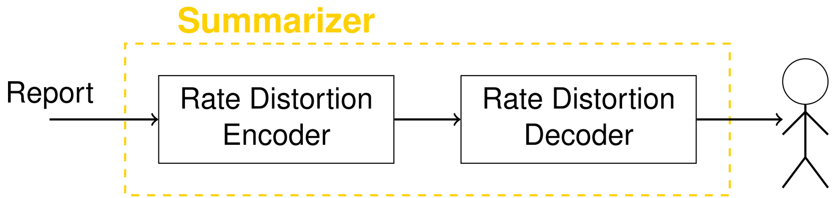
Model

Goal

Minimize $\mathbb{E}[d(T, S)]$, *where*

$$d(t, s) = \sum_{\mathcal{W}:t\in\mathcal{W}} c(t, \mathcal{W}) \left(1 - \frac{p_T(\mathcal{W} \cap \mathcal{T}(s))}{p_T(\mathcal{T}(s))}\right),$$

between the report, T*, and summary,* $S \subset T$.



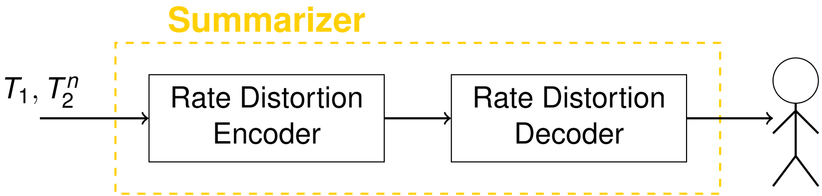
Model

Goal (Universal Summarization)

Minimize

$$\int_{\mathcal{P}(\mathcal{T})} \sum_{t^n} \prod_{i=1}^{n} p_T(t_i) \sum_{\mathcal{W}:t_1 \in \mathcal{W}} c(t_1, \mathcal{W}) \left(1 - \frac{p_T(\mathcal{W} \cap \mathcal{T}(s))}{p_T(\mathcal{T}(s))}\right) \mathrm{d}r$$

where $\frac{\mathrm{d}r}{\mathrm{d}p_T}$ *is constant and* $\int_{\mathcal{P}(\mathcal{T})} \mathrm{d}r = 1$.



Results

Theorem (Optimal universal summarizer)

The optimal summarizer chooses the summary $s(t_1|t_2^n)$ that minimizes

$$\sum_{\mathcal{W}:t_1\in\mathcal{W}} c(t_1,\mathcal{W})\,[1-q(\mathcal{W}\cap\mathcal{T}(s))\eta_{t^n,s}]$$

where

$$\eta_{t^n,s}=\sum_{k=0}^{\infty}\frac{(n+|\mathcal{T}|+k-\pi(\mathcal{T}(s)|t^n)-2)!\,(n+|\mathcal{T}|)!}{(n+|\mathcal{T}|+k)!\,(n+|\mathcal{T}|-\pi(\mathcal{T}(s)|t^n)-2)!},$$

$$q(a)=\frac{\pi(a|t^n)+1}{n+|\mathcal{T}|}. \quad (1)$$

Analysis

Proof.

Calculate

$$\int_{\mathcal{P}(\mathcal{T})} \prod_{i=1}^{n} p_T(t_i) \left(1 - \frac{p_T(\mathcal{W} \cap \mathcal{T}(s))}{p_T(\mathcal{T}(s))}\right) \mathrm{d}r,$$

by writing p_T as a $|\mathcal{T}|$-dimensional vector, using taylor's theorem, and then applying $\int_0^y (y-x)^a x^b \;\mathrm{d}x = \frac{a!b!}{(a+b+1)!} y^{a+b+1}$ recursively. □

Goal (Universal Summarization)

Minimize

$$\int_{\mathcal{P}(\mathcal{T})} \sum_{t^n} \prod_{i=1}^{n} p_T(t_i) \sum_{\mathcal{W}: t_1 \in \mathcal{W}} c(t_1, \mathcal{W}) \left(1 - \frac{p_T(\mathcal{W} \cap \mathcal{T}(s))}{p_T(\mathcal{T}(s))}\right) \mathrm{d}r$$

where $\frac{\mathrm{d}r}{\mathrm{d}p_T}$ *is constant and* $\int_{\mathcal{P}(\mathcal{T})} \mathrm{d}r = 1$.

Results

Lemma

For positive integers b, a such that $1 \leq b < b+2 \leq a$,

$$\frac{a+1}{a-b} < \sum_{y=0}^{\infty} \frac{(b+y)!a!}{(a+y)!b!} \leq \frac{a+1}{a-b}\left(1+\varepsilon(a-b)\right)$$

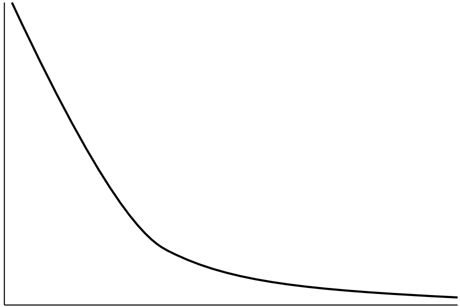
where

$$\varepsilon(x) = 3\frac{1+\ln(x)}{x} + 4e^{\frac{1}{12}} \cdot 2^{-x/2}.$$

Analysis

Proof.

Plotting $y = \frac{(b+x)!a!}{(a+x)!b!}$

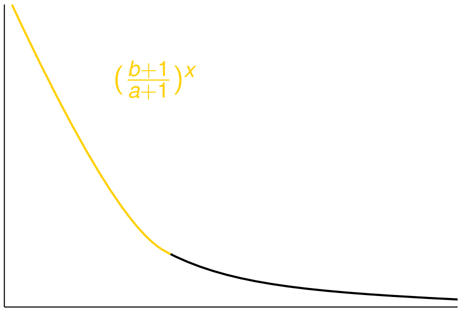


□

Analysis

Proof.

Plotting $y = \frac{(b+x)!a!}{(a+x)!b!}$

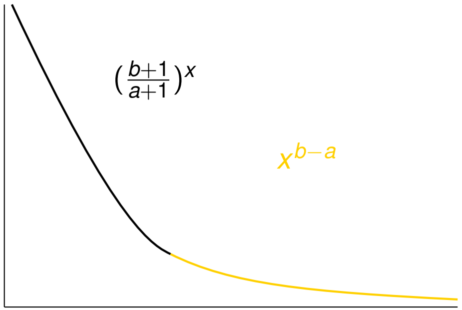


□

Analysis

Proof.

Plotting $y = \frac{(b+x)!a!}{(a+x)!b!}$



□

Best results

Theorem

The (nearly) optimal summarizer chooses the summary $s(t_1|t_2^n)$ that minimizes

$$\sum_{\mathcal{W}:t_1\in\mathcal{W}} c(t_1,\mathcal{W})\left[1-\frac{q(\mathcal{W}\cap\mathcal{T}(s))}{\hat{q}(\mathcal{T}(s))}\right]$$

where

$$q(a)=\frac{\pi(a|t^n)+1}{n+|\mathcal{T}|},\quad \hat{q}(a)=\begin{cases}\frac{\pi(a|t^n)+2}{n+|\mathcal{T}|+1} & \text{if } a=t_1\\ \frac{\pi(a|t^n)+1}{n+|\mathcal{T}|+1} & \text{else}\end{cases}.$$

Conclusions

Any questions?