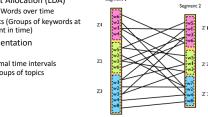
ThemeDelta: Dynamic Segmentations over Temporal Topic Models

Paper By: Samah Gad, Wagas Javed, Sohaib Ghani, Niklas Elmqvist, Tom Ewing, Keith N. Hampton, and Naren Ramakrishnan

Published: IEEE Transactions on Visualization and Computer Graphics 21(5) 2015 Presentation By: Yasha Pushak



Latent Dirichlet Allocation (LDA) Input: Bag of Words over time Output: Topics (Groups of keywords at a specific point in time) **Timeline Segmentation**



What: Text Dataset with Timestamps

"I can't believe I have so many racist friends...'

"Terror in Paris..."

"These are the faces of the Syrian refugees." Men, women, and children who's homes were destroyed and were forced to flee."

"Canada extends it's condolences to France. "Shouldn't we help our

"Diversity is our strength. We strongly condemn the acts aimed at certain Canadians after the Paris attacks.

"Canada is full, Sav

"Canada stands with Paris."



"My great grand parents immigrated

from Russia to escape violence.. No one told them 'We're full'."

What: Data

Timestamped text dataset

What: Derived

Bag of Words over time

Topics (Groups of keywords at a specific point in time)

Why: Tasks

Identify scatter/gather relationships

Why: Identify Scatter/Gather Relationships "These are the faces of the Syrian refugee:



But what if we have lots of data?

What: Derived Bag of word representation

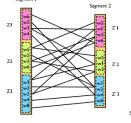
"Canada extends it's condolences to France."

Canada: 1 Condolences: 1 France: 1

Syria: 0

Processing the bags of words

Input: Topics Output: Optimal time intervals containing groups of topics



Time intervals containing groups of topics

Identify changes in topics over time

What? Why? How?

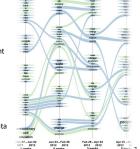
How: Encode

Parallel axes for time segments Spatially partition topics along a segment Label keywords within topics Linked keywords across time intervals

Segment labels for dates and duration

How: Encode (Free Channels) Size of labels for quantitative data

Width of links for quantitative data Link colour for categorical or ordered data



What? Why? How?

How: Encode (Free Channels) Size of labels for quantitative data

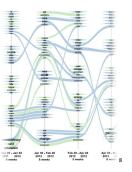
Width of links for quantitative data Link colour for categorical or ordered data

How: Manipulate

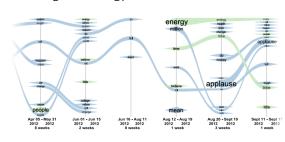
Navigate: geometric zooming and panning Select: highlight keywords Search: Select keywords by searching

How: Reduce

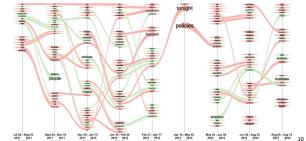
Filter: by selected keywords and resort



Filtering on "Energy"



Example: US Presidential Election 2012 - Mitt Romney



Spanish Flu in the News

Sep 09 - Oct 09

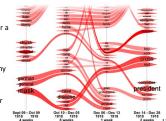
"German", "mask" - advisories from the first world war to wear a

Oct 10 - Dec 05

"home", "family", "son", "daughter" - men from the army were allowed to return home

Dec 06 - Dec 13

"German" disappears - The war was won on November 11.



Changed-Focused Questions

Expert User Study on Spanish Flu Data

How did the newspapers describe the spread of influenza? How does the description of the pandemic change over time?

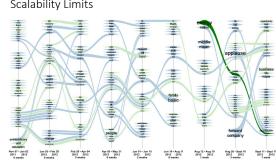
Are there different times when the influenza pandemic becomes less important? What are those time periods?

Connection-Focused Questions

What are the categories that appear to be associated with influenza in different

Was there a specific feeling that surrounded the influenza reporting in the newspapers?

Scalability Limits



Thank You

References

ThemeDelta: Dynamic Segmentations over Temporal Topic Models, by Samah Gad, Wagas Javed, Sohaib Ghani, Niklas Elmqvist, Tom Ewing, Keith N. Hampton, Naren Ramakrishnan, in IEEE Transactions on Visualization and Computer Graphics 21(5)

Visualization Analysis and Design, by Tamara Munzner, A K Peters Visualization Series, CRC Press, 2014.

What? Why? How?

What: Data

Timestamped text dataset What: Derived Bag of Words over time

Topics (Groups of words) Time intervals containing groups of topics Why: Tasks

Identify changes in topics over time Identify scatter/gather relationships How: Encode

> Spatially partition topics along a segment Label keywords within topics Linked keywords across time intervals

Segment labels for dates and duration

How: Manipulate Navigate: geometric zooming and panning Select: highlight keywords

Search: Select keywords by searching

How: Encode (Free Channels)

How: Reduce Filter: by selected keywords and resort

Size of labels for quantitative data

Width of links for quantitative data

Link colour for categorical or ordered data

14

