Outdated Posts about AWS on StackOverflow

2020-12-11

Team members

Students from DSI:

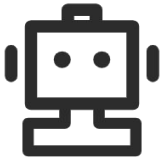
Huimin Jiang
Chenlu Jia
Ying Du
Jiaqi Tang
Jianing Li

Mentors from Amazon:

Tapodipta Ghosh
Marvin Dong

Instructor from Columbia:

Sining Chen



Our motivation?

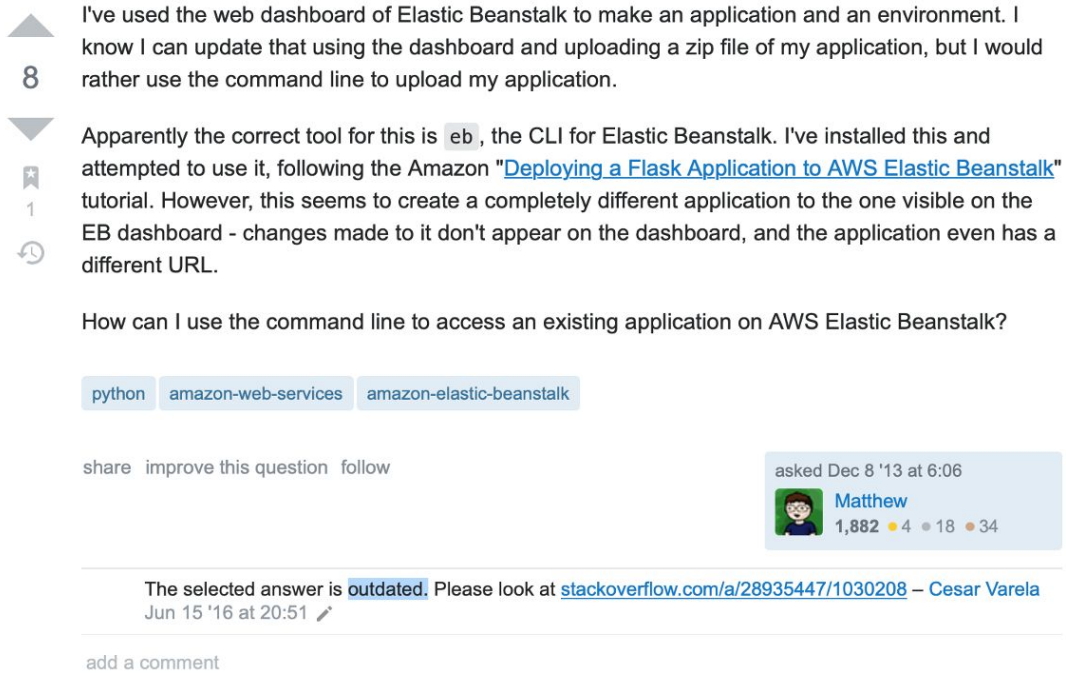
AWS users go to **Stackoverflow** for advice on their use of Amazon Web Services quite often**.** The community Q&A are super helpful, but…

AWS keeps evolving.

New versions were released and old ones are deprecated.

Outdated posts can be misleading to new users & cause frustrations.

An example of outdated post



8

1

I've used the web dashboard of Elastic Beanstalk to make an application and an environment. I know I can update that using the dashboard and uploading a zip file of my application, but I would rather use the command line to upload my application.

Apparently the correct tool for this is `eb`, the CLI for Elastic Beanstalk. I've installed this and attempted to use it, following the Amazon "Deploying a Flask Application to AWS Elastic Beanstalk" tutorial. However, this seems to create a completely different application to the one visible on the EB dashboard - changes made to it don't appear on the dashboard, and the application even has a different URL.

How can I use the command line to access an existing application on AWS Elastic Beanstalk?

python amazon-web-services amazon-elastic-beanstalk

share improve this question follow

asked Dec 8 '13 at 6:06

Matthew

1,882 ● 4 ● 18 ● 34

The selected answer is outdated. Please look at stackoverflow.com/a/28935447/1030208 – Cesar Varela

Jun 15 '16 at 20:51

add a comment

Our goals?

Our mentors from AWS and we want to know...

How many questions and answers on Stackoverflow about AWS are outdated?

Which AWS services tend to have more outdated posts?

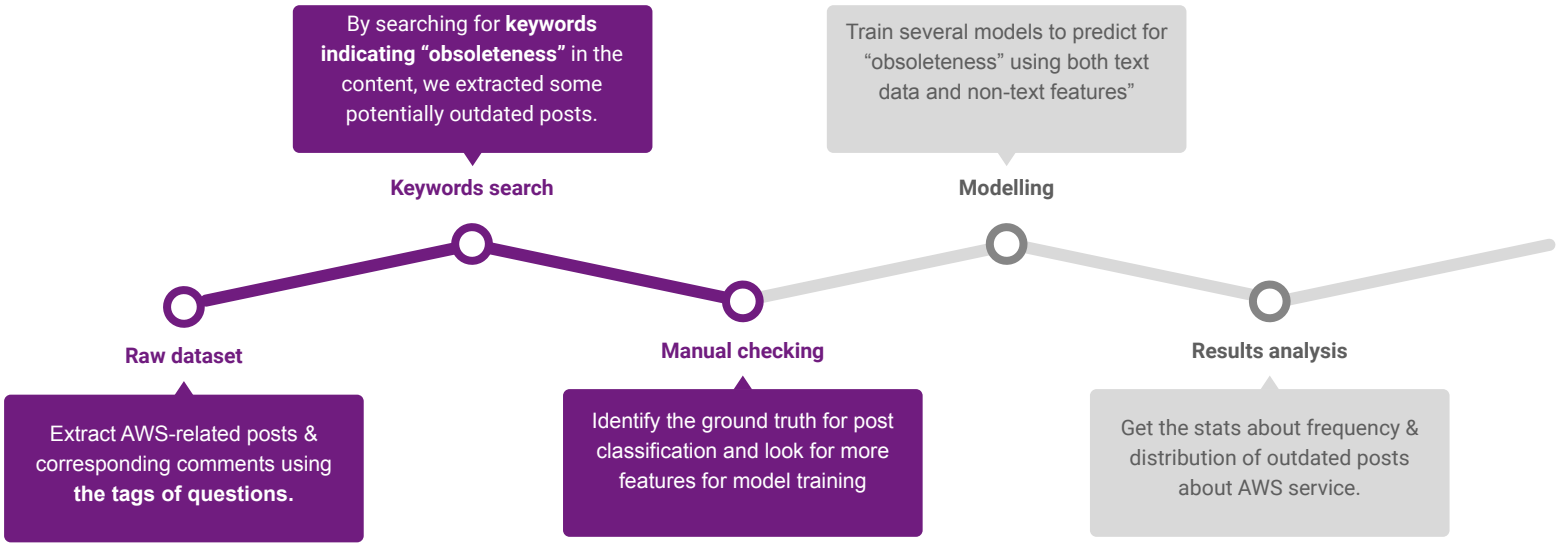
➔ A model that identifies outdated posts from StackOverflow dataset.

The Overview

1. We use the Stack Overflow archive dataset: content & metadata about comments, questions, answers.
2. A major **challenge**: there exists **no ground truth** about whether a subset of any posts are outdated or not.
3. We made use of an observation that **comments on an outdated post would sometimes point out its staleness**.
4. We extracted a subset of potentially outdated posts by keywords.
5. With some manual labels, we start our modellings.

The Overview

An example of false positive:
“query the database on the server only when the local database is out of date”,
where the comments actually mean some “outdated data”



Source dataset

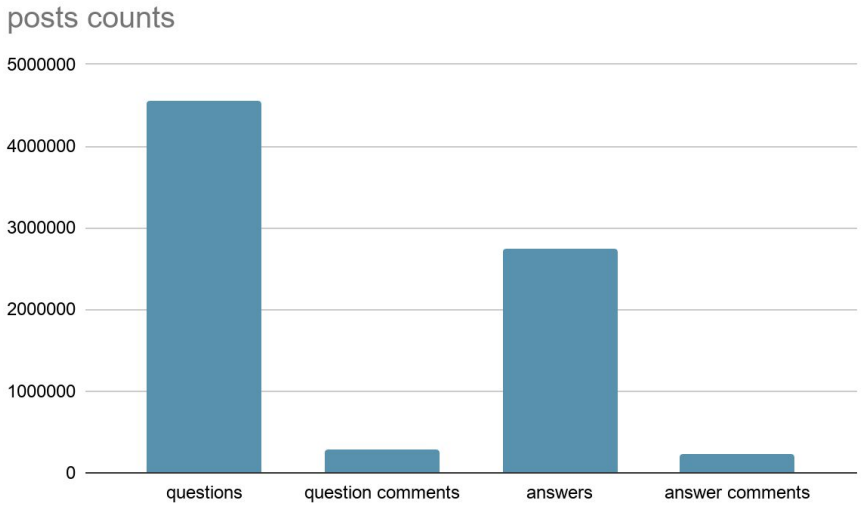
Download Posts.xml (82G) + Comments.xml (24G) from StackExchange Archive

↓

Use keyword “amazon” to search for aws-related posts in “tags” field.

↓

Collect aws-related questions & attached comments of 300MB, related answers & attached comments of 200MB



Heuristic method

From a sample, we observed keywords in comments that indicate the posts' staleness:

Then we extract a subset of comments / answers that contain those keywords:

	Comments of ans.	**Comments of que.**	**Answers content**
outdated	148	107	154
deprecated	257	168	285
obsolete	40	34	45
out of date	38	17	35
discouraged	79	48	84
stale	44	34	114
total	**927**		**598**

Analysis of heuristic results

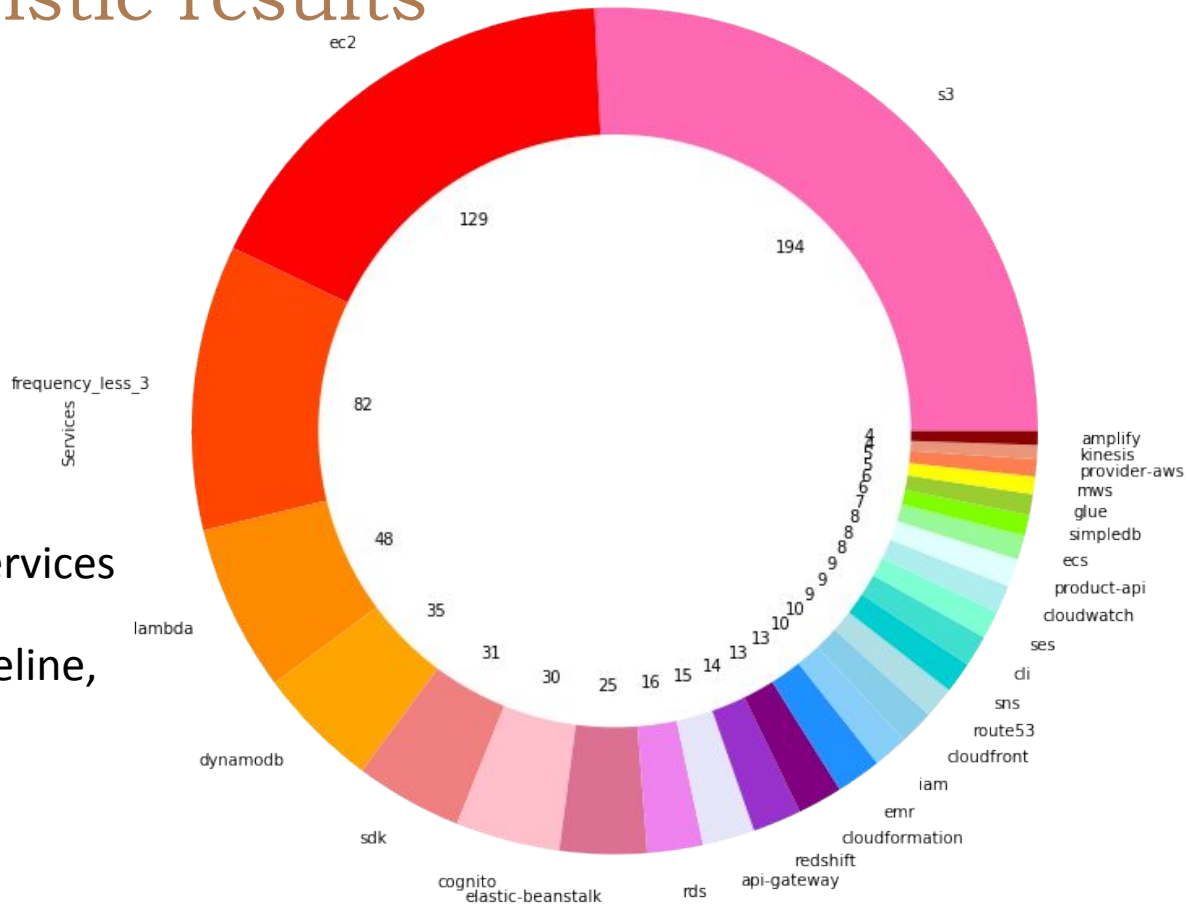
- Most Frequent services:

  S3, ec2, lambda, dynamodb

- Frequency less than 3: 52 services

  Including: silk, sqs, data-pipeline,

  fps, fire-tv, elb...



Manual checking process

link outdated	answer outdated	service in questior	other reasons
1	0	0	0
0	1	0	0
0	0	0	1
0	0	1	0
0	0	1	0

- **Goal:**
  - Get the ground truth for classification result
  - Define “false positive” posts for the heuristic method
  - Extract more features and patterns for model training
- **Extracted Features:**
  - Types of obsoleteness (ex: link outdated)
  - Aws-related or not
  - Keywords (ex: outdated, out of date)
  - Text content (ex: provide code)

One-Hot Encoding
(1 for yes, 0 for no)

Research on modelling (text data)

- **Data (after manual checking):**
  - 294 examples (194 comments & 100 answers)
  - 80% training, 20% testing
- **Models:**
  - Multinomial Naive Bayes
  - Logistic regression
    - GridSearch for parameter C
    - Class_weight = ‘balanced’
  - Fasttext

} Preprocessing: CountVectorizer & Tfidf

Research on modelling (text data)

- **Model result comparison:**

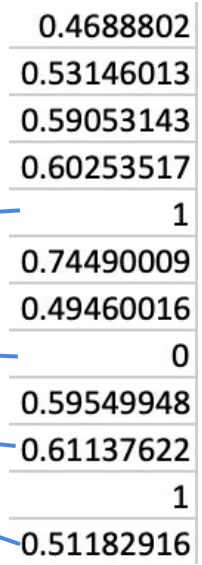
model choice

	Naïve Bayes	Logistic Regression	Fast text
Accuracy	0.6923	0.7115	0.6923
F1 score	0.8182	0.8046	0.8182

Research on modelling (text data)

- **Model scores:**
  - Treated as one of the features for next model
  - Score: probability of being “True”
    - Training data (True): 1
    - Training data (False): 0
    - Test data & other unchecked data:

probability given by model prediction



Modelling pipeline

1. Load data & **extract potential outdated** comments / answers
    - The models are to **identify "false positives"** from these suspects
    - A **classification** problem with insufficient data
2. Extract **heuristic features** using linguistic analysis methods
3. Merge comments / answers tables with columns from parent posts table.
4. Apply scalable features we identified in the manual checking process and join the featured data with manual labels.
5. Prediction scores with mere text model
6. Ultimate model combining the structured features & text model results

Feature engineering

- **Linguistic features:**
  Linguistic analysis with Spacy to recognize **subject and negative words in the sentences** containing target keywords (i.e outdated, obsolete).

  Extracted features:

Subject of sentence	the subject described as ‘outdated’
Punctuation	exclamation mark ! / question mark ? / period .
Negative statement	whether it is a negative statement
Subject irrelevance	whether the subject indicates “false positive”

Feature engineering

- **Parent posts' features:**
  Some of the parent posts' properties are also built into our models that classifies whether the comment indicates parents' staleness or not.

Date difference	The span between comment's creation date and the answer's last edit time.
Answer count	The number of answers on that parent question.
Score	The upvote count minus downvote count.
Comment count	The number of comments on the parent post.

Feature engineering

- **Content features:**
  - We collected a few features about the content of comments during the manual checking process & some of them are scalable.
  - Not all the features are scalable. (Such as True or False)
  - Extract link and code structure from the ‘text’ feature by using regular expression operator.
  - Use keyword search to get keywords contained in text.
  - As a result, we scale 4 manually checking features from the text directly.
    - ‘**Share_link**’, ’**share_code**’, ’**get_keywords**’, ’**aws_related_tags**’

Research on modelling (All Info)

- **Data:**
  - 1550 samples
  - Use manually checked data as the training data
    - 130 TRUE, 57 FALSE
  - Many missing values from the original data

Research on modelling (All Info)

- **Preprocess:**
  - Calculate date difference
    - CreationDate - (LastActivtyDate or LastEditDate or Today)
  - Remove unuseful columns (Ids, Dates, Texts, too many missing values)
    - Eventually 17 columns of features
  - Numeric:
    - Fill in missing values with mean
    - Standard Scaler
  - Boolean:
    - Passthrough

Research on modelling (All Info)

- **Model result comparison:**

	Logistic Regression (baseline)	Random Forest	SVC	Stochastic Gradient Descent	KNN
Accuracy	0.835	0.8453	0.883	0.788	0.85
F1 score	0.882	0.89	0.92	0.842	0.902
Precision	0.869	0.866	0.87	0.854	0.83
Recall	0.9	0.923	0.977	0.846	0.992
ROC_AUC	0.804	0.794	0.831	0.776	0.797

Research on modelling (All Info)

- **Feature Selection**
  - Select from model
    - ViewCount, model_score, Date_diff
  - Result:

Accuracy	F1	Precision	Recall	ROC_AUC
0.823	0.868	0.877	0.877	0.853

Analysis about the “all info” model

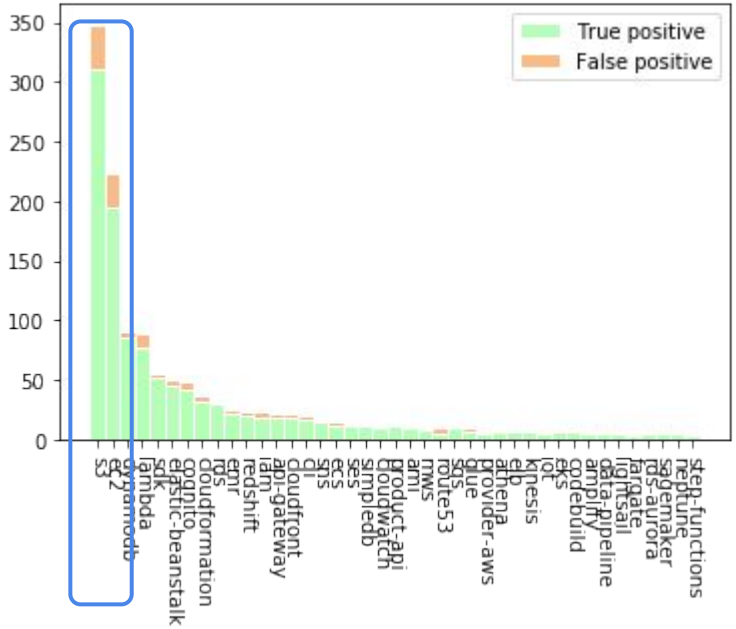
- **Confusion matrix**

  “All info” model’s prediction results on the comments.

	Predicted as outdated	Predicted as non-outdated
Confirmed outdated	38	19
Checked non-outdated	3	127

Final results

- **How many AWS-related posts contain outdated information?**
  - True (Outdated): 1,354
  - False (Not outdated): 196
- **Which AWS's tend to have more outdated posts?**



Discussion

- **Challenges during research:**
  - Insufficient data size for accurate model
  - Manual labeling process can be inaccurate and time-consuming
  - Difficult to identify scalable features
- **Future works:**
  - Identify more “outdated” keywords
  - Further improve the quantity and quality of manually labeled data
  - Try more accurate models for larger dataset

Acknowledgement

1. Haoxiang Zhang, Shaowei Wang, Peter Chen and Ahmed E. Hassan. “An Empirical Study of Obsolete Answers on Stack Overflow”.
2. Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”
3. https://fasttext.cc/docs/en/supervised-tutorial.html
4. https://github.com/slundberg/shap
5. https://scikit-learn.org/stable/modules/naive_bayes.html
6. https://scikit-learn.org/stable/tutorial/text_analytics/working_with_text_data.html