Ranking a Stream of News

Gianna M. Del Corso
joint work with Antonio Gullì and Francesco Romani

Dipartimento di Informatica, Università di Pisa, Italy

The 14th International World Wide Web Conference
May 10-14 2005, Chiba, Japan

The Problem

Lot of interest around news engines

News browsing is one of the most important Internet activities.

The Problem

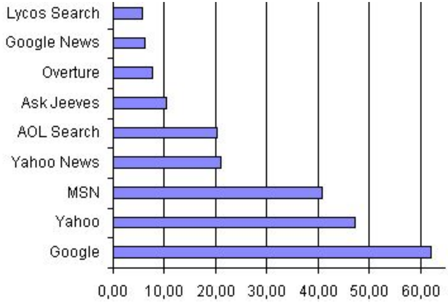
Lot of interest around news engines

News browsing is one of the most important Internet activities.

The Problem

Lot of interest around news engines

News browsing is one of the most important Internet activities.



October 2004 [Nielsen/Netratings]

The commercial scenario

Many commercial news engines available: `Google News`, `Yahoo News`, `MSNBot`, `AllTheWeb News`, `AltaVista News`, `Daypop`, `Ananova` etc ... with news provided by many news agencies.

Personalized services as in `NewsBot` and `Findory`

No public information about how do they work!

From our observation they take into account criteria as freshness of a piece of news, authoritativness of sources, replication/aggregation of pieces of news.

The commercial scenario

Many commercial news engines available: Google News, Yahoo News, MSNBot, AllTheWeb News, AltaVista News, Daypop, Ananova etc ... with news provided by many news agencies.

Personalized services as in NewsBot and Findory

No public information about how do they work!

From our observation they take into account criteria as freshness of a piece of news, authoritativness of sources, replication/aggregation of pieces of news.

The commercial scenario

Many commercial news engines available: `Google News`, `Yahoo News`, `MSNBot`, `AllTheWeb News`, `AltaVista News`, `Daypop`, `Ananova` etc ... with news provided by many news agencies.

Personalized services as in `NewsBot` and `Findory`

No public information about how do they work!

From our observation they take into account criteria as freshness of a piece of news, authoritativness of sources, replication/aggregation of pieces of news.

The commercial scenario

Many commercial news engines available: `Google News`, `Yahoo News`, `MSNBot`, `AllTheWeb News`, `AltaVista News`, `Daypop`, `Ananova` etc ... with news provided by many news agencies.

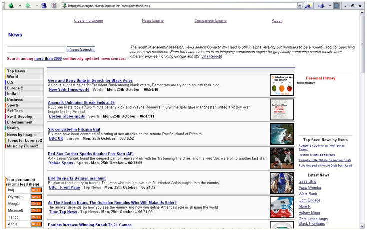
Personalized services as in `NewsBot` and `Findory`

No public information about how do they work!

From our observation they take into account criteria as freshness of a piece of news, authoritativness of sources, replication/aggregation of pieces of news.

The News Engine

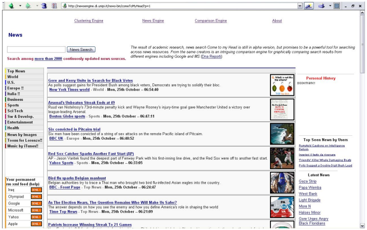
COMETOMYHEAD is an academic news search engine developed by A. Gullì (available at http://newsengine.di.unipi.it) for gathering, indexing, searching and classifying news information.



We have used this engine to gather news articles from about 2,000 sources over a period of two months. The pieces of news collected was about 300,000. The news are classified into 13 categories

The News Engine

COMETOMYHEAD is an academic news search engine developed by A. Gullì (available at http://newsengine.di.unipi.it) for gathering, indexing, searching and classifying news information.



We have used this engine to gather news articles from about 2,000 sources over a period of two months. The pieces of news collected was about 300,000. The news are classified into 13 categories

The Goal

Our goal:

Propose ranking strategies for a stream of news information and a set of news sources.

We do not know other algorithms for comparing our results

. . . moreover . . .

. . . ranking news articles is a different task than ranking web pages

The Goal

Our goal:

Propose ranking strategies for a stream of news information and a set of news sources.

We do not know other algorithms for comparing our results

. . . moreover . . .

. . . ranking news articles is a different task than ranking web pages

The Goal

Our goal:

Propose ranking strategies for a stream of news information and a set of news sources.

We do not know other algorithms for comparing our results

... moreover ...

... ranking news articles is a different task than ranking web pages

The Goal

Our goal:

Propose ranking strategies for a stream of news information and a set of news sources.

We do not know other algorithms for comparing our results

. . . moreover . . .

. . . ranking news articles is a different task than ranking web pages

Some Desiderata

- Property P1: Ranking for News posting and News sources
- Property P2: Important News articles are Clustered
- Property P3: Mutual Reinforcement between News Articles and News Sources
- Property P4: Time awareness
- Property P5: Online processing

Some Desiderata

- Property P1: Ranking for News posting and News sources
- Property P2: Important News articles are Clustered
- Property P3: Mutual Reinforcement between News Articles and News Sources
- Property P4: Time awareness
- Property P5: Online processing

Some Desiderata

- Property P1: Ranking for News posting and News sources
- Property P2: Important News articles are Clustered
- Property P3: Mutual Reinforcement between News Articles and News Sources
- Property P4: Time awareness
- Property P5: Online processing

Some Desiderata

- Property P1: Ranking for News posting and News sources
- Property P2: Important News articles are Clustered
- Property P3: Mutual Reinforcement between News Articles and News Sources
- Property P4: Time awareness
- Property P5: Online processing

Some Desiderata

- Property P1: Ranking for News posting and News sources
- Property P2: Important News articles are Clustered
- Property P3: Mutual Reinforcement between News Articles and News Sources
- Property P4: Time awareness
- Property P5: Online processing

Some Desiderata

- Property P1: Ranking for News posting and News sources
- Property P2: Important News articles are Clustered
- Property P3: Mutual Reinforcement between News Articles and News Sources
- Property P4: Time awareness
- Property P5: Online processing

The Model for News Articles

Let N be the news stream observed and let S be a set of news sources; the news creation process can be represented by means of a undirected graph

$$G = (N \cup S, E)$$

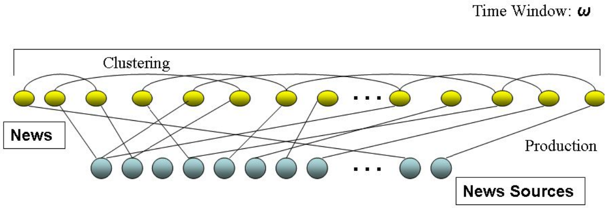
Two kinds of edges:

1. undirected edges from nodes in S and nodes in N
2. undirected edges with both endpoints in N representing the results of the clustering process

The Model for News Articles

Let N be the news stream observed and let S be a set of news sources; the news creation process can be represented by means of a undirected graph

$$G = (N \cup S, E)$$



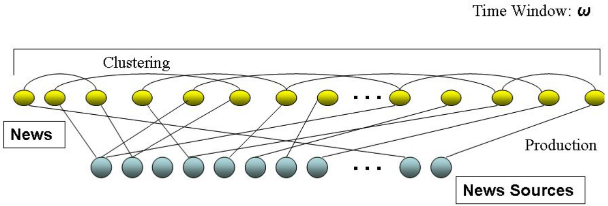
Two kinds of edges:

1. undirected edges from nodes in S and nodes in N
2. undirected edges with both endpoints in N representing the results of the clustering process

The Model for News Articles

Let N be the news stream observed and let S be a set of news sources; the news creation process can be represented by means of a undirected graph

$$G = (N \cup S, E)$$



Two kinds of edges:

1. undirected edges from nodes in S and nodes in N
2. undirected edges with both endpoints in N representing the results of the clustering process

Clustering Technique

We adopt a continuous measure of the lexical similarity between news posting

In our current implementation

- Similarity between news abstracts is represented using the canonical bag of words paradigm
- The abstracts are filtered out against a list of stop words
- The lexical similarity is expressed as a function of the number of common words

Clustering Technique

We adopt a continuous measure of the lexical similarity between news posting

In our current implementation

- Similarity between news abstracts is represented using the canonical bag of words paradigm
- The abstracts are filtered out against a list of stop words
- The lexical similarity is expressed as a function of the number of common words

Clustering Technique

We adopt a continuous measure of the lexical similarity between news posting

In our current implementation

- Similarity between news abstracts is represented using the canonical bag of words paradigm
- The abstracts are filtered out against a list of stop words
- The lexical similarity is expressed as a function of the number of common words

Clustering Technique

We adopt a continuous measure of the lexical similarity between news posting

In our current implementation

- Similarity between news abstracts is represented using the canonical bag of words paradigm
- The abstracts are filtered out against a list of stop words
- The lexical similarity is expressed as a function of the number of common words

Clustering Technique

We adopt a continuous measure of the lexical similarity between news posting

In our current implementation

- Similarity between news abstracts is represented using the canonical bag of words paradigm
- The abstracts are filtered out against a list of stop words
- The lexical similarity is expressed as a function of the number of common words

Minimal Requirements

We have designed our ranking algorithm requiring that it will handle correctly two limit cases.

1. LC1 Source s_1 emits a stream of news articles with emission rate $1/\Delta$.
   We expect s_1 to have a stationary mean rank μ independent of time, but increasing with $1/\Delta$.
2. LC2 Two sources s_1 and s_2, one mirroring the other.

Minimal Requirements

We have designed our ranking algorithm requiring that it will handle correctly two limit cases.

1. LC1 Source s_1 emits a stream of news articles with emission rate $1/\Delta$.
   We expect s_1 to have a stationary mean rank μ independent of time, but increasing with $1/\Delta$
2. LC2 Two sources s_1 and s_2, one mirroring the other.
   We expect s_1 and s_2 to have a similar rank.

Minimal Requirements

We have designed our ranking algorithm requiring that it will handle correctly two limit cases.

1. LC1 Source s_1 emits a stream of news articles with emission rate $1/\Delta$.
   We expect s_1 to have a stationary mean rank μ independent of time, but increasing with $1/\Delta$
2. LC2 Two sources s_1 and s_2, one mirroring the other.
   We expect s_1 and s_2 to have a similar rank.

Minimal Requirements

We have designed our ranking algorithm requiring that it will handle correctly two limit cases.

1. LC1 Source s_1 emits a stream of news articles with emission rate $1/\Delta$.
   We expect s_1 to have a stationary mean rank μ independent of time, but increasing with $1/\Delta$
2. LC2 Two sources s_1 and s_2, one mirroring the other.
   We expect s_1 and s_2 to have a similar rank.

Minimal Requirements

We have designed our ranking algorithm requiring that it will handle correctly two limit cases.

1. LC1 Source s_1 emits a stream of news articles with emission rate $1/\Delta$.
   We expect s_1 to have a stationary mean rank μ independent of time, but increasing with $1/\Delta$
2. LC2 Two sources s_1 and s_2, one mirroring the other.
   We expect s_1 and s_2 to have a similar rank.

Non-time-aware algorithms

The naive approach

A news source has a rank proportional to the number of pieces of news released

Behaves poorly for the limit case LC1. The rank of a single source will increase unboundedly

Mutual reinforcement

The second algorithm exploits the mutual reinforcement property between news articles and news sources. Defines the rank as the right eigenvalue of the adjacency matrix of the graph G.

Behaves poorly for the limit case LC1. The rank $\mathbf{r} = \mathbf{0}$ is the only solution.

Non-time-aware algorithms

The naive approach

A news source has a rank proportional to the number of pieces of news released

Behaves poorly for the limit case LC1. The rank of a single source will increase unboundedly

Mutual reinforcement

The second algorithm exploits the mutual reinforcement property between news articles and news sources. Defines the rank as the right eigenvalue of the adjacency matrix of the graph G.

Behaves poorly for the limit case LC1. The rank $\mathbf{r} = \mathbf{0}$ is the only solution.

Non-time-aware algorithms

The naive approach

A news source has a rank proportional to the number of pieces of news released

Behaves poorly for the limit case LC1. The rank of a single source will increase unboundedly

Mutual reinforcement

The second algorithm exploits the mutual reinforcement property between news articles and news sources. Defines the rank as the right eigenvalue of the adjacency matrix of the graph G.

Behaves poorly for the limit case LC1. The rank $\mathbf{r} = \mathbf{0}$ is the only solution.

Non-time-aware algorithms

The naive approach

A news source has a rank proportional to the number of pieces of news released

Behaves poorly for the limit case LC1. The rank of a single source will increase unboundedly

Mutual reinforcement

The second algorithm exploits the mutual reinforcement property between news articles and news sources. Defines the rank as the right eigenvalue of the adjacency matrix of the graph G.

Behaves poorly for the limit case LC1. The rank $\mathbf{r} = \mathbf{0}$ is the only solution.

Time-aware algorithms

1. To deal with news data stream we have to design time aware mechanism.
2. The importance of a piece of news depends also on the time of its posting.
3. We introduce α, a parameter accounting for the decay in freshness of a news story.
4. The value of α depends on the category the pieces of news belong.

Time-aware algorithms

1. To deal with news data stream we have to design time aware mechanism.
2. The importance of a piece of news depends also on the time of its posting.
3. We introduce α, a parameter accounting for the decay in freshness of a news story.
4. The value of α depends on the category the pieces of news belong.

Time-aware algorithms

1. To deal with news data stream we have to design time aware mechanism.
2. The importance of a piece of news depends also on the time of its posting.
3. We introduce α, a parameter accounting for the decay in freshness of a news story.
4. The value of α depends on the category the pieces of news belong.

Time-aware algorithms

1. To deal with news data stream we have to design time aware mechanism.
2. The importance of a piece of news depends also on the time of its posting.
3. We introduce α, a parameter accounting for the decay in freshness of a news story.
4. The value of α depends on the category the pieces of news belong.

Decay Rule

$R(n, t)$ is the rank of news n at time t.

Decay Rule:

$$R(n, t + \tau) = e^{-\alpha \tau} R(n, t), \qquad t > t_i,$$

t_i is the time n_i was posted

Decay Rule

$R(n, t)$ is the rank of news n at time t.

Decay Rule:

$$R(n, t+\tau) = e^{-\alpha\,\tau} R(n, t), \qquad t > t_i,$$

t_i is the time n_i was posted

Algorithms TA1

This class of algorithms assigns to every source the sum of the ranks assigned to the articles emitted by that source in the past.

Let $R(s,t)$ be the rank of source s at time t

With $S(n_i) = s_k$ we denote that n_i has been posted by s_k.

$$R(s_k,t) = \sum_{S(n_i)=s_k} R(n_i,t),$$

Possible definitions for the rank of pieces of news:

- $R(n_i, t_i) = 1$

Algorithms TA1

This class of algorithms assigns to every source the sum of the ranks assigned to the articles emitted by that source in the past.
Let $R(s,t)$ be the rank of source s at time t
With $S(n_i) = s_k$ we denote that n_i has been posted by s_k.

$$R(s_k, t) = \sum_{S(n_i)=s_k} R(n_i, t),$$

Possible definitions for the rank of pieces of news:

- $R(n_i, t_i) = 1$

However...

... the rank of n_i does not depend on the rank of the generating source

Algorithms TA1

This class of algorithms assigns to every source the sum of the ranks assigned to the articles emitted by that source in the past.
Let $R(s, t)$ be the rank of source s at time t
With $S(n_i) = s_k$ we denote that n_i has been posted by s_k.

$$R(s_k, t) = \sum_{S(n_i)=s_k} R(n_i, t),$$

Possible definitions for the rank of pieces of news:

- $R(n_i, t_i) = 1$

However...

... the rank of n_i does not depend on the rank of the generating source

Algorithms TA1

This class of algorithms assigns to every source the sum of the ranks assigned to the articles emitted by that source in the past.
Let $R(s,t)$ be the rank of source s at time t
With $S(n_i) = s_k$ we denote that n_i has been posted by s_k.

$$R(s_k, t) = \sum_{S(n_i)=s_k} R(n_i, t),$$

Possible definitions for the rank of pieces of news:

- $R(n_i, t_i) = 1$

However...

... the rank of n_i does not depend on the rank of the generating source

Algorithms TA1

This class of algorithms assigns to every source the sum of the ranks assigned to the news emitted by that source in the past. Let $R(s,t)$ be the rank of source s at time t
With $S(n_i) = s_k$ we denote that n_i has been posted by s_k.

$$R(s_k, t) = \sum_{S(n_i)=s_k} R(n_i, t),$$

Possible definitions for the rank of pieces of news:

- $R(n_i, t_i) = c \lim_{\tau \to 0^+} R(S(n_i), t_i - \tau), \quad 0 < c < 1,$

It doesn't work for the limit case LC1

Algorithms TA1

This class of algorithms assigns to every source the sum of the ranks assigned to the news emitted by that source in the past. Let $R(s,t)$ be the rank of source s at time t

With $S(n_i) = s_k$ we denote that n_i has been posted by s_k.

$$R(s_k, t) = \sum_{S(n_i)=s_k} R(n_i, t),$$

Possible definitions for the rank of pieces of news:

- $R(n_i, t_i) = c \lim_{\tau \to 0^+} R(S(n_i), t_i - \tau), \quad 0 < c < 1,$

It doesn't work for the limit case LC1

Algorithms TA1

This class of algorithms assigns to every source the sum of the ranks assigned to the news emitted by that source in the past. Let $R(s,t)$ be the rank of source s at time t
With $S(n_i) = s_k$ we denote that n_i has been posted by s_k.

$$R(s_k, t) = \sum_{S(n_i)=s_k} R(n_i, t),$$

Possible definitions for the rank of pieces of news:

- $R(n_i, t_i) = [\lim_{\tau \to 0^+} R(S(n_i), t_i - \tau)]^{\beta}, \quad 0 < \beta < 1.$

β is similar to the magic ε in PageRank.

Algorithms TA1

This class of algorithms assigns to every source the sum of the ranks assigned to the news emitted by that source in the past. Let $R(s,t)$ be the rank of source s at time t
With $S(n_i) = s_k$ we denote that n_i has been posted by s_k.

$$R(s_k, t) = \sum_{S(n_i)=s_k} R(n_i, t),$$

Possible definitions for the rank of pieces of news:

- $R(n_i, t_i) = [\lim_{\tau \to 0^+} R(S(n_i), t_i - \tau)]^{\beta}, \quad 0 < \beta < 1.$

β is similar to the magic ε in PageRank.

Algorithms TA1

Summarizing

Algorithm TA1

$$
\begin{aligned}
R(s_k, t) &= \sum_{S(n_i)=s_k} R(n_i, t), \\
R(n_i, t_i) &= \left[\lim_{\tau \to 0^+} R(S(n_i), t_i - \tau) \right]^{\beta}
\end{aligned}
$$

Desiderata

- Property P1: Ranking for News posting and News sources
- Property P2: Important News articles are Clustered
- Property P3: Mutual Reinforcement between News Articles and News Sources
- Property P4: Time awareness
- Property P5: Online processing

Algorithms TA1

A possible algorithm is

Algorithm TA1

$$
\begin{aligned}
R(s_k, t) &= \sum_{S(n_i)=s_k} R(n_i, t), \\
R(n_i, t_i) &= \left[\lim_{\tau \to 0^+} R(S(n_i), t_i - \tau) \right]^{\beta}
\end{aligned}
$$

It doesn't take into account the clustering process of news!

Algorithms TA2

A good news ranking algorithm working on a stream of information should exploit some data stream clustering technique.

$$R(n_i, t_i) = \left[\lim_{\tau \to 0^+} R\left(S(n_i), t_i - \tau\right)\right]^\beta + \\ + \sum_{t_j < t_i} e^{-\alpha(t_i - t_j)} \sigma_{ij} R(n_j, t_j)^\beta,$$

Algorithms TA2

A good news ranking algorithm working on a stream of information should exploit some data stream clustering technique.

$$
\begin{aligned}
R(n_i, t_i) &= \left[\lim_{\tau \to 0^+} R\left(S(n_i), t_i - \tau\right)\right]^{\beta} + \\
&+ \sum_{t_j < t_i} e^{-\alpha(t_i - t_j)} \sigma_{ij} R(n_j, t_j)^{\beta},
\end{aligned}
$$

Algorithms TA2

A good news ranking algorithm working on a stream of information should exploit some data stream clustering technique.

$$
\begin{aligned}
R(n_i, t_i) &= \left[\lim_{\tau \to 0^+} R\left(S(n_i), t_i - \tau\right)\right]^{\beta} + \\
&+ \sum_{t_j < t_i} e^{-\alpha(t_i - t_j)} \sigma_{ij} R(n_j, t_j)^{\beta},
\end{aligned}
$$

The rank of a piece of news depends on

- the rank of the source
- the rank of "similar" pieces of news

Algorithms TA2

A good news ranking algorithm working on a stream of information should exploit some data stream clustering technique.

$$\begin{aligned} R(n_i, t_i) &= \left[\lim_{\tau \to 0^+} R\left(S(n_i), t_i - \tau\right) \right]^\beta + \\ &+ \sum_{t_j < t_i} e^{-\alpha(t_i - t_j)} \sigma_{ij} R(n_j, t_j)^\beta, \end{aligned}$$

The rank of a piece of news depends on

- the rank of the source
- the rank of "similar" pieces of news

Algorithms TA2

A good news ranking algorithm working on a stream of information should exploit some data stream clustering technique.

$$R(n_i, t_i) = \left[\lim_{\tau \to 0^+} R\left(S(n_i), t_i - \tau\right)\right]^\beta + \sum_{t_j < t_i} e^{-\alpha(t_i - t_j)} \sigma_{ij} R(n_j, t_j)^\beta,$$

The rank of a piece of news depends on

- the rank of the source
- the rank of "similar" pieces of news

Algorithms TA2

Too bad!

LC2 case

A news source mirroring another, gets a finite rank significantly greater than the rank of the mirrored one!

The Final TA algorithm: TA3

Idea

Modify a posteriori the rank of a source

A source which has emitted in the past news stories highly mirrored in the future, will receive a “bonus” acknowledging the importance

The Final TA algorithm: TA3

Idea

Modify a posteriori the rank of a source

A source which has emitted in the past news stories highly mirrored in the future, will receive a “bonus” acknowledging the importance

Algorithms TA3

The algorithm is

$$
\begin{aligned}
R(s_k, t) &= \sum_{S(n_i)=s_k} R(n_i, t) + \\
&+ \sum_{S(n_i)=s_k} e^{-\alpha(t-t_i)} \sum_{\substack{t_j \in [t_i, t] \\ S(n_j) \neq s_k}} \sigma_{ij} R(n_j, t_j)^{\beta}, \\
R(n_i, t_i) &= \left[\lim_{\tau \to 0^+} R\left(S(n_i), t_i - \tau\right) \right]^{\beta} + \\
&+ \sum_{t_j < t_i} e^{-\alpha(t_i - t_j)} \sigma_{ij} R(n_j, t_j)^{\beta}.
\end{aligned}
$$

Algorithms TA3

The rank of s_k is

$$R(s_k, t) = \sum_{S(n_i)=s_k} R(n_i, t) + \sum_{S(n_i)=s_k} e^{-\alpha(t-t_i)} \sum_{\substack{t_j \in [t_i, t] \\ S(n_j) \neq s_k}} \sigma_{ij} R(n_j, t_j)^{\beta}$$

- the ranks of the pieces of news generated in the past
- a factor of the rank of news articles similar and posted later on by other sources

Algorithms TA3

The rank of s_k is

$$R(s_k, t) = \sum_{S(n_i)=s_k} R(n_i, t) + \sum_{S(n_i)=s_k} e^{-\alpha(t-t_i)} \sum_{\substack{t_j \in [t_i, t] \\ S(n_j) \neq s_k}} \sigma_{ij} R(n_j, t_j)^{\beta}$$

- the ranks of the pieces of news generated in the past
- a factor of the rank of news articles similar and posted later on by other sources

Algorithms TA3

The rank of s_k is

$$R(s_k, t) = \sum_{S(n_i)=s_k} R(n_i, t) + \sum_{S(n_i)=s_k} e^{-\alpha(t-t_i)} \sum_{\substack{t_j \in [t_i, t] \\ S(n_j) \neq s_k}} \sigma_{ij} R(n_j, t_j)^{\beta}$$

- the ranks of the pieces of news generated in the past
- a factor of the rank of news articles similar and posted later on by other sources

Algorithms TA3

The rank of a single piece of news n_i is still

$$\begin{aligned} R(n_i, t_i) &= \left[\lim_{\tau \to 0^+} R\left(S(n_i), t_i - \tau\right)\right]^\beta + \\ &+ \sum_{t_j < t_i} e^{-\alpha(t_i - t_j)} \sigma_{ij} R(n_j, t_j)^\beta. \end{aligned}$$

Note:

If n_i aggregates with a set stories posted in the future, we do not assign to n_i an extra bonus

The idea is that we want to privilege the freshness of a news article rather than its clustering importance

Algorithms TA3

The rank of a single piece of news n_i is still

$$
\begin{aligned}
R(n_i, t_i) &= \left[\lim_{\tau \to 0^+} R\left(S(n_i), t_i - \tau\right)\right]^\beta + \\
&+ \sum_{t_j < t_i} e^{-\alpha(t_i - t_j)} \sigma_{ij} R(n_j, t_j)^\beta.
\end{aligned}
$$

Note:

If n_i aggregates with a set stories posted in the future, we do not assign to n_i an extra bonus

The idea is that we want to privilege the freshness of a news article rather than its clustering importance

Algorithms TA3

1. TA3 is coherent with all the desirable properties P1–P5
2. unfortunately is more complicated than the others
3. it is not easy to write down a formula for the stationary mean value of the sources for the limit cases LC1 and LC2

Algorithms TA3

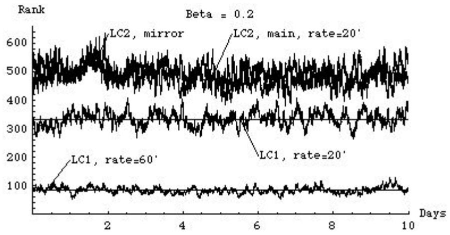
1. TA3 is coherent with all the desirable properties P1–P5
2. unfortunately is more complicated than the others
3. it is not easy to write down a formula for the stationary mean value of the sources for the limit cases LC1 and LC2

Algorithms TA3

1. TA3 is coherent with all the desirable properties P1–P5
2. unfortunately is more complicated than the others
3. it is not easy to write down a formula for the stationary mean value of the sources for the limit cases LC1 and LC2

TA3 behavior on the limit cases

Limit cases LC1 and LC2 are satisfied.



Experimental Settings

- Experiments were performed on a PC with a Pentium IV 3GHz, 2.0GB of memory and 512Kb of L2 cache
- The Java code requires few minutes for ranking about 20,000 pieces of news
- The all computation including the clustering of the articles is done online

Sensitivity to the parameters

A first group of experiments address the sensitivity at changes of the parameters ρ and β

Algorithm TA3 is not much sensitive to changes in the parameters involved

Sensitivity to the parameters

A first group of experiments address the sensitivity at changes of the parameters ρ and β

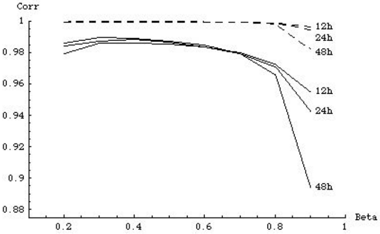
ρ is the half-life decay time, that is

$$e^{-\alpha\rho} = \frac{1}{2}$$

Algorithm TA3 is not much sensitive to changes in the parameters involved

Sensitivity to the parameters

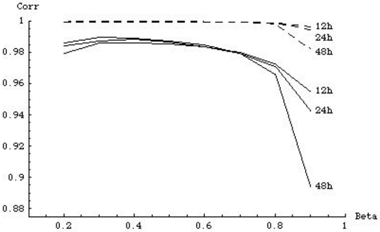
A first group of experiments address the sensitivity at changes of the parameters ρ and β



Algorithm TA3 is not much sensitive to changes in the parameters involved

Sensitivity to the parameters

A first group of experiments address the sensitivity at changes of the parameters ρ and β



Algorithm TA3 is not much sensitive to changes in the parameters involved

TA3 vs NTA1

It is necessary to have such a complicate algorithm?

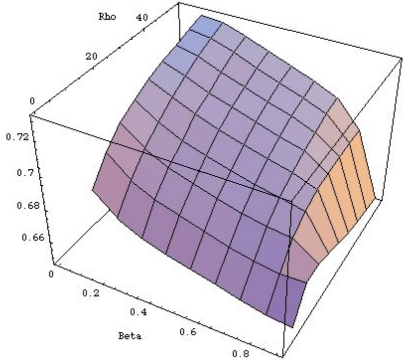
TA3 vs NTA1

It is necessary to have such a complicate algorithm?

Yes

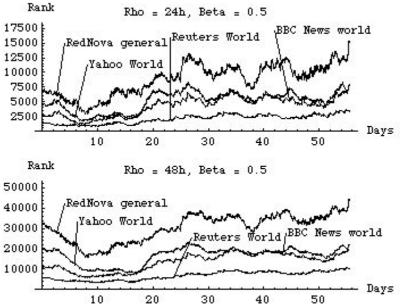
TA3 vs NTA1

It is necessary to have such a complicate algorithm?



Lower correlation for large values of β ad low values of ρ

Rank evolution over 55 days of the top 4 sources in the category `World`



Top News Sources

Source	# Postings
RedNova general	3154
Yahoo World	1924
Reuters World	1363
Yahoo Politics	900
BBC News world	1368
Reuters	555
Xinhua	339
New York Times world	549

Remark

Some news agency are considered more important than others even if they release a lower number of pieces of news

Top News Sources

Source	# Postings
RedNova general	3154
Yahoo World	1924
Reuters World	1363
Yahoo Politics	900
BBC News world	1368
Reuters	555
Xinhua	339
New York Times world	549

Remark

Some news agency are considered more important than others even if they release a lower number of pieces of news

Top News Articles

Posted	News Source	News Abstract
8/17	Reuters	Argentina Wins First Olympic Gold for 52 Years
8/18	Reuters	British Stun US in Sprint Relay
8/18	NBCOlympics	Argentina wins first basketball gold
9/9	Reuters Sports	Monty Seals Record Ryder Cup Triumph for Europe
8/18	Reuters Sports	Men's Basketball: Argentina Beats Italy, Takes Gold
10/11	Yahoo Sports	Pot Charge May Be Dropped Against Anthony (AP)
10/10	Reuters Sports	Record-Breaking Red Sox Reach World Series
8/17	China Daily	China's Xing Huina wins Olympic women's 10,000m gold
8/17	Reuters Sports	El Guerrouj, Holmes Stride Into Olympic History

Top News Articles

Note

For top pieces of news it is common to recognize the same piece of information re-posted by other agencies

The rank of a singular news article is deeply dependent on the rank of the source posting it

Top News Articles

Note

For top pieces of news it is common to recognize the same piece of information re-posted by other agencies

The rank of a singular news article is deeply dependent on the rank of the source posting it

Conclusions

- Algorithms for ranking News articles and news agencies
- Step-by-step construction
- Extensive testing on more than 300,000 pieces of news and 2,000 news sources
- The ranking is done online
- Same ideas for ranking publications, authors and scientific journals etc

Conclusions

- Algorithms for ranking News articles and news agencies
- Step-by-step construction
- Extensive testing on more than 300,000 pieces of news and 2,000 news sources
- The ranking is done online
- Same ideas for ranking publications, authors and scientific journals etc

Conclusions

- Algorithms for ranking News articles and news agencies
- Step-by-step construction
- Extensive testing on more than 300,000 pieces of news and 2,000 news sources
- The ranking is done online
- Same ideas for ranking publications, authors and scientific journals etc

Conclusions

- Algorithms for ranking News articles and news agencies
- Step-by-step construction
- Extensive testing on more than 300,000 pieces of news and 2,000 news sources
- The ranking is done online
- Same ideas for ranking publications, authors and scientific journals etc

Conclusions

- Algorithms for ranking News articles and news agencies
- Step-by-step construction
- Extensive testing on more than 300,000 pieces of news and 2,000 news sources
- The ranking is done online
- Same ideas for ranking publications, authors and scientific journals etc