



Feature LDA: a Supervised Topic Model for Automatic Detection of Web API Documentations from the Web

Chenghua Lin, Yulan He, Carlos Pedrinaci, and
John Domingue

Knowledge Media Institute, The Open University

Milton Keynes, United Kingdom

11th International Semantic Web Conference. Boston, USA,
2012



Finding a Web API

- Dedicated registries, e.g. ProgrammableWeb
 - Contain out of date or incorrect information, e.g. invalid pages or incorrect links to APIs documentation pages
 - Only a limited number of Web APIs listed, left out a large number of third party Web APIs
 - General search engine, e.g. Google
 - Not optimized for Web API discovery
 - Mix up with pages that are not (so) relevant, e.g. blogs and advertisement about Web APIs.
-

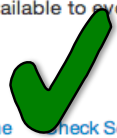
- Welcome
- Register
- SUBSCRIBER SERVICES**
- Profiles
- Statistics
- Data
- FREE DATA SERVICES**
- Sentiment

Developer Center

Sentiment Analysis API Methods

Viralheat is a social media measurement platform that allows you to track topics, brands, television shows, and movies across many different mediums all in one place using the best technology. The users of Viralheat choose to share a subset of their profiles with the world. Sentiment Analysis API is freely available to everyone with a developer account.

Available Methods



[Get Sentiment](#) [Train the Sentiment Analysis Engine](#) [Check Sentiment API usage quota](#)

Method 1: Get Sentiment

URL (JSON): `http://www.viralheat.com/api/sentiment/review.json?api_key=API_KEY&text=i%20dont%20like`

URL (XML): `http://www.viralheat.com/api/sentiment/review.xml?api_key=API_KEY&text=i%20dont%20like`

Format: XML, JSON

HTTP Method: GET

Requires Authentication: Yes

Parameters:

| Name | Description |
|---------|---|
| api_key | The authenticated account key. |
| text | Text for which you want the sentiment. The limit on the text is 360 characters. |

Return Values:

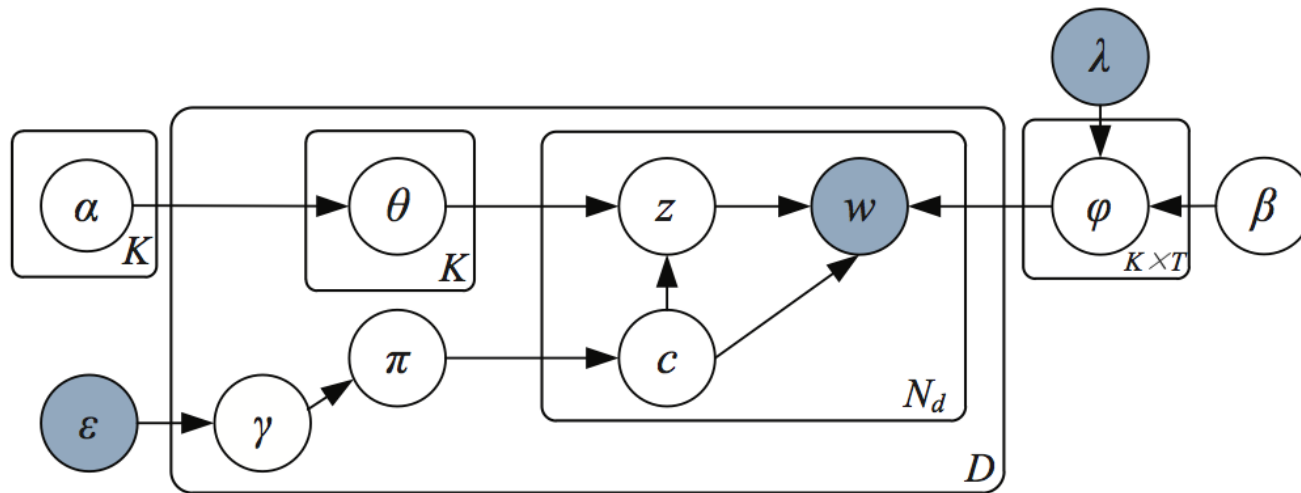
| Name | Description |
|------|---|
| prob | The probability with which the system thinks the given text has the output sentiment. |
| mood | Sentiment output for the given text. |
| text | Input text for which the sentiment was generated. |



Our Goal ...

- **Goal:** To build a customized search engine for detecting third party Web APIs on the Web scale
 - Assume every Web API provides public documentation page(s)
 - These pages provide the most relevant information for developers
 - Approached as a binary classification problem, i.e. distinguishing API documentation VS. normal pages
-

- **Feature LDA model:** a generic probabilistic framework for text classification.
 - A supervised four-layer hierarchical Bayesian model
 - Accommodate supervisions from both labelled instance and labelled features for training
 - Able to extract meaningful class specific topics





Conclusions

- Discovering Web APIs is becoming increasingly important and existing support is not optimal
 - Treat Web API discovery as a classification problem
 - Presented a supervised topic model called feaLDA
 - Outperforms SVM, NB and MaxEnt and supervised topic models labeledLDA and pLDA
 - Offers very high precision which is crucial for reducing false positive when mining from the Web
 - Future Work
 - web-scale processing for discovering Web API is in progress.
-