Learning to Label Documents

Timothy Baldwin

THE UNIVERSITY OF MELBOURNE

Talk Outline

1. Introduction
2. Labelling by Classification
3. Labelling by Topic
4. Labelling by Summarisation: Unstructured
5. Labelling by Summarisation: Structured
6. Summary

Outline

- What do I mean by “labelling” documents?
  - classification
  - topic modelling + labelling of topic (mixtures)
  - summarisation
- Why should you care?
  - user-in-the-loop document filtering
  - document fusion/summarisation
  - document collection navigation/“gisting”
- Recurring themes:
  - where is current research at?
  - what issues have we swept under the carpet?
  - what is the true nature of the beast?

Talk Outline

Labelling by Classification

- At its simplest, **labelling by classification** = document categorisation [Lewis and Ringuette, 1994, Sebastiani, 2002]:
  - label = semantic class, captured in the form of a static label (e.g. FINANCE, SPORTS)
  - fixed label set; single- or multi-label classification

Labelling by Classification

- At its simplest, **labelling by classification** = document categorisation [Lewis and Ringuette, 1994, Sebastiani, 2002]:
    - label = semantic class, captured in the form of a static label (e.g. FINANCE, SPORTS)
    - fixed label set; single- or multi-label classification
- In practice, the semantics of the label set can be high-dimensional and highly multi-dimensional, e.g. MeSH [Lipscomb, 2000]:
    - anatomy, organisms, diseases, chemicals, techniques, ...
    - constantly evolving over time (4400 in 1960 $\rightarrow$ 27455 in 2015), with expectation of missing categories at any given time [Nam et al., 2016]

MeSH Example

The role of coenzyme Q10 in heart failure

OBJECTIVE: To review the clinical data demonstrating the safety and efficacy of coenzyme Q10 (CoQ10) in heart failure (HF).

DATA SOURCES: Pertinent literature was identified ...

DATA SYNTHESIS: HF impairs the ability of the heart ...

CONCLUSIONS: Large, well-designed studies ...

MeSH Terms: (1) Antioxidants/therapeutic use; (2) Coenzymes; (3) Heart Failure/drug therapy; (4) Heart Failure/pathology; (5) Heart Failure/physiopathology; (6) Humans; (7) Oxidative Stress/drug effects; (8) Ubiquinone/analogs & derivatives; (9) Ubiquinone/therapeutic use; (10) Ventricular Remodeling/drug effects

Talk Outline

Topic Modelling

- Topic model = (unsupervised) latent variable model for capturing the semantics of a document collection, usually in the form of:
    - set of topics = multinomial distribution over vocab terms
    - topic allocations to documents = multinomial distribution over topics

Source(s): Deerwester et al. [1990], Hofmann [1999], Blei et al. [2003]

Topic Model Example

Full text of March 5 UK monetary minutes
The following is the complete text of the minutes of the March 5 monthly monetary policy meeting between Chancellor of the Exchequer Kenneth Clarke and Bank of England Governor Eddie George. The Chancellor of the Exchequer and Governor met, together with officials, ...

Topic allocation:

- 0.30 ⟨*economic, people, country, years, economy, ...*⟩
- 0.21 ⟨*percent, unemployment, rate, year, growth, ...*⟩
- 0.05 ⟨*party, government, parliament, minister, opposition, ...*⟩
- 0.00 ⟨*peru, hostages, rebels, fujimori, residence, ...*⟩

Topic Labelling

- While the topics and topic allocation for a given document provide a common "API" over the document collection, they are far from user friendly:
  - interpreting individual topics can be hard
  - making sense of the mixture of topics can be perplexing
- Ideal = automatic method for labelling topics with **succinct labels**, and documents with descriptions that capture the mixture of topics

Automatic Topic Labelling

stock, market, investor, fund, trading, investment, firm, exchange, company, share

⇓

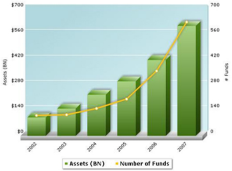
STOCK MARKET

vs.

MARKET

vs.

EXCHANGE-TRADED FUND



Related Work

- A few methods have been proposed to automatically label topics, for example using:
  - textual labels from topic model documents (e.g. bigrams) [Mei et al., 2007]
  - textual labels from external knowledge bases (e.g. Wikipedia) [Lau et al., 2011]
  - image labels from external knowledge bases [Aletras and Stevenson, 2013]

Wikipedia Titles as Labels

- We propose using Wikipedia article titles to label topics
- That is, given a topic, we want to find the most relevant Wikipedia article, and use its title as the topic label
- Idea borrowed from our previous work [Lau et al., 2011]: search Wikipedia using topic terms and train a support vector regression model to rank top titles based on a number of lexical association features
- In this work, we propose using neural embeddings of words and documents to generate topic labels

Source(s): Bhatia et al. [2016]

Word and Document Emebddings

- First generate word embeddings using `word2vec` [Mikolov et al., 2013]:
    - use `skip-gram` to generate embeddings for topic terms and Wikipedia titles

Word and Document Emebddings

- First generate word embeddings using `word2vec` [Mikolov et al., 2013]:
  - use `skip-gram` to generate embeddings for topic terms and Wikipedia titles
- Next, use `doc2vec` (or paragraph vectors) to learn embeddings for word sequences (e.g. paragraphs or documents) [Le and Mikolov, 2014]:
  - use `dbow` as an alternative means of generating embeddings for topic terms and Wikipedia titles

Context Representation for *Operating System*

`doc2vec` context:

Operating system

From Wikipedia, the free encyclopedia

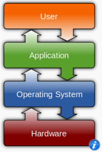
An **operating system** (**OS**) is system software that manages computer hardware and software resources and provides common services for computer programs. All computer programs, excluding firmware, require an operating system to function.

Time-sharing operating systems schedule tasks for efficient use of the system and may also include accounting software for cost allocation of processor time, mass storage, printing, and other resources.

For hardware functions such as input and output and memory allocation, the operating system acts as an intermediary between programs and the computer hardware,[1][2] although the application code is usually executed directly by the hardware and frequently makes system calls to an OS function or is interrupted by it. Operating systems are found on many devices that contain a computer – from cellular phones and video game consoles to web servers and supercomputers.

The dominant desktop operating system is Microsoft Windows with a market share of around 82%. OS X by Apple Inc. is in second place (9.8%), and Linux is in third position (1.5%).[3] In the mobile (smartphone and tablet combined) sector, based on Strategy Analytics Q3 2016 data, Android by Google is dominant with 87.5 percent or growth by 10.3 percent in one year and iOS by Apple is placed second with 12.1 percent or decrease by 5.2 percent in one year, while other operating systems amount to just 0.3 percent.[4] Linux is dominant in the server and supercomputing sectors. Other specialized classes of operating systems, such as embedded and real-time systems, exist for many applications.

Operating systems



Common features

Process management · Interrupts ·

`word2vec` context:

- ... a multi-tasking **operating system** allows more ...
- ... Single-user **operating systems** have no ...
- ... while time-sharing **operating systems** switch tasks ...

Finding Relevant Articles

Given the embeddings, we find the most relevant title by computing the cosine similarity between topic terms and titles

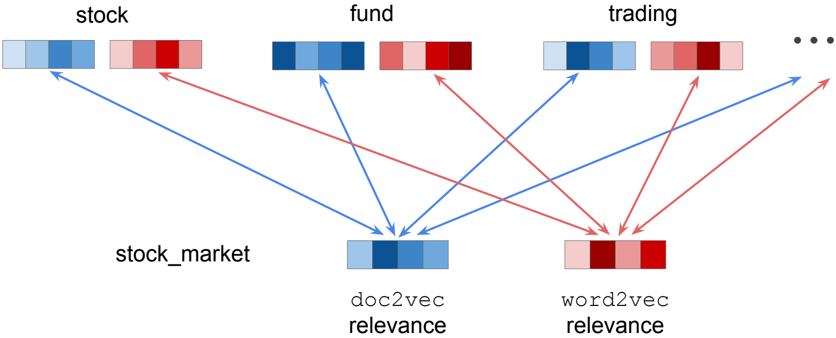
Formally, the relevance of a title a and a topic T is as follows:

$$rel_{d2v}(a, T) = \frac{1}{|T|} \sum_{v \in T} \cos\left(E^{d}_{d2v}(a), E^{w}_{d2v}(v)\right)$$

$$rel_{w2v}(a, T) = \frac{1}{|T|} \sum_{v \in T} \cos\left(E^{w}_{w2v}(a), E^{w}_{w2v}(v)\right)$$

$$rel_{d2v+w2v}(a, T) = rel_{d2v}(a, T) + rel_{w2v}(a, T)$$

Illustration



relevance(stock_market, <stock, fund, trading, …>) =
doc2vec_relevance + word2vec_relevance

word2vec vs. doc2vec

Topic terms	blogs, vmware, server, virtual, oracle, update, virtualization, application, desktop, infrastructure, management
word2vec **labels**	software, desktop, operating system, virtualization, middleware
doc2vec **labels**	microsoft visual studio, desktop virtualization, microsoft exchange server, cloud computing, windows server 2008

- word2vec labels tend to be general;
- doc2vec labels are more specific

Methodology Overview

- Our method consists of two steps:

  label generation: using `doc2vec` and `word2vec` embeddings to match related article titles with topics;
  label ranking: given a set of candidate titles for a topic, we train a support vector regression model to rank them to select the best label
- The candidates generated from step 1 are technically ordered (by the combined relevance score)
- The idea of step 2 is to re-rank the candidates using additional features and with supervision to improve performance

Label Ranking Features

Letter Trigram [Kou et al., 2015]

- Overlap of letter trigrams between a label and topic terms
- Also our unsupervised baseline

PageRank

- Uses directed links to estimate the significance of a document
- We construct a directed graph from Wikipedia based on hyperlinks within the article text
- Compute a PageRank value for each Wikipedia article/title

Num Words

- Number of words in the label

Topic Overlap

- Lexical overlap between a label and topic terms

Datasets

- BLOGS: 120K blog articles from Spinn3r dataset;
- BOOKS: 1K books from Internet Archive American libraries;
- NEWS: 29K New York Times articles from English Gigaword;
- PUBMED: 77K PubMed biomedical abstracts.

LDA is run on each domain to learn 100 topics; incoherent topics are automatically removed based on NPMI [Lau et al., 2014]

Gold Standard Judgements

- To evaluate our method and train the label ranking model, gold-standard ratings of the candidates are required
- We used CrowdFlower to collect human judgements
- We followed previous work in asking judges to rate a label given a topic on an ordinal scale of 0–3, where 0 = inappropriate label and 3 = perfect label
- Post-filtered (quality control), we have an average of 6.4 annotations per label
- We aggregate ratings of a label by taking its mean rating
- This produces a gold-standard ranking of labels for each topic

Evaluation Metric

Top 1 average rating

- mean rating of top-ranked label;
- provides an evaluation of the absolute utlity of the top labels

Normalised discounted cumulative gain (nDCG)

- measures relative quality of the ranking, relative to gold standard

Benchmark System [Lau et al., 2011]

- Uses Wikipedia titles to label topics.
- Label generation:
  - Query Wikipedia using top topic terms using Google's search API and Wikipedia's native search API
  - Top articles from both sources are pooled
  - Create additional labels by generating component *n*-grams from the original labels
  - Filter labels using RACO (based on article category overlap)
- Label ranking:
  - Train a support vector regression model over a number of lexical association features, e.g. PMI, Dice, to rank the candidate labels

Findings

- Cross-domain performance similar to in-domain; unsurprising since it has more training data: 3 out-of-domain data vs. 9 folds of cross-validation single-domain data
- Our system achieves substantial improvements over PUBMED and BOOKS
- Upper bound of top-1 average rating is much higher compared to benchmark system; this indicates we are generating better label candidates

Sample of Topics and Generated Labels

Domain	Topic Terms	Label Candidate
BLOGS	vmware server virtual oracle update virtualization application infrastructure management microsoft	virtualization
BOOKS	church archway building window gothic nave side value tower	church architecture
NEWS	investigation fbi official department federal agent investigator charge attorney evidence	criminal investigation
PUBMED	rate population prevalence study incidence datum increase mortality age death	mortality rate

Summary

- We proposed a neural embedding approach to automatically label topics using Wikipedia titles
- Our system combines document and word embeddings to select relevant titles
- Our model is simpler, more efficient and generates better labels than benchmark system

Summary

- We proposed a neural embedding approach to automatically label topics using Wikipedia titles
- Our system combines document and word embeddings to select relevant titles
- Our model is simpler, more efficient and generates better labels than benchmark system
- But what about document labels?

  ... document label = weighted combination of topic labels OR label generated from vector representing weighted sum over topics

Summary

- We proposed a neural embedding approach to automatically label topics using Wikipedia titles
- Our system combines document and word embeddings to select relevant titles
- Our model is simpler, more efficient and generates better labels than benchmark system
- But what about document labels?

  ... document label = weighted combination of topic labels OR label generated from vector representing weighted sum over topics
- And are textual labels always *really* the best way to go?

  ... no, images are sometimes much better than text ... see Sorodoc et al. [2017]

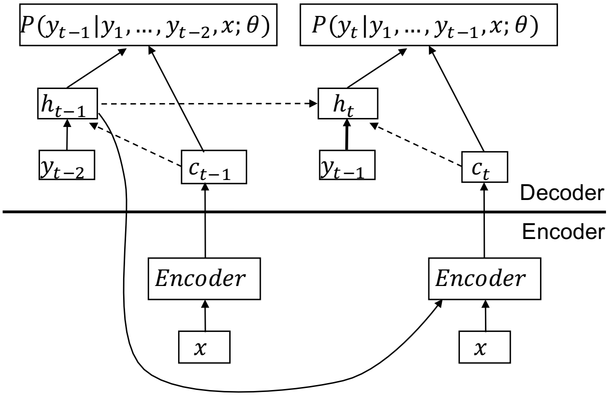
Talk Outline

Labelling by Summarisation

- We consider (single-document) summarisation in two primary forms:
  (1) summarisation over unstructured text
  (2) summarisation over a structured document representation

Summarisation over Unstructured Text

- Summarisation over unstructured text has focused on headline generation [Rush et al., 2015, Chopra et al., 2016] and machine reading tasks [Hermann et al., 2015]:



Headline Generation Example

- Given the following article:

  US President Donald Trump has warned North Korea "will be met with fire and fury like the world has never seen" if it continues to threaten the United States.

  But within hours of Mr Trump's threat North Korea's military said it was "carefully examining" a plan to strike the US Pacific territory of Guam with missiles. ...

  generate a headline ...

Headline Generation Example

- Given the following article:

  US President Donald Trump has warned North Korea "will be met with fire and fury like the world has never seen" if it continues to threaten the United States.

  But within hours of Mr Trump's threat North Korea's military said it was "carefully examining" a plan to strike the US Pacific territory of Guam with missiles. ...

  generate a headline ...

Donald Trump threatens North Korea with 'fire and fury', prompting threat to attack Guam [abc.net.au]

Headline Generation Example

- Given the following article:

  US President Donald Trump has warned North Korea "will be met with fire and fury like the world has never seen" if it continues to threaten the United States.

  But within hours of Mr Trump's threat North Korea's military said it was "carefully examining" a plan to strike the US Pacific territory of Guam with missiles. ...

  generate a headline ...

North Korea threatens missile strike on US Pacific territory dangerously close to Australia [news.com.au]

Or ... Summarising Proust

Proust Summaries

(1) **Harry:** Proust's novel ostensibly tells of the irrevocability of time lost, the forfeiture of innocence through experience, the reinstallment of extra-temporal values of time regained ... In the first volume, Swann, the family friend visits...

(2) **Ronald:** Er, well, Swann, Swann, there's this house, there's this house, and er, it's in the morning, it's in the morning — no, it's the evening, in the evening and er, there's a garden and er, this bloke comes in — bloke comes in — what's his name — what's his name, er just said it — big bloke — Swann, Swann

(3) **Bolton Choral Society:** Proust, in his first book wrote about ... fa la la ... Proust in his first book wrote about ... He wrote about ...

Source(s): `http://www.montypython.net/scripts/proust.php`

What's in a Summary/Headline?

- In practice, general-purpose summary/headline generation requires (at least) the following:

What's in a Summary/Headline?

- In practice, general-purpose summary/headline generation requires (at least) the following:
    - a notion of headline "style"

What's in a Summary/Headline?

- In practice, general-purpose summary/headline generation requires (at least) the following:
  - a notion of headline "style"
  - world knowledge (e.g. *US Pacific territory*)

What's in a Summary/Headline?

- In practice, general-purpose summary/headline generation requires (at least) the following:
  - a notion of headline "style"
  - world knowledge (e.g. *US Pacific territory*)
  - evaluation of the need for interpretation vs. retelling

What's in a Summary/Headline?

- In practice, general-purpose summary/headline generation requires (at least) the following:
  - a notion of headline "style"
  - world knowledge (e.g. *US Pacific territory*)
  - evaluation of the need for interpretation vs. retelling
  - some notion of importance/novelty ranking

What's in a Summary/Headline?

- In practice, general-purpose summary/headline generation requires (at least) the following:
  - a notion of headline "style"
  - world knowledge (e.g. *US Pacific territory*)
  - evaluation of the need for interpretation vs. retelling
  - some notion of importance/novelty ranking
  - some notion of surprise/spoilers

What's in a Summary/Headline?

- In practice, general-purpose summary/headline generation requires (at least) the following:
    - a notion of headline "style"
    - world knowledge (e.g. *US Pacific territory*)
    - evaluation of the need for interpretation vs. retelling
    - some notion of importance/novelty ranking
    - some notion of surprise/spoilers
    - chronology/fact-correctness

Fact Correctness #madashell

Source(s): `http://www.abc.net.au/tv/programs/shaun-micallefs-mad-as-hell/`

Where are we Currently at?

- **Style:** training data-driven (with careful choice of dataset ...), but some work on generation with style [Christensen et al., 2004, Ficler and Goldberg, to appear]

Where are we Currently at?

- **Style:** training data-driven (with careful choice of dataset ...), but some work on generation with style [Christensen et al., 2004, Ficler and Goldberg, to appear]
- **World knowledge:** training data-driven (with careful choice of dataset ...)

Where are we Currently at?

- **Style:** training data-driven (with careful choice of dataset ...), but some work on generation with style [Christensen et al., 2004, Ficler and Goldberg, to appear]
- **World knowledge:** training data-driven (with careful choice of dataset ...)
- **Interpretation vs. retelling:** no real need for headlines

Where are we Currently at?

- **Style:** training data-driven (with careful choice of dataset ...), but some work on generation with style [Christensen et al., 2004, Ficler and Goldberg, to appear]
- **World knowledge:** training data-driven (with careful choice of dataset ...)
- **Interpretation vs. retelling:** no real need for headlines
- **Importance/novelty ranking:** in headline case, document structure highly informative

Where are we Currently at?

- **Style:** training data-driven (with careful choice of dataset ...), but some work on generation with style [Christensen et al., 2004, Ficler and Goldberg, to appear]
- **World knowledge:** training data-driven (with careful choice of dataset ...)
- **Interpretation vs. retelling:** no real need for headlines
- **Importance/novelty ranking:** in headline case, document structure highly informative
- **Surprise/spoilers:** no real need for headlines

Where are we Currently at?

- **Style:** training data-driven (with careful choice of dataset ...), but some work on generation with style [Christensen et al., 2004, Ficler and Goldberg, to appear]
- **World knowledge:** training data-driven (with careful choice of dataset ...)
- **Interpretation vs. retelling:** no real need for headlines
- **Importance/novelty ranking:** in headline case, document structure highly informative
- **Surprise/spoilers:** no real need for headlines
- **Chronology/fact-correctness:** largely single-event, with little abstraction, so largely irrelevant

Talk Outline

Labelling by Summarisation: Structured

- In the case of summarisation from structured data, there is the assumption that there is a structured input (possibly in addition to source unstructured data), e.g.:
  - weather forecasts [Liang et al., 2009], movies [Gorinski and Lapata, 2015], basketball games [Wiseman et al., to appear]
- Promising new direction (esp. Wiseman et al. [to appear]), as importance/novelty ranking and (in-domain) world knowledge highly important, and chronology critical in terms of factual correctness

Example from Wiseman et al. [to appear]

TEAM	WIN	LOSS	PTS	FG_PCT	RB	AS ...
Heat	11	12	103	49	47	27
Hawks	7	15	95	43	33	20

PLAYER	AS	RB	PT	FG	FGA	CITY ...
Tyler Johnson	5	2	27	8	16	Miami
Dwight Howard	4	17	23	9	11	Atlanta
Paul Millsap	2	9	21	8	12	Atlanta
Goran Dragic	4	2	21	8	17	Miami
Wayne Ellington	2	3	19	7	15	Miami
Dennis Schroder	7	4	17	8	15	Atlanta
Rodney McGruder	5	5	11	3	8	Miami
Thabo Sefolosha	5	5	10	5	11	Atlanta
Kyle Korver	5	3	9	3	9	Atlanta
...						

The Atlanta Hawks defeated the Miami Heat , 103 - 95 , at Philips Arena on Wednesday . Atlanta was in desperate need of a win and they were able to take care of a shorthanded Miami team here . Defense was key for the Hawks , as they held the Heat to 42 percent shooting and forced them to commit 16 turnovers . Atlanta also dominated in the paint , winning the rebounding battle , 47 - 34 , and outscoring them in the paint 58 - 26.The Hawks shot 49 percent from the field and assisted on 27 of their 43 made baskets . This was a near wire - to - wire win for the Hawks , as Miami held just one lead in the first five minutes . Miami (7 - 15) are as beat - up as anyone right now and it 's taking a toll on the heavily used starters . Hassan Whiteside really struggled in this game , as he amassed eight points , 12 rebounds and one blocks on 4 - of - 12 shooting ...

Talk Outline

1. Introduction
2. Labelling by Classification
3. Labelling by Topic
4. Labelling by Summarisation: Unstructured
5. Labelling by Summarisation: Structured
6. Summary

Summary

- Basic flavours of document label generation:
  - classification
  - topic modelling/labelling
  - summarisation (unstructured or structured)
- Long, long way to go on document summarisation front, with real need for *truly* challenge datasets and realistic evaluations

References I

Nikolaos Aletras and Mark Stevenson. Representing topics using images. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT 2013)*, pages 158–167, Atlanta, USA, 2013.

Shraey Bhatia, Jey Han Lau, and Timothy Baldwin. Automatic labelling of topics with neural embeddings. In *Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016)*, pages 953–963, Osaka, Japan, 2016.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

Sumit Chopra, Michael Auli, and Alexander M Rush. Abstractive sentence summarization with attentive recurrent neural networks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics — Human Language Technologies (NAACL HLT 2016)*, pages 93–98, 2016.

Heidi Christensen, BalaKrishna Kolluru, Yoshihiko Gotoh, and Steve Renals. From text summarisation to style-specific summarisation for broadcast news. In *Proceedings of the European Conference on Information Retrieval 2004*, pages 223–237, Sunderland, UK, 2004.

Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the American Society of Information Science*, 41(6), 1990.

References II

Jessica Ficler and Yoav Goldberg. Controlling linguistic style aspects in neural language generation. In *Proceedings of the EMNLP 2017 Stylistic Variation Workshop*, Copenhagen, Denmark, to appear.

Philip John Gorinski and Mirella Lapata. Movie script summarization as graph-based scene extraction. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics — Human Language Technologies (NAACL HLT 2015)*, pages 1066–1076, Denver, USA, 2015.

Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. Teaching machines to read and comprehend. In *Advances in Neural Information Processing Systems 28 (NIPS-15)*, pages 1693–1701, 2015.

Thomas Hofmann. Probabilistic latent semantic indexing. In *Proceedings of 22nd International ACM-SIGIR Conference on Research and Development in Information Retrieval (SIGIR'99)*, pages 50–57, Berkeley, USA, 1999.

Wanqiu Kou, Li Fang, and Timothy Baldwin. Automatic labelling of topic models using word vectors and letter trigram vectors. In *Proceedings of the 11th Asian Information Retrieval Societies Conference (AIRS 2015)*, pages 229–240, Brisbane, Australia, 2015.

Jey Han Lau, Karl Grieser, David Newman, and Timothy Baldwin. Automatic labelling of topic models. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL HLT 2011)*, pages 1536–1545, Portland, USA, 2011.

References III

Jey Han Lau, David Newman, and Timothy Baldwin. Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. In *Proceedings of the 14th Conference of the EACL (EACL 2014)*, pages 530–539, Gothenburg, Sweden, 2014.

Quoc V. Le and Tomas Mikolov. Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning (ICML 2014)*, volume 14, pages 1188–1196, Beijing, China, 2014.

David D. Lewis and Marc Ringuette. A comparison of two learning algorithms for text categorization. In *Third Annual Symposium on Document Analysis and Information Retrieval*, pages 81–93, 1994.

Percy Liang, Michael I. Jordan, and Dan Klein. Learning semantic correspondences with less supervision. In *Proceedings of the Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP 2009)*, pages 91–99, Singapore, 2009.

Carolyn E. Lipscomb. Medical subject headings (MeSH). *Bulletin of the Medical Library Association*, 88(3):265, 2000.

Qiaozhu Mei, Xuehua Shen, and Chengxiang Zhai. Automatic labeling of multinomial topic models. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2007)*, pages 490–499, 2007.

References IV

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

Jinseok Nam, Eneldo Loza Mencía, and Johannes Fürnkranz. All-in text: Learning document, label, and word representations jointly. pages 1948–1954, Phoenix, USA, 2016.

Alexander M Rush, Sumit Chopra, and Jason Weston. A neural attention model for abstractive sentence summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)*, pages 379–389, Lisbon, Portugal, 2015.

Fabrizio Sebastiani. Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1):1–47, 2002.

Ionut Sorodoc, Jey Han Lau, Nikolaos Aletras, and Timothy Baldwin. Multimodal topic labelling. In *Proceedings of the 15th Conference of the EACL (EACL 2017)*, pages 701–706, Valencia, Spain, 2017.

Sam Wiseman, Stuart M. Shieber, and Alexander M. Rush. Challenges in data-to-document generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)*, Copenhagen, Denmark, to appear.