

Компьютерные шахматы, компьютерное го и нейронные сети

А.С. Трушечкин

Конференция «Декабрьские чтения»

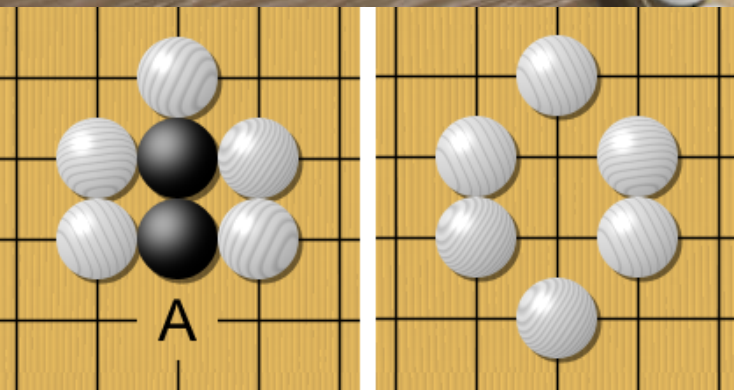
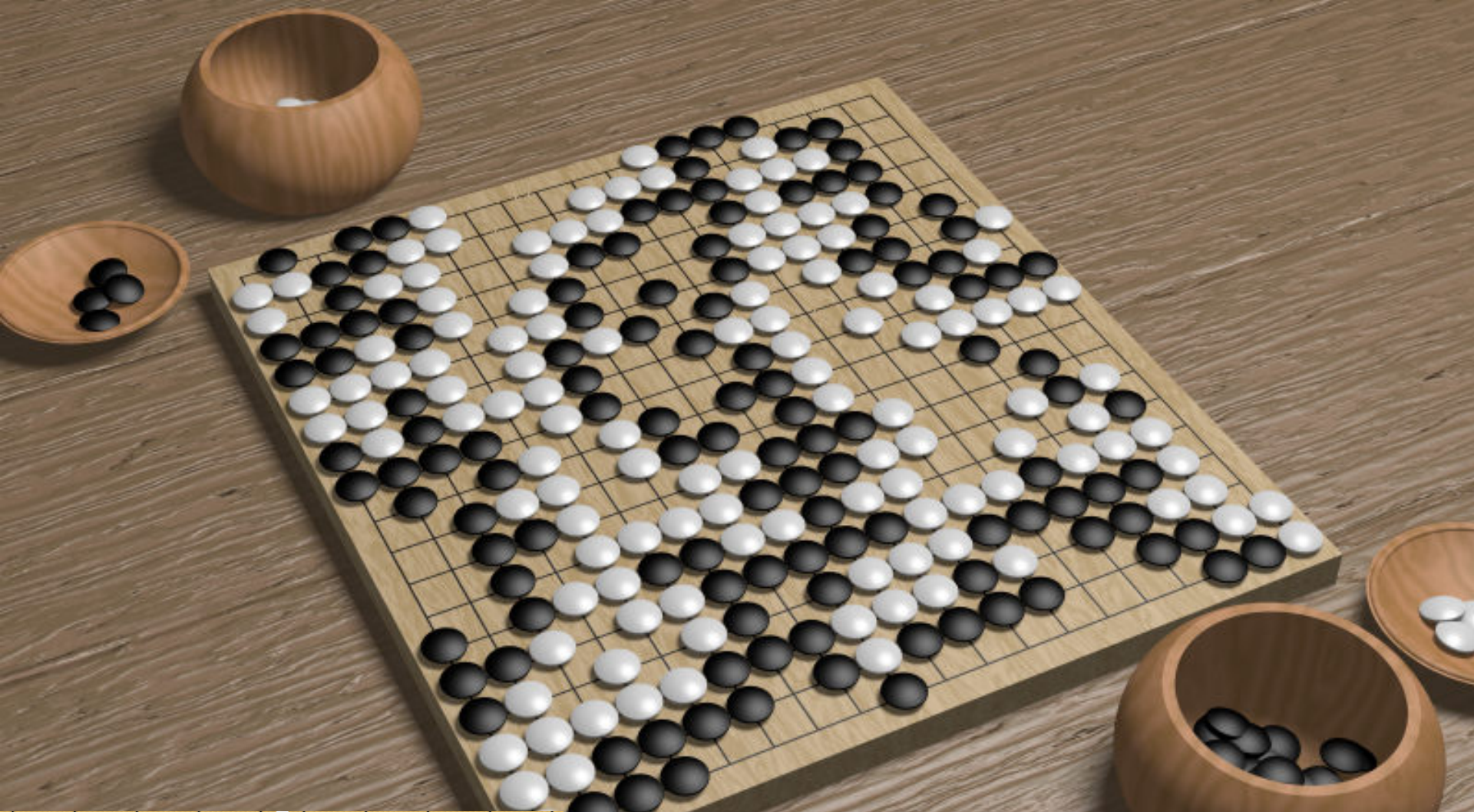
Новосибирск
22 декабря 2017 г.

- Октябрь 2015: победа AlphaGo (Google DeepMind, Лондон) в матче против трёхкратного чемпиона Европы Фань Хуэя (5:0)
- Март 2016: победа программы AlphaGo в матче по го против одного из сильнейших гоистов мира Ли Седоля (4:1)
- Январь 2017: «таинственный игрок» Master на онлайн-площадках: 60 побед из 60 против сильнейших игроков мира
- Май 2017: победа в матче против сильнейшего гоиста мира Кэ Цзе (3:0) и в партии против объединенной команды лучших китайских профессиональных игроков в го
- Октябрь 2017: полностью самообучаемая программа AlphaGo Zero превзошла все предыдущие версии

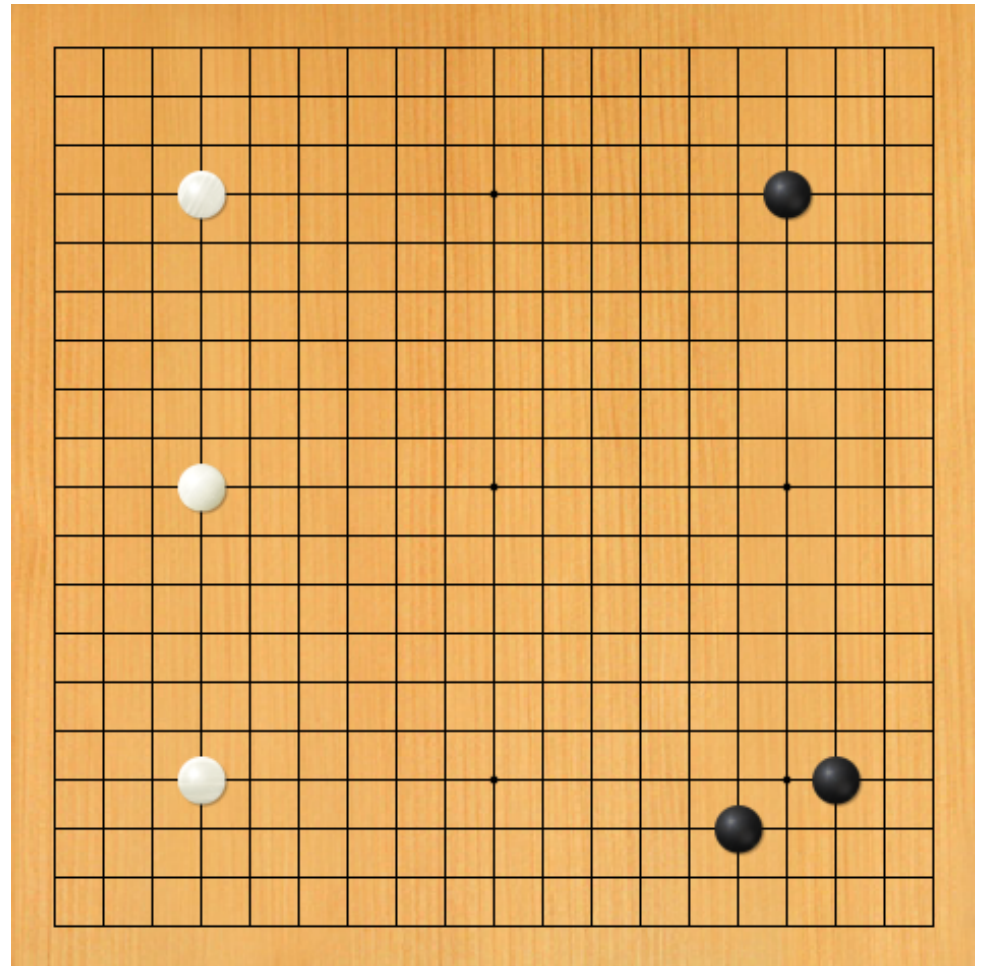
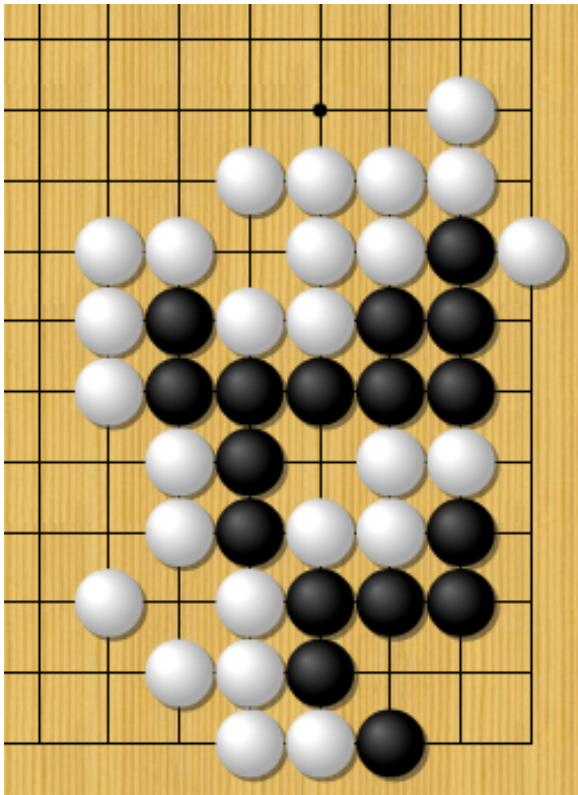
- 1997: победа компьютера DeepBlue (IBM) в матче по шахматам против Гарри Каспарова (3,5:2,5)
- Декабрь 2017: программа игры в шахматы AlphaZero после 4 часа обучения победила одну из сильнейших программ Stockfish (64:36), которая, в свою очередь, превосходила сильнейших шахматистов-людей.

Game	White	Black	Win	Draw	Loss
Chess	<i>AlphaZero</i>	<i>Stockfish</i>	25	25	0
	<i>Stockfish</i>	<i>AlphaZero</i>	3	47	0

Total games: w 242/353/5, b 48/533/19



Расчёт vs Интуиция



Александр Динерштейн, 7-кратный чемпион Европы по го:

«Для меня это остаётся большим вопросом – как будет действовать программа, если с первых же ходов свернуть с дебютных справочников. На пустой доске вариантов столько, что никаким методом Монте-Карло их не просчитать. В этом го выгодно отличается от шахмат. В шахматах всё давно изучено на глубину 20-30 ходов, а в го, при желании, уже первым ходом можно создать позицию, которая не встречалась в истории профессионального го. Программе придётся играть самостоятельно, а не вытаскивать варианты из базы знаний. Посмотрим, сможет ли она это сделать. Я в этом сильно сомневаюсь и ставлю на Ли Седоля».

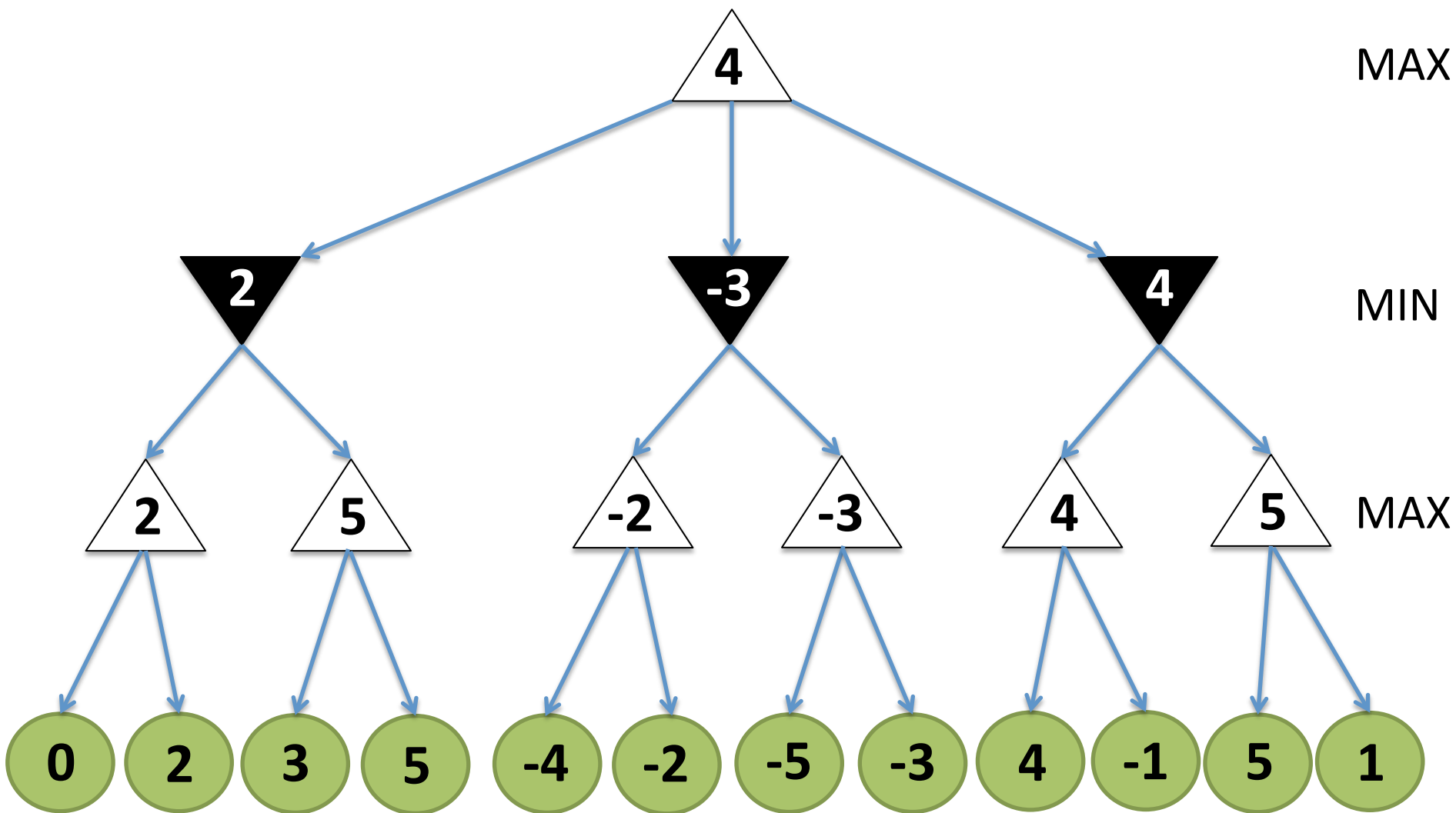
Стоя на плечах гигантов...

- Первая программа игры в го – 1960
- Непосредственно перед AlphaGo: программы на уровне мастеров-любителей
- Но был большой разрыв в силе игры с профессионалами
- А. Динерштейн: «Я был уверен, что у нас есть хотя бы 10 лет в запасе. Ещё пару месяцев назад мы играли с программами на форе в четыре камня – это примерно как фора в ладью в шахматах. И тут – бац! – и сразу Ли Седоль повержен».

Шашки, шахматы, го – это игры со следующими свойствами:

- С полной информацией
- Антагонистические (с нулевой суммой)
- Конечные

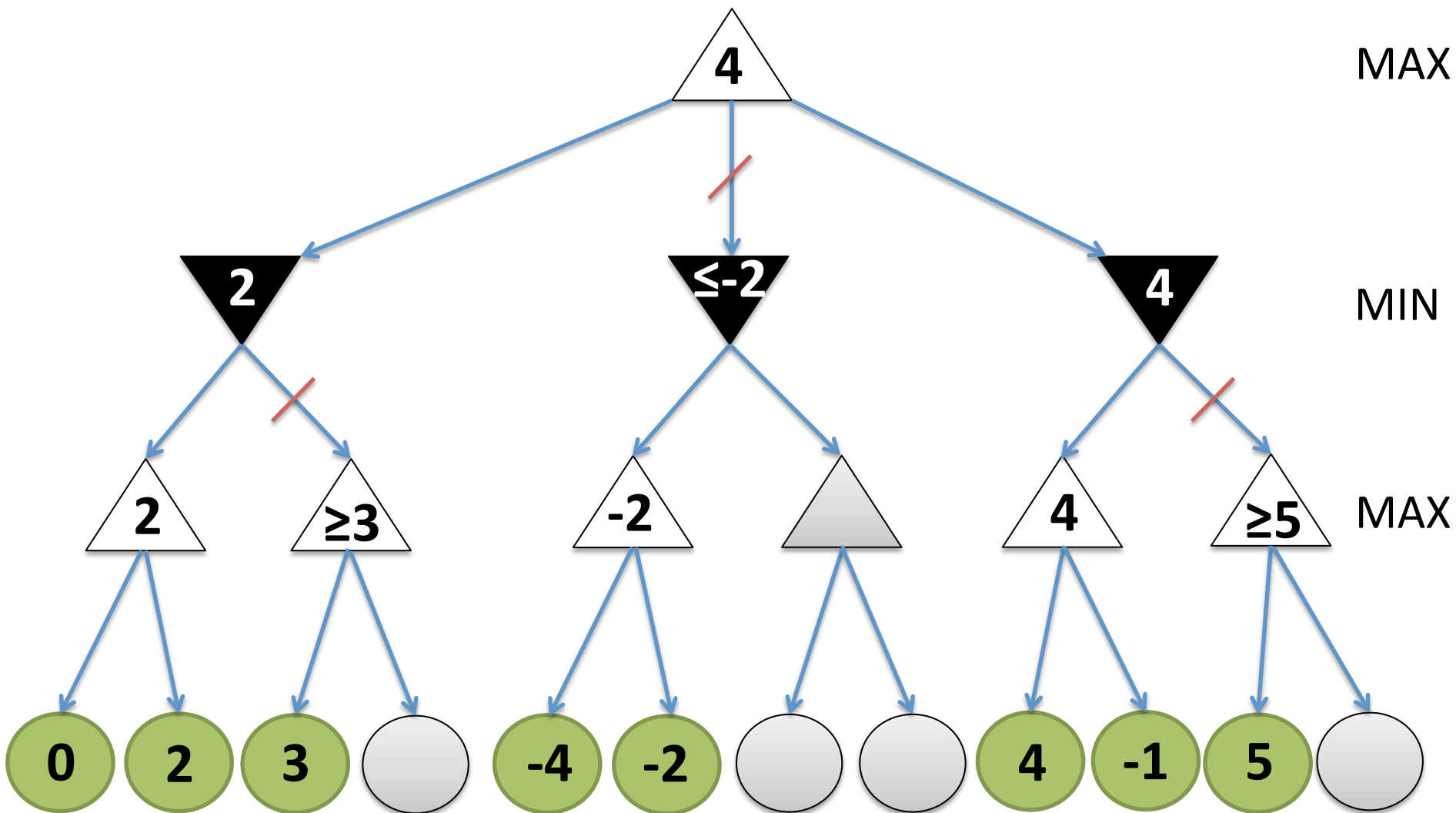
Дерево игры



Вариативность шахмат и го

- Число всевозможных партий $\approx b^d$
 - b – средний коэффициент ветвления (число возможных ходов в фиксированной позиции)
 - d – средняя глубина дерева (длина игры)
- Шахматы: $b \approx 35$, $d \approx 80$, $b^d \approx 10^{124}$
- Го: $b \approx 250$, $d \approx 150$, $b^d \approx 10^{544}$
- Число атомов во Вселенной: 10^{80}

Альфа-бета-отсечение



Применение альфа-бета-отсечения

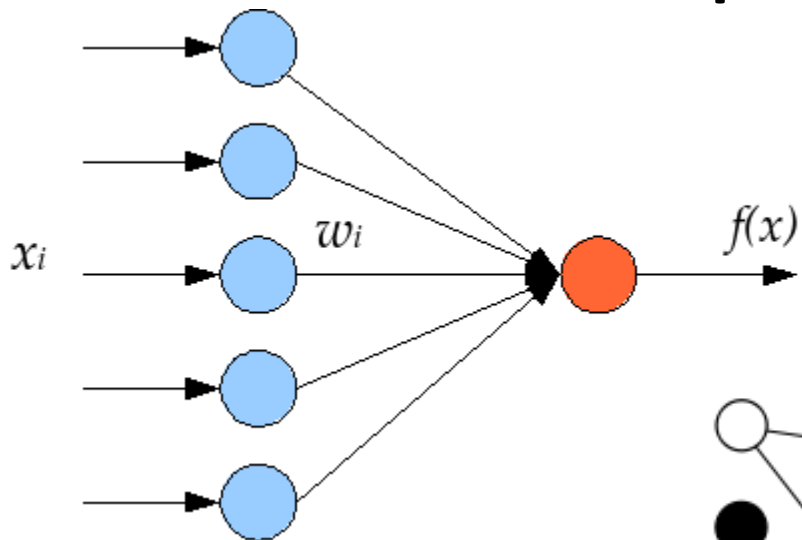
- Итак, каждая вершина (состояние игры) s имеет свою ценность $v^*(s)$ – выигрыш одного из игроков при оптимальной игре обоих
- Рассчитать на несколько ходов и вычислить *оценку* $v(s)$ ценности терминальных позиций
 - Пример в шахматах: оценка материала (пешка – 1 очко, конь и слон – по 3 очка и т.д.)
 - В DeerBlue – сложная экспертная формула оценки позиции (8000 параметров)
- Итеративное углубление
 - В DeerBlue: углубление на от 6 до 16, в отдельных случаях – до 40 уровней
- **Проблема:** в го, помимо большего объёма перебора, сложно формализовать качество позиции

Машинное обучение

«Область исследований, изучающая
возможность компьютеров обучаться без
непосредственного программирования»
(Артур Сэмюэль)

В.Н. Вапник, А.Я. Червоненкис, Ю.И. Журавлёв

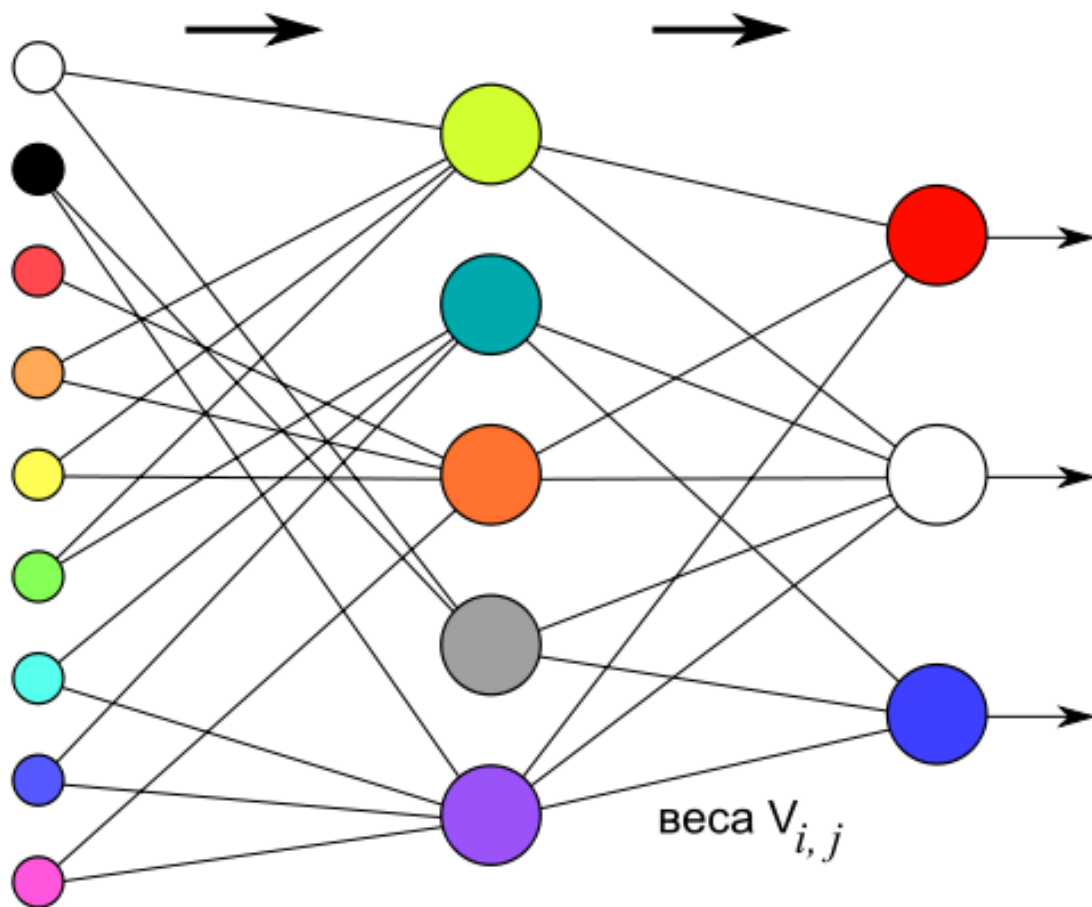
Нейронные сети



$$y = f\left(\sum_i w_i x_i - b\right)$$

При большом
количестве слоёв –
«проклятие
размерности»

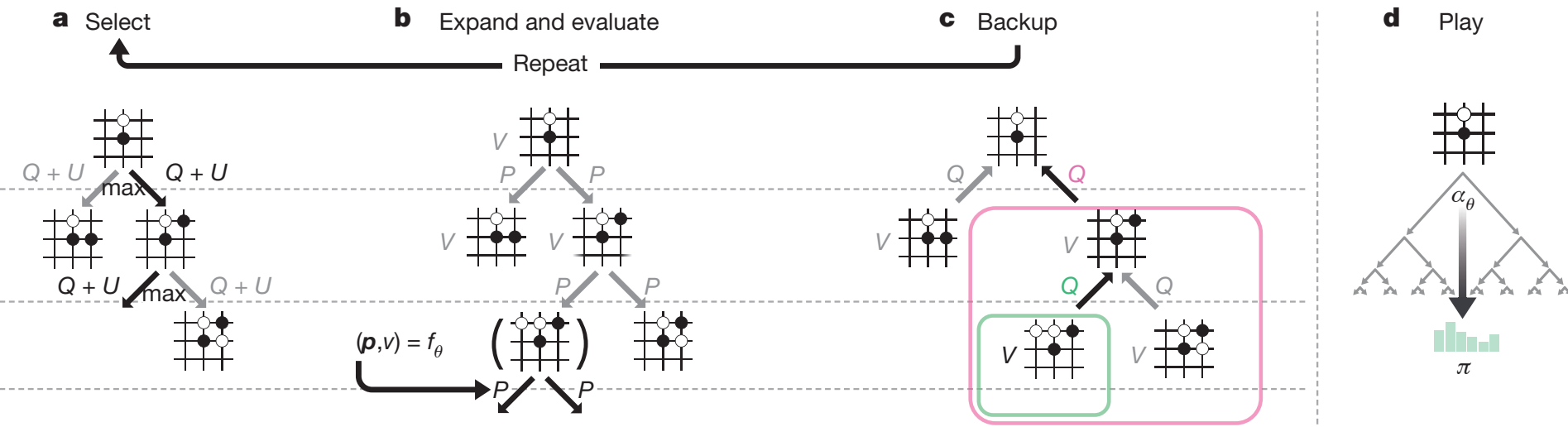
Нейронные сети
глубокого обучения
(deep learning),
с середины 2000-х



Устройство AlphaGo Zero

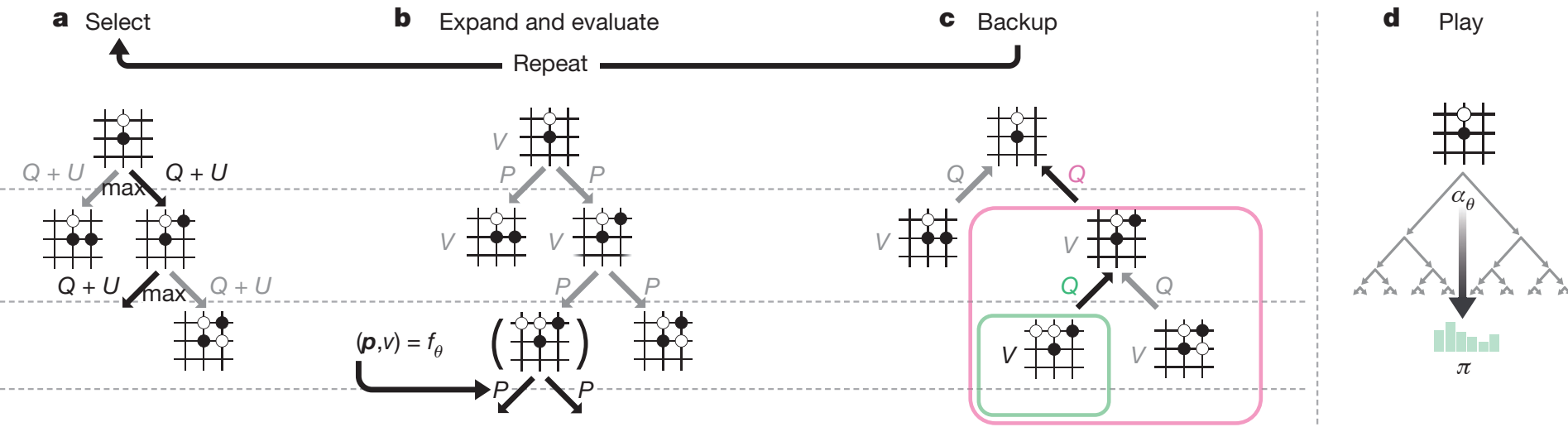
- Анализ дерева методом Монте-Карло: рассматривать не всё дерево, а только перспективные ходы.
- Перспективные с точки зрения нейронной сети
- Оценка позиции также дается нейронной сетью
- Более глубокое рассмотрение перспективных вариантов

Метод Монте-Карло для поиска в дереве



- Дерево формируется итеративно
- Начальное дерево – текущая позиция
- На каждой i -й итерации дерево проходится от корня до некоторого листа, который затем разветвляется.
- Нейронная сеть выполняет оценки априорного качества каждого нового хода $0 \leq P \leq 1$ и оценку ценности $-1 \leq V \leq 1$ каждого нового листа ($V = 1$ – 100% выигрыш, $V = -1$ – 100% проигрыш).

Метод Монте-Карло для поиска в дереве

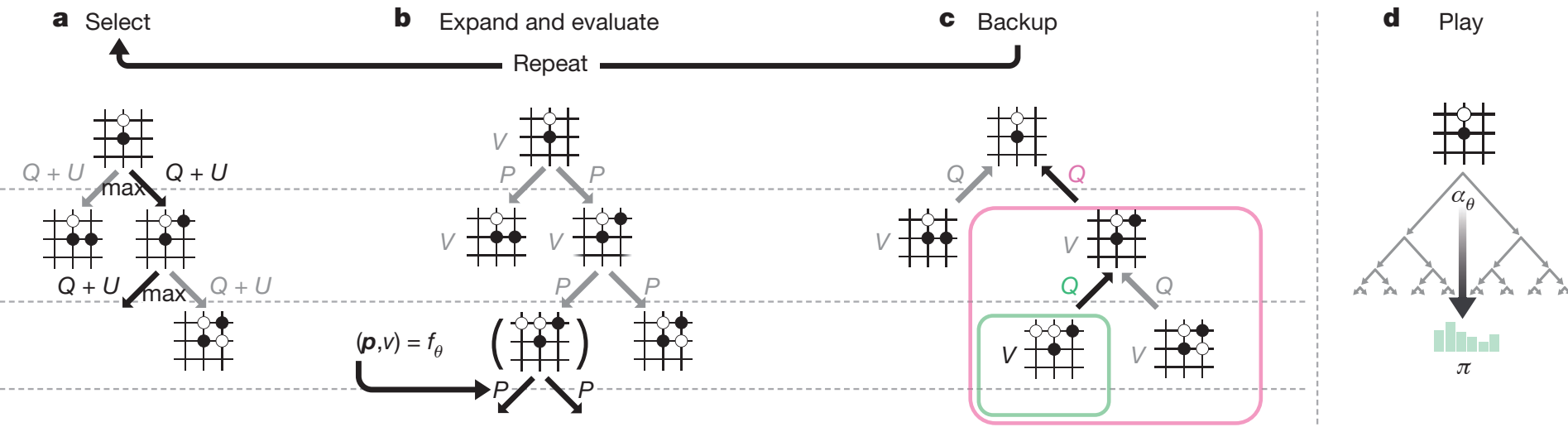


Пусть сделано n итераций. В результате сформировано некоторое дерево. Каждое ребро (s, a) хранит переменные:

- Априорная вероятность $P(s, a)$ (от нейросети – «насколько данный ход априори перспективен»)
- Счётчик посещений $N(s, a)$
- Среднее качество

$Q(s, a) = \text{сумма } V \text{ при прохождении через ребро } (s, a) / N(s, a)$

Метод Монте-Карло для поиска в дереве



- Правило прохода по дереву: следующий ход из позиции s_t выбирается по правилу

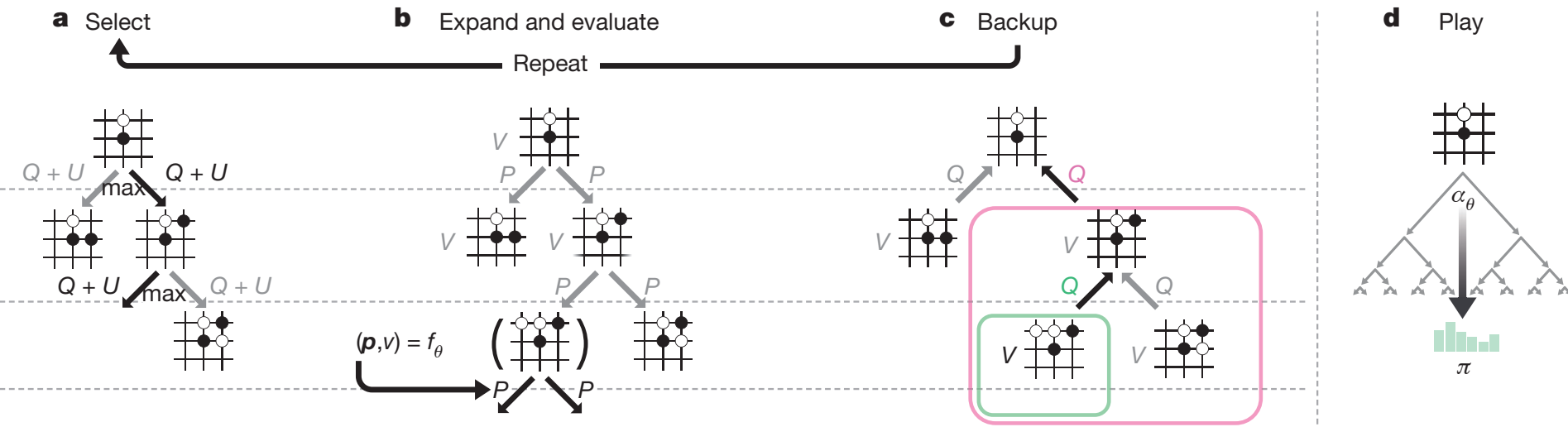
$$a_t = \arg \max_a (Q(s_t, a) + u(s_t, a)), \quad u(s_t, a) \sim \frac{P(s_t, a)}{1 + N(s_t, a)}$$

Среднее качество

Бонус

- Разветвление: как только доходим до листа, он разветвляется и подсчитываются априорные вероятности следующих ходов

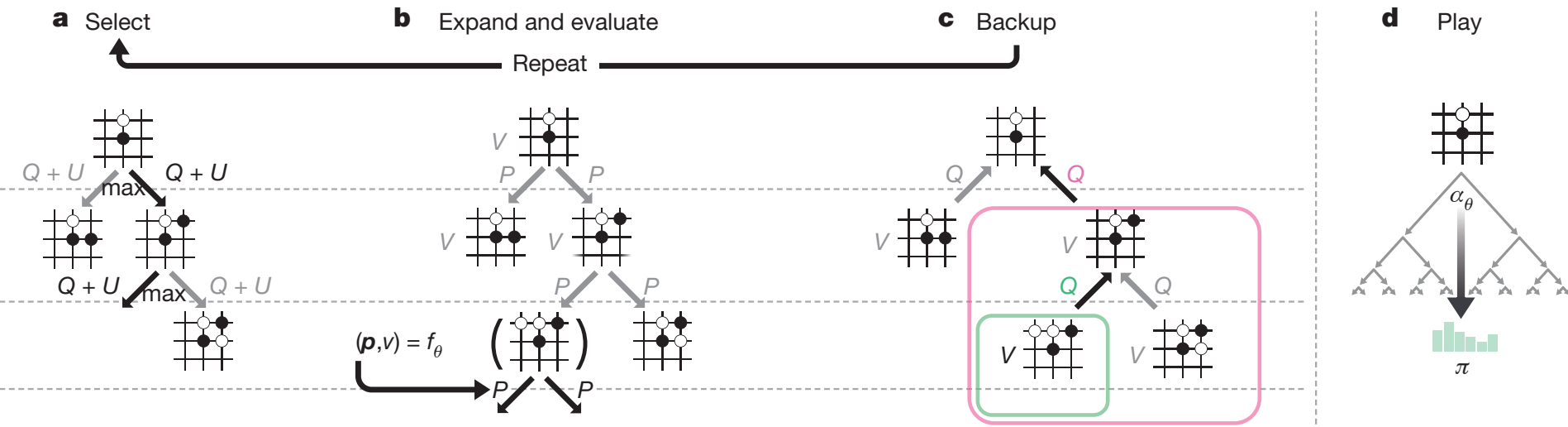
Метод Монте-Карло для поиска в дереве



- 1) Выбор пути до листовой вершины
- 2) Оценивание листовой вершины
- 3) Обратное распространение
- 4) Разветвление и вычисление априорных вероятностей следующих ходов

После всех итераций выбирается случайный ход согласно распределению $\pi_a \sim N(S, a)^{1/\tau}$ (τ – температура)

Метод Монте-Карло для поиска в дереве



Для хорошей работы метода необходима хорошо обученная нейронная сеть:

- Хорошая функция априорной вероятности $P(s, a)$ для начального отбора ходов-кандидатов
- Хорошая функция оценивания $V(s)$

Применения метода Монте-Карло для поиска в дереве

- Игры
- Планирование
- Составление расписаний
- Задача удовлетворения ограничений (выполнимость булевых функций, раскраска графа, раскраска карты)

Нейронная сеть AlphaGo Zero

- Вход: характеристика позиции – 17 двоичных матриц (карт признаков) размерности 19x19 (доска го) – история 8 последних своих ходов, 8 последних ходов противника и карта признака «сейчас ходят черные или белые»?
- Выход: априорные вероятности всех ходов из заданной позиции $P(s,a)$ и её оценка $V(s)$
- Задача: $P(s,a)$ должно как можно лучше аппроксимировать распределение вероятностей π_a , выдаваемое методом M-K, а $V(s)$ – результат игры по M-K.

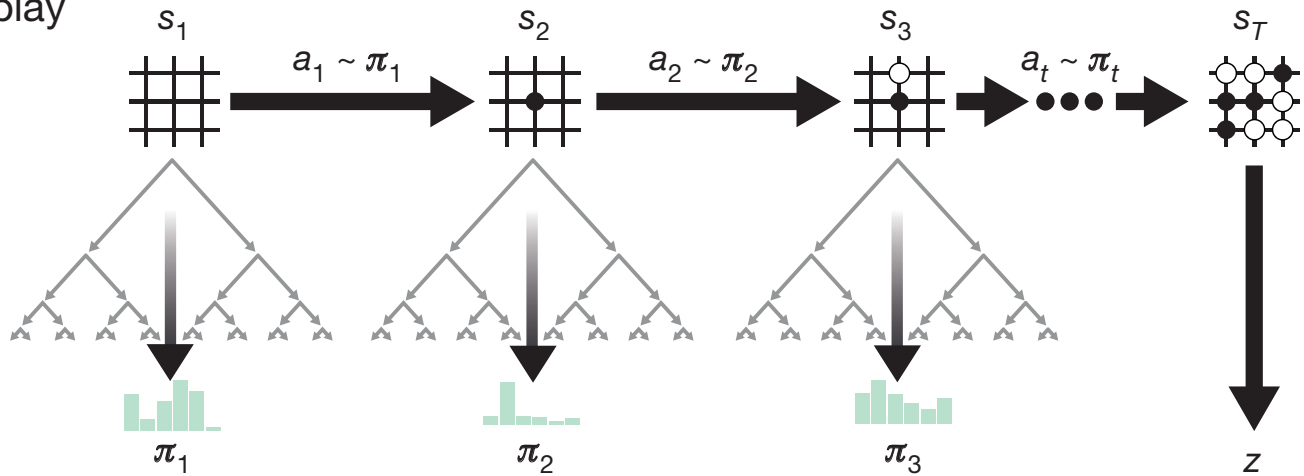
Нейронная сеть AlphaGo Zero

- Минимизация функционала (θ – веса сети):

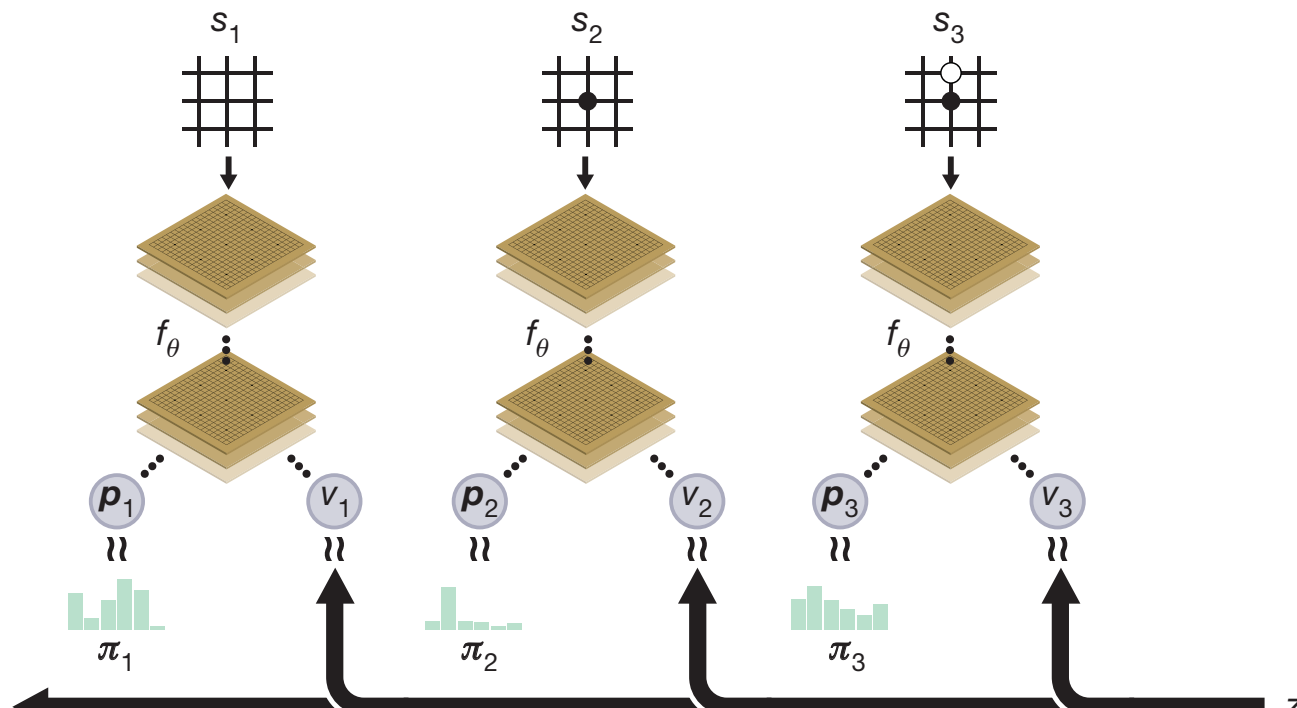
$$l = (z - v)^2 - \sum_a \pi_a \log p_a + c \|\theta\|^2$$

Итерационный процесс

a Self-play



b Neural network training



Обучение

- Первые 3 дня: 4,9 млн. игр, 1600 симуляций Монте-Карло для каждого хода
- Обновление параметров нейросети: каждые 700 000 мини-партий по 2048 позиций
- 36 часов: превзойдён уровень AlphaGo Lee (которая обучалась несколько месяцев)
- 72 часа: победа 100:0 над AlphaGo Lee
- 40 дней: 29 млн. игр.
- 21 день: превзойдён уровень AlphaGo Master
- 40 дней: победа 89:11 над AlphaGo Master

Stockfish анализировала 70 млн. позиций в сек.
AlphaZero – 80 тыс.

«AlphaZero компенсирует меньшее число анализируемых позиций тем, что с помощью нейронных сетей глубокого обучения концентрируется на наиболее перспективных вариантах. Пожалуй, это более «человекоподобный» подход к поиску, что предлагал ранее К. Шеннон».

Особенно удивительно

- Необычные дебюты в го
- Программа AlphaZero сама нашла многие известные дебюты, в том числе те, которые находятся на передовом крае современной шахматной теории и даже за его пределами

AlphaZero vs «ПИОНЕР»

М.М. Ботвинника

- Известны работы М.М. Ботвинника по созданию программы игры в шахматы «ПИОНЕР», воспроизводящей человеческое мышление
- М.М. Ботвинник: «Я исхожу из гипотезы, что шахматист-человек мыслит по определённой программе. Эта идея легла в основу программы ПИОНЕР»; «Наш компьютер будет делать ход так же и по тому же методу, как это делает шахматный мастер. Это в отличие от всех тех программ, которые уже играют и используют очень сильные компьютеры».
- AlphaZero реализовало его идеи
- Но: машинное обучение вместо логического описания алгоритма работы шахматного мастера

Что дальше?

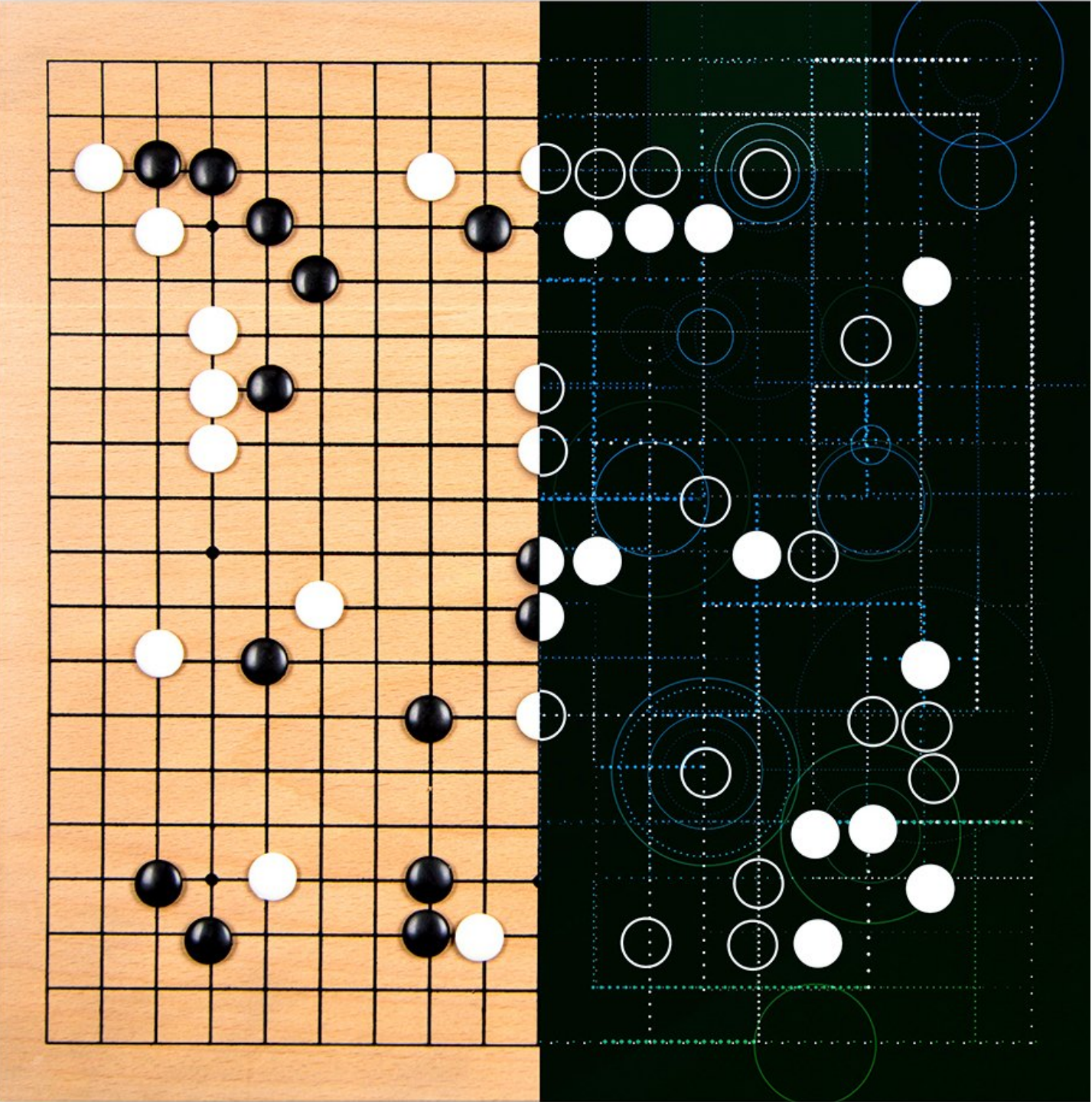
- Человек обучается на много меньшем числе игр
- Можно ли эти технологии перенести за пределы игр, «в реальный мир»?
- Достоинство AlphaGo: алгоритмы довольно общие, особенности игры го почти не используются
- Объявлено об изучении сворачивания белка с потенциальными применениями в фармакологии

Заключение

Новизна подхода AlphaGo и AlphaZero заключается в **совмещении**

- 1) методов машинного обучения (нейронных сети глубокого обучения) и
- 2) традиционных методов информатики (метода Монте-Карло для поиска на дереве).

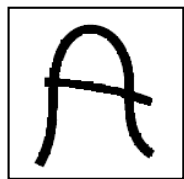
Спасибо



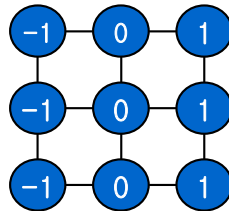
За внимание!

Свёрточные нейронные сети

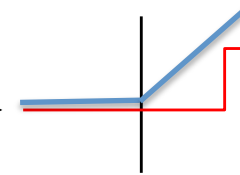
Convolutional Neural Network



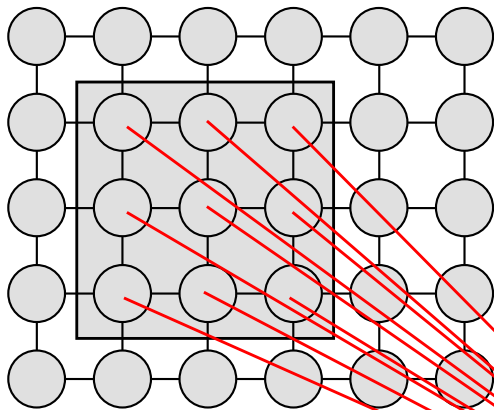
Convolve with



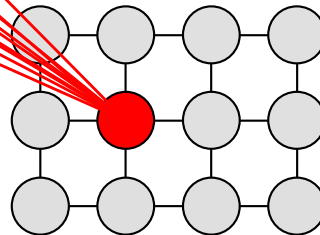
Threshold



Input Layer



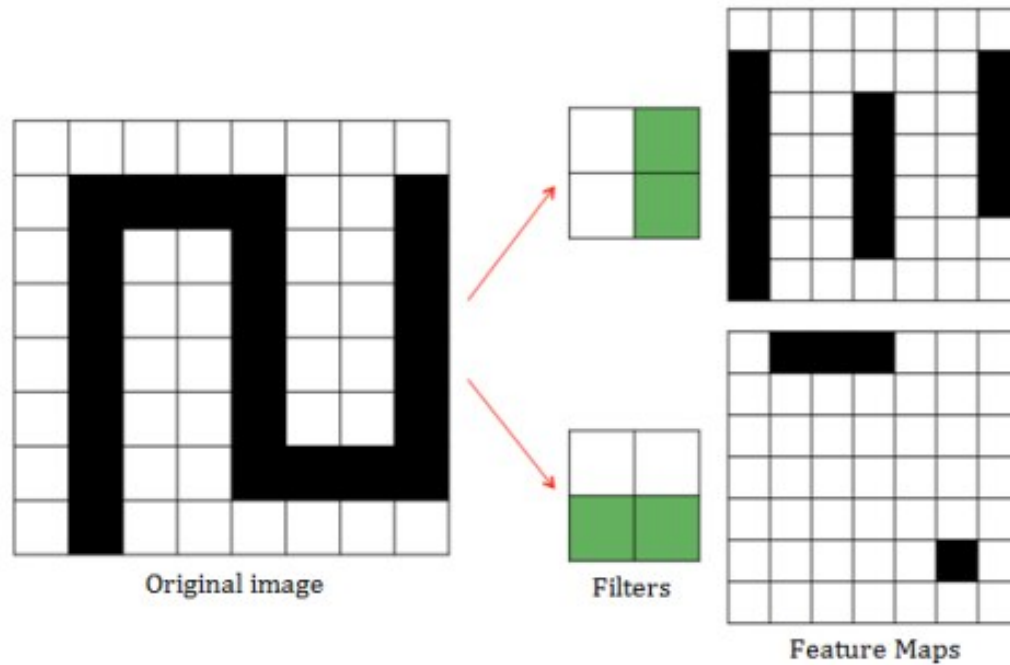
Output Layer



$$y_{ij} = \theta \left(\sum_{k,l=1}^3 w_{kl} x_{i+k-1, j+l-1} - b \right)$$

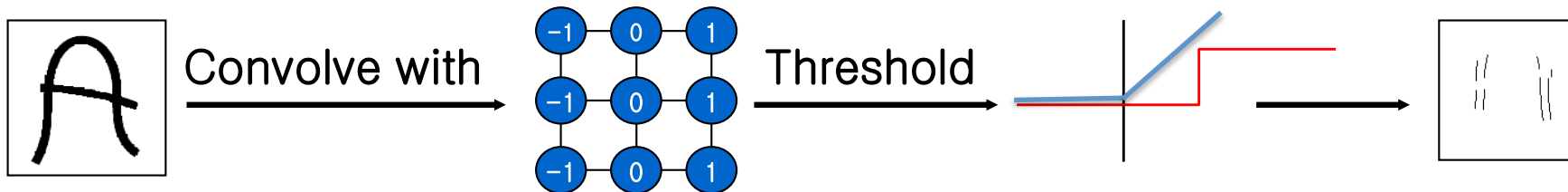
$$\theta(x) = \max(0, x)$$

Фильтр: вычленение элементов

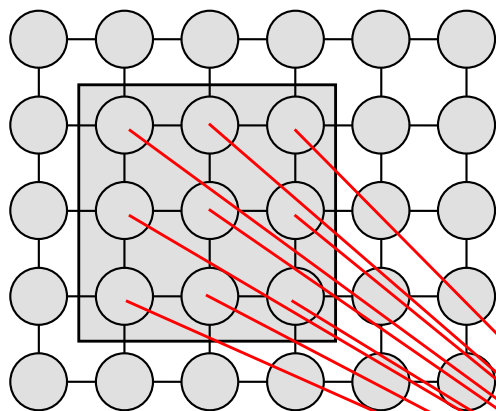


Свёрточные нейронные сети

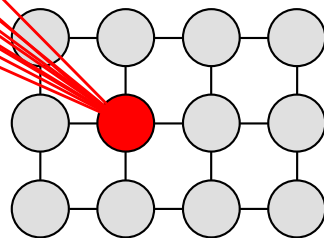
Convolutional Neural Network



Input Layer



Output Layer



$$y_{ij} = \theta \left(\sum_{k,l=1}^3 w_{kl} x_{i+k-1, j+l-1} - b \right)$$

Многослойная сеть:
выход одного слоя
– вход следующего

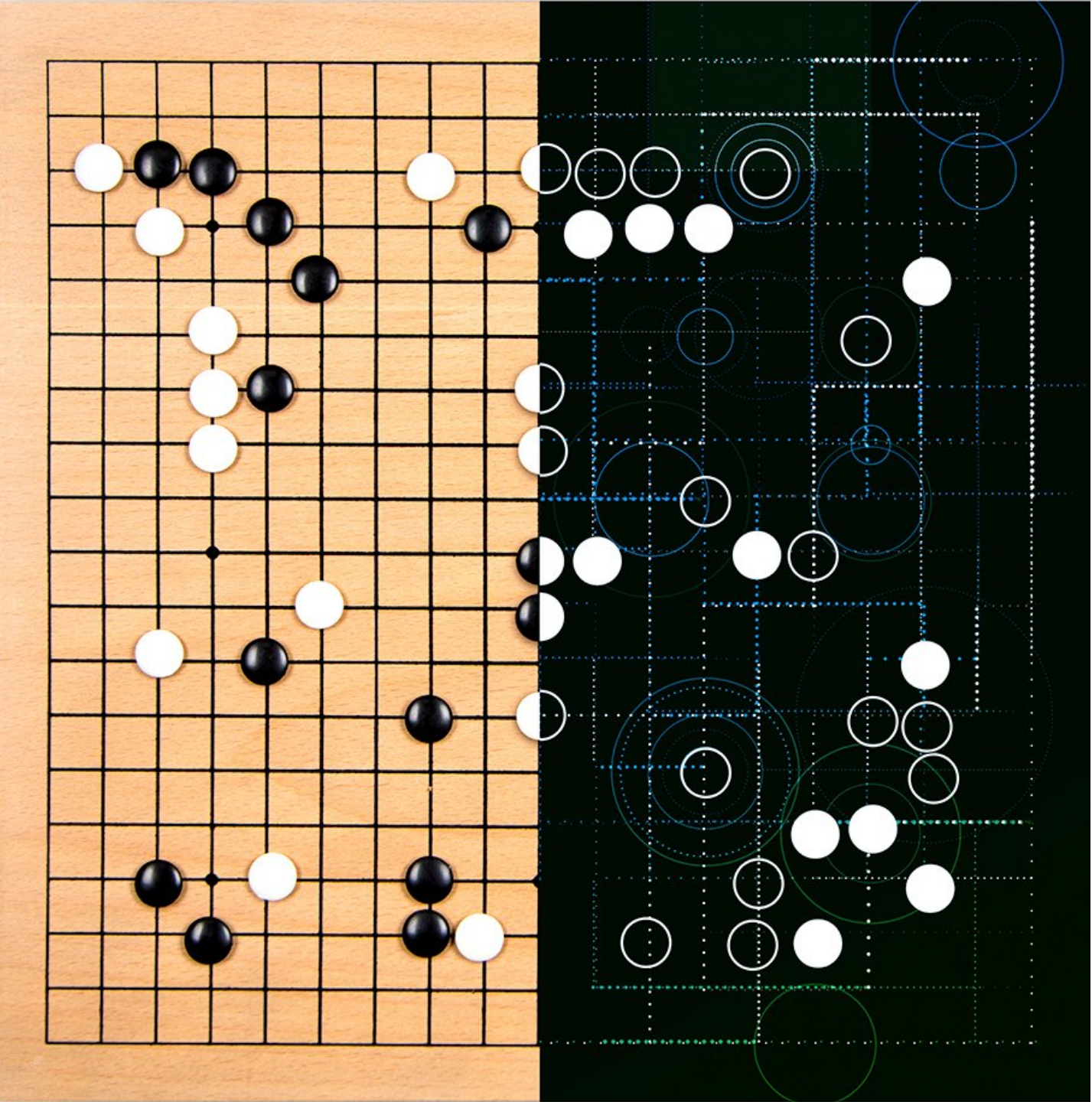
Обучение – подбор
Коэффициентов w_{kl} и b
для каждого фильтра и слоя

slide by Abi-Roozgard

Нейронная сеть AlphaGo Zero

- Свёрточный слой: 256 фильтров 3×3 , нормализация, функция выпрямителя
- 19 слоёв остаточной нейронной сети: 256 фильтров 3×3 , нормализация, функция выпрямителя
- Выход $P(s, a)$: свёрточный слой с 2 фильтрами 1×1 , нормализация, функция выпрямителя, полносвязный слой, сигмоидная функция, выход – матрица 19×19
- Выход $V(s)$: свёрточный слой с 2 фильтрами 1×1 , нормализация, функция выпрямителя, полносвязный слой, выпрямитель, ещё один полносвязный слой, функция th .

Спасибо



За внимание!