Выявление мошенничества с кликами в интернет-рекламе

Рысьмятова Анастасия

ВМК МГУ 317 группа

16.02.2015

Содержание

Кликфрод

Кликфрод — один из видов сетевого мошенничества, представляющий собой обманные клики на рекламную ссылку лицом, не заинтересованным в рекламном объявлении. Может осуществляться с помощью автоматизированных скриптов или программ, имитирующих клик пользователя по рекламным объявлениям Pay per click. Скликивать объявления могут недобросовестные вебмастера со своих же сайтов, либо конкуренты рекламодателей.

Примеры

- Клики рекламодателей — переходы по ссылкам, совершаемые рекламодателями по собственным объявлениям с целью поднятия CTR.
- Клики конкурентов — переходы по ссылкам, совершаемые сотрудниками конкурирующих структур.
- Клики со стороны недобросовестных вебмастеров — переходы по ссылкам, совершаемые вебмастерами или созданными ими системами с целью увеличения доходности рекламной площадки

Содержание

Цель конкурса

В 2012 году была предложена задача, целью которой являлось обнаружение площадок - мошенников

Данные

id	numericip	deviceua	publisherid	campaignid	usercountry	clicktime	channel	referredurl
13417867	3648406743	GT-I9100	8iaxj	8fj2j	ru	2012-02-09 00:00:00	ad	26okyx5i82hws84o
13417870	3756963656	Samsung_S5233	8jljr	8geyk	in	2012-02-09 00:00:00	es	15vynjr7rm00gw0g
13417872	693232332	SonyEricsson_K70	8jljr	8gkkx	ke	2012-02-09 00:00:00	es	
13417893	2884200452	Nokia_6300	8jljr	8gp95	vn	2012-02-09 00:00:01	es	
13418096	3648406743	GT-I9100	8iaxj	8fj2m	ru	2012-02-09 00:00:08	ad	24w9x4d25ts00400
13418395	781347853	GT-I9003	8iaxj	8fj2j	ru	2012-02-09 00:00:20	ad	4im401arl30gc0gk

id - Уникальный идентификатор клика
numericip - IP адресс кликера
deviceua - Модель телефона с которого сделан клик
publisherid -Уникальный идентификатор площадки
campaignid - Уникальный идентификатор рекламной компании
usercountry - Страна в которой был сделан клик
clicktime - Время клика
referredurl - URL клика
channel - тип сайта

Данные

publisherid	bankaccount	address	status
8iaxj		14vxbyt6sao00s84	Fraud
8jljr			OK

publisherid - Уникальный идентификатор площадки
bankaccount - Счет площадки
status - Статус площадки, который имеет три категории:

- OK
- Observation
- Fraud

Лучшие результаты

Rank	Team	Average precision		Affiliation
		Validation set	Test set	
1	starrystarrynight	59.38%	51.55%	Institute of Infocomm Research
2	TeamMasdar	59.39%	46.42%	Masdar Institute of Science & Technology
3	DB2	62.21%	46.15%	National University of Singapore
4	Tea	51.55%	42.01%	Tokyo Institute of Technology
(*)	LARC	57.79%	55.64%	Singapore Management University

Содержание

Особенности в данных

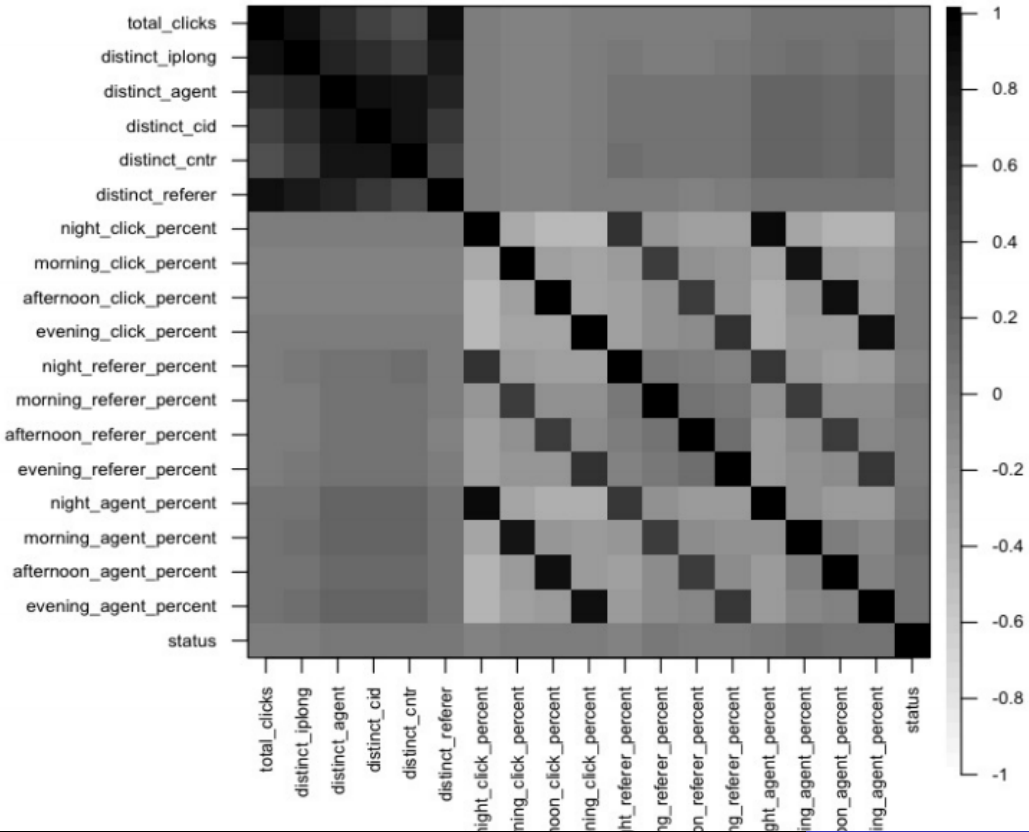
Разделили день на 4 периода: утро, день, вечер и ночь
Разделить час на 4 части
Заметили что в различных категориях сайтов мошеннические клики осуществляются в разное время

Channel	Publisher count	Fraud clicks (fraud %)	Night Fraud clicks (fraud %)	Morning Fraud clicks (fraud %)	Afternoon Fraud clicks (fraud %)	Evening Fraud clicks (fraud %)
Adult (ad)	10	**47226 (37%)**	**15435 (12%)**	6439 (5%)	**11299 (9%)**	**14053 (11%)**
Mobile content (mc)	**23**	41941 (33%)	13589 (11%)	**9284 (7%)**	9623 (8%)	9445 (7%)
Community (co)	12	16411 (13%)	7218 (6%)	3301 (3%)	2612 (2%)	3280 (3%)
Entertainment and lifestyle (es)	14	14433 (11%)	2649 (2%)	3265 (3%)	3573 (3%)	4946 (4%)
Search, portal, services (se)	4	3180 (3%)	682 (1%)	572 (0%)	689 (1%)	1568 (1%)
Premium portal (pp)	6	2926 (2%)	351 (0%)	608 (0%)	732 (1%)	904 (1%)
Information (in)	3	893 (1%)	49 (0%)	284 (0%)	428 (0%)	132 (0%)
Total	72	127010 (100%)	**39973 (31%)**	23753 (19%)	28956 (23%)	34328 (27%)

Заметили, что процент мошеннических кликов сильно зависит от страны.

Признаки

На основе замеченных особенностей в данных было выделено 118 признаков и составлен график корреляций для некоторых из них, чтобы исключать добавление признаков, сильно похожих на уже добавленные.



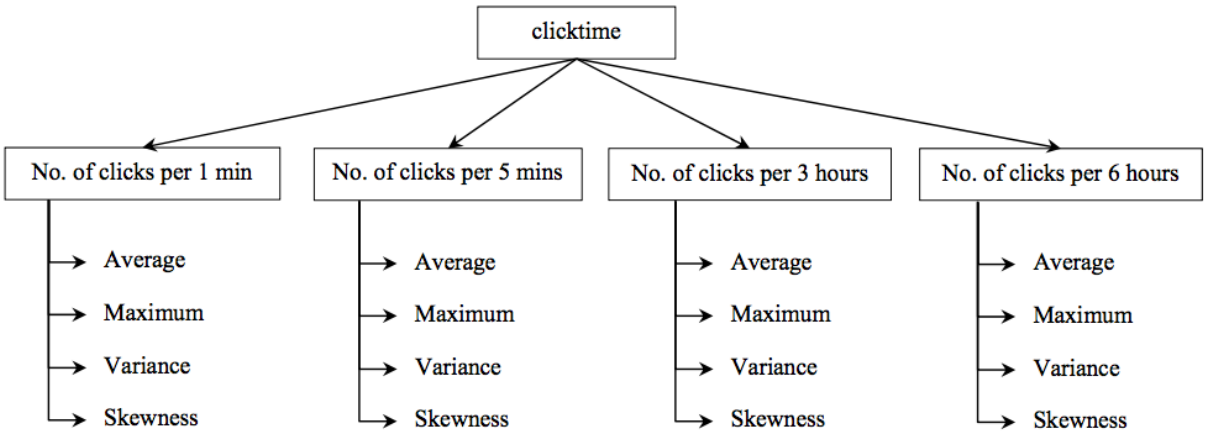
Метод

Была использована boosted regression model (GBM) на 118 признаках, которые были получены.
Также пытались использовать randomForest, но не смогли получить хорошего результата
Параметры для GBM:

- distribution (loss function): "bernoulli"
- n.trees (number of iterations): 5000
- shrinkage (learning rate): 0.001
- interaction.depth (tree depth): 5
- n.minobsinnode (minimum observations in terminal node): 5

Особенности в данных

Наблюдают число кликов для каждой площадки в различные промежутки времени. Пытаются найти постоянные клики с одинаковыми интервалами времени



Многократное нажатие с одного и того же IP осуществляет мошенник

Добавили признаки:

- Максимальное количество кликов от всех уникальных IP
- Отношение числа кликов к числу уникальных IP

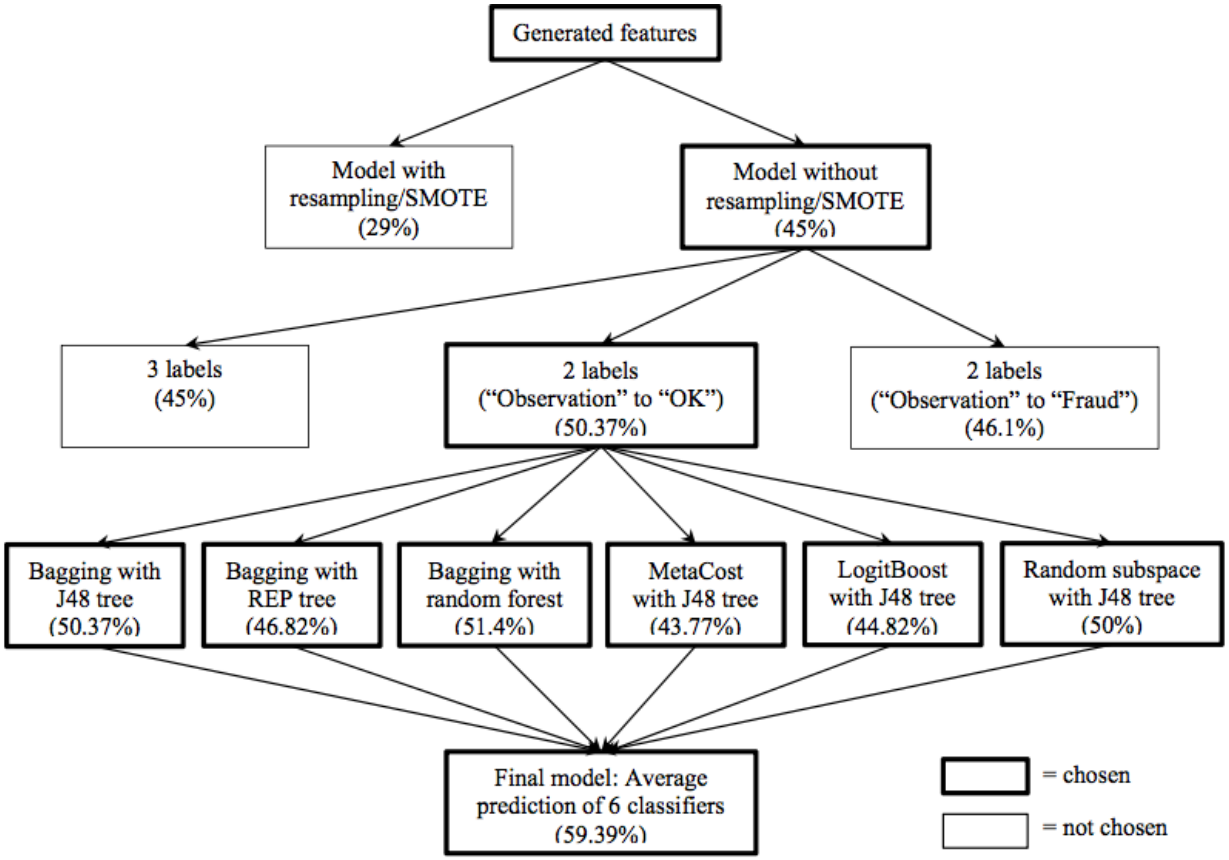
Часто мошенники используют одну и ту же модель телефона

Метод

- Создали три набора данных, заменяя Observation на Ok, Fraud и обучались на всех трех
- Из-за того, что количество мошеннических площадок много меньше честных, пылались убирать из данных хорошие площадки и дублировать плохие, чтобы их стало примерно равное число.
  Но это не дало хорошего результата на тестовой выборке

Метод

Финальный метод



Особенности в данных

Предполагают, что мошенники будут посещать с одного и того же IP
Добавляют признаки:
- avg_IP_sec - среднее число визитов с этого IP в секунду;
- std_IP_sec - стандартное отклонение среднего числа визитов с этого IP в секунду;
- count_IP_sec - суммарное число визитов с этого IP за секунду;
Аналогично вычислили для каждой минуты, часа и дня
Проделали то же самое для одной и той же подсети

Особенности в данных

Заметили, что мошенники могут использовать различные IP, но кликать с одного мобильного устройства

id	numericip	deviceua	campaignid	usercountry	clicktime	channel	referredurl
8jk0d	1,917,852,952	MSIE_6.0	8gp6q	cn	2012-02-11 02:55:50.0	se	?
8jk0d	1,917,853,022	MSIE_6.0	8gp6q	cn	2012-02-11 02:56:36.0	se	?
8jk0d	1,917,853,060	MSIE_6.0	8gp6q	cn	2012-02-11 03:53:12.0	se	?
8jk0d	1,917,852,993	MSIE_6.0	8gp6q	cn	2012-02-11 04:49:42.0	se	?
8jk0d	701,380,683	Nokia2600c	8k7xb	ng	2012-02-11 04:51:58.0	se	?
8jk0d	1,917,852,993	MSIE_6.0	8gp6q	cn	2012-02-11 05:33:51.0	se	?
8jk0d	1,917,853,114	MSIE_6.0	8gp6q	cn	2012-02-11 06:30:02.0	se	?
8jk0d	1,917,853,146	MSIE_6.0	8gp6q	cn	2012-02-11 07:09:23.0	se	?
8jk0d	1,917,853,146	MSIE_6.0	8gp6q	cn	2012-02-11 07:31:21.0	se	?
8jk0d	1,917,852,993	MSIE_6.0	8gp6q	cn	2012-02-11 07:53:16.0	se	?
8jk0d	1,917,852,952	MSIE_6.0	8gp6q	cn	2012-02-11 07:55:14.0	se	?

Признаки

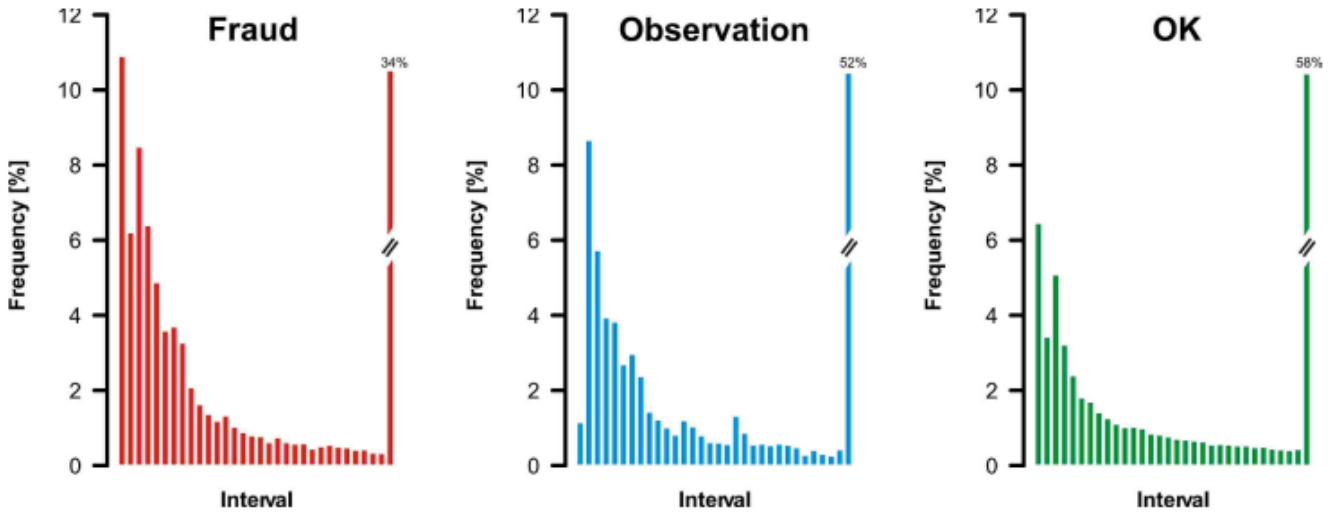
Feature	Description
unique_count(referredurl)	Unique count of referredurl
unique_count(campaignid)	Unique count of campaignid
unique_count(country)	Unique count of country
total_visit	Count of the click log's row
count_ip_hour	Sum of visit count (> 2) by numericip per hour
count_ip_ag_sec	Sum of visit count (> 2) by numericip+deviceua per second
count_ip_ag_day	Sum of visit count (> 2) by numericip+deviceua per day
count_sip2_sec	Sum of visit count (> 2) by subnetwork (divided by 1,000,000) per second
count_sip2_min	Sum of visit count (> 2) by subnetwork (divided by 1,000,000) per minute
count_sip2_hour	Sum of visit count (> 2) by subnetwork (divided by 1,000,000) per hour
count_sip_day	Sum of visit count (> 2) by subnetwork (divided by 1,000) per day
avg_sip2_day	Average visit by subnetwork (divided by 1,000,000) per day
avg_ip_ag_min	Average visit by numericip+deviceua per minute
avg_ip_ag_day	Average visit by numericip+deviceua per day
avg_campaignid_min	Average visit by campaignid per minute
deviceua1	Statistics for click data sorted by time and deviceua, as discussed in Section 5.1.3
deviceua2	Statistics for click data sorted by deviceua, as discussed in Section 5.1.3

Методы

Type	Method	Average Precision
Single	FT tree	36.3%
	REP tree	35.8%
	Bayes network	33.7%
	RPROP	48.3%
Ensemble	LAD tree	37.0%
	NB tree	37.9%
	Random forest	47.7%
	Random subspace	38.9%
	Rotation forest	42.9%
	Tree ensemble	49.3%
Ensemble of ensemble	Blending	52.3%

Особенности в данных

Заметили, что при многократные клики с одного IP - мошеннические



(a)

Для каждой площадки: проверили, если на нее кликнули хотя бы 5 раз с одного IP адреса с промежутком менее 20 секунд, то увеличиваем переменную redflag.

Метод

Рассмотрели две модели:
Для первой использовали основные признаки
+ многократные клики с одного IP
Обучали с помощью random forests with up- and down-sampling

#	ntree	nodesize	Fraud	Observation	OK	OOB error
1	250	5	90%	63%	41%	4.64%
2	250	5	90%	63%	41%	4.67%
3	250	3	83%	75%	51%	4.64%
4	250	3	69%	38%	34%	4.48%
5	250	3	69%	38%	34%	4.74%
6	250	3	90%	44%	51%	4.54%
7	250	4	83%	63%	68%	4.45%

Из всех моделей выбрали семь, а затем объединили их в одну

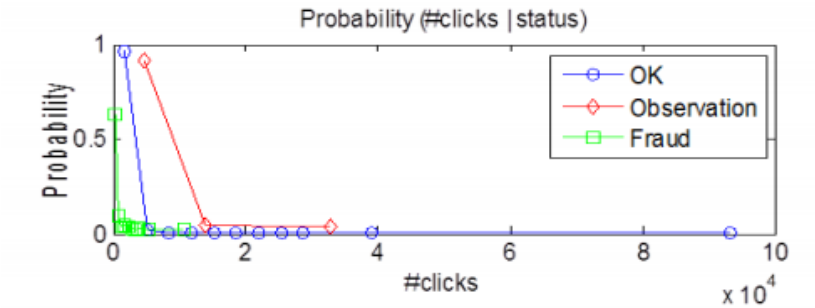
Метод

Для второй модели использовали основные признаки
+ многократные клики с одного URL
+ многократные клики с одного IP
+ redflag
Обучали с помощью случайных лесов.
В итоге данные модели показали следующие результаты
model #1 (ensemble of random forests, 1750 trees)
model #2 (single random forest, 50 trees)

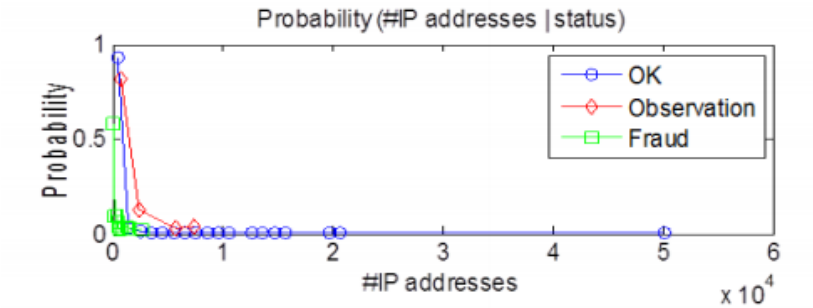
Model #	OOB error	Average precision (validation)	Average precision (test)
1	4.61%	49.99%	42.01%
2	3.66%	51.55%	36.94%

Особенности в данных

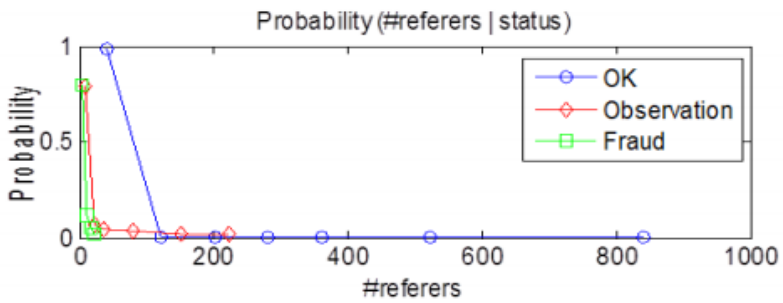
Для каждого издателя вычислили:
- вероятность числа кликов
- число посетителей
- количество referredurls
- отношение числа кликов к количеству посетителей.

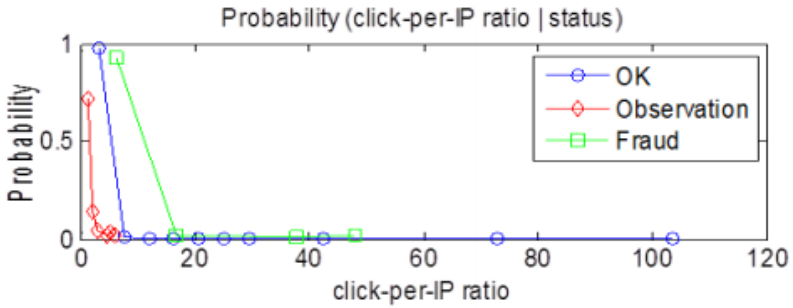


(a)



(b)





Особенности в данных

Заметили, что:
- в некоторых странах процент мошеннических кликов больше чем в остальных
- отсутствие признака referredurl чаще всего у мошенников
- использовали различные временные промежутки
(например, количесво кликов в каждую минуту и. т. д.)
В итоге получили 257 признаков, из которых потом удаляли неинформативные.

Метод

		Full features		Reduced features		
Type	Method	Validation AP	Test AP	#features	Validation AP	Test AP
Single	Logistic regression	41.20%	29.65%	46 / 257	46.02%	31.18%
	SVM (Linear)	30.45%	21.89%	38 / 257	36.75%	26.91%
	SVM (Polynomial)	22.69%	16.72%	256 / 257	28.42%	20.23%
	SVM (Radial basis)	34.32%	23.38%	255 / 257	39.10%	23.66%
	k-NN	28.78%	33.46%	257 / 257	28.78%	33.46%
Ensemble	Random forest	57.53%	51.44%	59 / 257	58.84%	52.17%
	GTB	48.78%	49.25%	235 / 257	58.33%	49.90%
	Extra trees	55.36%	54.04%	118 / 257	57.79%	55.64%