Prioritizing What to Work On

Building a Spam Classifier

Advice for Applying Machine Learning:

Building a spam classifier

From: cheapsales@buystufffromme.com
To: ang@cs.stanford.edu
Subject: Buy now!

Deal of the week! Buy now!
Rolex w4tchs - $100
Med1cine (any kind) - $50
Also low cost M0rgages
available.

Spam

From: Alfred Ng
To: ang@cs.stanford.edu
Subject: Christmas dates?

Hey Andrew,
Was talking to Mom about plans
for Xmas. When do you get off
work. Meet Dec 22?
Alf

Non-spa

Andrew Ng

Building a spam classifier

From: cheapsales@buystufffromme.com
To: ang@cs.stanford.edu
Subject: Buy now!

Deal of the week! Buy now!
Rolex w4tchs - $100
Med1cine (any kind) - $50
Also low cost M0rgages
available.

Spam

From: Alfred Ng
To: ang@cs.stanford.edu
Subject: Christmas dates?

Hey Andrew,
Was talking to Mom about plans
for Xmas. When do you get off
work. Meet Dec 22?
Alf

Non-spam

Andrew Ng

Building a spam classifier

From: cheapsales@buystufffromme.com
To: ang@cs.stanford.edu
Subject: Buy now!

Deal of the week! Buy now!
Rolex w4tchs - $100
Med1cine (any kind) - $50
Also low cost M0rgages
available.

Spam (1)

From: Alfred Ng
To: ang@cs.stanford.edu
Subject: Christmas dates?

Hey Andrew,
Was talking to Mom about plans
for Xmas. When do you get off
work. Meet Dec 22?
Alf

Non-spam (0)

Andrew Ng

Building a spam classifier

Supervised learning. $x =$ features of email. $y =$ spam (1) or not spam (0).
Features x: Choose 100 words indicative of spam/not spam.

Andrew Ng

Building a spam classifier

Supervised learning. $x =$ features of email. $y =$ spam (1) or not spam (0).

Features x: Choose 100 words indicative of spam/not spam.

E.g. deal, buy, discont, andrew, now, ...

Andrew Ng

Building a spam classifier

Supervised learning. $x =$ features of email. $y =$ spam (1) or not spam (0).

Features x: Choose 100 words indicative of spam/not spam.

E.g. deal, buy, discont, andrew, now, ...

From: cheapsales@buystufffromme.com
To: ang@cs.stanford.edu
Subject: Buy now!

Deal of the week! Buy now!

Andrew Ng

Building a spam classifier

Supervised learning. $x = $ features of email. $y = $ spam (1) or not spam (0).

Features x: Choose 100 words indicative of spam/not spam.

E.g. deal, buy, discont, andrew, now, ...

andrew
buy
deal
discont
⋮
now
⋮

From: cheapsales@buystufffromme.com
To: ang@cs.stanford.edu
Subject: Buy now!

Deal of the week! Buy now!

Andrew Ng

Building a spam classifier

Supervised learning. $x = $ features of email. $y = $ spam (1) or not spam (0).
Features x: Choose 100 words indicative of spam/not spam.

E.g. deal, buy, discont, andrew, now, ...

$$X = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ \vdots \\ 1 \\ \vdots \end{bmatrix} \begin{matrix} \text{andrew} \\ \text{buy} \\ \text{deal} \\ \text{discont} \\ \vdots \\ \text{now} \\ \vdots \end{matrix}$$

From: cheapsales@buystufffromme.com
To: ang@cs.stanford.edu
Subject: Buy now!

Deal of the week! Buy now!

Andrew Ng

Building a spam classifier

Supervised learning. x = features of email. y = spam (1) or not spam (0).

Features x: Choose 100 words indicative of spam/not spam.

E.g. deal, buy, discont, andrew, now, ...

$$x = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ \vdots \\ 1 \\ \vdots \end{bmatrix} \begin{matrix} \text{andrew} \\ \text{buy} \\ \text{deal} \\ \text{discont} \\ \vdots \\ \text{now} \\ \vdots \end{matrix}$$

$x \in \mathbb{R}^{100}$

$$x_j = \begin{cases} 1 & \text{if word } j \text{ appears in email} \\ 0 & \text{otherwise.} \end{cases}$$

```
From: cheapsales@buystufffromme.com
To: ang@cs.stanford.edu
Subject: Buy now!

Deal of the week! Buy now!
```

Andrew Ng

Building a spam classifier

Supervised learning. $x =$ features of email. $y =$ spam (1) or not spam (0).

Features x: Choose 100 words indicative of spam/not spam.

E.g. deal, buy, discont, andrew, now, ...

$$x_j = \begin{cases} 1 & \text{if word } j \text{ appears in email} \\ 0 & \text{otherwise.} \end{cases}$$

$$x = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ \vdots \\ 1 \\ \vdots \end{bmatrix} \begin{matrix} \text{andrew} \\ \text{buy} \\ \text{deal} \\ \text{discont} \\ \vdots \\ \text{now} \\ \vdots \end{matrix}$$

$x \in \mathbb{R}^{100}$

```
From: cheapsales@buystufffromme.com
To: ang@cs.stanford.edu
Subject: Buy now!

Deal of the week! Buy now!
```

Note: In practice, take most frequently occurring n words (10,000 to 50,000)
in training set, rather than manually pick 100 words.

Building a spam classifier

How to spend your time to make it have low error?

- Collect lots of data

  - E.g. "honeypot" project.

Andrew Ng

Building a spam classifier

How to spend your time to make it have low error?

- Collect lots of data
  - E.g. "honeypot" project.
- Develop sophisticated features based on email routing information (from email header).

Andrew Ng

Giriş × Posta-Gele... × Re: Bitirme ... × Online regis... ×

Mehmet Guray GULER
daisy/yildiz/tr

Gelen Kutusu (6778)
Taslaklar
Gönderilenler
İzleme
Tüm Belgeler
Yığın Posta
Çöp Kutusu

Görünümler

Klasörler

Arşiv

Araçlar

Diğer Posta

Yeni Klasör · Yanıtla · Tümünü Yanıtla · İlet

Online registration is opened

3rd WORLD CONFERENCE ON TECHNOLOGY, INNOVATI

Kime: mgguler@yildiz.edu.tr

Güvenlik: Gizliliği korumak için, uzak sitelerdeki resimlerin yüklenm

---

MIME İleti Bilgisi ×

Received: from spamgw.cc.yildiz.edu.tr ([193.140.1.10])
by daisy.cc.yildiz.edu.tr (Lotus Domino Release 8.5.3FP6HF1222)
with ESMTP id 2018120714590240-50910 ;
Fri, 7 Dec 2018 14:59:02 +0200
X-Quarantine-ID: <hFxcZCbYUrcf>
X-Virus-Scanned: by SpamTitan at cc.yildiz.edu.tr
X-Spam-Flag: NO
X-Spam-Score: 1.777
X-Spam-Level: *
X-Spam-Status: No, score=1.777 tagged_above=-999 required=5
tests=[BAYES_00=-1.9, DKIM_SIGNED=0.1, DKIM_VALID=-0.1,
DKIM_VALID_AU=-0.1, HTML_IMAGE_ONLY_24=1.618,
HTML_MESSAGE=0.001,
HTML_SHORT_LINK_IMG_3=0.148, SPF_PASS=-0.001,
ST_RCVD_IN_SPAMRATS_DYNA=1, ST_URIBL_SWINOG=1,
T_REMOTE_IMAGE=0.01,
URIBL_BLOCKED=0.001] autolearn=no autolearn_force=no
Authentication-Results: spamgw.cc.yildiz.edu.tr (amavisd-new);
dkim=pass (1024-bit key) header.d=e.figurnews.net;
domainkeys=pass (1024-bit key)
header.from=bulten@e.figurnews.net
header.d=e.figurnews.net
Received-SPF: pass (e.figurnews.net: Sender is authorized to use
'bulten@e.figurnews.net' in 'mfrom' identity (mechanism
'include:subspf3.rvs0.com' matched)) receiver=spamgw;
identity=mailfrom; envelope-from="bulten@e.figurnews.net";
helo=mail312108311.revomailtr.com; client-ip=31.210.83.11
Received: from mail312108311.revomailtr.com

Building a spam classifier

How to spend your time to make it have low error?

- Collect lots of data
  - E.g. "honeypot" project.
- Develop sophisticated features based on email routing information (from email header).
- Develop sophisticated features for message body, e.g. should "discount" and "discounts" be treated as the same word? How about "deal" and "Dealer"? Features about punctuation?

Andrew Ng

Building a spam classifier

How to spend your time to make it have low error?

- Collect lots of data
  - E.g. "honeypot" project.
- Develop sophisticated features based on email routing information (from email header).
- Develop sophisticated features for message body, e.g. should "discount" and "discounts" be treated as the same word? How about "deal" and "Dealer"? Features about punctuation?
- Develop sophisticated algorithm to detect misspellings (e.g. m0rtgage, med1cine, w4tches.)

Andrew Ng