

Šírenie správ

Random walk simulácie sú jedny z najpopulárnejších simulácií, ktoré sa využívajú na modelovanie rôznych javov. Jedným z týchto javov je simulácia šírenia správ v populácii ľudí, čo môže byť využité novinárom pre optimalizáciu obsahu článkov, alebo pri boji proti dezinformáciám a poplašným správam. V rámci druhého zadania využijete prístup random walk pre skúmanie dynamiky šírenia správ. Samozrejme pritom zavedieme zjednodušenia pre zvládnuteľnosť úlohy.

Naša simulácia bude pracovať s nasledovnými výpočtovými modelmi (každý bude reprezentovaný vlastnou triedou):

- správa – príklad článku, ktorý si ľudia v rámci populácie preposielajú; správa bude mať kategóriu obsahu a bude popísaná mierou prekvapenia (ako veľmi zaujme čitateľa), dĺžkou aktuálnosti a časom vytvorenia.
- človek – člen populácie, ktorý má svoje záujmy, číta a preposiela správy svojim priateľom; okrem záujmov bude mať každý človek prah prekvapenia, teda ako veľmi musí byť správa prekvapujúca, aby ju prečítal a preposlal.
- populácia – skupina ľudí, v ktorej sa správa šíri.

Pomocou našich simulácií budeme hľadať odpoveď na otázky:

1. Ako sa šíria rôzne správy v danej populácii?
2. Ku koľkým ľuďom sa v priemere dostane správa s istou mierou prekvapenia?
3. Akú mieru prekvapenia potrebuje správa, aby sa dostala k istému podielu všetkých ľudí zaujímavujúcich sa o tému správy v rámci populácie?
4. Akú mieru prekvapenia potrebuje správa, aby sa dostala k istému podielu všetkých ľudí zaujímavujúcich sa o tému správy v rámci populácie, kde každý sa zaujíma o danú tému?

Simulácia bude pracovať s diskretným časom, teda bude sa skladať z niekoľkých kôl, kde v jednom kole ľudia, ktorí správu dostali, ju môžu prečítať a poslať ďalej, príjemcovia ale správu prečítajú až v ďalšom kole.

Poznámka: Jednotlivé triedy a simulácie stavajú na predošlé koncepty, takže pri vypracovaní zadania odporúčame postupovať podľa poradia úloh popísaných v tomto dokumente. Vzhľadom na prítomnosť elementu náhodnosti budú mať testy inú podobu ako v prípade prvého zadania (výsledok nie je úplne deterministický). Potrebujete implementovať minimálne metódy, ktoré sú popísané v tomto návode, ak ale chcete niektorú funkcionálnosť rozdeliť do rôznych metód, prípadne potrebujete pomocné metódy, môžete ich zdefinovať, hodnotené ale nebudú.

1. úloha – trieda News

Trieda News definuje reprezentáciu správy v našej simulácii. Každá správa je popísaná nasledovnými hodnotami:

- `category` – kategória obsahu správy; v súbore `news.py` nájdete aj konštantu `CATEGORIES`, ktorá definuje zoznam dovolených kategórií v riešení
- `excitement_rate` – miera prekvapenia; desatinné číslo medzi 0 a 1, ktoré vyjadruje, ako veľmi je správa schopná prekvapiť čitateľa: čím je táto hodnota vyššia, tým je viac pravdepodobné, že si ju človek prečíta a následne prepošle svojim kamarátom
- `validity_length` – vyjadruje čas v počte kôl, v ktorom je správa relevantná a aktuálna; ľudí po uplynutí tohto času už správa v žiadnom prípade nebude zaujímať a teda ju ani neprečítajú (bez ohľadu na mieru prekvapenia)
- `created` – kedy bola správa vytvorená. Je to celočíselná hodnota, ktorá vyjadruje, v ktorom kole simulácie sa správa vytvorila.

Trieda už obsahuje hotový konštruktor, ktorý najprv zavolá metódu pre kontrolu správnosti hodnôt parametrov, a následne nastaví členské premenné triedy. Do triedy potrebujete implementovať metódy:

`check_data(category, excitement_rate, validity_length, created)`
– 0,5b

Metóda slúži na kontrolu platnosti hodnôt parametrov pri vytváraní nového objektu. V tejto metóde potrebujete vykonať rôzne kontroly a vygenerovať rôzne výnimky ak niektorý predpoklad nie je splnený. Parametre pritom musia mať nasledovné hodnoty:

- `category` – reťazec zo zoznamu `CATEGORIES`
- `excitement_rate` – desatinné číslo (float) z intervalu $<0, 1>$ - ak nie je správny typ, vygeneruje sa `TypeError`, ak neseďí hodnota, tak `ValueError`
- `validity_length` – celé číslo z intervalu $<1, 10>$ - ak nie je správny typ, vygeneruje sa `TypeError`, ak neseďí hodnota, tak `ValueError`
- `created` – celé číslo nie menšie ako 1 - ak nie je správny typ, vygeneruje `TypeError`, ak neseďí hodnota, tak `ValueError`

`get_excitement(time_step)` – 0,5b

Metóda vypočíta a vráti aktuálnu mieru prekvapenia správy v danom časovom okamihu, ktorý dostane ako parameter `time_step`. Miera sa vypočíta na základe uplynutého času (počtu uplynutých kôl) v aktuálnom okamihu od času vytvorenia správy. Platia pritom pravidlá:

- ak uplynulo viac času ako je dĺžka aktuálnosti správy (`validity_length`), tak metóda vracia hodnotu 0
- v opačnom prípade sa aktuálna miera prekvapenia vypočíta podľa vzorca $ER^{\Delta t}$, kde ER je pôvodná miera prekvapenia správy (`excitement_rate`), a Δt je čas uplynutý od momentu vytvorenia správy vyjadrený ako počet kôl od vytvorenia správy. Vzorec vyjadruje pokles v aktuálnosti a teda v miere prekvapenia správy časom.

2. úloha – trieda Person

Trieda Person reprezentuje jedného človeka z našej skupiny potenciálnych čitateľov. Každý človek je popísaný nasledovnými vlastnosťami:

- `threshold` – prah, ktorý vyjadruje, ako veľmi je ťažké získať pozornosť človeka – správa musí mať aktuálnu mieru prekvapenia vyššiu ako je táto hodnota, aby ju človek prečítal
- `interested_in` – zoznam kategórií, o ktoré sa daný človek zaujíma – prečíta iba správy, ktoré zapadajú do jednej z týchto kategórií, iné správy nikdy neprečíta, bez ohľadu na mieru ich prekvapenia
- `friends_list` – zoznam kamarátov človeka – pri inicializácii to bude prázdny zoznam, do ktorého neskôr pridáme ľudí (objekty typu Person), ktorým daný človek môže preposlať správu
- `has_read` – zoznam správ, ktoré človek už prečítal – pri inicializácii to bude prázdny zoznam, do ktorého neskôr pridávame správy (objekty typu News), ktoré človek prečíta; zoznam potrebujeme kvôli tomu, aby človek viackrát neprečítal (a nepreposlal) tú istú správu.

Trieda obsahuje hotový konštruktor, potrebujete ale doplniť zvyšné metódy:

`is_interested_in(category)` – 0,1b

Vráti informáciu o tom, či sa človek zaujíma o kategóriu, ktorú dostane ako parameter `category`. Návrátová hodnota metódy je teda `True` alebo `False`.

`has_read_news(news)` – 0,1b

Vráti informáciu o tom, či už človek prečítal správu, ktorú dostane ako parameter `news`. Návrátová hodnota metódy je teda `True` alebo `False`.

`make_friends(population, n)` – 0,3b

Metóda naplní zoznam priateľov daného človeka, pričom dostane dva parametre:

- `population` – zoznam ľudí, z ktorých môže vyberať priateľov
- `n` – počet priateľov človeka.

Po vykonaní metódy bude zoznam priateľov človeka (`friends_list`) naplnený *n* objektmi typu Person zo zoznamu `population`. Výber priateľov je úplne náhodný, priatelia musia ale byť jedinečné hodnoty (v zozname priateľov bude každý objekt maximálne raz).

`process_news(news, time_step)` – 1b

Metóda predstavuje simuláciu toho, čo človek urobí, ak dostane nejakú správu. Metóda vždy vráti zoznam ľudí, ktorým človek správu prepošle. Metóda má dva parametre:

- `news` – objekt typu News – správa, ktorú človek dostal
- `time_step` – celé číslo – časový okamih, v ktorom správu dostal.

Človek po obdržaní správy sa správa nasledovne:

- ak správu už pred tým prečítal, neprečíta ju znova a teda nikomu ju neprepošle
- ak správa má kategóriu, ktorá človeka nezaujíma, správu neprečíta a nikomu ju neprepošle
- ak ho správa nezaujíma svojou aktuálnou mierou prekvapenia, neprečíta ju a nikomu ju neprepošle
- v opačnom prípade správu prečíta a prepošle im všetkým svojim priateľom, ktorí sa zaujímajú o tému správy.

Poznámka: V prípade, že človek správu prečíta, nezabudnite aktualizovať príslušnú členskú premennú objektu `Person`.

3. úloha – triedy `Population` a `HomogeneousPopulation`

Trieda `Population` predstavuje skupinu ľudí, ktorí čítajú a posielajú si správy. Konštruktor triedy je už hotový, nastavujú sa v ňom nasledovné členské premenné:

- `people` – zoznam ľudí v populácii – po inicializácii prázdny zoznam, napĺňa sa v metóde `generate_population`
- `active_news` – zoznam správ, ktoré sa šíria v populácii – po inicializácii prázdny, obsah sa aktualizuje v metódach `introduce_news` a `update_news`.

Okrem toho konštruktor obsahuje volanie metódy `generate_population`, ktorú už ale musíte implementovať vy spolu s ďalšími metódami:

`generate_population(n, friends_count)` – 1b

Metóda vygeneruje zoznam ľudí v populácii a vytvorí priateľstvá medzi nimi. Metóda má dva parametre, a nemá žiadnu návratovú hodnotu – aktualizuje priamo zoznam ľudí v populácii. Parametre metódy majú nasledovný význam:

- `n` – celé číslo – počet ľudí v populácii – toľko objektov potrebujete vytvoriť a pridať do zoznamu ľudí
- `friends_count` – celé číslo – každý člen populácie bude mať po vygenerovaní populácie toľko priateľov.

Ľudí generujte s náhodnou hodnotou prahu z intervalu $<0, 1)$ a štyrmi náhodnými oblasťami záujmu (zo zoznamu dovolených kategórií).

`introduce_news(news, time_step)` – 0,5b

Metóda predstavuje moment, v ktorom sa správa začne šíriť v populácii. Parameter `news` je objekt typu `News`, teda je to správa, ktorú chceme dať kolovať v populácii. Metóda vráti zoznam prvých piatich ľudí, ktoré správu dostanú, pričom môžu to byť iba ľudia, ktorých daná téma bude zaujímať.

Poznámka: pre správnu funkcionálnosť nezabudnite aktualizovať zoznam `active_news`.

`update_news(time_step)` – 0,5b

Metóda aktualizuje zoznam aktívnych správ a je potrebná z toho dôvodu, aby sme sa zbytočne nesnažili posielat' správy, ktoré už nie sú aktuálne (ich lehota aktuálnosti vypršala). Metóda má jeden parameter – `time_step` (celé číslo) –, ktorý vyjadruje časový okamih resp. kolo, v ktorom chceme aktualizovať zoznam aktívnych správ. Metóda nemá žiadnu návratovú hodnotu, iba vymaže zo zoznamu aktívnych správ tie správy, ktoré už nemajú šancu byť prečítané členmi populácie (ich aktuálna miera prekvapenia je nulová).

`count_readers(news)` – 0,25b

Metóda vypočíta a vráti počet ľudí v populácii, ktorí správu `news` (objekt typu `News`, parameter metódy) prečítali. V metóde ošetríte aj prípad, ak `news` nie je aktívna správa, mohla ale byť aktívna v minulosti. Návratový typ metódy je celé číslo.

get_number_of_interested(category) – 0,25b

Metóda vráti počet ľudí, ktorí sa zaujímajú o kategóriu správ danú ako parameter metódy `category`. Návrátový typ metódy je celé číslo.

Súbor `population.py` obsahuje ešte definíciu metódy `HomogeneousPopulation`, ktorá je bližšou špecifikáciou (a podtriedou) triedy `Population`, a predstavuje populáciu, v ktorej sa každý člen zaujíma o istú kategóriu správ. Rozhranie a funkcionálnosť triedy sú nezmenené, avšak konštruktor triedy je rozšírený o ďalší parameter `category`, ktorý predstavuje kategóriu, o ktorú sa zaujímajú všetci. V konštruktoze sa následne táto hodnota uloží do členskej premennej `self.category`.

V triede potrebujete znova definovať metódu **`generate_population`** (hodnotenie **0,5b**), pričom môžete vychádzať z implementácie v triede `Population`, dbajte ale na to, aby každý člen populácie mal v zozname pre neho zaujímavých tém spoločnú kategóriu záujmu pre danú populáciu.

4. úloha – simulácie

Definície funkcií implementujúce rôzne simulácie sa nachádzajú v súbore `simulation.py`. Vašou úlohou je implementovať tieto funkcie.

`simulate_spread(all_news, population)` – 2b

Funkcia slúži na skúmanie dynamiky šírenia správ (v zozname `all_news`) v populácii (`population`). Parametre funkcie majú nasledovnú formu a význam:

- `all_news` – zoznam inštancií `News`, ktoré sa budú šíriť v populácii
- `population` – objekt typu `Population` (resp. `HomogeneousPopulation`), teda skupina ľudí.

Funkcia sa snaží zodpovedať otázku, koľko ľudí si prečíta správy zo zoznamu správ v populácii po jednotlivých krokoch (`time_step`) simulácie. Simulácia pritom bude mať nasledovnú štruktúru:

1. na začiatku simulácie pridajte správy do populácie a získajte zoznam ľudí, ktorí správu prečítajú ako prví
2. opakovane nasimulujte jeden krok simulácie, pričom v jednom kroku sa majú vykonať nasledovné operácie:
 - zaznamenajte počet ľudí, ktorí správu prečítali pre každú správu zo zoznamu `all_news`
 - nasimulujte moment, keď si ľudia, ktorí správu dostali, ju prečítajú, a získajte zoznam ľudí, ktorým správu prepošlú
 - zapamätajte si zoznam ľudí, ktorým sa jednotlivé správy prepošlú v ďalšom kole simulácie
 - aktualizujte zoznam správ v populácii (vymažte tie, ktoré už nie sú aktuálne)
3. simulácia trvá dovtedy, kým v populácii je niektorá správa aktívna

Návratová hodnota funkcie je informácie o zmene počte ľudí, ktorí jednotlivé správy prečítali a to v nasledovnej štruktúre:

- návratová hodnota je slovník (`dictionary`)
- kľúčmi slovníka sú objekty typu `News` zo zoznamu `all_news`
- pod kľúčmi sú uložené zoznamy celých čísel, ktoré reprezentujú počet ľudí, ktorí si danú správu prečítali po jednotlivých kolách simulácie.

Príklad návratovej hodnoty môžete vidieť nižšie, pričom prvá správa má mieru prekvapenia 0,9 a druhá správa 0,5 – práve preto ju prečíta menej ľudí:

```
{  
<news.News object at 0x000002BCFA430A58>: [0, 5, 36, 190, 439,  
511, 513, 513, 513, 513, 513, 513, 513],  
<news.News object at 0x000002BCFA477898>: [0, 5, 17, 41, 59, 65,  
65, 65, 65, 65, 65, 65]  
}
```

average_spread_with_excitement_rate(excitement_rate, pop_size, friends_count, test_count) – 0,5b

Ďalšia funkcia hľadá priemerný počet ľudí, ktorí si prečítajú správu s danou mierou prekvapenia počas simulácie. Parametre funkcie sú nasledovné:

- excitement_rate – float – miera prekvapenia správy
- pop_size – int – počet ľudí v populácii
- friends_count – int – počet priateľov každého člena populácie
- test_count – int – nepovinný parameter, vyjadruje počet simulácií, z ktorých sa vypočíta priemerný počet čitateľov.

Vo funkcii vykonajte opätovné simulácie (počet daný parametrom test_count), pričom v každej iterácii potrebujete:

1. vytvoriť správu s náhodnou kategóriou, zadanou mierou prekvapenia, s dĺžkou aktuálnosti 10 a časom vytvorenia 1
2. vytvoriť populáciu danej veľkosti a počtom priateľov za člena (každá iterácia bude pracovať s inou populáciou)
3. nasimulovať šírenie vygenerovanej správy a získať finálny počet ľudí, ktorí si danú správu prečítali.

Funkcia má dve návratové hodnoty:

- prvá hodnota je zoznam finálneho počtu čitateľov za jednotlivé testovacie iterácie (dĺžka zoznamu bude zodpovedať parametru test_count)
- priemerná hodnota prvej návratovej hodnoty.

excitement_to_reach_percentage(percentage, pop_size, friends_count) – 0,5b

Funkcia hľadá odpoveď na otázku, minimálne akú mieru prekvapenia potrebuje mať správa, aby sa dostala k istému podielu čitateľov zaujímajúcich sa o jej kategóriu. Parametre funkcie sú nasledovné:

- percentage – float medzi 0 a 1 – podiel ľudí, ktorí sa môžu zaujímať o správu a chceme ich osloviť, napr. ak má hodnotu 0,5, znamená to, že správu chceme dostať minimálne k polovici všetkých ľudí, ktorí sa o danú tému zaujímajú
- pop_size – integer – veľkosť populácie (počet ľudí)
- friends_count – integer – počet priateľov jedného člena populácie.

Vo funkcii potrebujete postupne vyskúšať rôzne miery prekvapenia správ a zistiť, či sa s danou hodnotou dosiahne potrebný podiel čitateľov. Vzhľadom na náhodnosť simulácií táto hodnota sa bude trochu líšiť pri rôznych vykonávaníach funkcie, stačí ale ak s každou mierou urobíte jednu simuláciu, odchýlka bude len minimálna.

Pokus so správou s istou mierou prekvapenia bude mať štruktúru:

1. vytvorte populáciu s danými parametrami
2. vytvorte správu s náhodnou kategóriou, danou mierou prekvapenia, s dĺžkou aktuálnosti 10 a časom vytvorenia 1
3. získajte počet ľudí, ktorí si správu prečítajú do konca jednej simulácie
4. získajte počet ľudí z populácie, ktorí sa o tému vygenerovanej správy zaujímajú
5. zistite, či správu prečítala požadovaná časť podskupiny ľudí
6. ak podiel bol dosiahnutý, funkcia vráti mieru prekvapenia; ak so žiadnou mierou prekvapenia sa nedosiahne požadovaný podiel, funkcia vráti hodnotu None.

Poznámka: Pri skúšaní rôznych mier prekvapenia inkrementujte hodnotu postupne o 0,01, odporúčame použiť numpy-ovskú reprezentáciu floatov pre vyššiu presnosť a ľahšiu prácu s počítadlom.

`excitement_to_reach_percentage_special_interest(percentage, pop_size, friends_count, news_category) – 0,5b`

Posledná funkcia rieši podobnú otázku ako funkcia `excitement_to_reach_percentage`, avšak robí to v populácii typu `HomogeneousPopulation` a pre správu so zadanou témou (definovaná v parametri `news_category`).

Rozdiely oproti predošlej funkcii sú teda nasledovné:

- vygenerujete populáciu typu `HomogeneousPopulation`, kde sa každý člen zaujíma o tému definovanú v parametri `news_category`
- vygenerovaná správa musí mať kategóriu podľa parametra `news_category`.

Štruktúra simulácie a návratová hodnota funkcie sa nemenia.

Súbor `simulation.py` obsahuje aj `main` funkciu, ktorú môžete využiť pre ľubovoľné pokusy. Vaše riešenia ďalej môžete otestovať pomocou ukázkových testov, ktoré sú vytvorené pre jednotlivé súbory projektu riešenia.

Za implementáciu potrebných metód a funkcií môžete získať spolu 9 bodov. Ak dosiahnete minimálne 1 bod, 1 bod navyše dostanete automaticky.