TD Learning - Ein Modell für das Lernverhalten von Tieren/Menschen?

Seminar Biologisch Motivierte Lernverfahren
Sommersemester 2012

Maria Dost, MB Biophysik

Gliederung

- 1. Was ist TD-Learning?
- 2. Value Functions
- 3. Beispielprogramm
- 4. Lernverhalten von Tieren
- 4a. TD-Learning in Neurowissenschaften

1. Was ist TD-Learning?

- Temporal Difference Learning
- Bewertungsfunktion für mögliche Aktionen
- Beim Erreichen des Ziels gibt es Belohnung
- Bewertung wird aus Erfahrung geschätzt

1. Was ist TD-Learning?

- Bewertungsaktion wird nach jeder Aktion/ jedem Zeitschritt angepasst -> Temporal
- Geändert wird aufgrund von Differenzen von erwartetem und tatsächlichem Wert -> Difference
- Es wird eine Bewertungsfunktion erlernt -> Learning

1. Was ist TD-Learning?

- Vorteil: Lerner weiß nicht erst am Ende, ob eine Aktion gut war oder nicht
- jeder Aktion wird ein Wert zugewiesen
- Verfahren konvergiert nach vielen Episoden
- schlechte Aktionen werden seltener positiv verstärkt, gute Aktionen werden öfter positiv verstärkt

1. Was ist TD-Learning?

- benötigt kein Modell, sondern nur Erfahrung
- Erfahrung wirkt sich schon während einer Episode aus
- Zur Zeit eine der meist genutzten Lernmethoden
- nicht beschränkt auf Reinforcement Learning

1. Was ist TD-Learning?

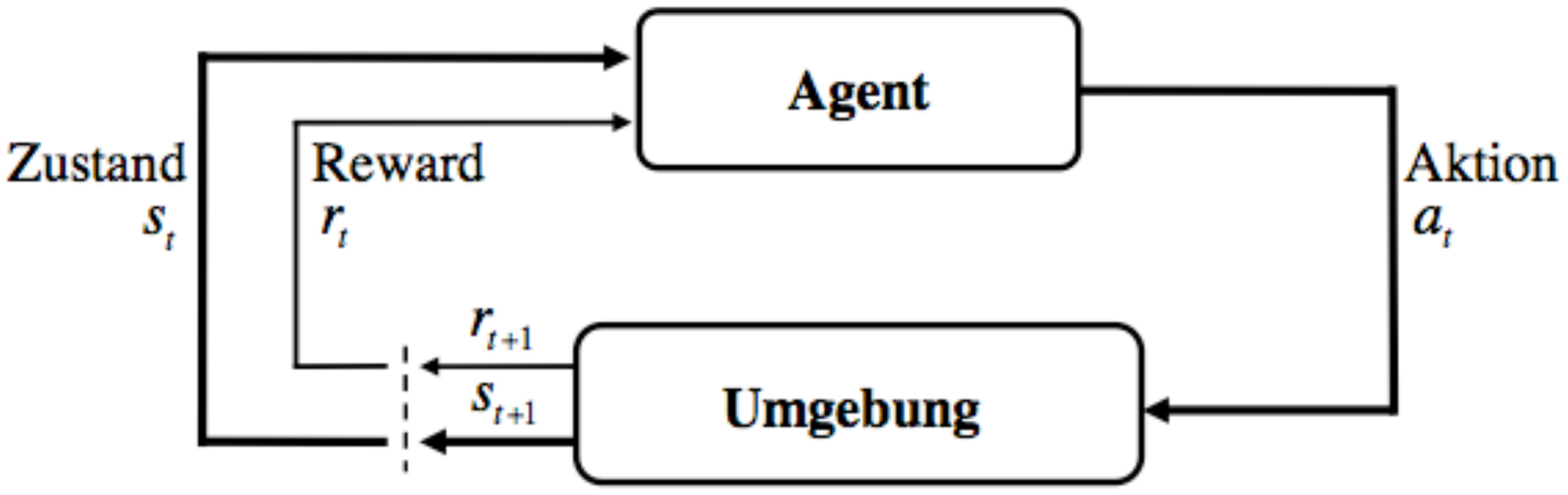


Abbildung 2: Das Interaktionsschema zwischen Agent und Umgebung

2. Value Functions

- Value Function wird erlernt
- Bildlich: alle möglichen Situation als Knoten eines Baumes
- normalerweise Belohnung am Ende eines Astes (Blatt)
- Value Function weist jedem Knoten einen Wert zu

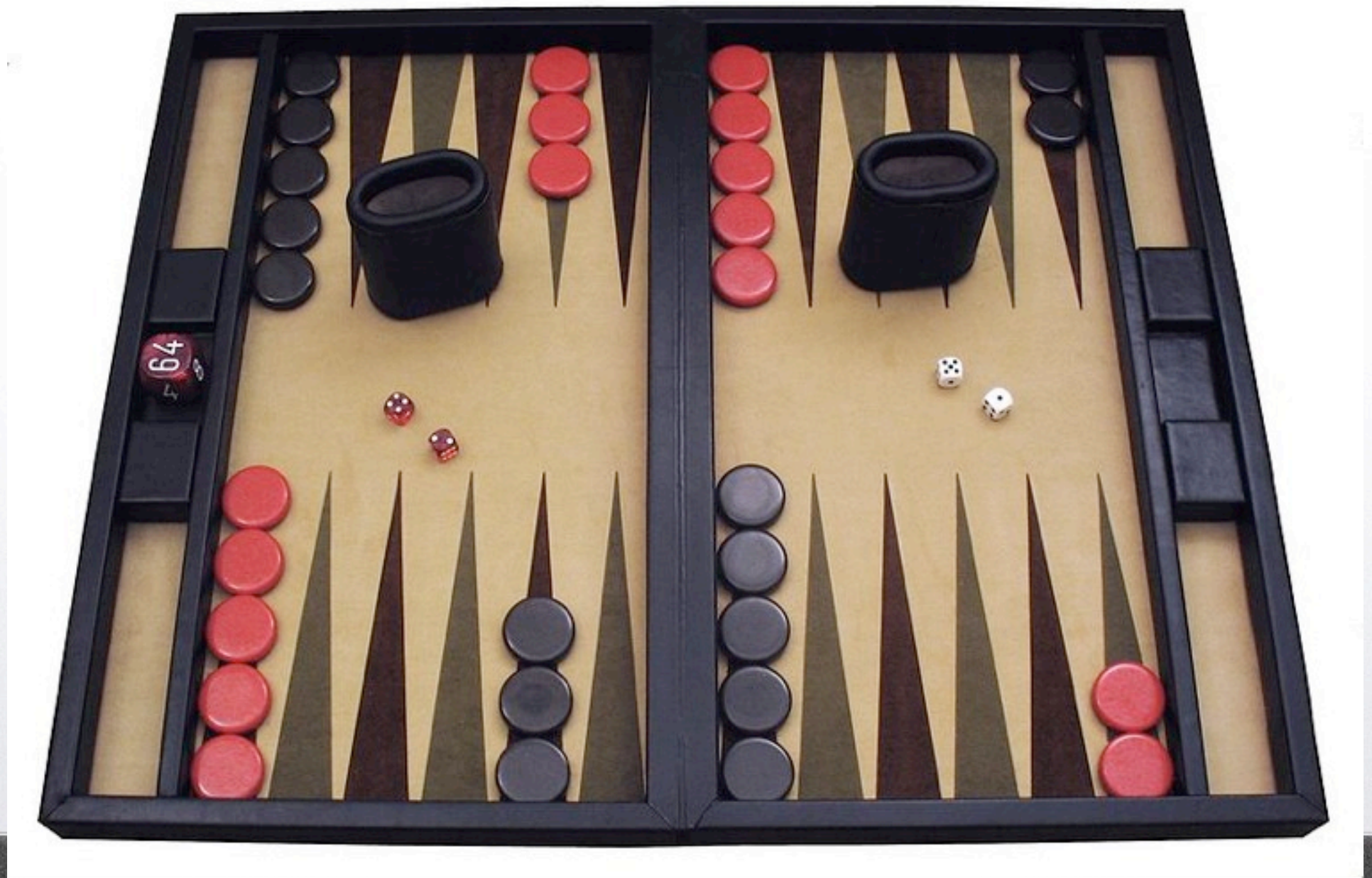
2. Value Function

- Es muss erst Erfahrung gemacht werden, bevor das Programm lernt
- Bewertungsfunktion (reward function)
- bewertet Zustände
- error function gibt Wert zwischen erwarteter Belohnung uns tatsächlicher Belohnung an
- Vorschrift A -> B

3. Beispielprogramm

- Beispiel: TD-Backgammon
- lernte Spieltaktik in 1.500.000 Spielen gegen sich selbst
- verlor 1998 die Weltmeisterschaft mit acht Punkten
- hat keinerlei Vorwissen über das Spiel, keine Beeinflussung durch menschliche Spieler

3. Beispielprogramm



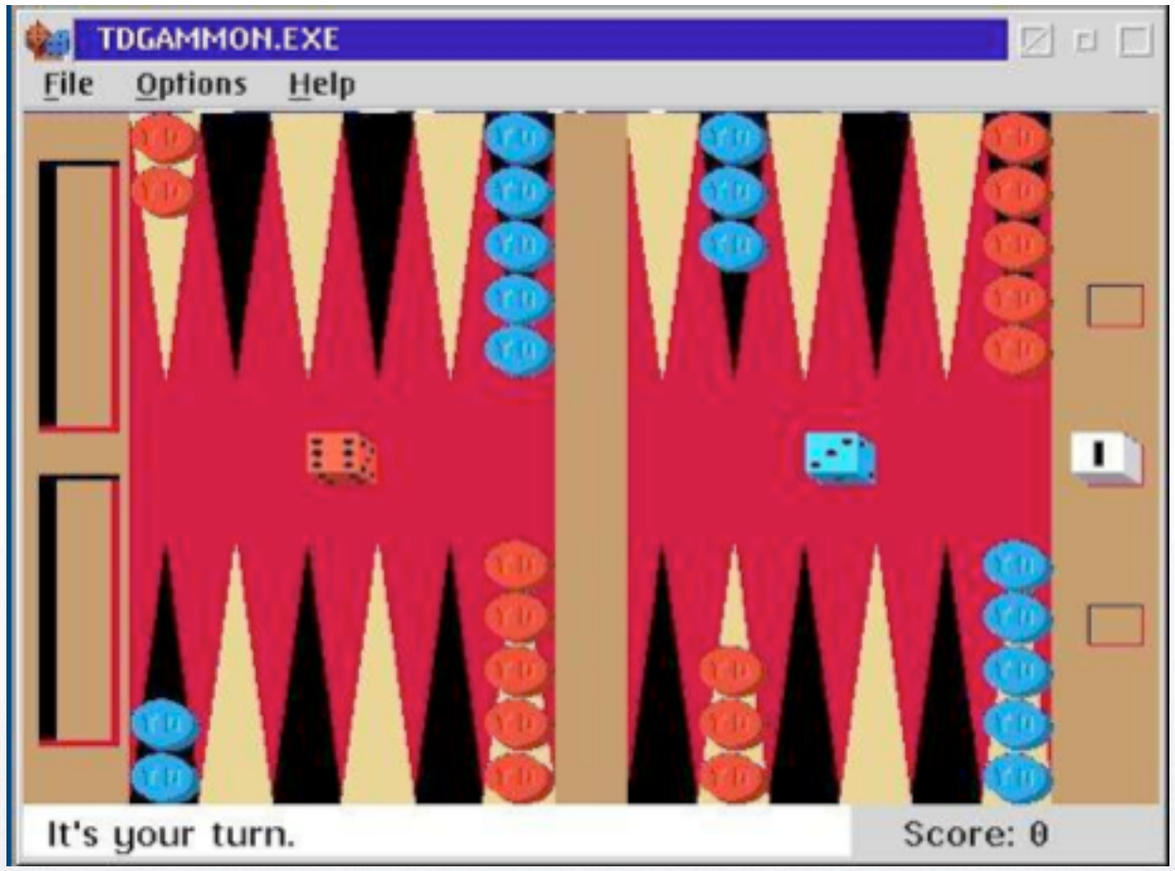
3. Beispielprogramm

- Backgammon ist ein sehr komplexes Spiel
- nach einer Wertetabelle zu spielen ist nicht möglich
- sehr hoher Verzweigungsfaktor, weit komplexer als Schach

3. Beispielprogramm

- Ein neuronales Netz, welches lernt
- nach jedem Spielzug wird TD-Algorithmus angepasst
- Würfeln beim Backgammon fördert den Lernprozess

3. Beispielprogramm



4. Lernverhalten von Tieren

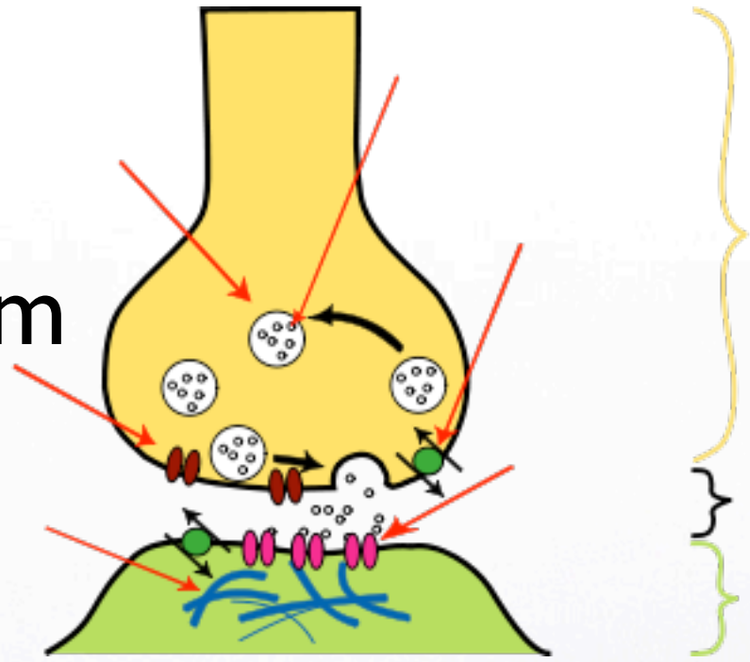
- bekannter Zusammenhang: Reinforcement Learning
- „gute“ Handlungen werden belohnt, „schlechte“ bestraft
- z.B. Training von Tieren, Schmerz
- Erklärt Lernverhalten nur unzureichend

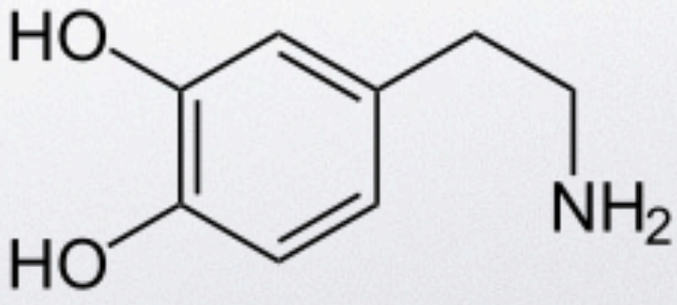
4a. TD-Learning in Neurowissenschaften

- Lernen ist nur unzureichend verstanden -> Modellvorstellungen
- Zusammenhang zwischen Dopamin Neuronen und TD-Learning
- Dopamin-Signal als reward prediction error, welches durch TD-Learning Algorithmus berechnet werden kann

Dopamin

- Neurotransmitter
- in verschiedenen Vorgängen im Gehirn beteiligt
- bekannt als „Glückshormon“
- wichtig für alle Lernprozesse

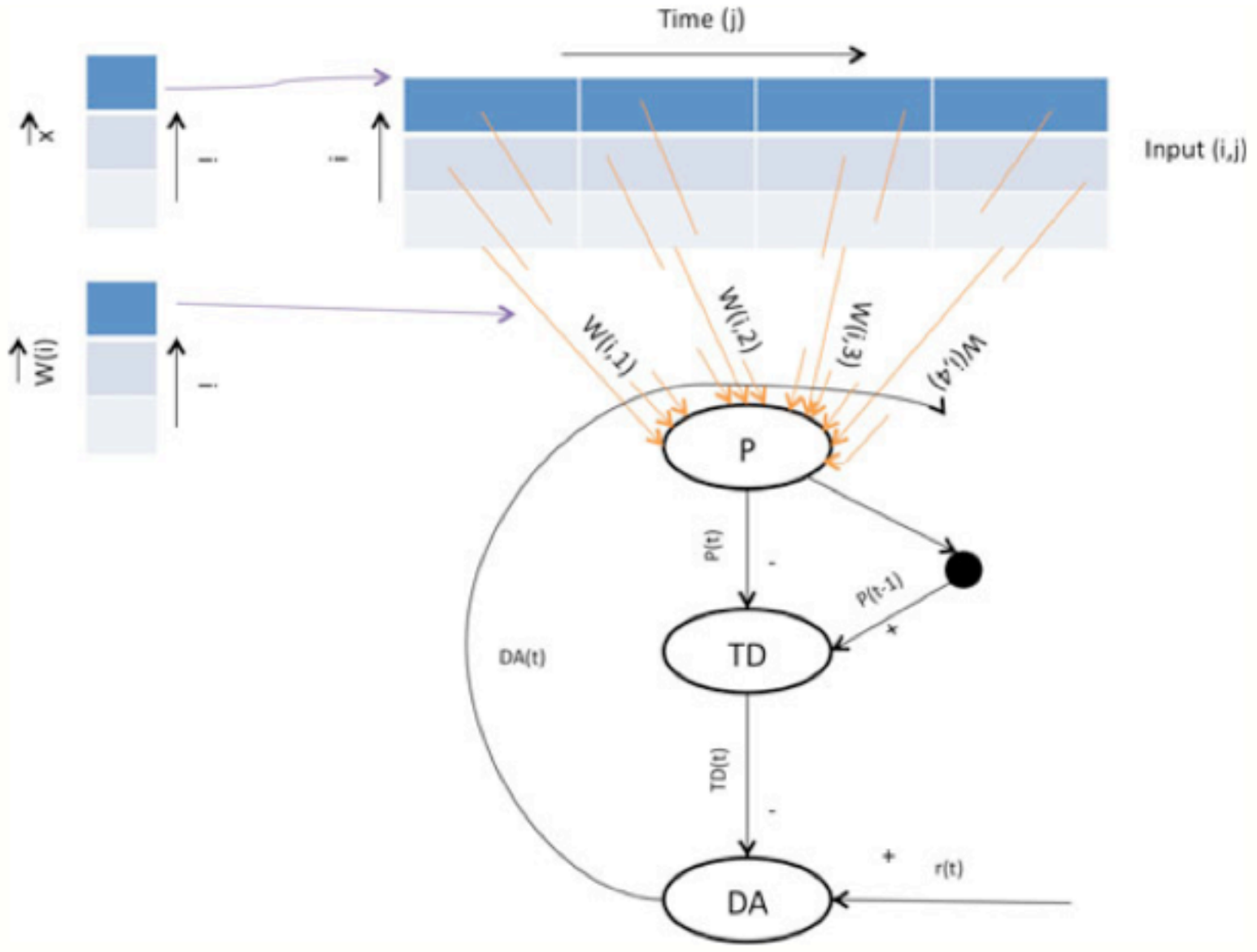




4a. TD-Learning in Neurowissenschaften

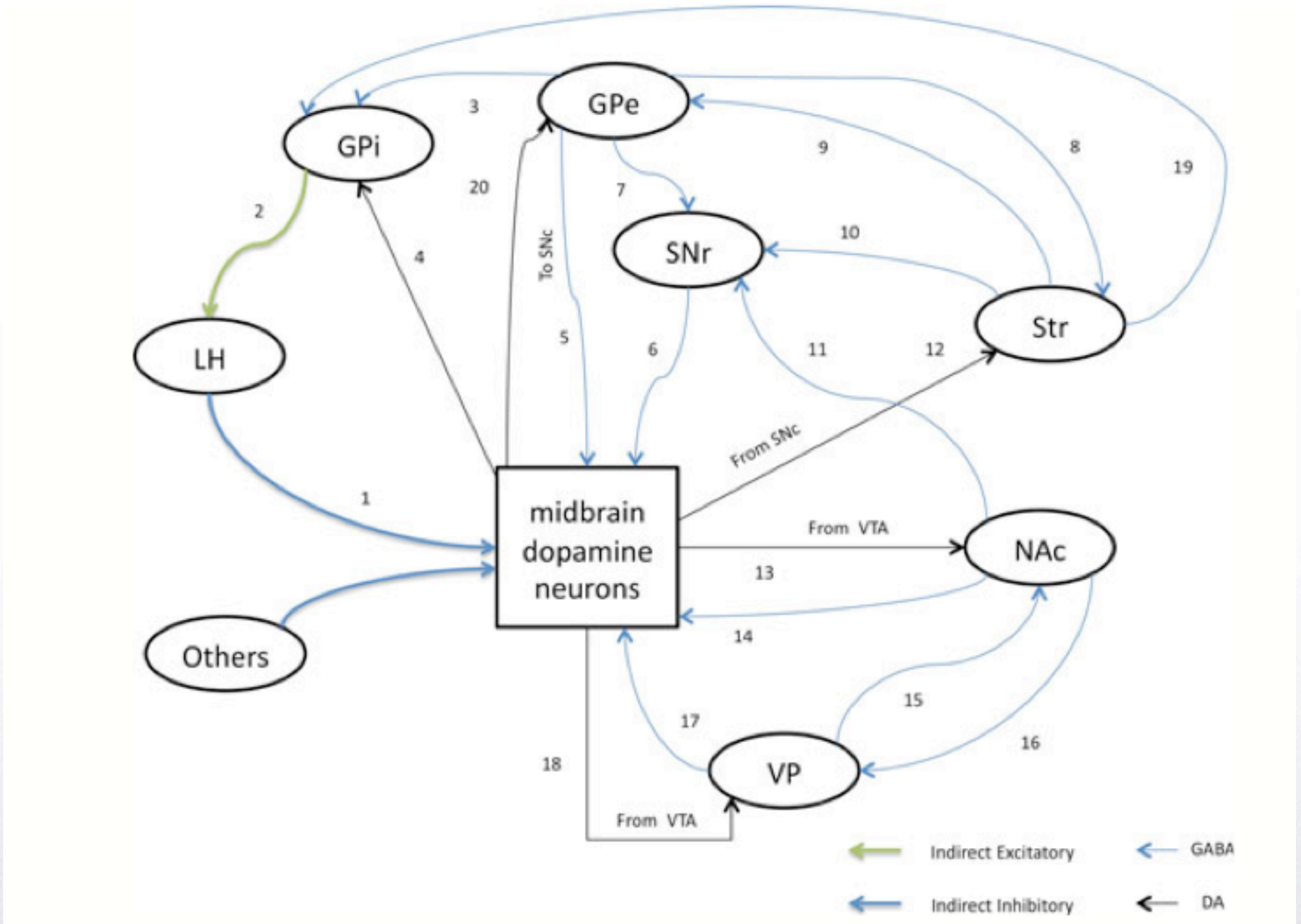
- TD-Learning kann als Beschreibung der Feuerungsrate der Neuronen verstanden werden -> Synapsen/Neuronen sind Knoten/Funktionen im TD-Modell
- Modell für Neuronen Kontrolle -> Knotenpunkte sind biologische Einheiten

4a. TD-Learning in Neurowissenschaften



4a. TD-Learning in Neurowissenschaften

- Für TD-Learning sind folgende Annahmen nötig:
- Dopamin-Signal ist ein Skalar
- Signal wird überall gleich aufgenommen und ist unabhängig von Zeit
- Fehler der erwarteten Belohnung kann positiv und negativ sein
- TD-Signal selbst kann positiv oder negativ sein
- eligibility traces müssen angenommen werden
- Um Gewichtung zu ändern, ist nur Input und Dopamin nötig
- Es gibt eine Möglichkeit alle Zustände zu speichern


GPe
GPi
SNr
Str
LH
midbrain
dopamine
neurons
Others
NAc
VP
To SNc
From SNc
From VTA
From VTA
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Indirect Excitatory
Indirect Inhibitory
GABA
DA

Zusammnenfassung

- TD = Time Difference
- eines der meist verwendeten Lernverfahren
- Zusammenhang mit Dopamin Neuronen
- könnte einzelnes Neuron gut darstellen
- um komplexe Lernprozesse darzustellen ist es nur eine Modellvorstellung

- Wirklich gutes Paper zum Thema:
  Neural control of dopamine neurotransmission: implications for reinforcement learning
  European Journal of Neuroscience, Vol. 35, pp. 1115–1123, 2012
  http://onlinelibrary.wiley.com/store/10.1111/j.1460-9568.2012.08055.x/asset/j.1460-9568.2012.08055.x.pdf?

  v=1&t=h31hbvv8&s=89fbf20f8a526d822c78d1b795579d2f2ba2deab

Quellen

- http://www.informatik.uni-ulm.de/ni/Lehre/SS07/RL/vorlesung/rl06.pdf am 17.05.2012
- http://www.techfak.uni-bielefeld.de/~afinke/DopaminNeuronen.pdf am 17.05.2012
- http://en.wikipedia.org/wiki/Backgammon 17.05.2012
- http://www-staff.informatik.uni-frankfurt.de/asa/seminare/PSemWS04/SchmidAusarbeitung.pdf 17.05.2012
- http://books.nips.cc/papers/files/nips14/CS01.pdf am 3.06.2012
- Learning and Computational Neuroscience: Foundations of Adaptive Networks, M. Gabriel and J. Moore, Eds., pp. 497–537. MIT Press, 1990.
- http://en.wikipedia.org/wiki/File:Dopamine_and_serotonin_pathways.gif 04.06.2012
- http://en.wikipedia.org/wiki/Temporal_difference_learning 04.06.2012
- http://books.nips.cc/papers/files/nips14/CS01.pdf 04.06.2012
- http://onlinelibrary.wiley.com/store/10.1111/j.1460-9568.2012.08055.x/asset/j.1460-9568.2012.08055.x.pdf?v=1&t=h31hbvv8&s=89fbf20f8a526d822c78d1b795579d2f2ba2deab 04.06.2012
- http://de.wikipedia.org/w/index.php?title=Datei:Synapse_Illustration_unlabeled.svg&filetimestamp=20091231181331 04.06.2012

Danke!!!

- # Fragen???