

ЛЕКЦИЯ 14. ПОИСК НОВЫХ ТЕРМОЭЛЕКТРИЧЕСКИХ МАТЕРИАЛОВ. МЕТОДЫ ПОИСКА ДЕСКРИПТОРОВ.

В начале изучения новой темы рассмотрим пример использования символьной регрессии для нахождения дескрипторов, связанных с материалами – топологическими изоляторами. Данные материалы имеют необычные физические свойства: за счет топологии на их поверхности могут существовать состояния, стабилизируемые определенной электронной и атомной структурами. Данные состояния могут иметь уникальные свойства, и поэтому эти материалы в настоящее время активно исследуются, например, с целью использования в катализе.

На рис. 1 представлена работа Г. Као и соавторов, которые исследовали класс топологических изоляторов – двумерные системы с различными элементами.

• Проектирование топологических изоляторов (материалы для спинтроники, катализа, термоэлектриков)

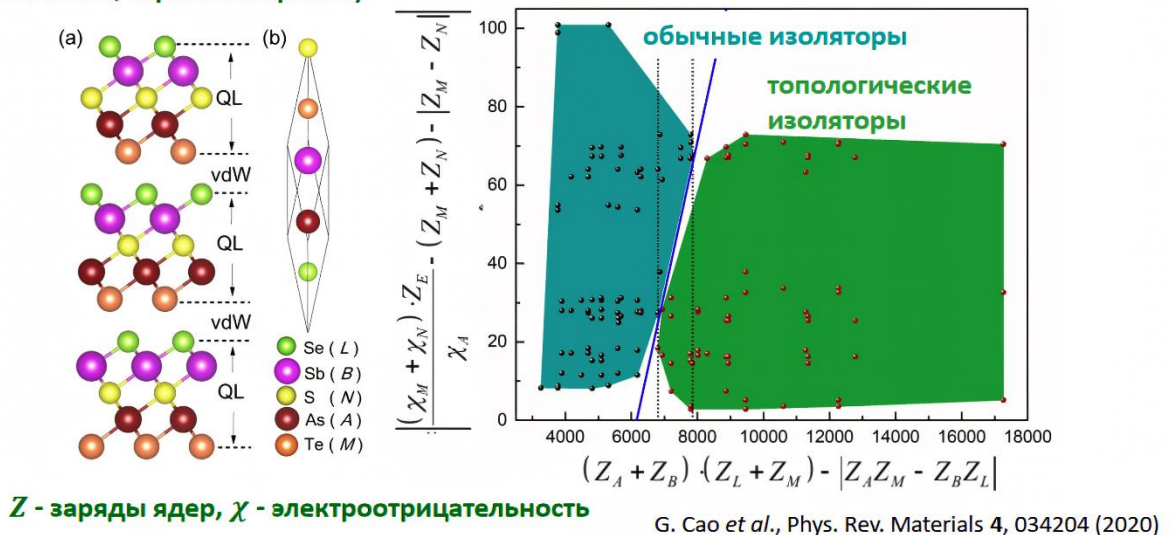


Рисунок 1. Проектирование топологических изоляторов

С помощью SISO были определены дескрипторы, которые, в основном, зависят от зарядов ядер и электроотрицательностей элементов. Карта, представленная на рис. 1, демонстрирует разделение с помощью этого

дескриптора на области, в которых живут топологические и обычные изоляторы.

Следующий, не менее интересный метод – это метод обнаружения подгрупп, который также называется Data Mining, что дословно означает «добыча данных». Идея использования данного метода следующая: имея дескрипторы – свойства материалов, мы имеем данные, на основе которых формируем глобальную модель. Однако данные нельзя описать линейно. Необходимо найти способ выделения подгрупп, и для каждой из них найти свою собственную модель (рис. 2).

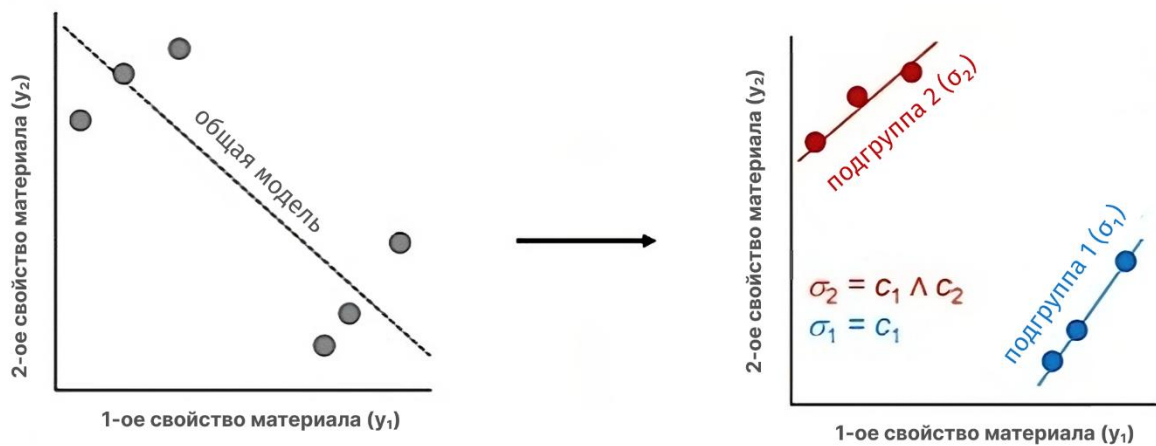


Рисунок 2. Модель обнаружение подгрупп (Data Mining)

Данный метод позволяет находить подгруппы, определяющиеся, например, в виде сочетаний типа «AND» («И»), утверждений типа «ширина запрещенной зоны $< 2\text{eV}$ » или «атомный радиус $> 1,4\text{\AA}$ » и др., т.е. в виде простых утверждений в сочетании с «И».

Алгоритм метода обнаружения подгрупп SGD (метод стохастического градиентного спуска (Stochastic Gradient Descent – SGD)) состоит в следующем: вводится функция, называемая функцией качества. Она состоит из нескольких множителей или компонентов. Первый компонент – это отношение числа элементов подгруппы к полному числу элементов, второй – разность между средним значением целевого свойства в подгруппе и средним значением целевого свойства для всего набора данных, третий – отношение

вариации целевого свойства в подгруппе к отношению вариации во всем наборе данных. Данная функция (формула 1) – это стандартный подход SGD. Она применяется в статистическом анализе Data Mining.

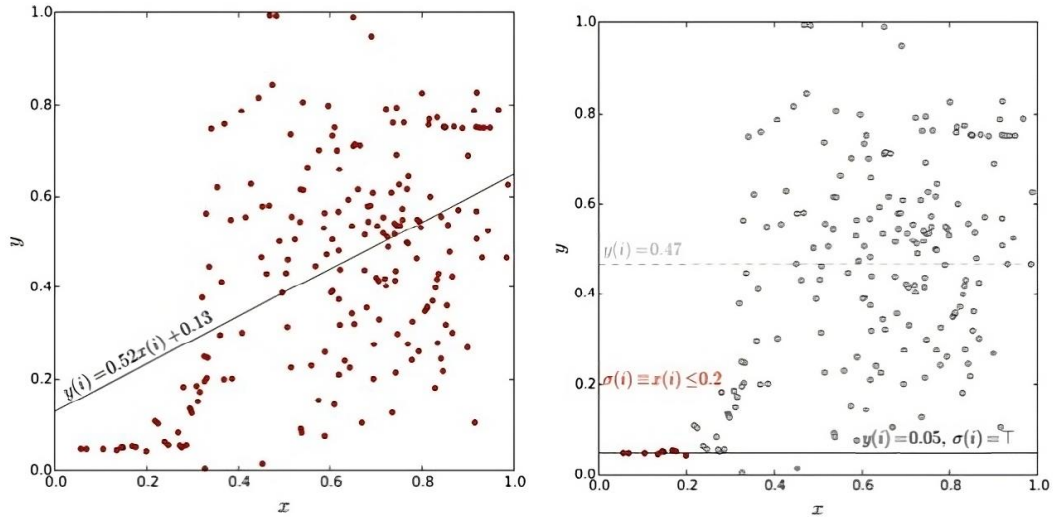
$$f = N_{\text{subgroup}}/N_{\text{all}} \times |\text{mean}_{\text{subgroup}} - \text{mean}_{\text{all}}| \times (1 - \text{variance}_{\text{subgroup}}/\text{variance}_{\text{all}}) \quad (1)$$

В целом, эта функция является простым способом найти подгруппу, которая используется для нахождения данных, объединенных определенными свойствами, отличающими их от общего набора.

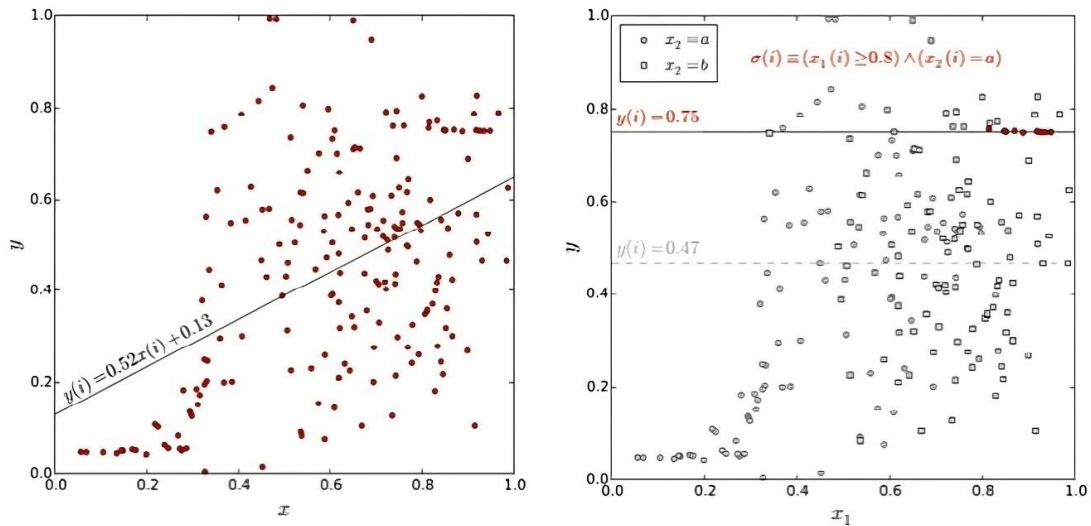
Числовые разделители в этих утверждениях (например, «атомный радиус > 1,4А») находятся с помощью метода машинного обучения, называемого кластеризацией по k-средним.

Непосредственно алгоритм SGD находит наиболее значимую подгруппу путем сравнения функций качества для каждого из сочетаний. Поиск подгруппы ведется либо методом Монте-Карло (статистический метод), либо алгоритмом ветвления и привязки.

В чем же заключается суть обнаружения подгрупп? М. Болей (M. Voley), приспособивший данный метод для материаловедения, продемонстрировал его действие в пространстве X и Y (рис. 3).



M. Boley *et al.*, Data Min. Knowl. Disc. 31, 1391 (2017); B. Goldsmith *et al.*, New J. Phys. 19, 013031 (2017)



M. Boley *et al.*, Data Min. Knowl. Disc. 31, 1391 (2017); B. Goldsmith *et al.*, New J. Phys. 19, 013031 (2017)

Рисунок 3. Метод обнаружение подгрупп в пространстве X и Y на примере работ М. Болея (M. Boley)

Допустим у нас есть два свойства - x и y , и линейная корреляция между ними такая, как показано на рисунке 3. Нам необходимо найти подгруппы в данном наборе данных.

Методом обнаружения подгрупп были обнаружены две подгруппы и обозначены условия к ним. В подгруппе, отображенной справа (где $y(i)=0,75$), обнаруживается второе измерение, то есть данные характеризуются еще

третьим дескриптором, который может быть в двух значениях – *a* или *b* (отображены на рисунке квадратиками или кружочками). В таком многомерном пространстве сложно выделить подгруппы. Алгоритм обнаружения подгрупп находит лишь сочетание ($x_1 > 0,8$, а $x_2 = a$). Т.е. с ростом числа переменных (дескрипторов), задача нахождения подгрупп становится все более сложной.

На рис. 4 представлен пример обнаружения подгрупп при поиске материалов для каталитической активации CO_2 .

Диоксид углерода выделяется в атмосферу при сгорании топлива и других материалов. Молекула CO_2 относительно стабильна, поэтому углекислый газ накапливается в атмосфере, что приводит к парниковому эффекту, вызывающему глобальное потепление.

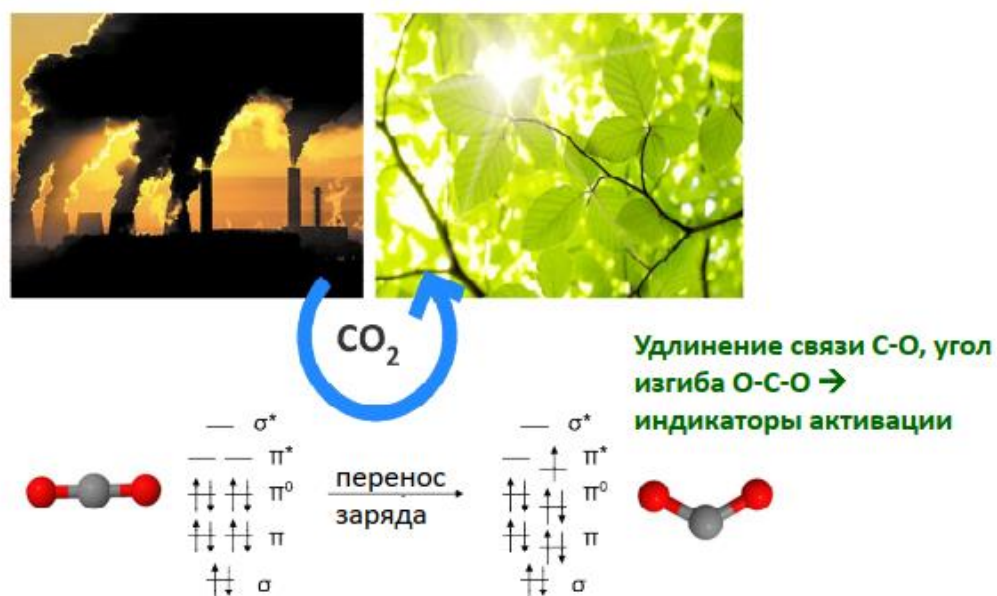


Рисунок 4. Молекулы CO_2 : электронная структура

Как преобразовать CO_2 ? Появилась идея гетерогенного катализа, заключающаяся в том, что CO_2 адсорбируются на поверхности и потом, путем химической реакции, преобразуется во что-нибудь полезное, например, обратно в топливо.

На рис. 4 представлена электронная структура молекулы CO_2 . Пришедший в эту молекулу электрон идет на разрыхляющую орбиталь π , что

приводит к ослаблению связей внутри молекулы. Это называется активацией – молекула подготавливается к дальнейшему преобразованию. Одновременно это приводит к изменению угла - изгибу. Именно поэтому удлинение связи С-О и изменение угла изгиба О-С-О рассматриваются как индикаторы активации.

На рисунке 5 отображены некоторые реакции с участием CO_2 . Например, сухая конверсия метана (в присутствии метана диоксид углерода можно преобразовать в СО и водород, данная смесь газов может использоваться для синтеза топлива), реакция Сабатье и частичное гидрирование CO_2 (например, для получения метанола).

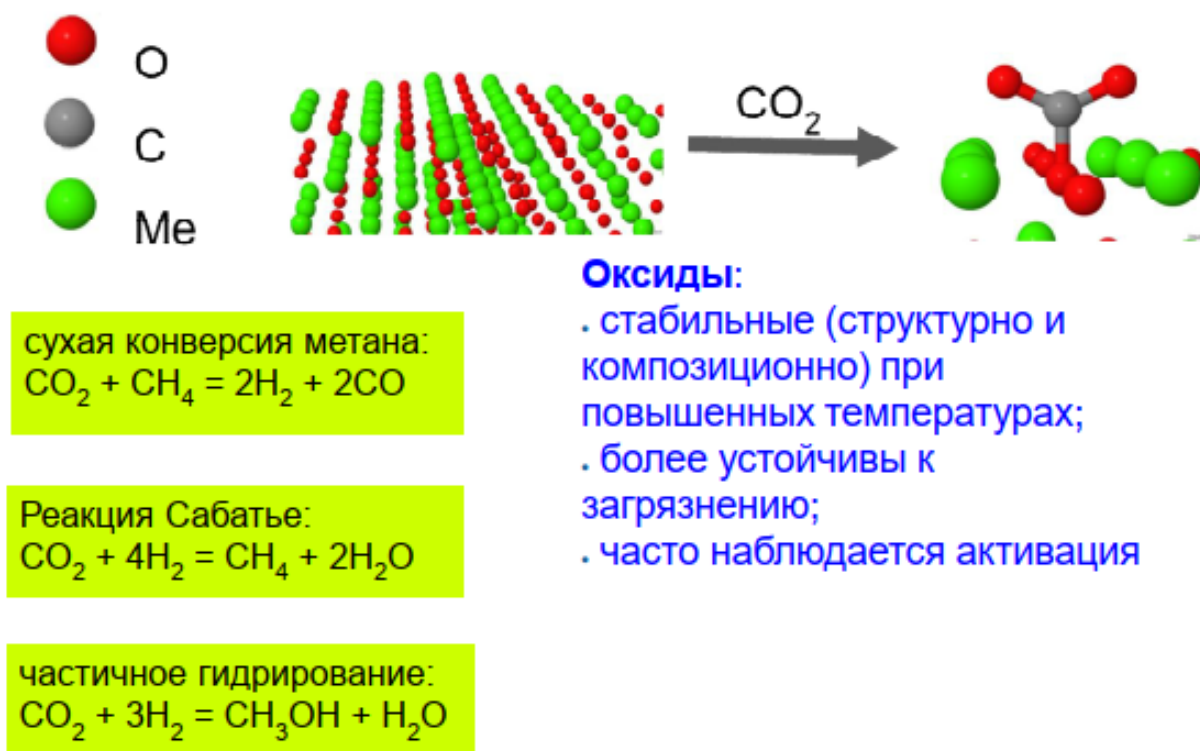
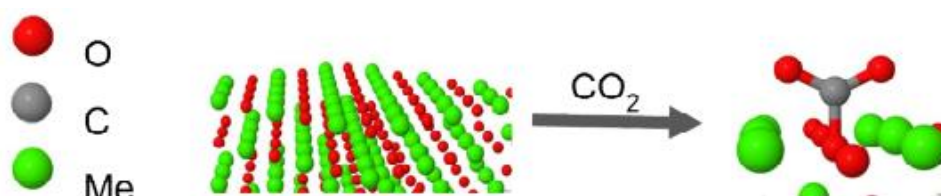


Рисунок 5. Примеры химических реакций с участием молекул CO_2

Существует множество типов материалов, которые могут участвовать в активации CO_2 , но это, в большинстве случаев, неэффективно или дорого. Наиболее широким классом, показывающим активность в реакциях с CO_2 , являются оксиды, которые при повышенных температурах стабильны структурно и композиционно.

Поверхности каких материалов обладают желаемыми значениями индикатора активации (молекула должна обладать длинной связью С-О (чем длиннее связь, тем она слабее), малым углом О-С-О (чем меньше угол, тем менее стабильна молекула, тем более она активирована))? Для решения данной задачи используется метод обнаружения подгрупп (рис. 6).



Удлинение связи С-О, угол О-С-О → индикаторы активации

Какие свойства поверхности приводят к желаемым значениям индикатора (длиннее С-О, меньше ОСО)?

Использование обнаружения подгрупп для поиска материалов, оптимизирующих показатели активации

$$f = \frac{N_{\text{subgroup}}}{N_{\text{all}}} \left(\frac{\text{mean}_{\text{subgroup}} - \text{mean}_{\text{all}}}{\text{min/max}_{\text{all}} - \text{mean}_{\text{all}}} \right) \left(1 - \frac{\text{variance}_{\text{subgroup}}}{\text{variance}_{\text{all}}} \right)$$

Максимизация длины связи С-О или изгиба О-С-О

Рисунок 6. Поиск материалов - активаторов CO_2 методом обнаружения подгрупп

В первой части функции (рис. 6) в знаменателе стоит разность между средним значением для подгруппы, характеризующейся минимумом угла/ максимумом длины связи и средним значением для всех материалов обучающих данных. Эта часть функции максимальна, когда значение длины связи СО в подгруппе как можно больше. Также, в формуле отражается вариация, которая не включает материалы с одной длинной связью С-О, а другой короткой, так как важно, чтобы все они были длинные. Если материал попадает в эту подгруппу, то он предсказывается как хороший материал активации CO_2 .

Изученные системы оксидов представлены на рис. 7.

$A^{2+}B^{4+}O_3, A^{3+}B^{3+}O_3, A^{1+}B^{5+}O_3, AO, BO_2, A_2O_3 (B_2O_3), A_2O, BO$

1 H 1.008	2 He 4.0026											13 Al 26.982	14 Si 28.085	15 P 30.974	16 S 32.06	17 Cl 35.45	18 Ar 39.948
3 Li 6.94	4 Be 9.0122											5 B 10.81	6 C 12.011	7 N 14.007	8 O 15.999	9 F 18.998	10 Ne 20.180
11 Na 22.990	12 Mg 24.305	3	4	5	6	7	8	9	10	11 Cu 63.546	12 Zn 65.38	13 Al 26.982	14 Si 28.085	15 P 30.974	16 S 32.06	17 Cl 35.45	18 Ar 39.948
19 K 39.098	20 Ca 40.078	21 Sc 44.956	22 Ti 47.867	23 V 50.942	24 Cr 51.996	25 Mn 54.938	26 Fe 55.845	27 Co 58.933	28 Ni 58.693	29 Cu 63.546	30 Zn 65.38	31 Ga 69.723	32 Ge 72.630	33 As 74.922	34 Se 78.97	35 Br 79.904	36 Kr 83.798
37 Rb 85.468	38 Sr 87.62	39 Y 88.906	40 Zr 91.224	41 Nb 92.906	42 Mo 95.95	43 Tc (98)	44 Ru 101.07	45 Rh 102.91	46 Pd 106.42	47 Ag 107.87	48 Cd 112.41	49 In 114.82	50 Sn 118.71	51 Sb 121.76	52 Te 127.60	53 I 126.90	54 Xe 131.29
55 Cs 132.91	56 Ba 137.33	57-71 *	72 Hf 178.49	73 Ta 180.95	74 W 183.84	75 Re 186.21	76 Os 190.23	77 Ir 192.22	78 Pt 195.08	79 Au 196.97	80 Hg 200.59	81 Tl 204.38	82 Pb 207.2	83 Bi 208.98	84 Po (209)	85 At (210)	86 Rn (222)
87 Fr (223)	88 Ra (226)	89-103 #	104 Rf (261)	105 Db (262)	106 Sg (263)	107 Bh (264)	108 Hs (265)	109 Mt (266)	110 Ds (267)	111 Rg (268)	112 Cn (269)	113 Nh (270)	114 Fl (271)	115 Mc (272)	116 Lv (273)	117 Ts (274)	118 Og (284)
* Lanthanide series			57 La 138.91	58 Ce 140.12	59 Pr 140.91	60 Nd 144.24	61 Pm (145)	62 Sm 150.36	63 Eu 151.96	64 Gd 157.25	65 Tb 158.93	66 Dy 162.50	67 Ho 164.93	68 Er 167.26	69 Tm 168.93	70 Yb 173.05	71 Lu 174.97
# Actinide series			89 Ac (227)	90 Th 232.04	91 Pa 231.04	92 U 238.03	93 Np (237)	94 Pu (244)	95 Am (243)	96 Cm (247)	97 Bk (247)	98 Cf (251)	99 Es (252)	100 Fm (257)	101 Md (258)	102 No (259)	103 Lr (262)

71 оксидов

141 поверхности с индексами Миллера ≤ 2

270 сайты адсорбции

Рисунок 7. Изученные системы оксидов- активаторов CO_2

Были исследованы 71 оксид, 141 поверхность (для каждого оксида разные поверхности с разными индексами Миллера), 270 сайтов (разные срезы, на каждом из них может быть больше одного неэквивалентного сайта адсорбции). Каждый сайт – это точка данных в обучающем наборе. Эти данные представлены на рис. 8 в координатах: угол O-C-O по оси X и длина связи C-O по оси Y. Все данные получены с помощью теории функционала плотности (это тренировочный набор). Красной линией представлено соотношение между длиной связи и углом O-C-O для заряженного газофазного CO_2 .

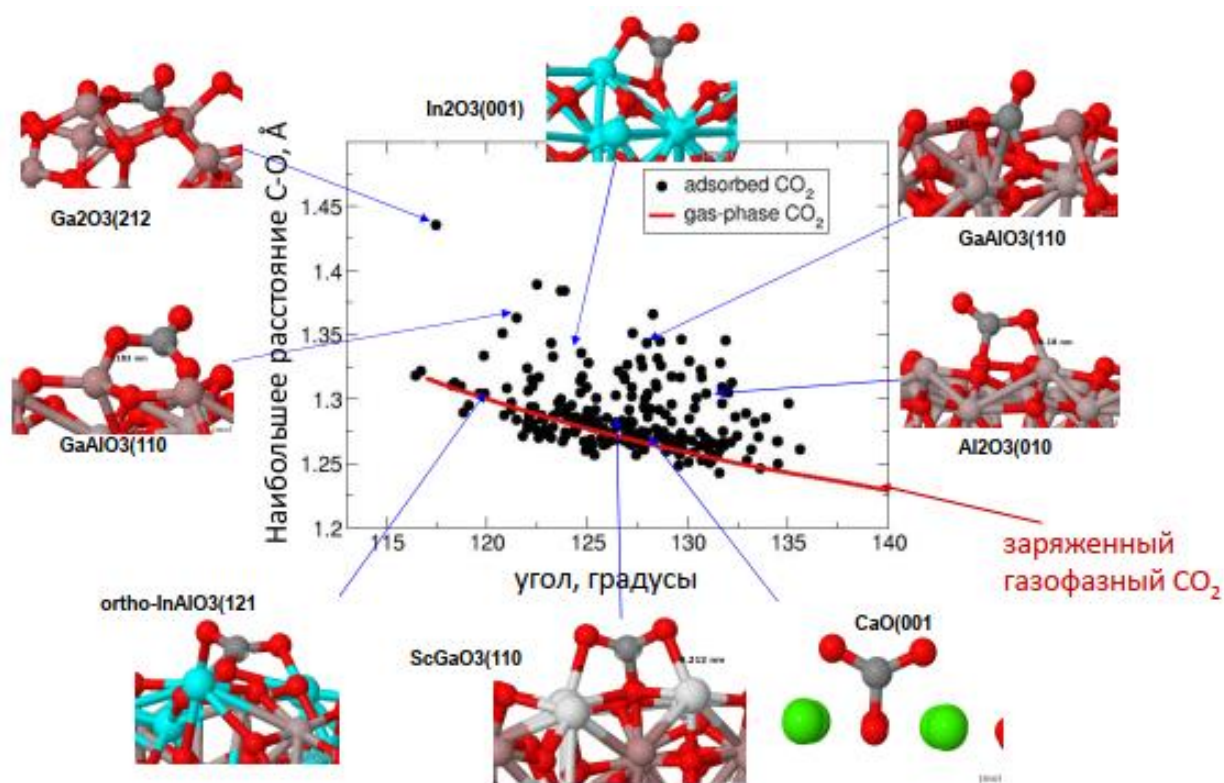


Рисунок 8. Обучающий набор данных в координатах угол (OCO) - длина связи ($C-O$)

Наблюдается, что для адсорбированных систем не все материалы следует тренду заряженного газофазного CO_2 . В основном, системы, в которых длина связи намного больше, чем ожидалось на основании угла, проявляют несимметричную адсорбцию, где CO_2 практически лежит на поверхности и образует CO_3 систему. Материалы, которые следуют тренду заряженного газофазного CO_2 , симметричны.

Какие величины первичных признаков ведут к наибольшей длине связи или наименьшему углу, и как это соотносится с каталитической активностью? Для ответа на этот вопрос необходимо сначала определить первичные признаки - следует рассмотреть свойства атомов: сродство к электрону, радиус орбиталей, потенциал ионизации, электроотрицательность и атомный номер. Также учитываются свойства объемного кристалла, такие как: работа выхода, ширина запрещенной зоны, энергия образования поверхности, а также сайт-

специфические свойства: электростатический потенциал, заряд по Хиршфельду, поляризуемость и др.

Синими точками (рис. 9) показаны материалы, входящие в подгруппу с минимальным углом; зелеными точками – входящие в подгруппу с максимальной длиной связи С-О; черными точками - остальные материалы. Подгруппа с минимальным углом следуют тренду заряженной газовой CO_2 ; подгруппа с длиной С-О связью отклоняется от этого тренда.

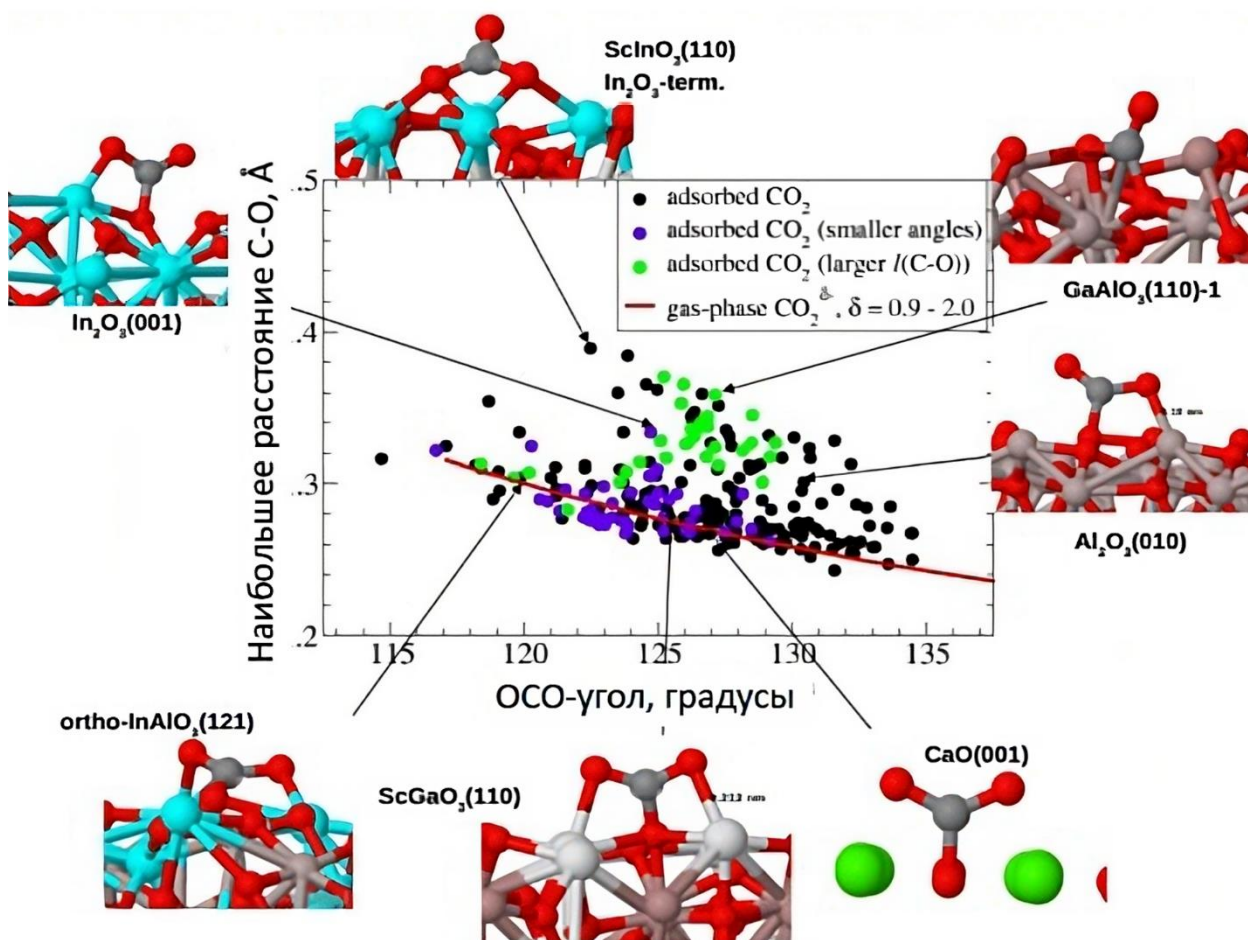


Рисунок 9. Обнаружение подгрупп с максимальной длиной связи (С-О) и минимальным углом (ОСО)

На рис. 10 представлена плотность энергии адсорбции, то есть число материалов с данной энергией адсорбции, для всего набора данных (черная линия) и для подгруппы с минимальным углом О-С-О (красная линия).

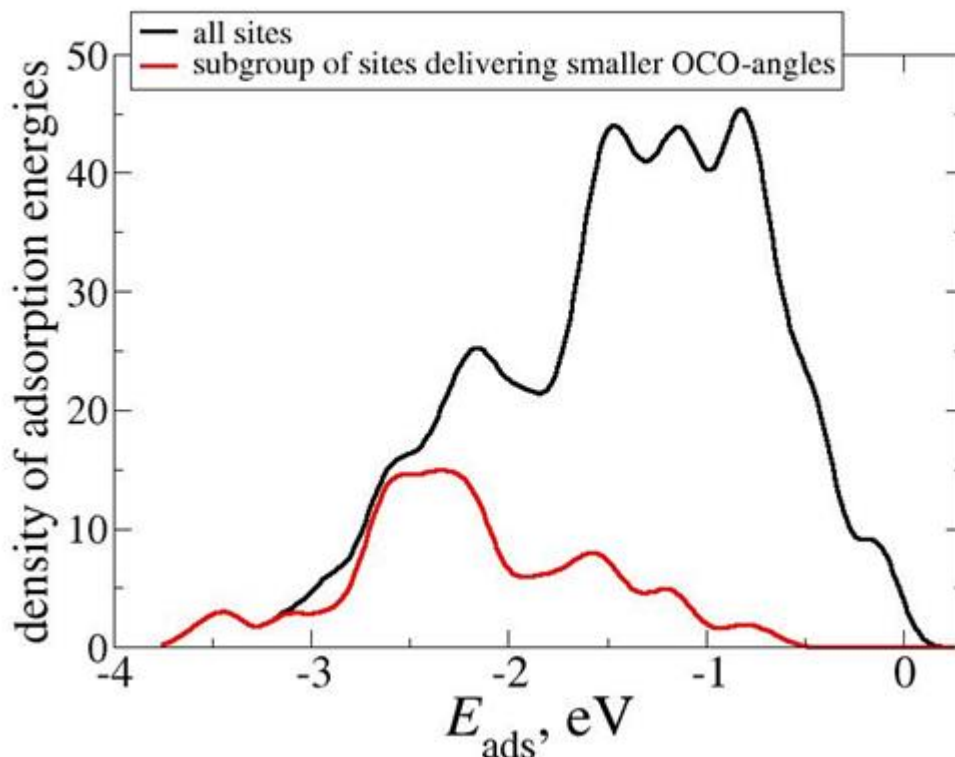


Рисунок 10. Плотность энергий адсорбции для полного набора данных и подгруппы с минимальными значениями угла OCO

Материалы подгруппы с минимальным углом $O-C-O$ характеризуются наиболее отрицательными значениями энергии адсорбции. Они могут быть определены следующими условиями:

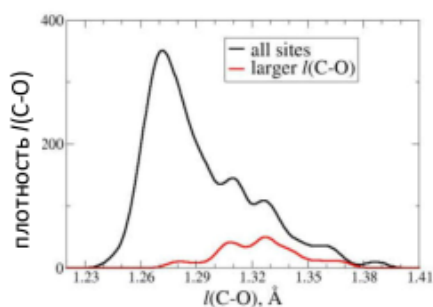
- Энергия максимума $O-2p$ зоны $>-6,0 eV$;
- Расстояние от O -сайта до первого ближайшего катиона $>1,8\text{\AA}$;
- Расстояние от O -сайта до второго ближайшего катиона $>2,1\text{\AA}$.

Данные условия можно интерпретировать следующим образом: на кислороде присутствует больше электронной плотности (с катионов больше электронной плотности перенеслось на кислороды), и эта электронная плотность доступна для переноса на CO_2 (она активирует CO_2 , и это приводит к уменьшению угла).

Т.е. методом обнаружения подгрупп был найден механизм активации CO_2 путем переноса электронов.

В подгруппе с минимальным углом O-C-O большинство сайтов, обеспечивающих малые углы, находятся на ионных материалах.

Далее рассмотрим подгруппу, с максимальной длиной C-O связи (рис. 11).



сайты, обеспечивающие большие $l(\text{C-O})$ (33 сайта):

(заряд катиона $< 0.5e$) AND
(работа выхода $\geq 5.2 \text{ eV}$) AND
(расстояние от O сайта до второго ближайшего катиона $\geq 2.14 \text{ \AA}$)

LaGaO_3 – катодный материал для высокотемпературного электрохимического восстановления CO_2 ;

KNbO_3 – фотокаталитическое восстановление CO_2 до CH_4 ;

NaNbO_3 – фотокатализатор восстановления CO_2 ;

NaSbO_3 – материал для улавливания и хранения CO_2

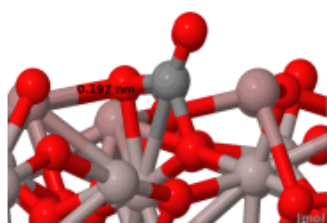


Рисунок 11. Характеристика подгруппы с максимальной длиной C-O связи

Данные можно интерпретировать следующим образом: материал неионный, присутствует ковалентная связь между ионом и атомом кислорода на поверхности. Получается, что CO_2 адсорбируется, на него переносится часть заряда. Также один из атомов кислорода может провзаимодействовать с поверхностью, что может привести к увеличению длины связи без увеличения угла. Многие материалы, известные как активаторы CO_2 , принадлежат к подгруппе с максимальной длиной C-O связи.

На рис. 12 представлено сравнение полученных данных.

Полному набору данных соответствует черная линия; подгруппе меньших углов – синяя линия, подгруппе с длинной связью C-O – зеленая линия.

Подгруппа с длинной связью C-O тоже демонстрирует маленькие углы, но большие, чем в подгруппе малых углов. Это приводит к тому, что энергия

адсорбции для этой подгруппы больше. Очень низкие значения энергии адсорбции говорят о том, что CO_2 сильно адсорбируется на поверхности (тяжело убрать даже высокими температурами), следовательно, поверхность загрязняется и перестает быть активной.

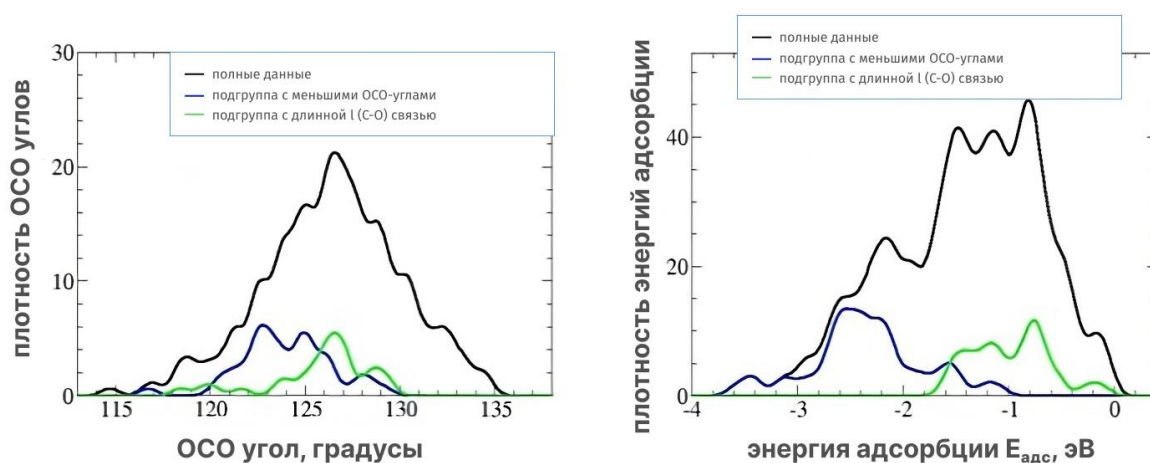


Рисунок 12. Плотности энергий адсорбции и ОСО-углов для полного набора данных и подгрупп с минимальными значениями угла ОСО/максимальными значениями длины С-О связи

В целевую функцию может быть добавлен принцип Сабатье - фактор, который пропорционален объему информации, получаемой при попадании материала в диапазон энергий, считающихся разумными (-2,5, -0,5). Желательно, чтобы материал подгруппы попадали в заданный предел, тогда максимальна информационная выгода (information gain).