

# Пробни тест I

## Самообучавајући и адаптивни алгоритми

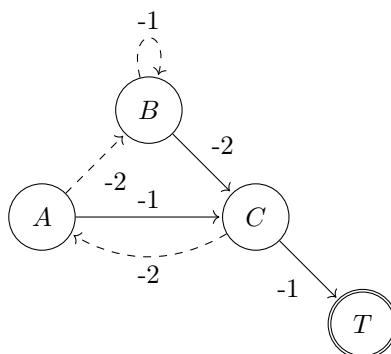
Катедра за аутоматско управљање • ФТН, Нови Сад  
14. новембар 2023.

1. Дефинисати детерминистички Марковљев процес одлучивања (МПО).
2. Дефинисати детерминистичку политику одлучивања.
3. У односу на претходно дефинисани процес одлучивања и претходно дефинисану политику, формирати Белманову једначину за одређивање вредности стања.
4. Дат је детерминистички МПО, са скупом могућих стања  $\mathcal{S} = \{A, B, C\}$ , скупом могућих акција  $\mathcal{A} = \{+, -\}$  и скупом могућих награда  $\mathcal{R} = \{-1, 0, 1\}$ . Над датим МПО делује детерминистички агент. Праћењем дејства агента, уочена је следећа секвенца стања, акција и награда:

$$A \xrightarrow[1]{+} B \xrightarrow[-1]{-} C \xrightarrow[0]{+} T,$$

где је  $A$  почетно стање, а свака стрелица представља један прелаз између стања. Ознака изнад стрелице означава акцију која је довела до прелаза, а број испод стрелице означава остварену награду.  $T$  је терминално "стање".

- Коју политику одлучивања користи агент?
  - Који је модел окружења? Да ли се комплетан модел може "научити" из приказане секвенце? (Уколико је одговор НЕ, исписати само онај део модела о коме се може судити.)
  - Која је вредност стања при политици одлучивања коју користи агент?
5. За дати МПО дата је табела оптималних Q-вредности за све парове стања и акција.



Слика 1: Пример детерминистичког МПО-а илустрованог графом. Стања су приказана чворовима графа, могуће акције врстом везе (акција "+" је приказана пуном, а акција "-" испрекиданом линијом). Награде су приказане бројевима поред одговарајућих прелаза.

стање	акција	награда
A	+	-1
A	-	+1
B	+	0
B	-	-1
C	+	2
C	-	1

- Одредити оптималне вредности свих стања (оптималне  $V$ -вредности).
- Одредите оптималну политику одлучивања.

6. МПО је приказан графом на слици 1. За дати МПО

- Записати систем Белманових једначина и срачунати вредности свих стања за политику "У СВАКОМ СТАЊУ КОРИСТИ АКЦИЈУ + (ПУНА ЛИНИЈА)". Користити фактор обезвређивања једнак јединици. Сматрати да терминално стање има вредност 0.
- Записати систем Белманових једначина за одређивање оптималне политике, и срачунати оптималне вредности свих стања. Користити фактор обезвређивања једнак јединици. Сматрати да терминално стање има вредност 0.
- Одредити оптималну политику.