Число сочетаний

Владимир Подольский

Факультет компьютерных наук, Высшая Школа Экономики

Число сочетаний

Неупорядоченные пары

Число сочетаний

Задача ранжирования

Задача ранжирования

Пусть у нас есть n документов и нам нужно отранжировать их по значимости. Один из стандартных подходов требует сравнения каждого текста с каждым из остальных. Сколько сравнений нам придется сделать в этом подходе?

Задача ранжирования

Попытка решения:

- У нас есть n документов

Задача ранжирования

Попытка решения:

- У нас есть n документов
- Нам нужно сравнить каждый документ с каждым из $n - 1$ остальных

Задача ранжирования

Попытка решения:

- У нас есть n документов
- Нам нужно сравнить каждый документ с каждым из $n-1$ остальных
- По правилу произведения у нас будет $n \times (n-1)$ сравнения

Задача ранжирования

Попытка решения:

- У нас есть n документов
- Нам нужно сравнить каждый документ с каждым из $n - 1$ остальных
- По правилу произведения у нас будет $n \times (n - 1)$ сравнения
- Все ли здесь правильно?

Задача ранжирования

Рассмотрим $n = 3$ и обозначим документы через a, b и c

- Наше решение дает $3 \times 2 = 6$ сравнений

Задача ранжирования

Рассмотрим $n = 3$ и обозначим документы через a, b и c

- Наше решение дает $3 \times 2 = 6$ сравнений
- В действительности их лишь 3: a vs. b, a vs. c и b vs. c

Задача ранжирования

Рассмотрим $n = 3$ и обозначим документы через a, b и c

- Наше решение дает $3 \times 2 = 6$ сравнений
- В действительности их лишь 3: a vs. b, a vs. c и b vs. c
- Посмотрим внимательнее на наше решение

Задача ранжирования

Рассмотрим $n = 3$ и обозначим документы через a, b и c

- Наше решение дает $3 \times 2 = 6$ сравнений
- В действительности их лишь 3: a vs. b, a vs. c и b vs. c
- Посмотрим внимательнее на наше решение
- Мы каждый текст сравниваем с двумя остальными

Задача ранжирования

Рассмотрим $n = 3$ и обозначим документы через a, b и c

- Наше решение дает $3 \times 2 = 6$ сравнений
- В действительности их лишь 3: a vs. b, a vs. c и b vs. c
- Посмотрим внимательнее на наше решение
- Мы каждый текст сравниваем с двумя остальными
- То есть, a сравнивается с b и c; b сравнивается с a и c; c сравнивается с a и b

Задача ранжирования

Рассмотрим $n = 3$ и обозначим документы через a, b и c

- Наше решение дает $3 \times 2 = 6$ сравнений
- В действительности их лишь 3: a vs. b, a vs. c и b vs. c
- Посмотрим внимательнее на наше решение
- Мы каждый текст сравниваем с двумя остальными
- То есть, a сравнивается с b и c; b сравнивается с a и c; c сравнивается с a и b
- Мы посчитали такие сравнения: ab, ac, ba, bc, ca, cb

Задача ранжирования

Рассмотрим $n = 3$ и обозначим документы через a, b и c

- Наше решение дает $3 \times 2 = 6$ сравнений
- В действительности их лишь 3: a vs. b, a vs. c и b vs. c
- Посмотрим внимательнее на наше решение
- Мы каждый текст сравниваем с двумя остальными
- То есть, a сравнивается с b и c; b сравнивается с a и c; c сравнивается с a и b
- Мы посчитали такие сравнения: ab, ac, ba, bc, ca, cb
- Мы посчитали каждое сравнение дважды!

Задача ранжирования

Теперь мы готовы исправить наше решение:

- У нас есть n документов

Задача ранжирования

Теперь мы готовы исправить наше решение:

- У нас есть n документов
- Нам нужно сравнить каждый с $n-1$ остальными

Задача ранжирования

Теперь мы готовы исправить наше решение:

- У нас есть n документов
- Нам нужно сравнить каждый с $n - 1$ остальными
- По правилу произведения у нас будет $n \times (n - 1)$ сравнение

Задача ранжирования

Теперь мы готовы исправить наше решение:

- У нас есть n документов
- Нам нужно сравнить каждый с $n-1$ остальными
- По правилу произведения у нас будет $n \times (n-1)$ сравнение
- Но мы посчитали каждое сравнение a vs. b дважды: как a vs. b и как b vs. a

Задача ранжирования

Теперь мы готовы исправить наше решение:

- У нас есть n документов
- Нам нужно сравнить каждый с $n-1$ остальными
- По правилу произведения у нас будет $n \times (n-1)$ сравнение
- Но мы посчитали каждое сравнение a vs. b дважды: как a vs. b и как b vs. a
- Давайте просто поделим результат на 2

Задача ранжирования

Теперь мы готовы исправить наше решение:

- У нас есть n документов
- Нам нужно сравнить каждый с $n-1$ остальными
- По правилу произведения у нас будет $n \times (n-1)$ сравнение
- Но мы посчитали каждое сравнение a vs. b дважды: как a vs. b и как b vs. a
- Давайте просто поделим результат на 2
- Получаем $n(n-1)/2$ сравнений

Что мы поняли

- Мы посчитали количество неупорядоченных пар из n объектов

Что мы поняли

- Мы посчитали количество неупорядоченных пар из n объектов
- Получилось $n(n-1)/2$ пар

Что мы поняли

- Мы посчитали количество неупорядоченных пар из n объектов
- Получилось $n(n-1)/2$ пар
- При подсчетах важно следить, что мы посчитали каждый объект только один раз

Что мы поняли

- Мы посчитали количество неупорядоченных пар из n объектов
- Получилось $n(n-1)/2$ пар
- При подсчетах важно следить, что мы посчитали каждый объект только один раз
- Если объект посчитан k раз, можно просто разделить результат на k

Число сочетаний

Неупорядоченные пары

Число сочетаний

Задача о поездке

Задача о поездке

Вы планируете автомобильную поездку. У Вас 5 друзей, но Вы можете посадить только 3 из них к себе в машину. Сколько есть различных способов это сделать?

Задача о поездке

Задача о поездке

Вы планируете автомобильную поездку. У Вас 5 друзей, но Вы можете посадить только 3 из них к себе в машину. Сколько есть различных способов это сделать?

- По существу мы хотим выбрать подмножество размера 3 в множестве из 5 друзей

Задача о поездке

Задача о поездке

Вы планируете автомобильную поездку. У Вас 5 друзей, но Вы можете посадить только 3 из них к себе в машину. Сколько есть различных способов это сделать?

- По существу мы хотим выбрать подмножество размера 3 в множестве из 5 друзей
- Мы хотим посчитать сколько всего есть различных подмножеств размера 3

Задача о поездке

Задача о поездке

Вы планируете автомобильную поездку. У Вас 5 друзей, но Вы можете посадить только 3 из них к себе в машину. Сколько есть различных способов это сделать?

- По существу мы хотим выбрать подмножество размера 3 в множестве из 5 друзей
- Мы хотим посчитать сколько всего есть различных подмножеств размера 3
- Напомним, что элементы подмножеств не упорядочены

Задача о поездке

Попытка решения:

- Мы можем выбрать первого друга 5 способами

Задача о поездке

Попытка решения:

- Мы можем выбрать первого друга 5 способами
- Мы можем выбрать второго друга 4 способами

Задача о поездке

Попытка решения:

- Мы можем выбрать первого друга 5 способами
- Мы можем выбрать второго друга 4 способами
- Мы можем выбрать третьего друга 3 способами

Задача о поездке

Попытка решения:

- Мы можем выбрать первого друга 5 способами
- Мы можем выбрать второго друга 4 способами
- Мы можем выбрать третьего друга 3 способами
- По правилу произведения получается
  $5 \times 4 \times 3 = 60$ вариантов

Задача о поездке

Попытка решения:

- Мы можем выбрать первого друга 5 способами
- Мы можем выбрать второго друга 4 способами
- Мы можем выбрать третьего друга 3 способами
- По правилу произведения получается $5 \times 4 \times 3 = 60$ вариантов
- Все ли тут правильно?

Задача о поездке

Попытка решения:

- Мы можем выбрать первого друга 5 способами
- Мы можем выбрать второго друга 4 способами
- Мы можем выбрать третьего друга 3 способами
- По правилу произведения получается $5 \times 4 \times 3 = 60$ вариантов
- Все ли тут правильно?
- Мы опять посчитали каждый вариант несколько раз

Задача о поездке

Исправление решения:

- Заметим, что мы посчитали упорядоченные последовательности друзей

Задача о поездке

Исправление решения:

- Заметим, что мы посчитали упорядоченные последовательности друзей
- Один из выбранных друзей первый, другой второй, и так далее

Задача о поездке

Исправление решения:

- Заметим, что мы посчитали упорядоченные последовательности друзей
- Один из выбранных друзей первый, другой второй, и так далее
- Каждая (неупорядоченная) группа друзей $\{a, b, c\}$ посчитана $3 \times 2 = 6$ раз: abc, acb, bac, bca, cab и cba

Задача о поездке

Исправление решения:

- Заметим, что мы посчитали упорядоченные последовательности друзей
- Один из выбранных друзей первый, другой второй, и так далее
- Каждая (неупорядоченная) группа друзей $\{a, b, c\}$ посчитана $3 \times 2 = 6$ раз: abc, acb, bac, bca, cab и cba
- Мы просто можем разделить прошлый результат на 6

Задача о поездке

Исправление решения:

- Заметим, что мы посчитали упорядоченные последовательности друзей
- Один из выбранных друзей первый, другой второй, и так далее
- Каждая (неупорядоченная) группа друзей $\{a, b, c\}$ посчитана $3 \times 2 = 6$ раз: abc, acb, bac, bca, cab и cba
- Мы просто можем разделить прошлый результат на 6
- В итоге получаем $5 \times 4 \times 3/(3 \times 2) = 10$ способов выбрать попутчиков

Задача о поездке

Друзья: a, b, c, d, e

Число способов выбрать друзей: X

abc	abd	abe	acd	ace	ade	bcd	bce	bde	cde
acb	adb	aeb	adc	aec	aed	bdc	bec	bed	ced
bac	bad	bae	cad	cae	dae	cbd	cbe	dbe	dce
bca	bda	bea	cda	cea	dea	cdb	ceb	deb	dec
cba	dba	eba	dca	eca	eda	dcb	ecb	edb	edc
cab	dab	eab	dac	eac	ead	dbc	ebc	ebd	ecd

Задача о поездке

Друзья: a, b, c, d, e

Число способов выбрать друзей: X

	abc	abd	abe	acd	ace	ade	bcd	bce	bde	cde
	acb	adb	aeb	adc	aec	aed	bdc	bec	bed	ced
	bac	bad	bae	cad	cae	dae	cbd	cbe	dbe	dce
$3!$	bca	bda	bea	cda	cea	dea	cdb	ceb	deb	dec
	cba	dba	eba	dca	eca	eda	dcb	ecb	edb	edc
	cab	dab	eab	dac	eac	ead	dbc	ebc	ebd	ecd

Задача о поездке

Друзья: a, b, c, d, e

Число способов выбрать друзей: X

X

$3!$

abc	abd	abe	acd	ace	ade	bcd	bce	bde	cde
acb	adb	aeb	adc	aec	aed	bdc	bec	bed	ced
bac	bad	bae	cad	cae	dae	cbd	cbe	dbe	dce
bca	bda	bea	cda	cea	dea	cdb	ceb	deb	dec
cba	dba	eba	dca	eca	eda	dcb	ecb	edb	edc
cab	dab	eab	dac	eac	ead	dbc	ebc	ebd	ecd

Задача о поездке

Друзья: a, b, c, d, e

Число способов выбрать друзей: X

X

$3!$

abc	abd	abe	acd	ace	ade	bcd	bce	bde	cde
acb	adb	aeb	adc	aec	aed	bdc	bec	bed	ced
bac	bad	bae	cad	cae	dae	cbd	cbe	dbe	dce
bca	bda	bea	cda	cea	dea	cdb	ceb	deb	dec
cba	dba	eba	dca	eca	eda	dcb	ecb	edb	edc
cab	dab	eab	dac	eac	ead	dbc	ebc	ebd	ecd

$$3! \cdot X = \frac{5!}{(5-3)!}$$

Число сочетаний

- Для множества S его k-сочетанием называется подмножество S размера k

Число сочетаний

- Для множества S его k-сочетанием называется подмножество S размера k
- Число k-сочетаний n-элементного множества обозначается через $\binom{n}{k}$ или C_n^k

Число сочетаний

- Для множества S его k-сочетанием называется подмножество S размера k
- Число k-сочетаний n-элементного множества обозначается через $\binom{n}{k}$ или C_n^k
- Читается как «число сочетаний из n по k» или «це из n по k»

Число сочетаний

- Для множества S его k-сочетанием называется подмножество S размера k
- Число k-сочетаний n-элементного множества обозначается через $\binom{n}{k}$ или C_n^k
- Читается как «число сочетаний из n по k» или «це из n по k»
- Мы доказали, что $\binom{5}{3} = 10$

Число сочетаний

Для числа k-сочетаний есть короткая формула:

Теорема

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Доказательство

- Нам нужно посчитать число подмножеств размера k в множестве размера n

Доказательство

- Нам нужно посчитать число подмножеств размера k в множестве размера n
- Первый элемент можно выбрать n способами, второй — $(n-1)$ способами, ..., k-ый элемент — $(n-k+1)$ способами

Доказательство

- Нам нужно посчитать число подмножеств размера k в множестве размера n
- Первый элемент можно выбрать n способами, второй — $(n-1)$ способами, …, k-ый элемент — $(n-k+1)$ способами
- Это дает $n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$ способов

Доказательство

- Нам нужно посчитать число подмножеств размера k в множестве размера n
- Первый элемент можно выбрать n способами, второй — $(n-1)$ способами, … , k-ый элемент — $(n-k+1)$ способами
- Это дает $n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$ способов
- Но мы посчитали число k-перестановок, а не количество k-сочетаний

Доказательство

- Другими словами, мы посчитали упорядоченные подмножества вместо неупорядоченных

Доказательство

- Другими словами, мы посчитали упорядоченные подмножества вместо неупорядоченных
- Каждое подмножество посчитано столько раз, сколькими способами можно его упорядочить

Доказательство

- Другими словами, мы посчитали упорядоченные подмножества вместо неупорядоченных
- Каждое подмножество посчитано столько раз, сколькими способами можно его упорядочить
- Есть $k!$ способов упорядочить k объектов (перестановки)

Доказательство

- Другими словами, мы посчитали упорядоченные подмножества вместо неупорядоченных
- Каждое подмножество посчитано столько раз, сколькими способами можно его упорядочить
- Есть $k!$ способов упорядочить k объектов (перестановки)
- Так что каждое подмножество у нас посчитано $k!$ раз

Доказательство

- Другими словами, мы посчитали упорядоченные подмножества вместо неупорядоченных
- Каждое подмножество посчитано столько раз, сколькими способами можно его упорядочить
- Есть $k!$ способов упорядочить k объектов (перестановки)
- Так что каждое подмножество у нас посчитано $k!$ раз
- В итоге получаем $\frac{n!}{(n-k)!} \times \frac{1}{k!} = \frac{n!}{k!(n-k)!}$

Сочетания

- Сочетания являются важным и часто встречающимся объектом

Сочетания

- Сочетания являются важным и часто встречающимся объектом
- Как они могут возникать в анализе данных? Давайте посмотрим на пример

Сочетания

- Сочетания являются важным и часто встречающимся объектом
- Как они могут возникать в анализе данных? Давайте посмотрим на пример
- Иногда нам нужно использовать слабую модель машинного обучения, которая в принципе не может дать хорошего предсказания на наших данных

Сочетания

- Сочетания являются важным и часто встречающимся объектом
- Как они могут возникать в анализе данных? Давайте посмотрим на пример
- Иногда нам нужно использовать слабую модель машинного обучения, которая в принципе не может дать хорошего предсказания на наших данных
- Возможный путь решения: добавление произведений признаков

Сочетания

- Пусть у нас есть 5 числовых признаков a, b, c, d и e на наших данных

Сочетания

- Пусть у нас есть 5 числовых признаков a, b, c, d и e на наших данных
- Для обогащения множества признаков мы можем добавить новые признаки, состоящие из произведений всевозможных троек из наших пяти признаков

Сочетания

- Пусть у нас есть 5 числовых признаков a, b, c, d и e на наших данных
- Для обогащения множества признаков мы можем добавить новые признаки, состоящие из произведений всевозможных троек из наших пяти признаков
- Например, мы добавляем $a \cdot b \cdot c$, $b \cdot c \cdot e$, $d \cdot a \cdot c$ и все остальные произведения троек

Сочетания

- Пусть у нас есть 5 числовых признаков a, b, c, d и e на наших данных
- Для обогащения множества признаков мы можем добавить новые признаки, состоящие из произведений всевозможных троек из наших пяти признаков
- Например, мы добавляем $a \cdot b \cdot c$, $b \cdot c \cdot e$, $d \cdot a \cdot c$ и все остальные произведения троек
- После этого мы можем запустить нашу слабую модель на новом множестве признаков

Сочетания

- Пусть у нас есть 5 числовых признаков a, b, c, d и e на наших данных
- Для обогащения множества признаков мы можем добавить новые признаки, состоящие из произведений всевозможных троек из наших пяти признаков
- Например, мы добавляем $a \cdot b \cdot c$, $b \cdot c \cdot e$, $d \cdot a \cdot c$ и все остальные произведения троек
- После этого мы можем запустить нашу слабую модель на новом множестве признаков
- Во многих случаях это дает значительно лучшие результаты

Сочетания

- Потенциальная проблема: число признаков выросло, эффективность работы может упасть

Сочетания

- Потенциальная проблема: число признаков выросло, эффективность работы может упасть
- Важно оценить сколько признаков мы добавили

Сочетания

- Потенциальная проблема: число признаков выросло, эффективность работы может упасть
- Важно оценить сколько признаков мы добавили
- По сути, каждый новый признак — это 3-сочетание изначального множества признаков

Что мы узнали за неделю?

- На этой неделе мы начали изучать подсчеты

Что мы узнали за неделю?

- На этой неделе мы начали изучать подсчеты
- Мы начали с очень простых наблюдений

Что мы узнали за неделю?

- На этой неделе мы начали изучать подсчеты
- Мы начали с очень простых наблюдений
- На их основе мы разобрались с более сложными конструкциями

Что мы узнали за неделю?

- На этой неделе мы начали изучать подсчеты
- Мы начали с очень простых наблюдений
- На их основе мы разобрались с более сложными конструкциями
- Мы дошли до уже достаточно непростой задачи о подсчете числа сочетаний

Что мы узнали за неделю?

- На этой неделе мы начали изучать подсчеты
- Мы начали с очень простых наблюдений
- На их основе мы разобрались с более сложными конструкциями
- Мы дошли до уже достаточно непростой задачи о подсчете числа сочетаний
- На следующей неделе мы продолжим с числами сочетаний и увидим, что они обладают интересными и важными свойствами

Что мы узнали за неделю?

- На этой неделе мы начали изучать подсчеты
- Мы начали с очень простых наблюдений
- На их основе мы разобрались с более сложными конструкциями
- Мы дошли до уже достаточно непростой задачи о подсчете числа сочетаний
- На следующей неделе мы продолжим с числами сочетаний и увидим, что они обладают интересными и важными свойствами
- Мы также обсудим еще одну стандартную постановку задачи комбинаторики