

Применение машинного обучения
по ансамблю решающих правил
для вычисления прогноза
дополнительного КИН

Ф.В.Краснов, Н.Г.Главнов, А.Н.Ситников



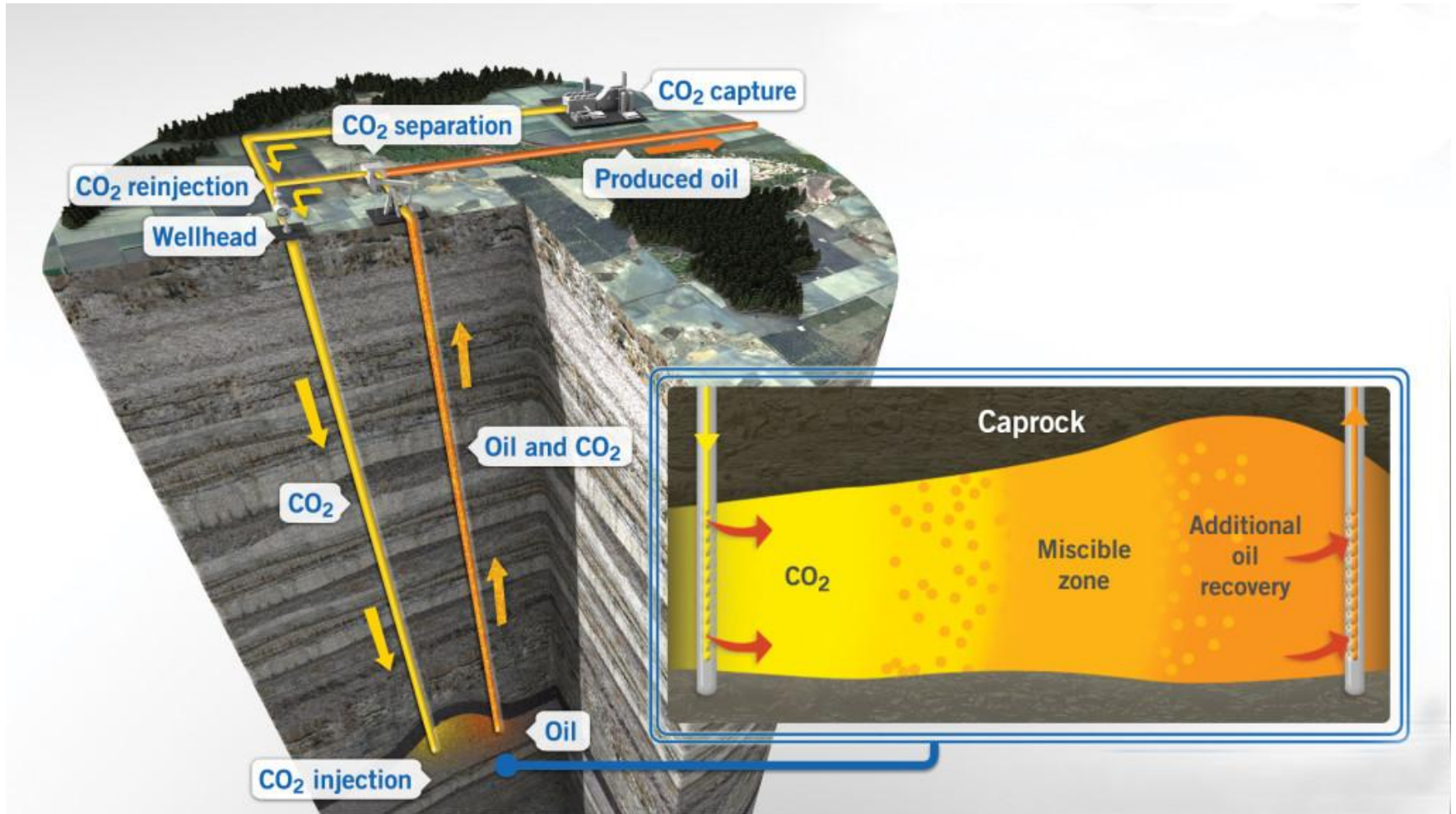
Нецифровая отрасль.



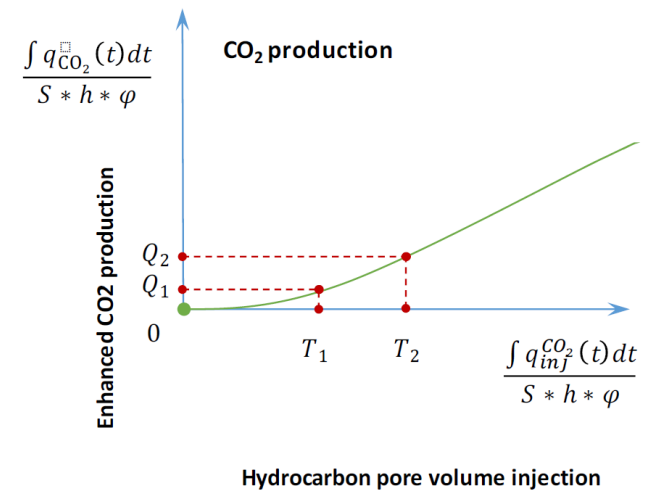
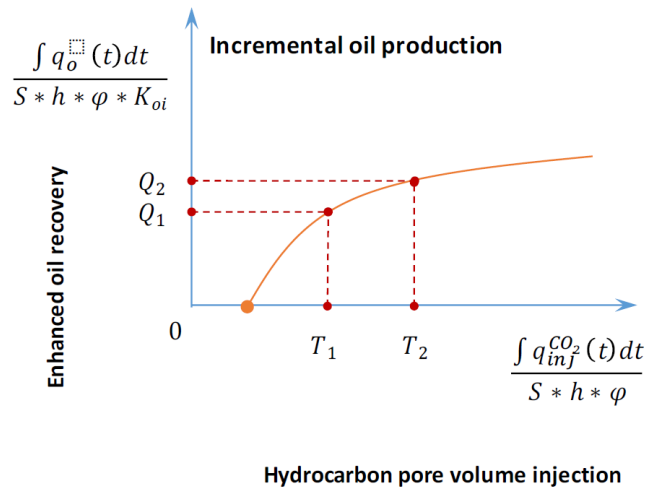
Трудноизвлекаемые запасы.



Геофизика. Природа моделируемого явления.



Геофизика. Моделируемые зависимости.



Применение МУН приводит к постепенному росту добычи, поэтому КИН зависит от времени (прокачанного порового объема).

Для того, чтобы рассчитать КИН для конкретного месторождения используют Гидродинамические симуляторы, на которых по 3D моделям распределения пористости, проницаемости и других параметров среды решают дифференциальные уравнения в частных производных для движения жидкости и газа в пористой среде. Это сложные уравнения и для их численного решения используют метод конечных элементов. Но и это не предмет моего доклада.

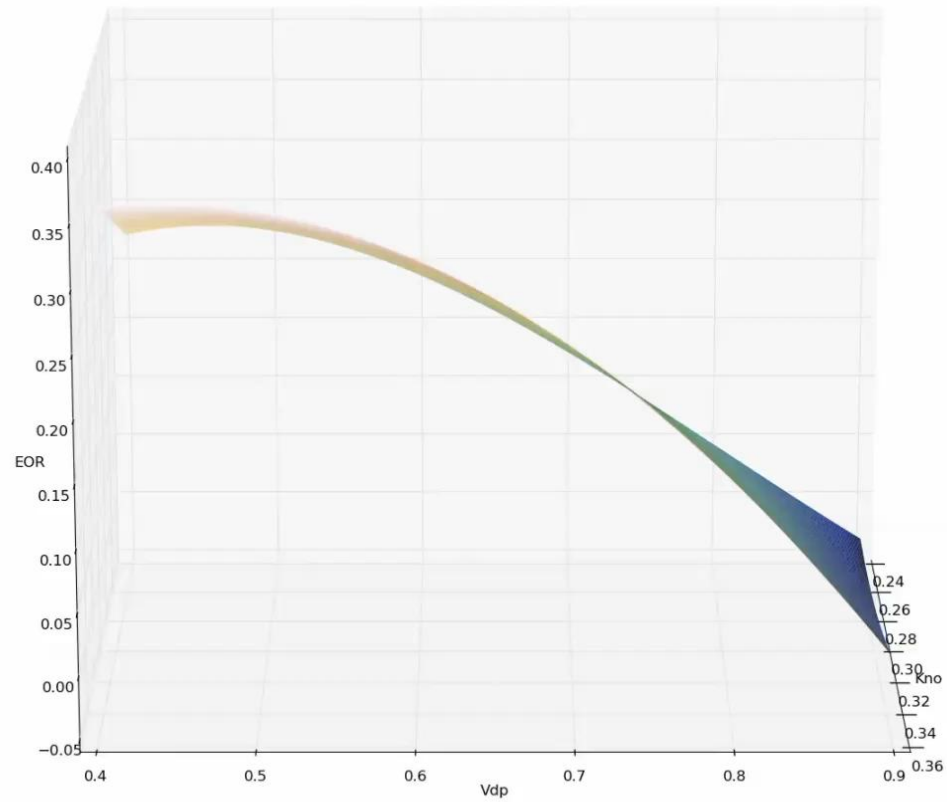
Многомерная полилинейная-интерполяция.

Таблица VIII. СИНУСЫ

А	0'	6'	12'	18'	24'	30'	36'	42'	48'	54'	60'		1'	2'	3'
0°	0,0000	0017	0035	0052	0070	0087	0105	0122	0140	0157	0,0000	90°	3	6	9
1°	0175	0192	0209	0227	0244	0262	0279	0297	0314	0332	0349	89°	3	6	9
2°	0349	0366	0384	0401	0419	0436	0454	0471	0488	0506	0523	87°	3	6	9
3°	0523	0541	0558	0576	0593	0610	0628	0645	0663	0680	0698	86°	3	6	9
4°	0698	0715	0732	0750	0767	0785	0802	0819	0837	0854	0,0872	85°	3	6	9
5°	0,0872	0889	0906	0924	0941	0958	0976	0993	1011	1028	1045	84°	3	6	9
6°	1045	1063	1080	1097	1115	1132	1149	1167	1184	1201	1219	83°	3	6	9
7°	1219	1236	1253	1271	1288	1305	1323	1340	1357	1374	1392	82°	3	6	9
8°	1392	1409	1426	1444	1461	1478	1495	1513	1530	1547	1564	81°	3	6	9
9°	1564	1582	1599	1616	1633	1650	1668	1685	1702	1719	0,1736	80°	3	6	9
10°	0,1736	1754	1771	1788	1805	1822	1840	1857	1874	1891	1908	79°	3	6	9
11°	1908	1925	1942	1959	1977	1994	2011	2028	2045	2062	2079	78°	3	6	9
12°	2079	2096	2113	2130	2147	2164	2181	2198	2215	2233	2250	77°	3	6	9
13°	2250	2267	2284	2300	2317	2334	2351	2368	2385	2402	2419	76°	3	6	8
14°	2419	2436	2453	2470	2487	2504	2521	2538	2554	2571	0,2588	75°	3	6	8
15°	0,2588	2605	2622	2639	2656	2672	2689	2706	2723	2740	2756	74°	3	6	8
16°	2756	2773	2790	2807	2823	2840	2857	2874	2890	2907	2924	73°	3	6	8
17°	2924	2940	2957	2974	2990	3007	3024	3040	3057	3074	3090	72°	3	6	8
18°	3090	3107	3123	3140	3156	3173	3190	3206	3223	3239	3256	71°	3	6	8
19°	3256	3272	3289	3305	3322	3338	3355	3371	3387	3404	0,3420	70°	3	5	8
20°	0,3420	3437	3453	3469	3486	3502	3518	3535	3551	3567	3584	69°	3	5	8
21°	3584	3600	3616	3633	3649	3665	3681	3697	3714	3730	3746	68°	3	5	8
22°	3746	3762	3778	3795	3811	3827	3843	3859	3875	3891	3907	67°	3	5	8
23°	3907	3923	3939	3955	3971	3987	4003	4019	4035	4051	4067	66°	3	5	8
24°	4067	4083	4099	4115	4131	4147	4163	4179	4195	4210	0,4226	65°	3	5	8
25°	0,4226	4242	4258	4274	4289	4305	4321	4337	4352	4368	4384	64°	3	5	8
26°	4384	4399	4415	4431	4446	4462	4478	4493	4509	4524	4540	63°	3	5	8
27°	4540	4555	4571	4586	4602	4617	4633	4648	4664	4679	4695	62°	3	5	8
28°	4695	4710	4726	4741	4756	4772	4787	4802	4818	4833	4848	61°	3	5	8
29°	4848	4863	4879	4894	4909	4924	4939	4955	4970	4985	0,5000	60°	3	5	8
30°	0,5000	5015	5030	5045	5060	5075	5090	5105	5120	5135	5150	59°	3	5	8
31°	5150	5165	5180	5195	5210	5225	5240	5255	5270	5284	5299	58°	2	5	7
32°	5299	5314	5329	5344	5358	5373	5388	5402	5417	5432	5446	57°	2	5	7
33°	5446	5461	5476	5490	5505	5519	5534	5548	5563	5577	5592	56°	2	5	7
34°	5592	5606	5621	5635	5650	5664	5678	5693	5707	5721	0,5736	55°	2	5	7
	60'	54'	48'	42'	36'	30'	24'	18'	12'	6'	0'	А	1'	2'	3'

Суррогатная модель.

$Pvt=1.00$, $KwKg=0.05$, $Knn=0.40$, $Hcpv=145$



Random Forest Regressor vs. Многомерная полилинейная-интерполяция

Особенности RFR :

1. Оценка чувствительности в RFR встроена
2. Увеличение детальности по параметру не приводит к степенному росту вычислительных ресурсов ("проклятие размерности". Размерность пространства параметров (фич) может быть большим
3. Требуется меньше памяти (RAM)
4. Оценка точности (по R^2)

Особенности МПИ:

1. Работает на равномерной сетке (кубе), поэтому есть требования к расчету полного набора узловых значений. Если в каком-то "узле" КИН не определен, то МПИ не работает.
2. Оценивать чувствительность необходимо отдельно
3. Точность не оценивается
4. Требования к памяти (RAM) растут при уменьшении "шага" (увеличении детальности) по параметру.
5. Размерность пространства параметров ограничена

ЗАКЛЮЧЕНИЕ

1. Методы машинного обучения используются в НГ отрасли для построения суррогатных моделей и моделирования КИН.
2. Необходимо исследовать возможность перехода интегральных фич к более простым. Это приведет к увеличению количества фич, но RFR должен с этим справиться.
3. Необходимо будет рассмотреть задачу снижения размерности параметров.
4. Наверное необходимо будет исследовать и другие методы, например, Boosted trees и CatBoost.