INFORMATION EXTRACTION ENGINE FOR SENTIMENT-TOPIC MATCHING

IN PRODUCT INTELLIGENCE APPLICATIONS

CORNELIA FERNER | INTERNATIONAL DATA SCIENCE CONFERENCE | SALZBURG 2017

WERNER POMWENGER | MARTIN SCHNÖLL | VERONIKA HAAF | ARNOLD KELLER | STEFAN WEGENKITTL

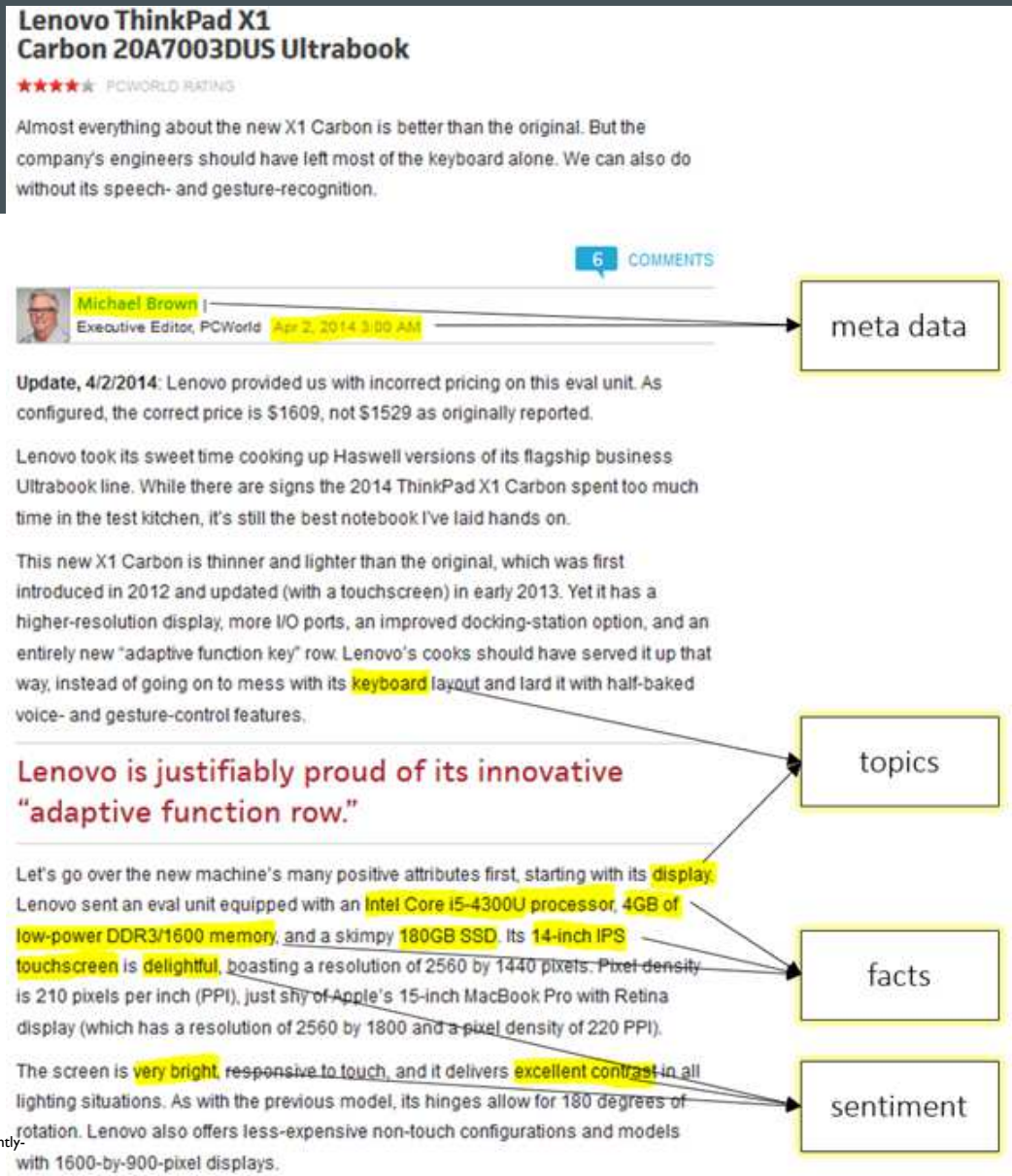
FFG

FACT AI

Applied Data Science Lab

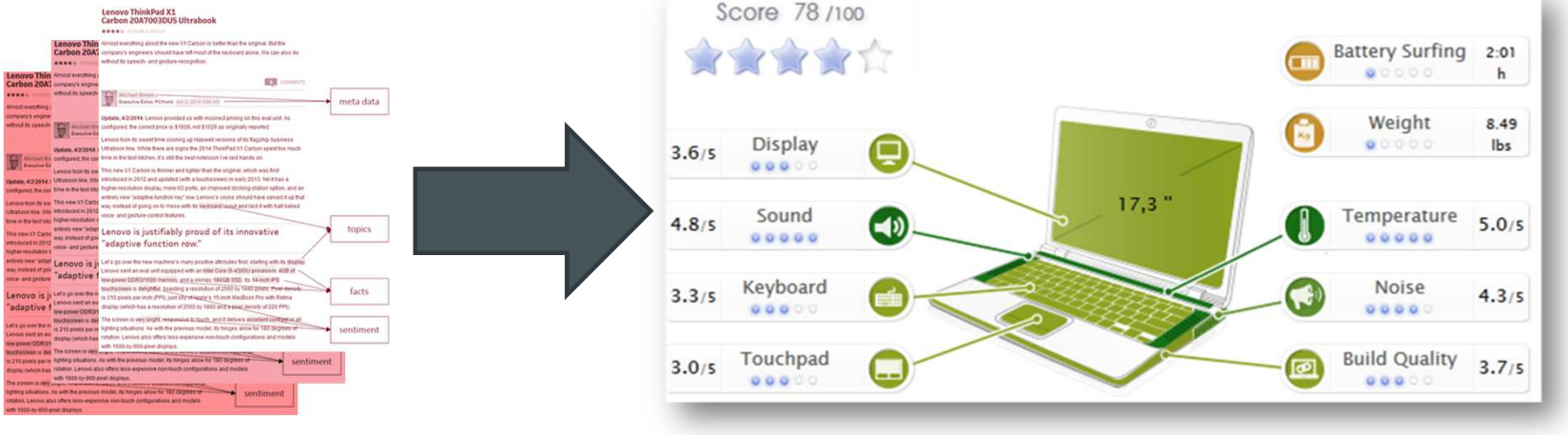
fhs Fachhochschule Salzburg University of Applied Sciences

MOTIVATION



http://www.pcworld.com/article/2138145/ultrabooks/lenovo-thinkpad-x1-carbon-review-slightly-overdone-but-plenty-tasty.html

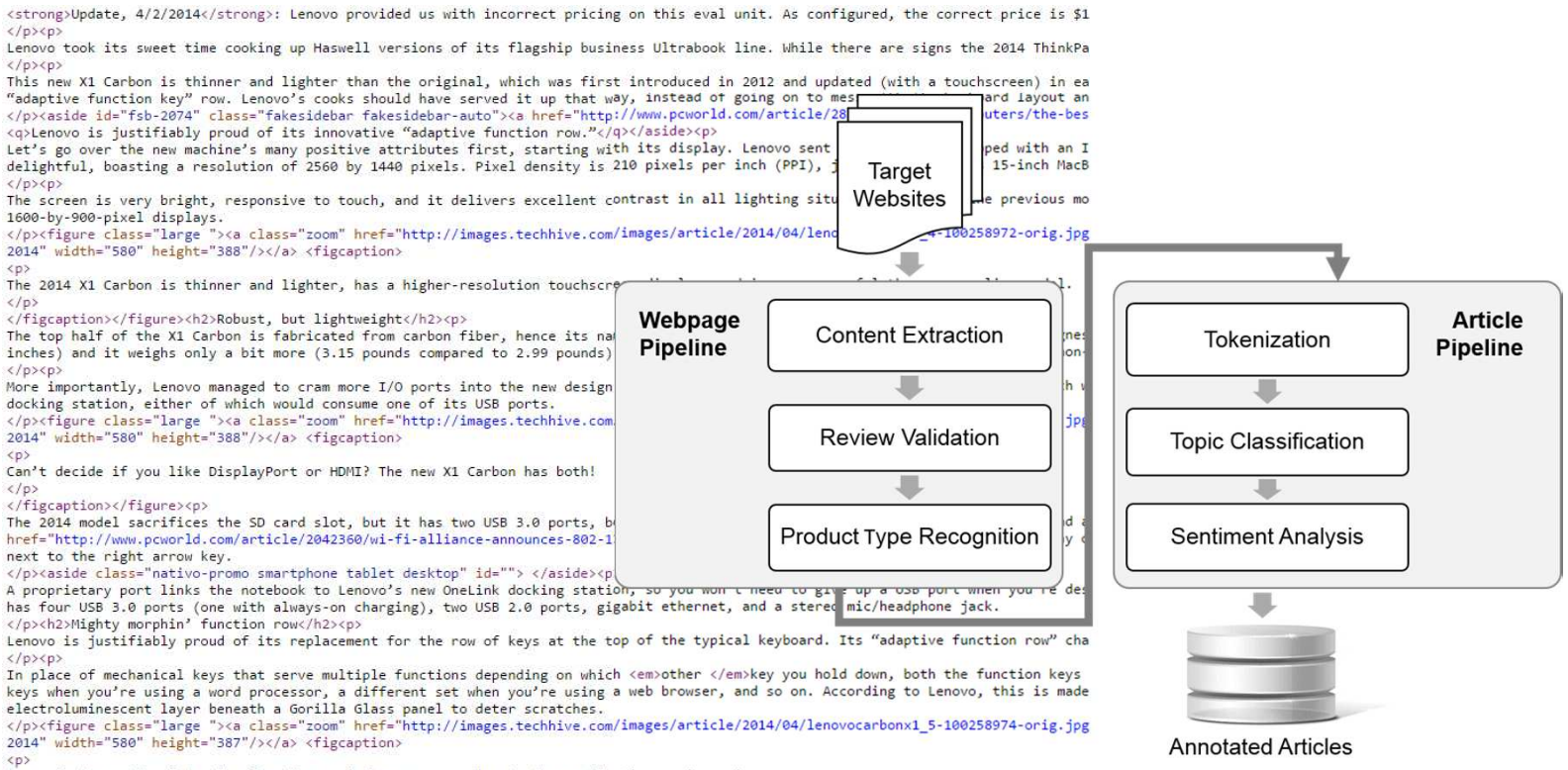
MOTIVATION


Lenovo ThinkPad X1
Carbon 20A7003DUS Ultrabook
meta data
topics
facts
sentiment
sentiment
sentiment
Score 78 /100
Battery Surfing 2:01 h
Weight 8.49 lbs
Temperature 5.0/5
Noise 4.3/5
Build Quality 3.7/5
Display 3.6/5
Sound 4.8/5
Keyboard 3.3/5
Touchpad 3.0/5
17,3 "

AGENDA

- ARIE – article and review information extraction engine
- Topic classification
- Sentiment analysis
- Recap

ARIE


Update, 4/2/2014: Lenovo provided us with incorrect pricing on this eval unit. As configured, the correct price is $1

Lenovo took its sweet time cooking up Haswell versions of its flagship business Ultrabook line. While there are signs the 2014 ThinkPa

This new X1 Carbon is thinner and lighter than the original, which was first introduced in 2012 and updated (with a touchscreen) in ea
“adaptive function key” row. Lenovo’s cooks should have served it up that way, instead of going on to mes ard layout an
<aside id="fsb-2074" class="fakesidebar fakesidebar-auto"><a href="http://www.pcworld.com/article/28 uters/the-bes
<q>Lenovo is justifiably proud of its innovative “adaptive function row.”</q></aside>
Let’s go over the new machine’s many positive attributes first, starting with its display. Lenovo sent ped with an I
delightful, boasting a resolution of 2560 by 1440 pixels. Pixel density is 210 pixels per inch (PPI), 15-inch MacB

The screen is very bright, responsive to touch, and it delivers excellent contrast in all lighting situ e previous mo
1600-by-900-pixel displays.
<figure class="large "><a class="zoom" href="http://images.techhive.com/images/article/2014/04/lenc -100258972-orig.jpg
2014" width="580" height="388"/> <figcaption>

The 2014 X1 Carbon is thinner and lighter, has a higher-resolution touchscre l.

</figcaption></figure><h2>Robust, but lightweight</h2>
The top half of the X1 Carbon is fabricated from carbon fiber, hence its na nes
inches) and it weighs only a bit more (3.15 pounds compared to 2.99 pounds) on

More importantly, Lenovo managed to cram more I/O ports into the new design h
docking station, either of which would consume one of its USB ports.
<figure class="large "><a class="zoom" href="http://images.techhive.com jpg
2014" width="580" height="388"/> <figcaption>

Can’t decide if you like DisplayPort or HDMI? The new X1 Carbon has both!

</figcaption></figure>
The 2014 model sacrifices the SD card slot, but it has two USB 3.0 ports, b d a
href="http://www.pcworld.com/article/2042360/wi-fi-alliance-announces-802-1 y c
next to the right arrow key.
<aside class="nativo-promo smartphone tablet desktop" id=""> </aside><p
A proprietary port links the notebook to Lenovo’s new OneLink docking station, so you won’t need to give up a USB port when you’re des
has four USB 3.0 ports (one with always-on charging), two USB 2.0 ports, gigabit ethernet, and a stereo mic/headphone jack.
<h2>Mighty morphin’ function row</h2>
Lenovo is justifiably proud of its replacement for the row of keys at the top of the typical keyboard. Its “adaptive function row” cha

In place of mechanical keys that serve multiple functions depending on which other key you hold down, both the function keys
keys when you’re using a word processor, a different set when you’re using a web browser, and so on. According to Lenovo, this is made
electroluminescent layer beneath a Gorilla Glass panel to deter scratches.
<figure class="large "><a class="zoom" href="http://images.techhive.com/images/article/2014/04/lenovocarbonx1_5-100258974-orig.jpg
2014" width="580" height="387"/> <figcaption>

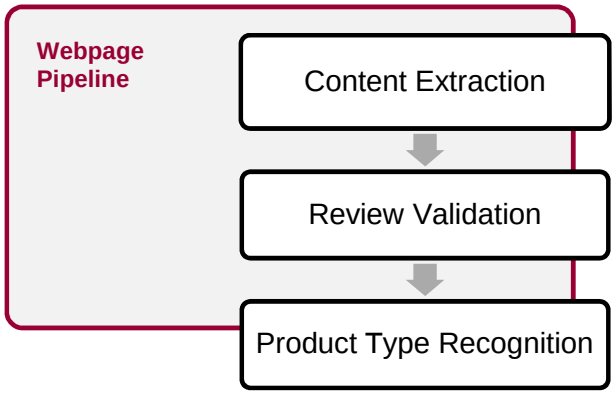
Lenovo’s innovative “adaptive function row” changes according to the application you’re using.

<figure class="large "><img sr
height="505"/> <figcaption>

This mock-up shows all four of the function rows and which applications activate them.

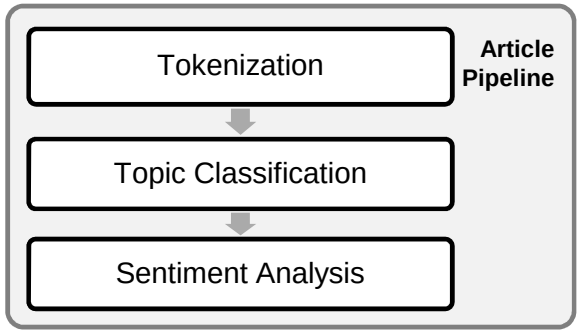
</figcaption></figure>
Target Websites
Webpage Pipeline
Content Extraction
Review Validation
Product Type Recognition
Tokenization
Topic Classification
Sentiment Analysis
Article Pipeline
Annotated Articles

WEBPAGE PIPELINE

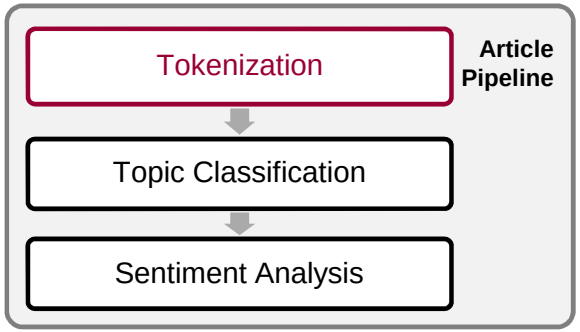


- Content extraction:
  - Boilerplate removal (comments, ads, teasers etc.)
  - Raw text extraction (without html tags)
  - Store meta data
- Review validation:
  - Only expert reviews are needed
  - Sort out ads, comparisons etc.
  - Latent Dirichlet Allocation + SVM
- Product type recognition:
  - Only laptops (e.g. ultrabooks, convertibles) are needed
  - Sort out reviews on displays, speakers etc.
  - Maximum Entropy (logistic regression for multiclass problems)

ARTICLE PIPELINE

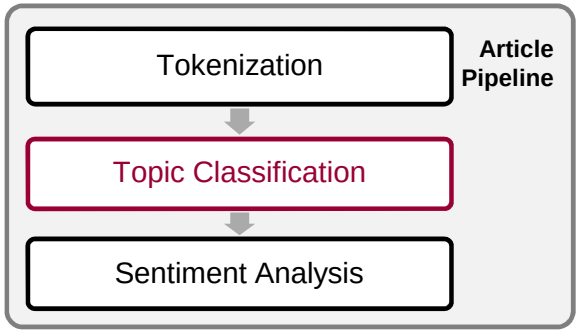


ARTICLE PIPELINE



- Tokenization:
  - Words and sentences
  - Sentence-level annotations for topic classification and sentiment analysis
- Prepocessing:
  - Lowercasing
  - No stopwords, no stemming, no removal of infrequent/frequent words
- Features:
  - Bag-of-words (also with bigrams)
  - Word2Vec

ARTICLE PIPELINE

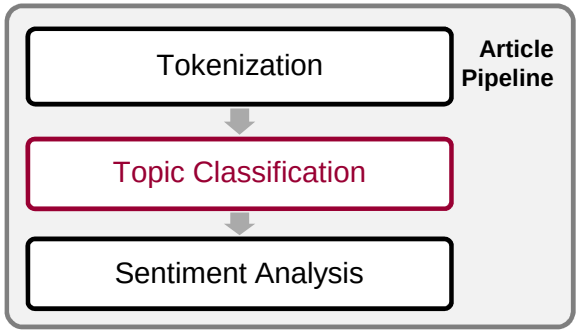


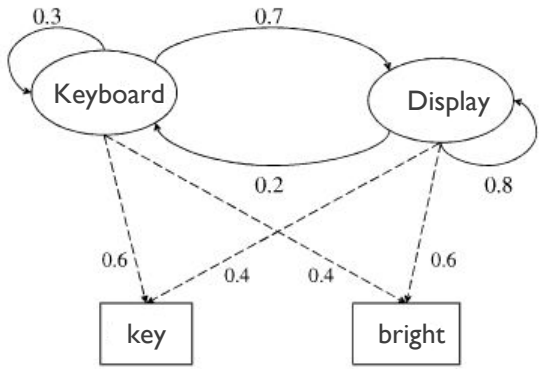
- $D = \{1, \dots, d\}$ … dictionary; set of all words
- $C = \{1, \dots, c\}$ … set of topics
- $S \in C$ and $W_t \in D$… topics and sequence of words, respectively

- MaxEnt (logistic regression, softmax):
  - $\vec{V} = (v_1, \dots, v_d)$ … count vector (absolute word counts in a sequence) with $v_i = \sum_{t=1}^{n} \mathbf{1}(W_t = i)$
  - $P[S = j \mid \vec{V}] = \frac{e^{\sum_{i \in D} v_i \cdot \lambda_{ij} + \mu_j}}{Z} = \prod_{t=1}^{n} \frac{e^{\lambda_{w_t j} + \mu_j}}{\sum_{l=1}^{c} e^{\lambda_{w_t l} + \mu_l}}$

$$= \prod_{t=1}^{n} P[S = j \mid W_t]$$

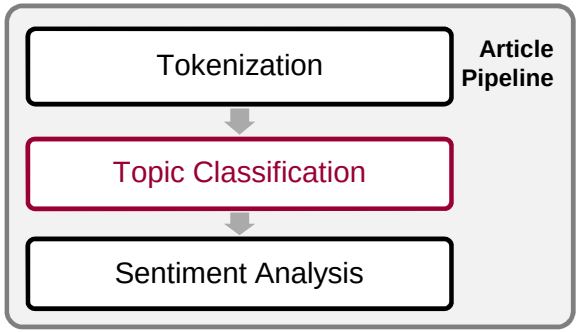
ARTICLE PIPELINE





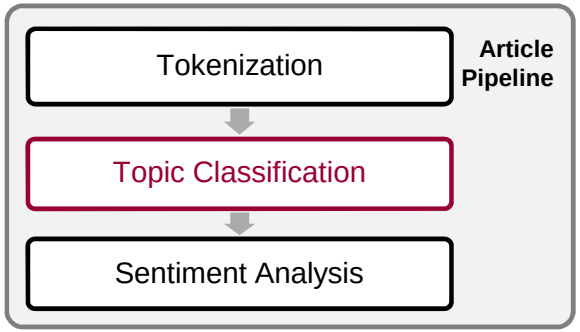
- Hidden Markov Model (HMM):
  - decoding: find the most probable sequence of hidden states (topics) given the model and a sequence of observations (words)?
  - $A = a_{ij} = P[S_{t+1} = j \mid S_t = i]$ … transition probabilities
  - $B = b_{jk} = P[W_t = k \mid S_t = j]$ … emission probabilities
  - $\pi_i = P[S_1 = 1]$ … initial state probabilities
  - $M = (C, D, A, B, \pi)$
- Combining MaxEnt and HMM (Bayes):
  - $b_{jk} = P[W_t = k \mid S_t = j] = \frac{P[S=j \mid W=k] \cdot P[W=k]}{\sum_{r=1}^{d} P[S=j \mid W=r] \cdot P[W=r]}$

ARTICLE PIPELINE



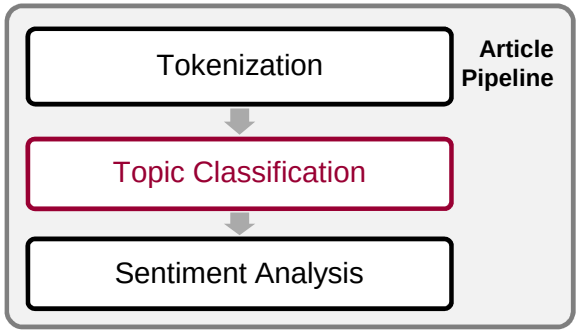
- Combining MaxEnt and HMM (Bayes):
  - $b_{jk} = \frac{P[S=j \mid W=k] \cdot P[W=k]}{\sum_{r=1}^{d} P[S=j \mid W=r] \cdot P[W=r]}$

ARTICLE PIPELINE



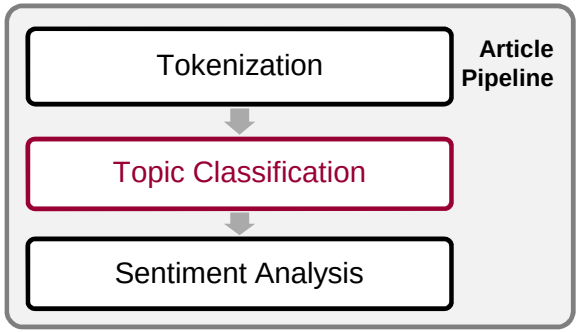
- Combining MaxEnt and HMM (Bayes):
  - $b_{jk} = \frac{P[S=j \mid W=k] \cdot \hat{p}_k}{\sum_{r=1}^{d} P[S=j \mid W=r] \cdot P[W=r]}$ $\quad \hat{p}_k \cong P[W = k]$

ARTICLE PIPELINE



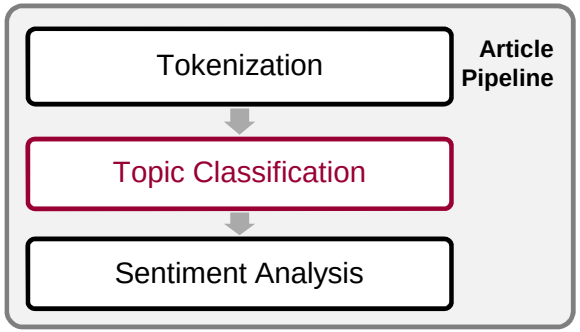
- Combining MaxEnt and HMM (Bayes):
  - $b_{jk} = \frac{P[S=j \mid W=k] \cdot \hat{p}_k}{f^{-1}(j)}$ $\qquad \hat{p}_k \cong P[W = k]$

ARTICLE PIPELINE



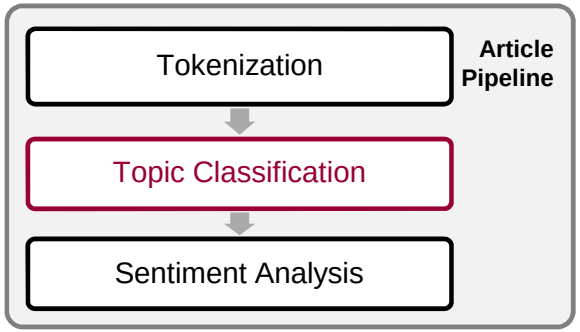
- Combining MaxEnt and HMM (Bayes):
  - $b_{jk} = \frac{P[S=j \mid W=k] \cdot \hat{p}_k}{f^{-1}(j)}$ $\qquad \hat{p}_k \cong P[W = k]$
  - $b_{jk} = \frac{e^{\lambda_{kj}+\mu_j} \cdot \hat{p}_k}{\sum_{l=1}^{C} e^{\lambda_{kl}+\mu_l}} \cdot f(j)$ $\qquad P[S = j \mid W_t] = \frac{e^{\lambda_{w_t j}+\mu_j}}{\sum_{l=1}^{C} e^{\lambda_{w_t l}+\mu_l}}$

ARTICLE PIPELINE



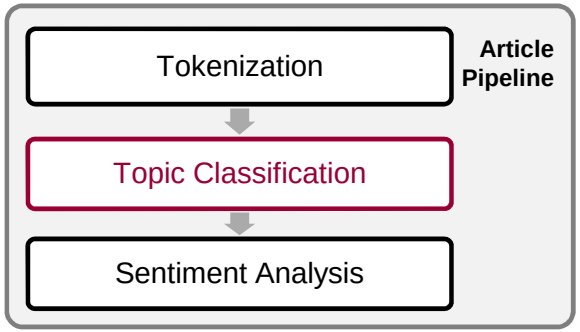
- Combining MaxEnt and HMM (Bayes):
  - $b_{jk} = \frac{P[S=j \mid W=k] \cdot \hat{p}_k}{f^{-1}(j)}$ $\qquad \hat{p}_k \cong P[W = k]$
  - $b_{jk} = \frac{e^{\lambda_{kj}+\mu_j} \cdot \hat{p}_k}{\sum_{l=1}^{c} e^{\lambda_{kl}+\mu_l}} \cdot f(j)$ $\qquad P[S = j \mid W_t] = \frac{e^{\lambda_{w_t j}+\mu_j}}{\sum_{l=1}^{c} e^{\lambda_{w_t l}+\mu_l}}$
  - $b_{jk} = g(j) \cdot f(j)$ $\qquad f(j) = \frac{1}{\sum_{i=1}^{d} g(i)}$

ARTICLE PIPELINE

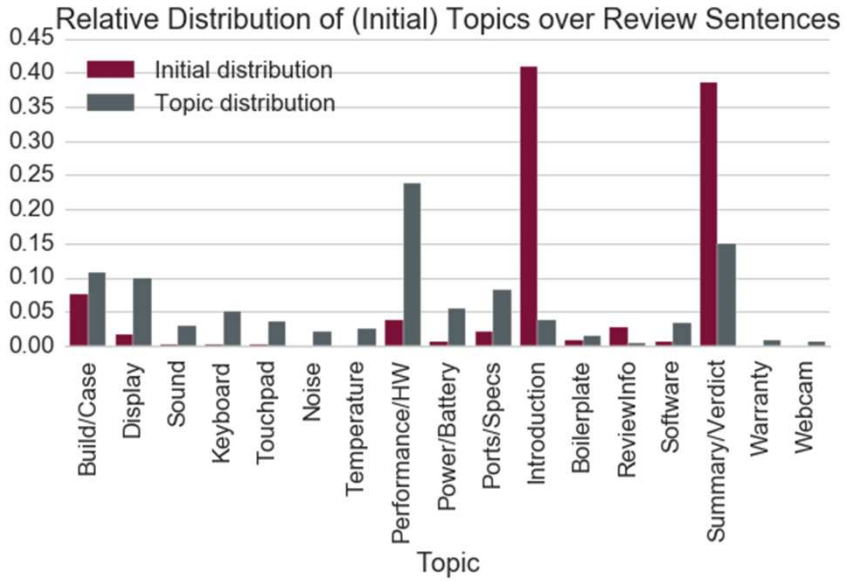


- Combining MaxEnt and HMM (Bayes):
  - $b_{jk} = \frac{P[S=j \mid W=k] \cdot \hat{p}_k}{f^{-1}(j)}$ $\qquad \hat{p}_k \cong P[W = k]$
  - $b_{jk} = \frac{e^{\lambda_{kj}+\mu_j} \cdot \hat{p}_k}{\sum_{l=1}^{c} e^{\lambda_{kl}+\mu_l}} \cdot f(j)$ $\qquad P[S = j \mid W_t] = \frac{e^{\lambda_{w_t j}+\mu_j}}{\sum_{l=1}^{c} e^{\lambda_{w_t l}+\mu_l}}$
  - $b_{jk} = g(j) \cdot f(j)$ $\qquad f(j) = \frac{1}{\sum_{i=1}^{d} g(i)}$

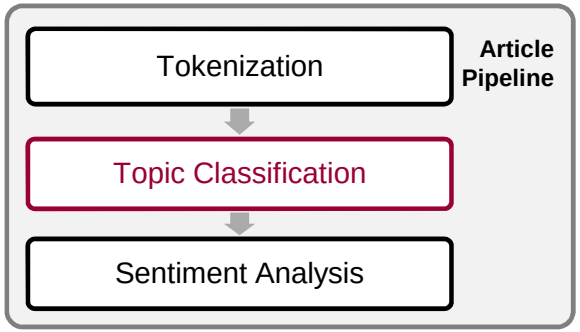
ARTICLE PIPELINE



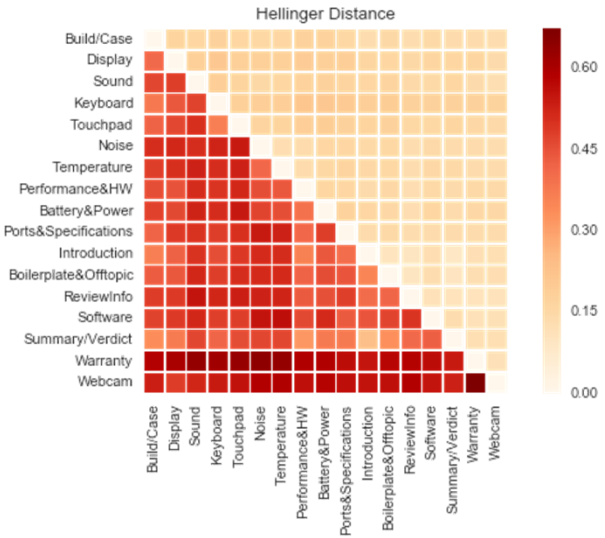
- 3152 reviews
- 240220 manually labelled sentences
- 17 predefined topics



ARTICLE PIPELINE



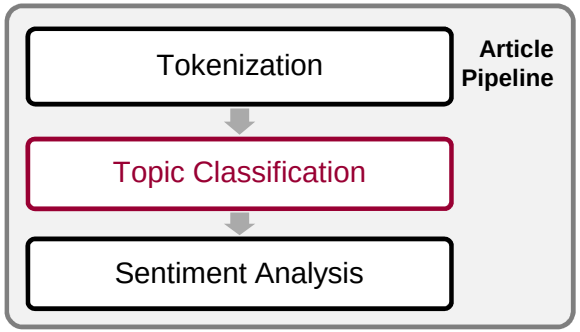
- Distance of the topics‘ MaxEnt distributions (bottom left) and the topics‘ word frequencies (top right).



- Hellinger distance:

$$H(S_1, S_2)^2 = \frac{1}{\sqrt{2}} \sqrt{\sum \left(\sqrt{S_1} - \sqrt{S_2}\right)^2}$$

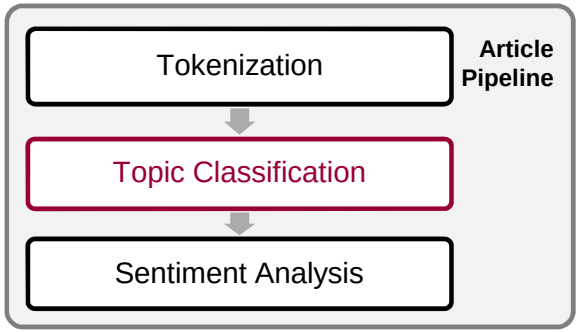
ARTICLE PIPELINE

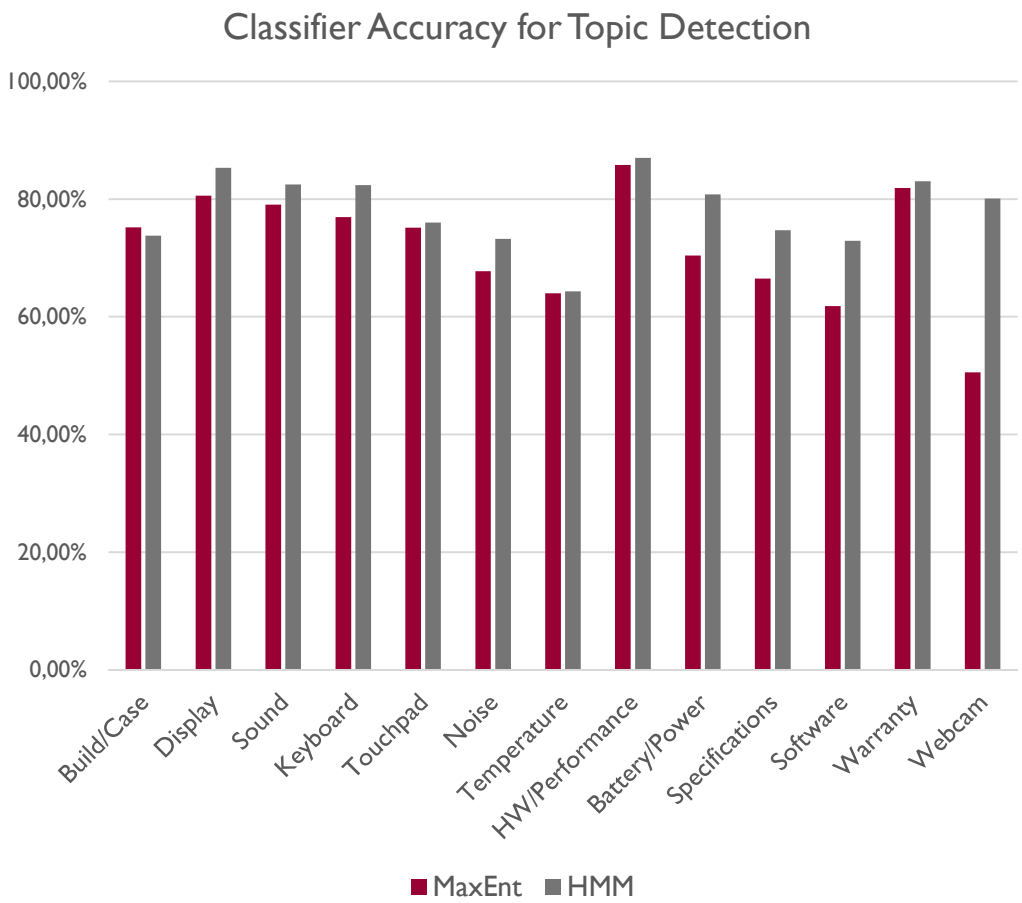


Sound	Noise	Temperature	Summary
sound	db	cool	verdict
music	quiet	heat	price
bass	noise	cooling	lasted
audio	fan	lap	watts
speakers	audible	temperature	read
and	db	and	of
the	to	to	and
speakers	and	degrees	for
sound	the	the	the
to	is	is	it

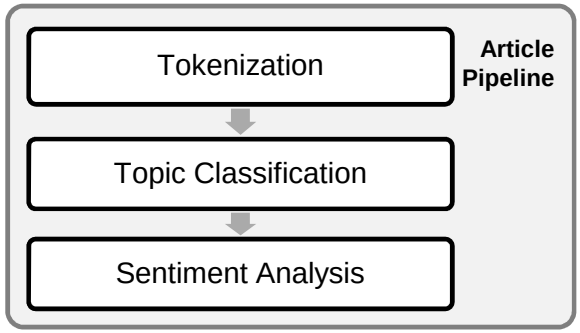
Table 1: Words with highest weights as trained by the MaxEnt classifier (above) and with maximum frequency (below) in four exemplary topics.

ARTICLE PIPELINE

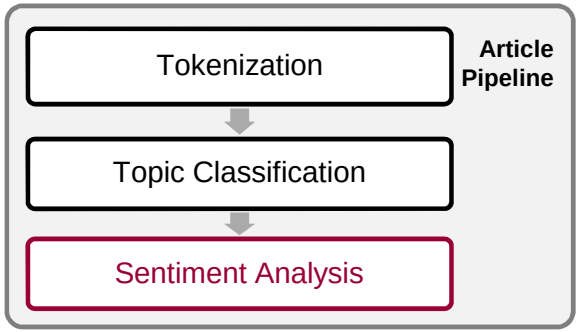




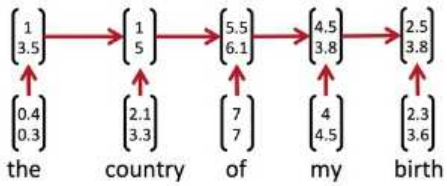
ARTICLE PIPELINE



ARTICLE PIPELINE



- MaxEnt
  - baseline
- Recurrent neural network (RNN)
  - Best performance on sentence level
- Recursive neural tensor network (RNTN)
  - Require parsed syntax tree
  - Good on word level



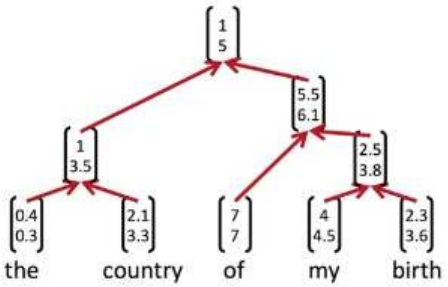
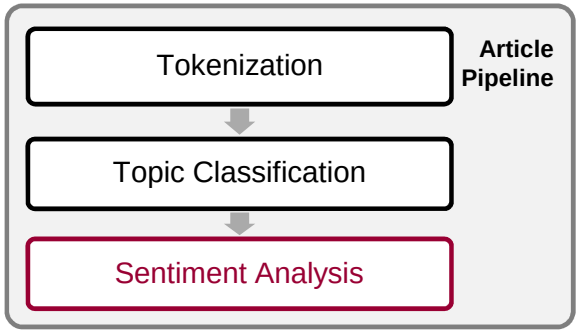


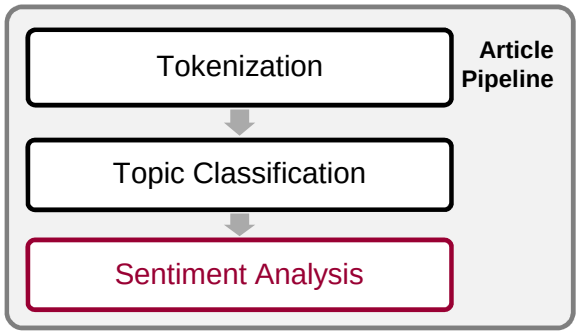
Image credit: Socher et al. https://www.slideshare.net/jiessiecao/parsing-natural-scenes-and-natural-language-with-recursive-neural-networks

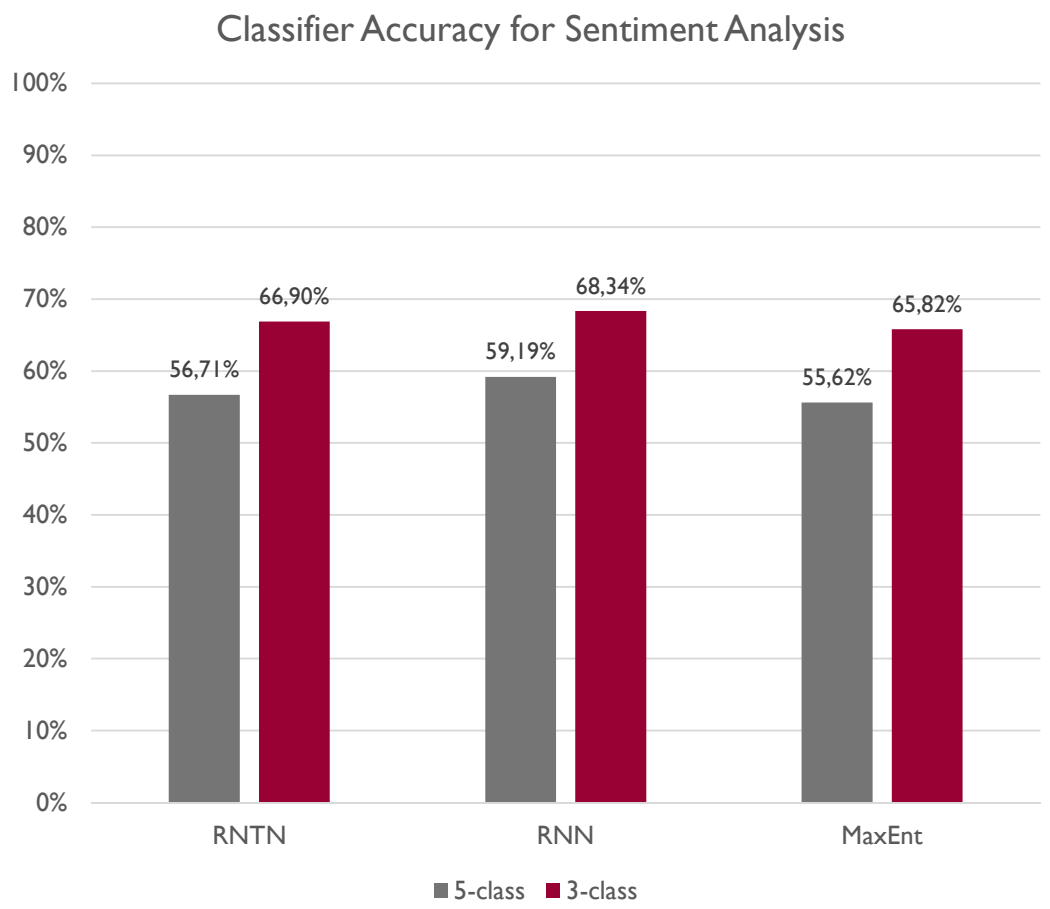
ARTICLE PIPELINE



- 21695 manually labelled sentences
- 5-class analysis:
  - very positive
  - positive
  - neutral
  - negative
  - very negative
- 3-class analysis:
  - positive
  - neutral
  - negative

ARTICLE PIPELINE



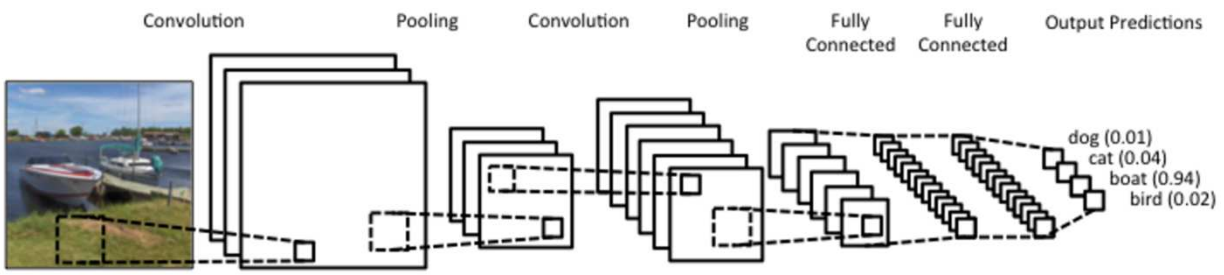


LESSONS LEARNT

- Language is ambiguous.
- Lack of standardized features or off-the-shelf (preprocessing) methods.
- There isn‘t only English.

ONGOING RESEARCH

- Representation learning with CNNs (deep neural network)
- Change fully connected layers to adapt net to new tasks
- Multiple languages: reuse network architecture
- Character-wise approach for reusability of representations



http://www.wildml.com

LITERATURE

[MaxEnt] A. Berger, S. Della Pietra and V. Della Pietra, "A maximum entropy approach to natural language processing," Computational Linguistics, vol. 22(1), pp. 39-71, 1996.

[HMM] Cappé, O., Moulines, E., and Ryden, T., "Inference in Hidden Markov Models," New York, Springer, 2005.

[RNN] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9(8), pp.1735-1780, 1997.

[RNTN] Socher, R., Perelygin, A., Wu. J., Chuang, J., Manning, C., Ng. A, and Potts, C., "Recursive deep models for semantic compositionality over a sentiment treebank," Conference on Empirical Methods in Natural Language Processing, 2013.

[UIMA] D. Ferrucci and A. Lally, "UIMA: An architectural approach to unstructured information processing in the corporate research environment," Natural Language Engineering, 10(3), pp. 237-348, 2004.