

Text classifier based on naive Bayes classifier

顾一多 王子扬 吴可斐 赵冬昊

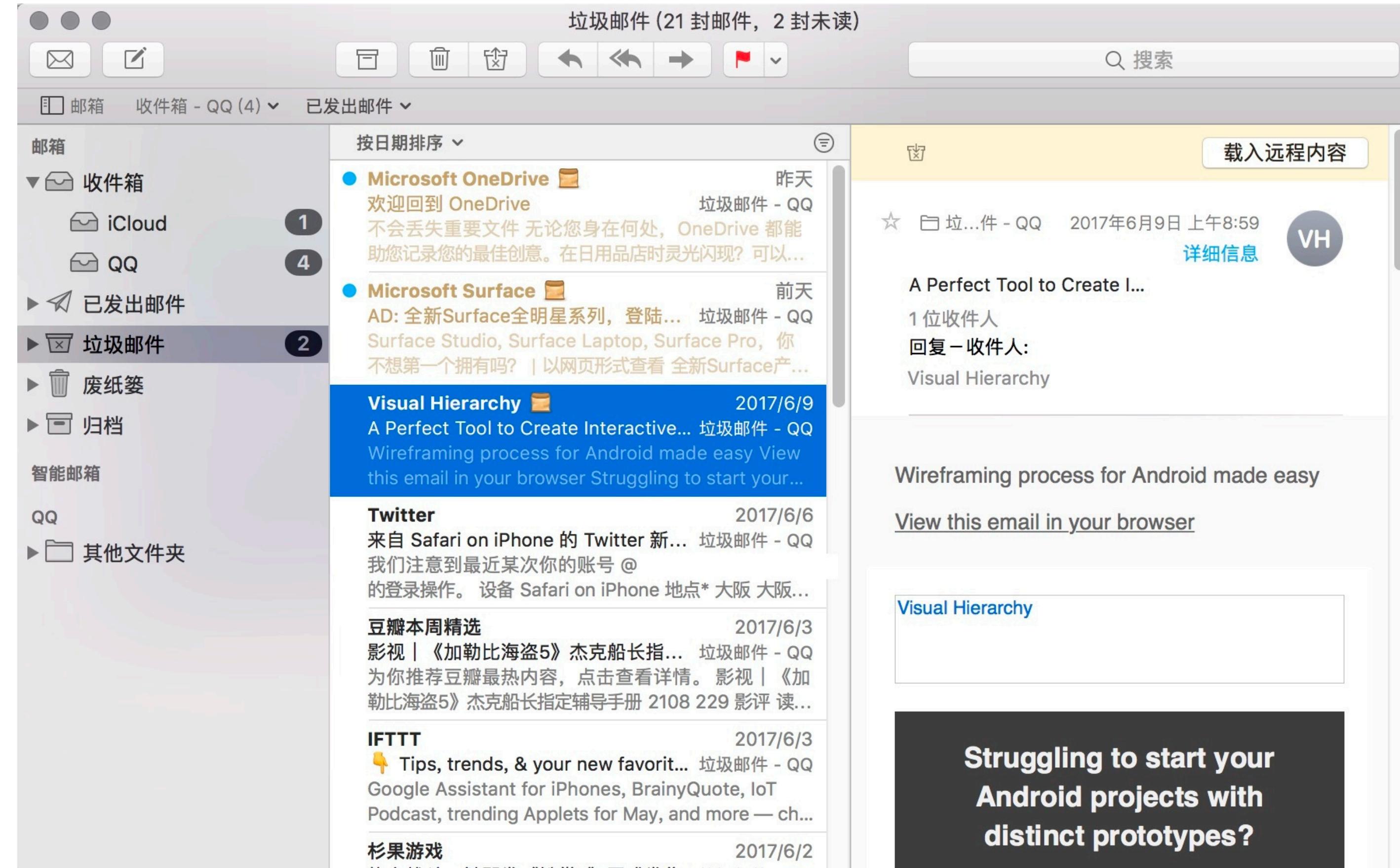
Motivation

We receive a lot of messages every day and they have different classifications, although most of them are spams, there're still some useful messages. Some of them are important, while some of them not.

Now there's a good spam filter, but we want to distinguish between more categories of messages and their importance.

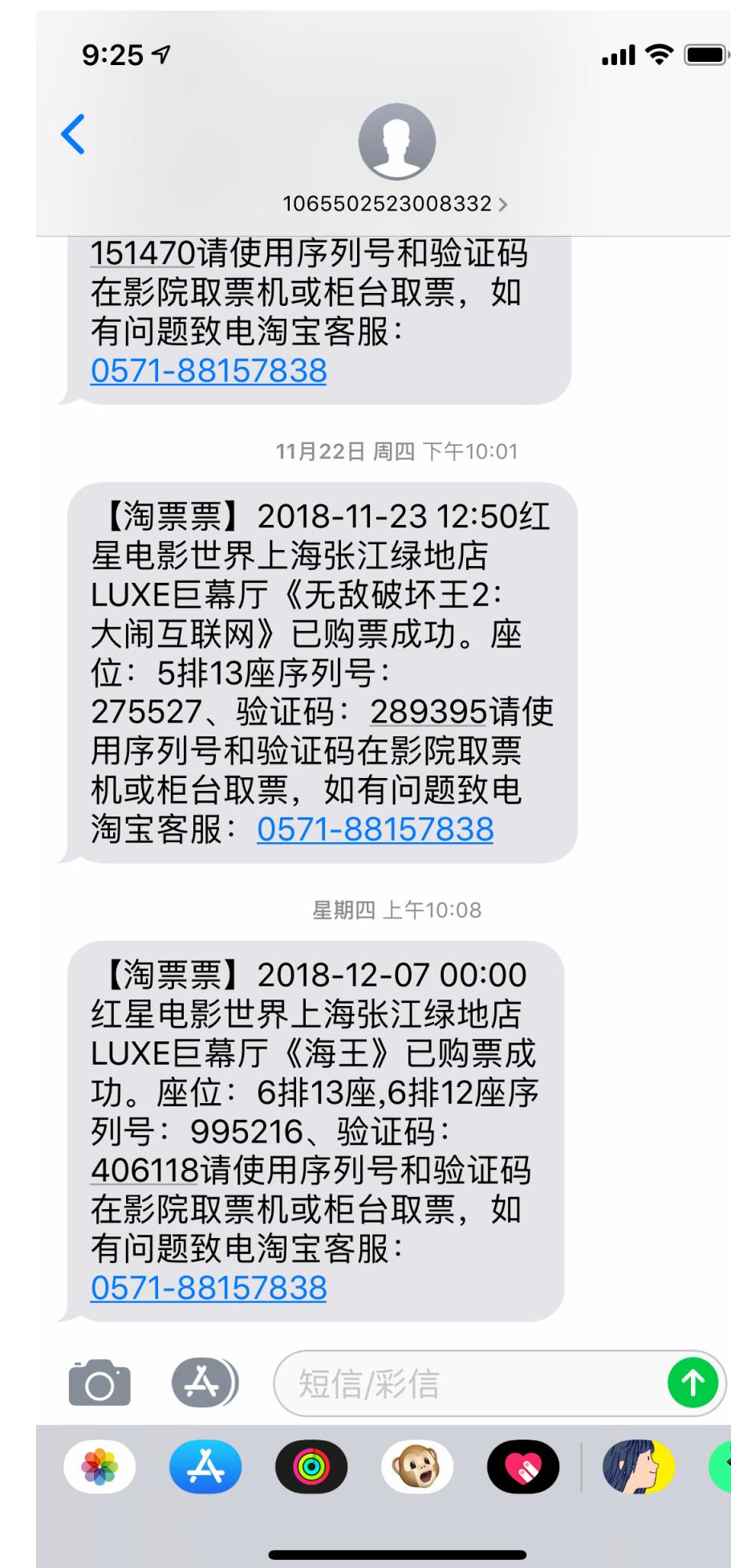
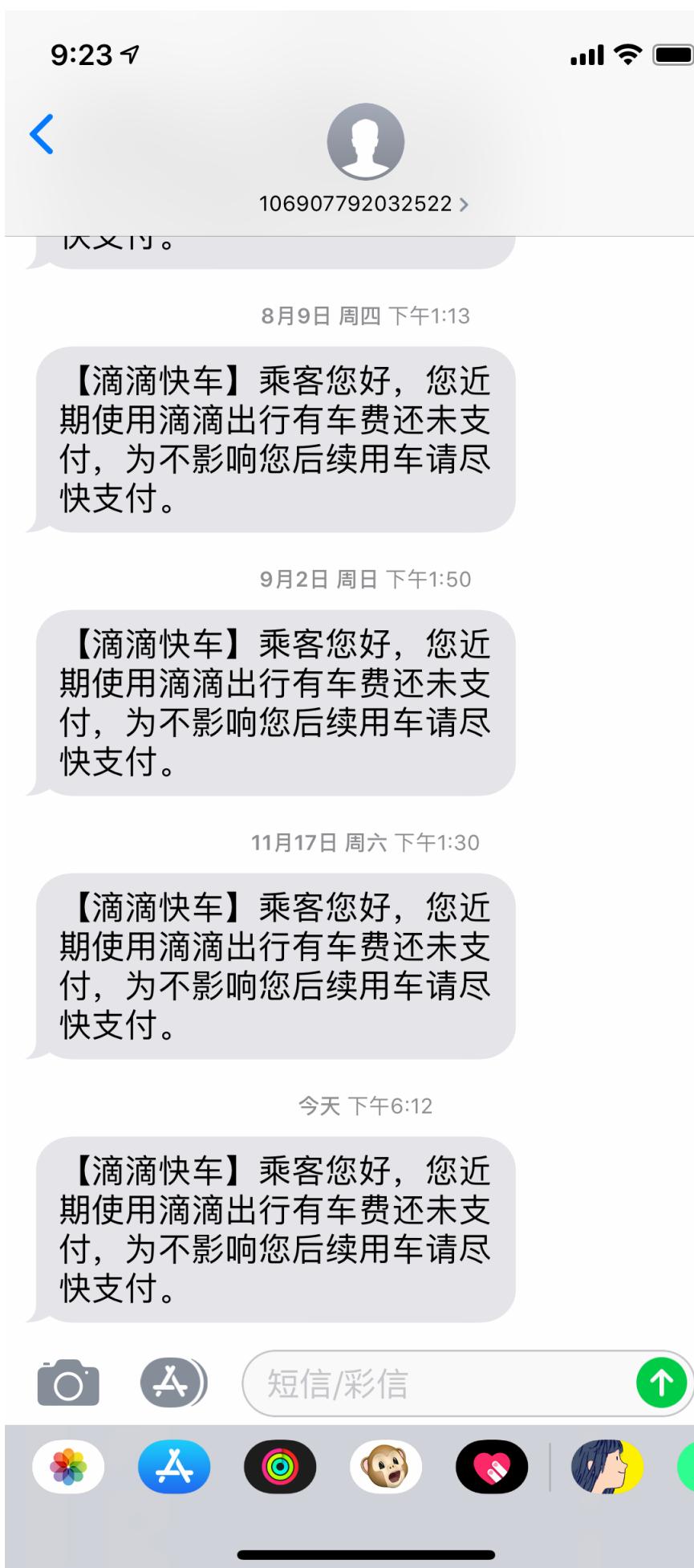


Spam SMS

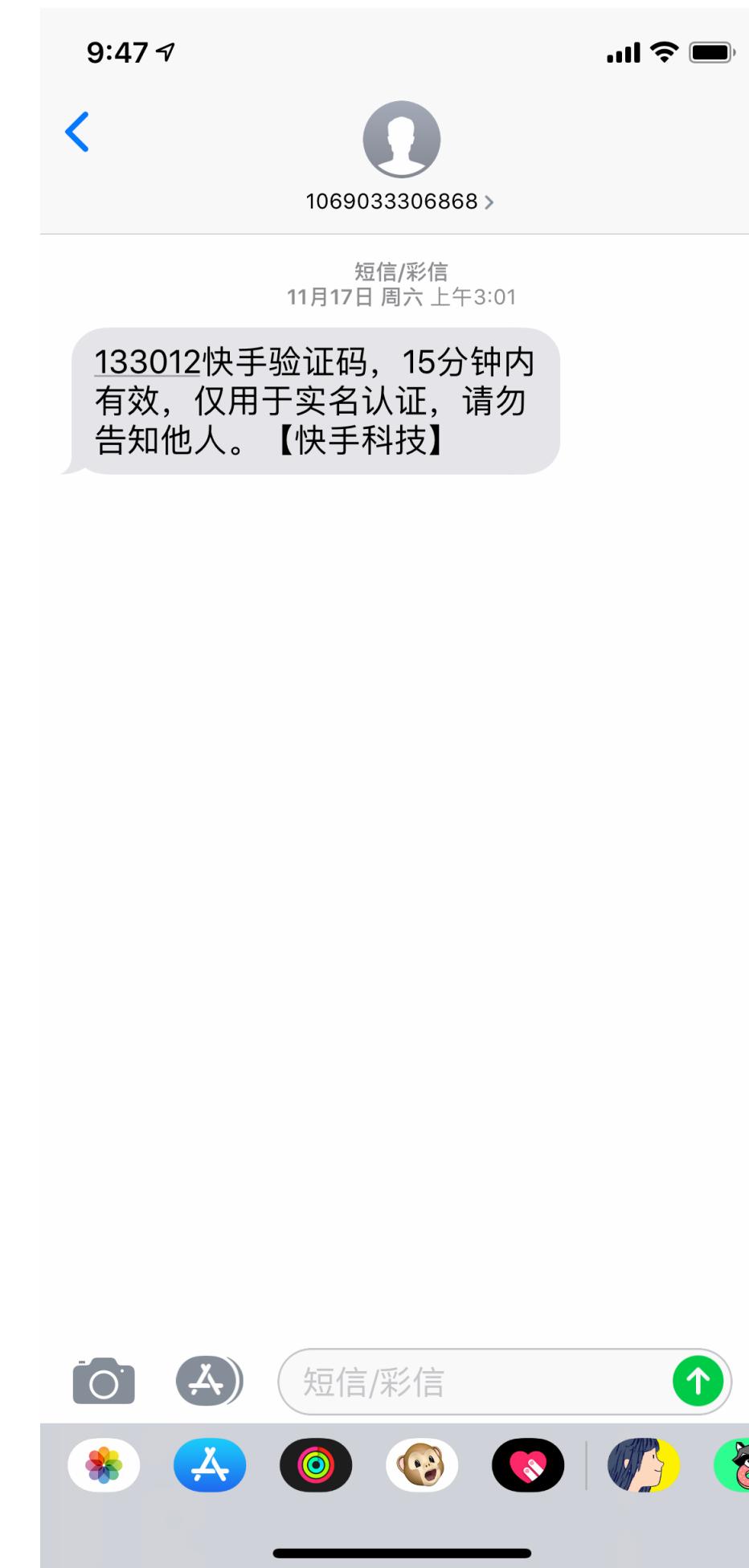


Spam Emails

Notifications



Verification code



Private messages



PandaEatsSMS is the most useful spam filter on iOS. It's implemented by naive Bayes classifier. We would also like to use naive Bayes to make our classifier.

Naive Bayes Classifier

Assume that the value of a particular feature is independent of the value of any other feature, given the class variable.

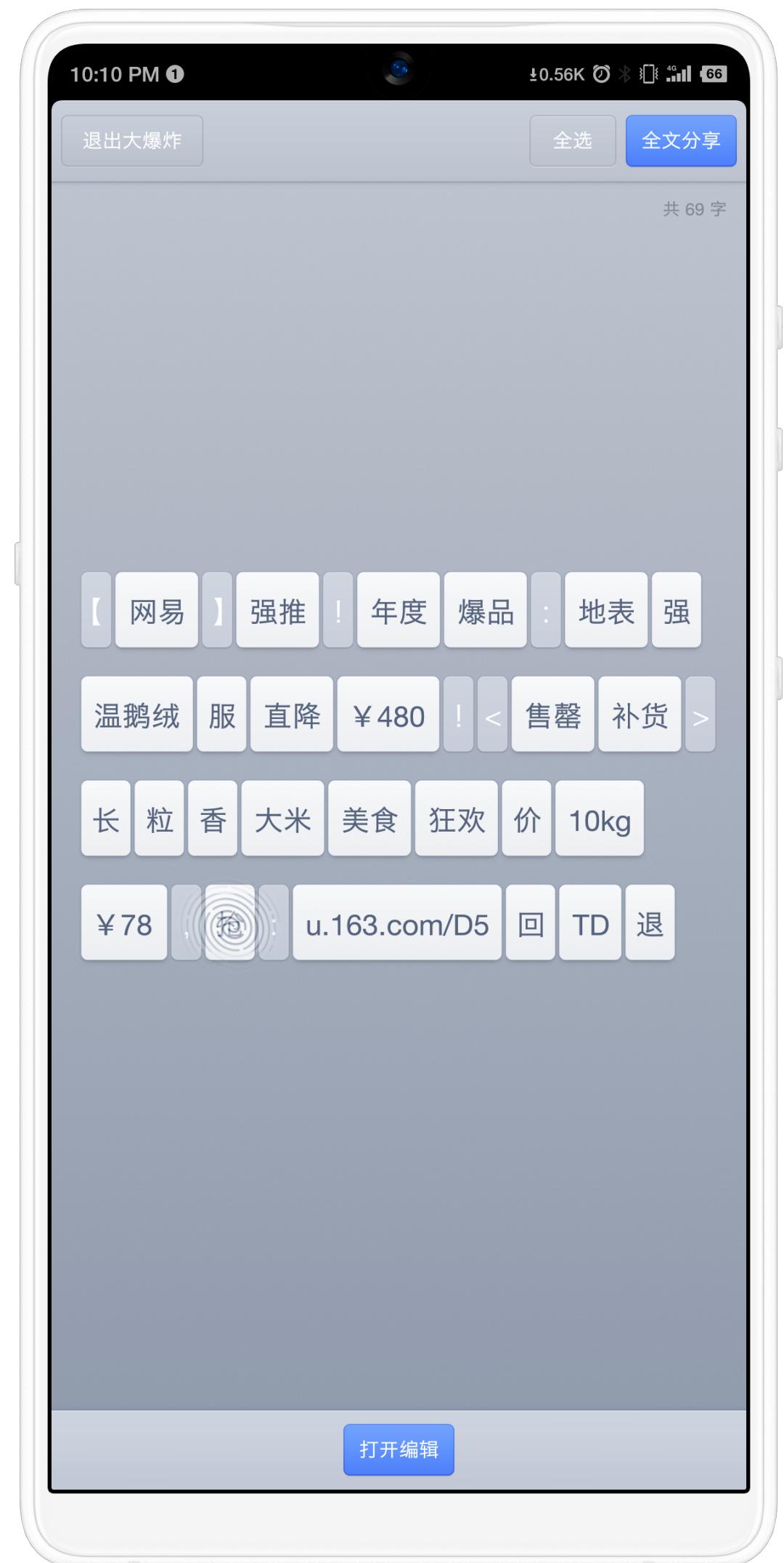
Bayes' theorem

$$P(A | B) = \frac{P(A) \times P(B | A)}{P(B)}$$

$$\textit{posterior} = \frac{\textit{prior} \times \textit{likelihood}}{\textit{evidence}}$$



Separate words as features



Collect enough labeled messages for training

Get the ideal model

What are we going to do?

1. Write a classifier that can separate multiple classifications.
2. Collect messages as more as possible.
3. Data processing.
4. Training and Testing

Thank you!