Automated Selection of Social Media Responses to News

Tadej Štajner[1], Bart Thomee [2], Ana-Maria Popescu[3], Marco Pennachiotti[4], Alejandro Jaimes[2]

[1]Jožef Stefan Institute

[2]Yahoo! Research Barcelona

[3]Research Consulting

[3]eBay, Inc.

KDD 2013, Chicago, Illinois, USA
August 12, 2013

Outline

1 Introduction

2 Proposed method

3 Related work

4 Evaluation

5 Conclusion

Motivation

- While news is produced through traditional channels, a lot of content related to these news events is produced through social media
- People post messages to share a particular news article, to express an opinion about the ongoing events, or to add or refute information about the topic or the mentioned article.
- Displaying social responses near a news article for an enhanced reading experience.

Motivation

- While news is produced through traditional channels, a lot of content related to these news events is produced through social media
- People post messages to share a particular news article, to express an opinion about the ongoing events, or to add or refute information about the topic or the mentioned article.
- Displaying social responses near a news article for an enhanced reading experience.

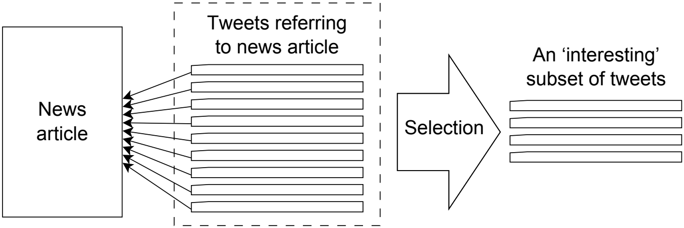
Motivation

- While news is produced through traditional channels, a lot of content related to these news events is produced through social media
- People post messages to share a particular news article, to express an opinion about the ongoing events, or to add or refute information about the topic or the mentioned article.
- Displaying social responses near a news article for an enhanced reading experience.

Situation

Given:

- a news article
- a set of tweets, responding to that news article



Hypothesis

Our hypothesis is that an interesting subset consists of diverse, informative, opinionated and popular messages composed by users with authority.

Problems

- There is a huge volume of tweets, but few actual distinct messages;
- Many tweets are repetitions of the title of the news article, having no additional information;
- Some of them have different content, but are nevertheless redundant;
- Yet, some people actually contribute novel facts or insightful remarks;

Problems

- There is a huge volume of tweets, but few actual distinct messages;
- Many tweets are repetitions of the title of the news article, having no additional information;
- Some of them have different content, but are nevertheless redundant;
- Yet, some people actually contribute novel facts or insightful remarks;

Problems

- There is a huge volume of tweets, but few actual distinct messages;
- Many tweets are repetitions of the title of the news article, having no additional information;
- Some of them have different content, but are nevertheless redundant;
- Yet, some people actually contribute novel facts or insightful remarks;

Problems

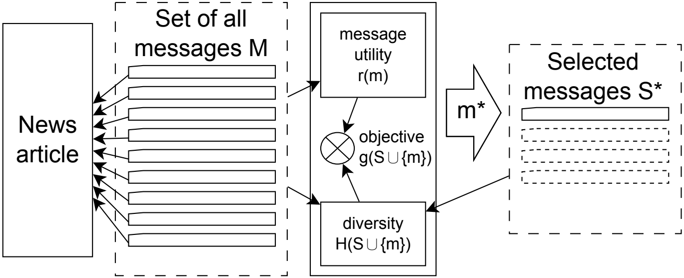
- There is a huge volume of tweets, but few actual distinct messages;
- Many tweets are repetitions of the title of the news article, having no additional information;
- Some of them have different content, but are nevertheless redundant;
- Yet, some people actually contribute novel facts or insightful remarks;

Problem statement

Given a news article and a set of related messages M, we seek a subset S of M, of size k, which are the most "interesting" to a typical reader in the context of the article

Proposed solution

We computationally model the four message-level indicators using an utility function r and the set-level diversity indicator using a normalized entropy function H_0, combined into a common objective function $g(S)$



Objective function

The solution to the social message selection problem is to optimize this objective function:

$$S^* = \underset{S \subset M, |S| = k}{\arg\max} \, (g(S)) \tag{1}$$

$$g(S) = \lambda \sum_{m \in S} r(m) + (1 - \lambda) H_0(S) \tag{2}$$

We consider the properties of our objective function to find a fast approximation to the problem.

Objective function

The solution to the social message selection problem is to optimize this objective function:

$$S^* = \underset{S \subset M, |S| = k}{\arg\max} \, (g(S)) \tag{1}$$

$$g(S) = \lambda \sum_{m \in S} r(m) + (1 - \lambda) H_0(S) \tag{2}$$

We consider the properties of our objective function to find a fast approximation to the problem.

Objective function

To this end, we demonstrate that $g(S)$ is a submodular set function, meaning that the difference of the value of the function that a single element addition makes decreases as the size of the input set increases.

Objective function

- $r(S) = \sum_{m \in S} r(m)$ - submodular, since

$$\begin{gathered} \forall X, Y \subseteq \Omega, X \subseteq Y, \forall x \in \Omega \setminus Y : \\ f(X \cup \{x\}) - f(X) \geq f(Y \cup \{x\}) - f(Y) \end{gathered} \tag{3}$$

- $H_0(S)$ - entropy is also known to be submodular
- A linear combination of submodular set functions is also submodular, which is why a greedy approach is at most $1 - \frac{1}{e}$ times worse than the optimal set of examples.

Objective function

- $r(S) = \sum_{m \in S} r(m)$ - submodular, since

$$\forall X, Y \subseteq \Omega, X \subseteq Y, \forall x \in \Omega \setminus Y : \\ f(X \cup \{x\}) - f(X) \geq f(Y \cup \{x\}) - f(Y) \tag{3}$$

- $H_0(S)$ - entropy is also known to be submodular
- A linear combination of submodular set functions is also submodular, which is why a greedy approach is at most $1 - \frac{1}{e}$ times worse than the optimal set of examples.

Objective function

- $r(S) = \sum_{m \in S} r(m)$ - submodular, since

$$\begin{gathered} \forall X, Y \subseteq \Omega, X \subseteq Y, \forall x \in \Omega \setminus Y: \\ f(X \cup \{x\}) - f(X) \geq f(Y \cup \{x\}) - f(Y) \end{gathered} \tag{3}$$

- $H_0(S)$ - entropy is also known to be submodular
- A linear combination of submodular set functions is also submodular, which is why a greedy approach is at most $1 - \frac{1}{e}$ times worse than the optimal set of examples.

Objective function - scoring

- Let r be a function, modeled by θ, that outputs a real value for a given tweet t : $y_i = r(t_i|\Theta), y_i \in \Re$
- We obtain Θ by training a support vector regression model on individual tweet judgments: (τ, y_{gold})

Objective function - scoring

- Let r be a function, modeled by θ, that outputs a real value for a given tweet t : $y_i = r(t_i|\Theta), y_i \in \Re$
- We obtain Θ by training a support vector regression model on individual tweet judgments: (τ, y_{gold})

Algorithm

Algorithm 1: Greedy iterative sampling algorithm

Input : A collection of messages M, the sample size k

Output: A sample set of messages S^*

$S \longleftarrow \{\arg\max_{m_i \in M}(r(m_i))\}$

while $|S| < k$ **do**

  $m^* \longleftarrow \arg\max_{m_i \in M \setminus S}(g(S \cup \{m_i\}))$

  $S \longleftarrow S \cup \{m^*\}$

end

$S^* \longleftarrow S$

Features

Both components of $g(S)$ have distinct features sets: features for $r(S)$ are scoring features, while features for $H_0(S)$ are selection features.

Selection features: n-grams, location, retweet, reply

Scoring features: average TF-IDF score, log-likelihood given a language model, length, me-information, question, quotation, quality, sentiment, intensity, controversy, retweet, reply, friends, followers, follower-friend ratio, number of retweets, tweet-retweet ratio, user is verified, user is spam, user authority, user topic authority

Features

Both components of $g(S)$ have distinct features sets: features for $r(S)$ are scoring features, while features for $H_0(S)$ are selection features.

Selection features: n-grams, location, retweet, reply

Scoring features: average TF-IDF score, log-likelihood given a language model, length, me-information, question, quotation, quality, sentiment, intensity, controversy, retweet, reply, friends, followers, follower-friend ratio, number of retweets, tweet-retweet ratio, user is verified, user is spam, user authority, user topic authority

Features

Both components of $g(S)$ have distinct features sets: features for $r(S)$ are scoring features, while features for $H_0(S)$ are selection features.

Selection features: n-grams, location, retweet, reply

Scoring features: average TF-IDF score, log-likelihood given a language model, length, me-information, question, quotation, quality, sentiment, intensity, controversy, retweet, reply, friends, followers, follower-friend ratio, number of retweets, tweet-retweet ratio, user is verified, user is spam, user authority, user topic authority

Related work: Diversity-based sampling[1]

Goal: sample tweets with respect to the desired diversity level

Method: greedy approach that iteratively picks tweets that minimize the distortion of entropy of the current subset using a probabilistic model of tweets given their diversity using social, content and network features.

Label: diversity

[1] M. De Choudhury, et al., Find me the right content! Diversity-based sampling of social media spaces for topic-centric search. ICWSM 2011

Related work: Social context summarization [2]

Goal: Joint summarization of news and tweets;

Method: based on conditional random fields that simultaneously addresses message and document summarization by modeling messages and sentences as connected 'wings' in a factor graph

Label: LR+, DWFG

[2] Zi Yang et al.: Social Context Summarization, SIGIR 2011

Our contribution

- We focus on a comprehensive notion of interestingness instead of solely optimizing for message diversity;
- We detect and avoid redundancy between the content of messages by including a richer set of textual features;
- We model the importance of messages beyond their literal content by including features that capture the wider social, conversational and linguistic context;
- We provide a theoretic guarantee of the sample quality.

Methods

DIVERSITY: de Choudhury et al, 2011
LR+: Yang et al, 2011
DWFG: Yang et al, 2011
SVR: message scoring only
ENTROPY: entropy selection only
SVR_ENTROPY: scoring, combined with entropy selection

Methods

DIVERSITY: de Choudhury et al, 2011
LR+: Yang et al, 2011
DWFG: Yang et al, 2011
SVR: message scoring only
ENTROPY: entropy selection only
SVR_ENTROPY: scoring, combined with entropy selection

Scenario 1: individual message labeling

Given a news article and a set of messages, judge each tweet on $\{-1, 0, +1\}$ on each of the folowing criteria:

Informative: the message brings new information to the news article and sample.

Opinionated: the message expresses a strong/controversial opinion or emotions.

Interesting: the message would serve as a good example of presenting the essence of the people's response to news.

Popularity and authority are computed mechanically.

Scenario 1: individual message labeling

Given a news article and a set of messages, judge each tweet on $\{-1, 0, +1\}$ on each of the folowing criteria:

Informative: the message brings new information to the news article and sample.

Opinionated: the message expresses a strong/controversial opinion or emotions.

Interesting: the message would serve as a good example of presenting the essence of the people's response to news.

Popularity and authority are computed mechanically.

Dataset

28,055 messages, gathered from Twitter, referring to **45 news articles** annotated with **3 indicators** by **14 professional editors**.

Individual message labeling - evaluation

We evaluate the individual message labeling modeling using F_1 and ROUGE, a widely used metric in text summarization.

These scores measure how closely the output of a method on a single tweet is to the "ideal" output, i.e. those scored positively by the annotators.

Individual message labeling - evaluation

We evaluate the individual message labeling modeling using F_1 and ROUGE, a widely used metric in text summarization.
These scores measure how closely the output of a method on a single tweet is to the "ideal" output, i.e. those scored positively by the annotators.

Individual message labeling

Table : Comparison of the tweet sets generated by the various methods in terms of their informativeness, opinionatedness and interestingness as measured by ROUGE-2 and F_1 at $k = 10$

	Informative		Opinionated		Interesting	
	R-2	F_1	R-2	F_1	R-2	F_1
DIVERSITY	0.178	0.050	0.270	0.108	0.280	0.109
DWFG	0.170	0.048	0.259	0.056	0.235	0.051
LR+	0.175	0.032	**0.355**	0.132	0.264	0.054
ENTROPY	**0.246***	0.124*	0.345	0.184*	**0.357***	0.216*
SVR	0.212*	**0.132***	0.326	**0.217***	0.336*	**0.239***
SVR_ENTROPY	0.232*	0.123*	0.343	0.211*	0.346*	0.224*

Decomposing interestingness of individual messages

We fitted a least-squares linear model to the indicators and obtained the following linear combination, supported by a coefficient of determination R^2 of 0.38:

$$int = 0.60 \cdot inf + 0.29 \cdot opi + 0.03 \cdot pop + 0.10 \cdot aut \tag{4}$$

Pairwise sample comparison

Given a news article and the associated tweets, we let evaluators choose between two message selections, generated by different approaches.

- Combining diversity and scoring?
- Comparison against baselines?

Pairwise sample comparison

Given a news article and the associated tweets, we let evaluators choose between two message selections, generated by different approaches.

- **Combining diversity and scoring?**
- **Comparison against baselines?**

Pairwise sample comparison

Table : Comparison of tweet selection methods based on annotator preference votes.

Method A	Method B	Votes A	Votes B	None	*p*-value
Combining entropy with scoring					
SVR	SVR_ENTROPY	151	**194**	16	0.02
SVR	ENTROPY	**193**	155	9	0.04
ENTROPY	SVR_ENTROPY	114	**147**	13	0.04
Comparison against baselines					
DIVERSITY	SVR_ENTROPY	79	**158**	3	0.00
DWFG	SVR_ENTROPY	99	**149**	5	0.00
DWFG	DIVERSITY	**118**	94	10	0.10

Distribution of preference reasons

Method A	Method B	inf.	opi.	aut.	pop.
SVR	SVR_ENTROPY	163	197	59	49
ENTROPY	SVR	139	226	43	54
ENTROPY	SVR_ENTROPY	160	125	24	13
DIVERSITY	SVR_ENTROPY	147	122	66	68
DIVERSITY	DWFG	166	109	66	50
DWFG	SVR_ENTROPY	146	134	85	68

Conclusion

- We show that SVR_ENTROPY, a near-optimal solution to the message selection optimization problem outperforms other baselines in a head-to-head comparison scenario.
- We show how various indicators affect interestingness: informativeness being most important, followed by opinionatedness and user authority.

Conclusion

- We show that SVR_ENTROPY, a near-optimal solution to the message selection optimization problem outperforms other baselines in a head-to-head comparison scenario.
- We show how various indicators affect interestingness: informativeness being most important, followed by opinionatedness and user authority.

Conclusion

Interestingness of a message and a message set:

- For individual messages, *informativeness* is most important but simultaneously the most difficult to judge, followed by *opinionatedness* and *popularity*.
- *Message set interestingness* was tied most strongly to *opinionatedness* and then to *informativeness*.

Future work

- Incporporating additional message- and author- level indicators
- Personalizing interestingness models
- Abstractive summarization
- Online sampling

Future work

- Incporporating additional message- and author- level indicators
- Personalizing interestingness models
- Abstractive summarization
- Online sampling

Future work

- Incporporating additional message- and author- level indicators
- Personalizing interestingness models
- Abstractive summarization
- Online sampling

Future work

- Incporporating additional message- and author- level indicators
- Personalizing interestingness models
- Abstractive summarization
- Online sampling

Questions?

arcomem

social sensor

RENDER
Reflecting Knowledge Diversity

SEVENTH FRAMEWORK PROGRAMME

SLOVENIAN RESEARCH AGENCY

social media