

Automatic Extraction of Opt-Out Choices from Privacy Policies

Kanthashree Mysore Sathyendra, Shomir Wilson, Florian Schaub and Norman Sadeh

Motivation

- Privacy policies are very long documents and it is difficult for users to identify if any choices are offered to them.
- Is it possible to automatically extract information about such “choice instances” from privacy policies?

Opt-Out Choices

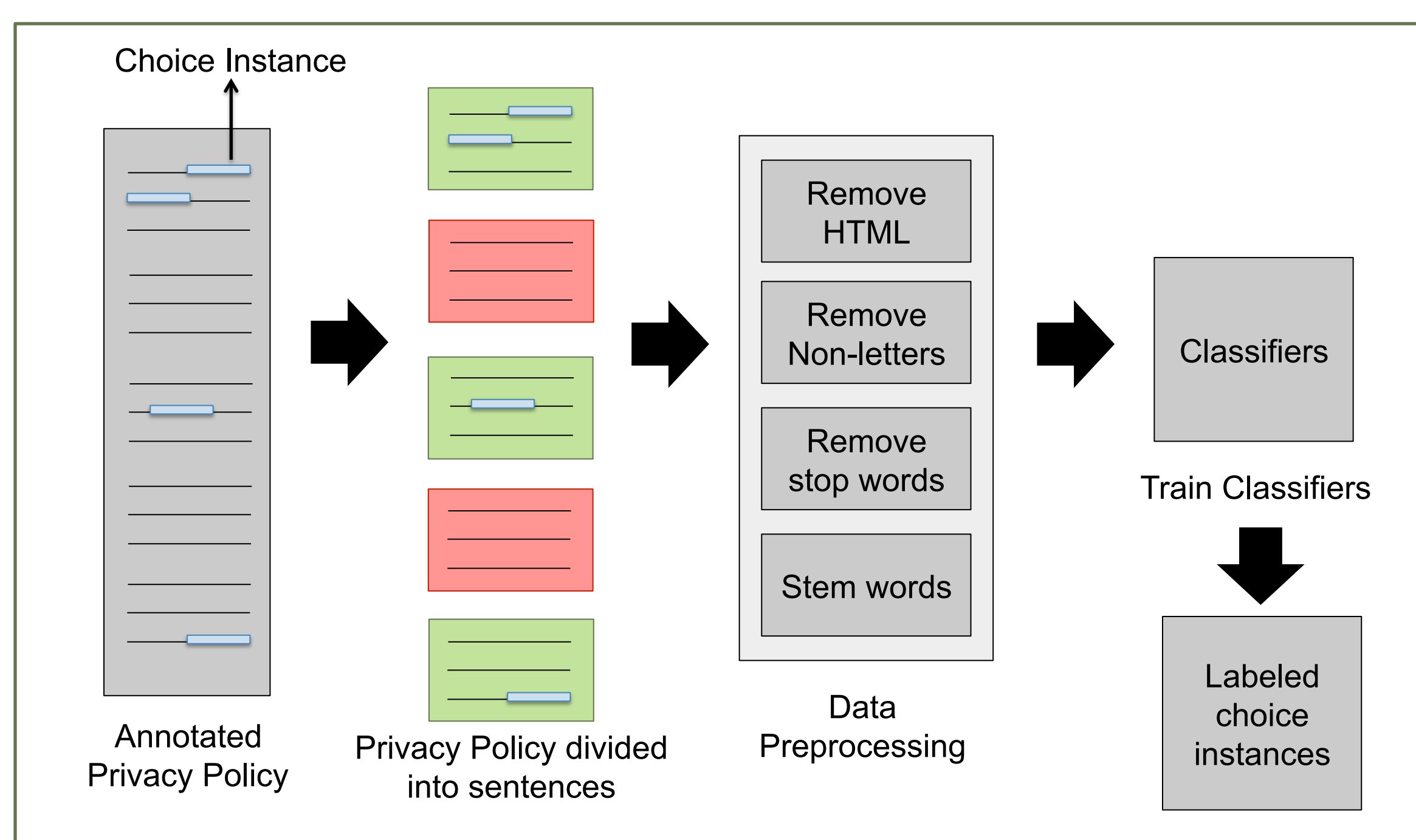
- a **choice instance** is a statement in a privacy policy that indicates the user has discretion over the collection, use, sharing, or retention of their data
- Examples of ‘Opt-Out’ Choices:
 - You are free to opt out of this data collection service by going to HasOffers Analytics' End User Opt-Out page at <https://www.optoutmobile.com/optout>.
 - Please click here to see a list of the third parties Military works with and to exercise your opt out.

OPP 115 Corpus

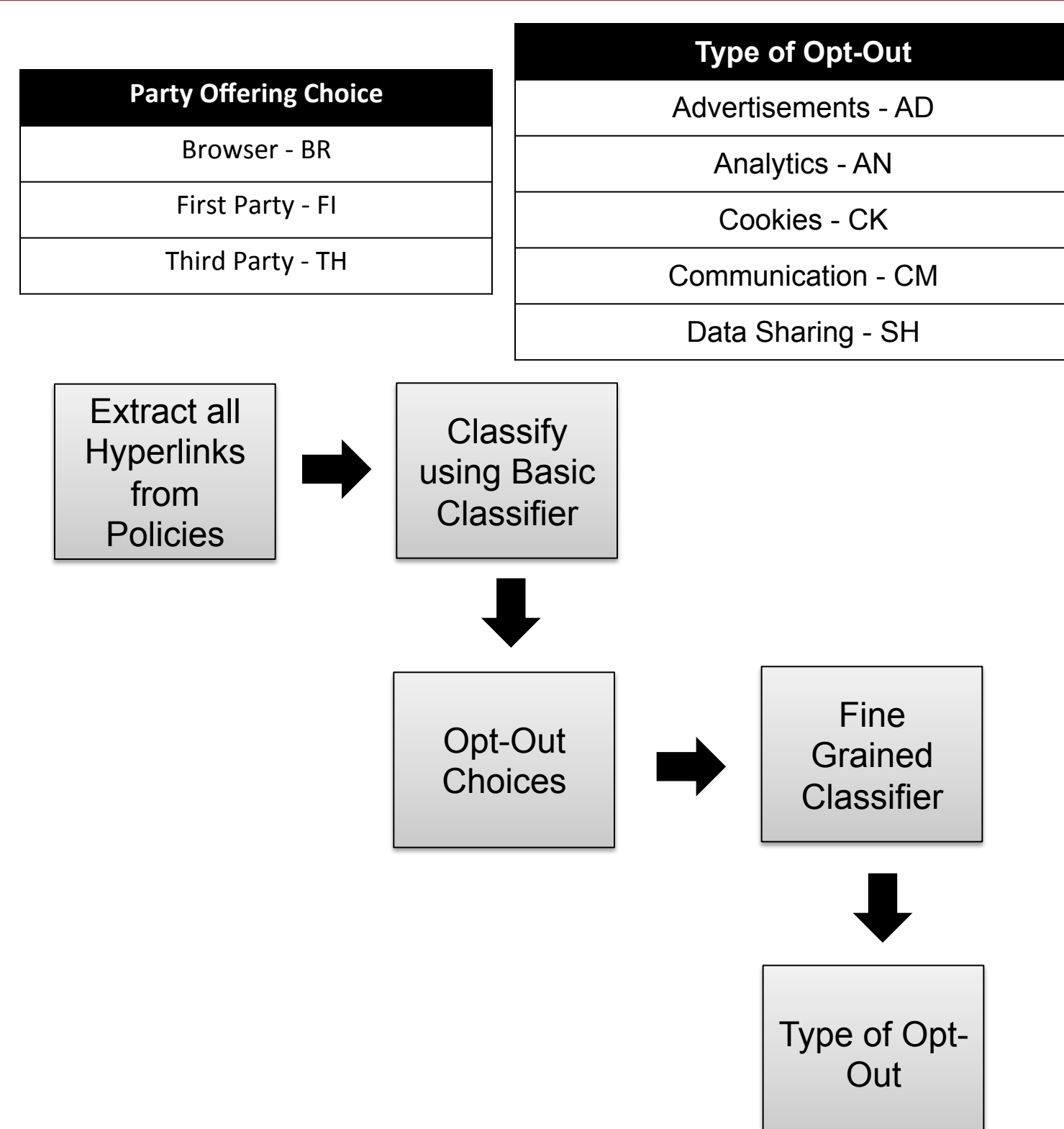
- For each paragraph in a privacy policy, the annotators identify data practices in the paragraphs (segments).
- 115 Privacy Policies
 - 3792 Segments

Approach

- Supervised Machine Learning



Two-Tier Model Architecture



Basic Model Model to identify opt-out choices. Features used for classification:

- Bag-of-ngrams
- Topic Distributions
- Modal Verb
- Opt-out Specific Words

Fine-Grained Model Model to identify the type of opt-out. Features used for classification:

- Bag-of-ngrams for sentences, segments
- Anchor Text
- Hyperlink
- Policy Url
- Url Similarity Measure: Measure of similarity of Hyperlink and policy Url

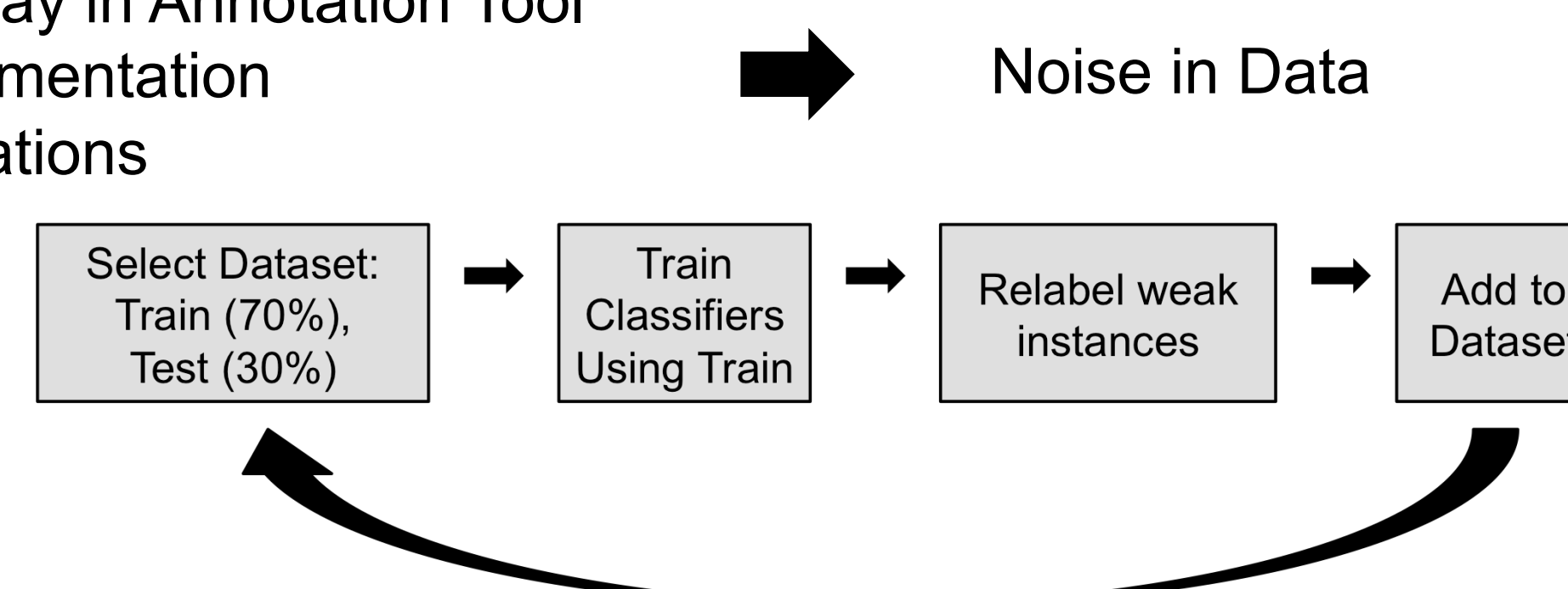
Modal Verbs

- A modal verb is a type of verb that is used to indicate modality – that is: likelihood, ability, permission, and obligation
- NLTK Part of Speech Tagger used to extract Modal verbs
- Examples of modal verbs: **may, might, can**
- Examples of positive instances containing Modal Verbs:
 - You **may** opt out of receiving these general communications by using one of the following methods: Select the email opt out or unsubscribe link, or follow the opt-out instructions included in each email communication.
- Example of POS Tagging:

You **may** unsubscribe to any of our online e-mail updates or newsletters by clicking here.
 PRP MD VB TO DT IN PRP JJ JJ NNS CC NNS IN VBG RB .

Handling Noise in Data

- Hyperlink display in Annotation Tool
- Sen tence segmentation
- Missing annotations



Topic Distributions as Features

- Each segment can be represented as a distribution of hidden topics
- Use topic distributions as features along with bag-of-ngram and modal verb features during classification
- Latent Dirichlet Allocation and Non-negative Matrix Factorization

Results

Feature Set	Model	Precision	Recall	F1	Accuracy
Unigram	Logistic Regression	0.574	0.493	0.530	0.987
	SVM	0.417	0.493	0.452	0.982
	Naïve Bayes	0.263	0.634	0.372	0.967
	Random Forest	0.667	0.254	0.367	0.987
Unigram + Bigram	Logistic Regression	0.59	0.507	0.545	0.987
	SVM	0.537	0.507	0.522	0.986
	Naïve Bayes	0.324	0.662	0.435	0.974
Unigram + Bigram + Custom Feature (Modal Verbs and opt-out specific phrases) *	Logistic Regression	0.727	0.686	0.705	0.989
	Logistic Regression	0.765	0.722	0.742	0.99

* After Noise Handling In Data