Tehtävä 2: Geneettiset algoritmit

On olemassa ongelmia, joiden ratkaisu joudutaan etsimään kokeilemalla erilaisia ratkaisuvaihtoehtoja. Tällaisia ongelmia ovat esimerkiksi edullisen matkustusreitin etsiminen, tavaroiden pakkaaminen pieneen tilaan tai pelien pelaamisstrategiat.

Ratkaisujen etsimiseen voidaan käyttää erilaisia menetelmiä. Erästä menetelmää kutsutaan geneettiseksi algoritmiksi. Menetelmän nimi juontaa alkunsa sen yhtäläisyyksistä biologiseen geneettiseen periytymiseen. Periytymisessä geenit siirtyvät vanhemmilta lapsille muokkautuen tiettyjen periaatteiden mukaan. Lapsen geenit ovat tulos vanhempien geenien risteytymisestä, jolloin osa geeneistä tulee toiselta vanhemmalta ja osa toiselta. Geenien periytymisessä voi myös tapahtua mutaatioita, joka tarkoittaa geenin satunnaista muuttumista toiseksi geeniksi - jollaista ei välttämättä ole kummallakaan vanhemmalla.

Geneettisessä algoritmissa sovelletaan biologisen geneettisen periytymisen periaatteita ja operaatioita ongelmanratkaisuun. Aluksi ongelma tai oikeastaan sen ratkaisuvaihtoehdot muotoillaan sopivaksi "geneettiseksi" merkistöksi - useimmiten luvuiksi. Näistä ratkaisuvaihtoehdoista etsitään parasta tai hyvää ratkaisua ongelmalle.

Yksinkertainen tapa esittää ongelman ratkaisu numeroina on binäärikoodaus, jossa kutakin ratkaisuvaihtoehdon osaa kuvaava luku voi saada arvot 0 ja 1. Jos esimerkiksi ongelman ratkaisuissa on neljä kyllä/ei -vaihtoehtoa, niin ne voidaan esittää neljällä merkillä. Kaikki mahdolliset vaihtoehdot ovat tällöin: 0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111, 1000, 1001, 1010, 1011, 1100, 1101, 1110 ja 1111.

Ratkaisun etsiminen aloitetaan tuottamalla joukko mahdollisia ratkaisuja jollain sopivalla tavalla. Seuraavaksi arvioidaan näiden ratkaisujen hyvyyttä. Hyvyyden arviointi tehdään ongelmaa varten mietityllä sopivalla funktiolla. Tuotetuista ratkaisuista valitaan hyvyysfunktion avulla parhaat jatkokäsittelyä varten. Hyvyysfunktio voi olla joko yksinkertainen summalauseke tai monimutkainen laskukaava. Esimerkiksi edellä esitetyn ongelman vaihtoehtoja voitaisiin arvioida siten, että jokaisesta ykkösestä saa 5 pistettä ja jokaisesta nollasta 2 pistettä. Silloin ratkaisuvaihtoehdon 0110 hyvyydeksi tulisi 14 ja ratkaisuvaihtoehdon 1101 hyvyydeksi 17.

Seuraavana askeleena algoritmissa on parhaiden ratkaisujen risteyttäminen keskenään, jolloin saadaan uusia ratkaisuvaihtoehtoja. Tarkastellaan tilannetta, jossa on saatu kaksi ratkaisuvaihtoehtoa 0100 ja 1010. Risteyttämällä ne ensimmäisen ja toisen merkin välistä (0:100 ja 1:010), saadaan kaksi uutta ratkaisuvaihtoehtoa: 0010 (ensimmäisen alkuosa ja jälkimmäisen loppuosa) ja 1100 (jälkimmäisen alkuosa ja ensimmäisen loppuosa). Risteytyksiä

voidaan tehdä useista kohdista, jolloin saadaan useita ratkaisuvaihtoehtoja arvioitavaksi.

Risteytyksen lisäksi geneettisessä algoritmissa käytetään satunnaisesti - yleensä varsin harvoin - mutaatio-operaatiota. Siinä ratkaisuvaihtoehdon johonkin kohtaan tehdään yksittäinen satunnainen muutos merkistä toiseksi. Esimerkiksi jos ratkaisuvaihtoehdossa 1001 tapahtuu mutaatio kolmannessa merkissä, niin tuloksena on uusi ratkaisuvaihtoehto 1011. Mutaation avulla voidaan löytää uusia hyviä ratkaisuvaihtoehtoja, jos pelkkä risteytyminen ei tuota niitä riittävän monipuolisesti. Mutaation tuloksena syntyneitä ratkaisuvaihtoehtoja arvioidaan hyvyysfunktiolla vastaavalla tavalla kuin risteytyksenkin tuloksia ja parhaat vaihtoehdot - tuottamistavasta riippumatta - otetaan jatkokäsittelyyn.Geneettinen algoritmi etenee siis seuraavasti:

1. Mietitään tapa esittää ongelman ratkaisut sopivalla merkkijonolla käyttäen sopivaa merkistöä. Kehitetään ratkaisujen hyvyyden arviointiin sopiva funktio.

2. Tuotetaan muutamia sopivia tai hyvältä tuntuvia ratkaisuja.

3. Arvioidaan näiden ratkaisujen hyvyys hyvyysfunktiolla.

4. Risteytetään parhaat ratkaisut keskenään, mahdollisesti useita kertoja eri kohdista. Satunnaisin harvoin väliajoin tehdään myös mutaatio ja otetaan sen tulos mukaan ratkaisuvaihtoehtoihin.

5. Arvioidaan uusia ratkaisuvaihtoehtoja tutkimalla onko löydetty paras mahdollinen tai riittävän hyvä ratkaisuvaihtoehto. Ellei tällaista ole löytynyt, niin jatketaan kohdasta 4.

6. Päätetään algoritmin toiminta kun paras tai riittävän hyvä ratkaisu on löytynyt.

Askeleeseen 5 sisältyy yleensä myös “ikuisen toiston estäminen”, jossa algoritmin suoritus lopetetaan ja lopputuloksena on paras siihen mennessä löydetty ratkaisu. Näin tehdään, jos algoritmi ei löydä määritellyssä ajassa uusia parempia ratkaisuja. Käytännön tilanteissa geneettinen algoritmi usein tutkii hyvin suuren joukon erilaisia ratkaisuvaihtoehtoja ennen parhaan vaihtoehdon löytymistä.

Tarkastellaan seuraavaksi pientä esimerkkiä geneettisen algoritmin käytöstä. Ongelman ratkaisuvaihtoehdot koodataan neljän merkin pituisella binäärikoodauksella. Hyvyysfunktion arvo saadaan lisäämällä jokaisesta ratkaisussa olevasta ykkösestä 5 pistettä ja jokaisesta ratkaisussa olevasta nollasta 2 pistettä.

Alkutilanteessa tuotetaan seuraavat 4 ratkaisua A, B, C, ja D:

Ratkaisut:	A: 0110	B: 1101	C: 0010	D: 0001
Hyvyysarvot:	14	17	11	11

Risteytetään A ja B toisen ja kolmannen merkin välistä jolloin saadaan ratkaisut E ja F

	A: 01:10	B: 11:01
Risteytyksen tulokset:	E: 0101	F: 1110
Hyvyysarvot:	14	17

Parhaat löydetyt ratkaisut tähän mennessä ovat A, B, E ja F:

	A: 0110	B: 1101	E: 0101	F: 1110
Hyvyysarvot:	14	17	14	17

Risteytetään B ja F kolmannen ja neljännen merkin välistä, saadaan ratkaisut G ja H:

	B: 110:1	F: 111:0
Risteytyksen tulokset:	G: 1100	H: 1111
Hyvyysarvot:	14	20

Löydettiin paras ratkaisu H: 1111 jonka hyvyysarvo on 20.

Samaan lopputulokseen olisi päädytty toisella tavalla, jos B:lle olisi tapahtunut mutaatio kolmannen merkin kohdalla ja olisi saatu ratkaisu I:

B: 11:0:1
I: 1111

Tehtävä 2: Kysymykset

1. Tarkastellaan geneettisen algoritmin soveltamista seuraavaan ongelmaan:

   Olet lähdössä kaveriporukalla juhannusfestareille yhdellä autolla. Auton tavaratilan kantavuus on 50 kg. Sinulle on tarjolla mukaan matkalle seitsemän kassia, jotka painavat 5, 12, 26, 10, 17, 8 ja 13 kg. Ongelmasi on valita mukaan otettavat kassit niin, että tavaratilaan saisi kasseissa mahdollisimman monta kilogrammaa tavaraa. Kassien kokoa ei tarvitse ottaa huomioon.

   (a) Millaisella merkistöllä ja merkkijonolla esittäisit ongelman ratkaisuvaihtoehtoja, jotta se voitaisiin ratkaista geneettisellä algoritmilla? (6 pistettä)

   (b) Esitä kohdassa a) kuvaamallasi tavalla seuraavat ratkaisuvaihtoehdot: (2 pistettä)

      i. Mukaan otetaan kassit jotka painavat 5, 26 ja 17 kiloa.

      ii. Mukaan otetaan kassit jotka painavat 12, 26, 17 ja 13 kiloa.

   (c) Millaisella funktiolla arvioisit a) kohdassa muotoilemiesi ratkaisuvaihtoehtojen hyvyyttä? (6 pistettä)

   (d) Mitkä ovat hyvyysfunktiosi arvot kohdan b) ratkaisuvaihtoehdoille? (2 pistettä)

2. Vektoripelissä pyritään arvaamaan mikä on vastustajan valitsema kuusinumeroinen binääriluku. Binääriluku on luku joka sisältää vain numeroita 0 ja 1. Ratkaisuvaihtoehtoja voisivat olla esim. luvut: 010110 ja 101101. Pelattaessa peliä kahdestaan toinen pelaaja valitsee arvattavan luvun, jonka toinen yrittää arvata. Kunkin arvauksen jälkeen hän kertoo kuinka monta numeroa arvauksessa on oikein, mutta hän ei kerro oikeiden numeroiden paikkoja. Näin jatketaan oikeaan arvaukseen saakka. Pelissä tavoitteena on keksiä oikea ratkaisu mahdollisimman vähillä arvauksilla.

   Tehtävänäsi on etsiä vektoripelissä oikea ratkaisu geneettisellä algoritmilla. Lähdet liikkeelle annetusta alkutilanteesta ja sinun tulee esittää kuinka päädyt geneettisellä algoritmilla oikeaan ratkaisuun. Esitä vaihe kerrallaan mitä geneettistä operaatiota (risteytys tai mutaatio) ja miten sovellat, mikä on operaation tulos ja hyvyysfunktion arvo. Esitä kaikki vaiheet, joiden kautta etenet geneettisten operaatioiden avulla lopulliseen ratkaisuun.

Aloitat pelin seuraavista neljästä arvauksesta: 110100, 111101, 011011 ja 101100. Oikea vastaus, johon sinun tulee päätyä geneettisiä operaatioita soveltamalla, on 001010. Mutaatio-operaatiota saat käyttää vain yhden kerran ja voit muuttaa sillä vain yhtä merkkiä. Ratkaisuvaihtoehdon hyvyyttä arvioiva hyvyysfunktion arvo lasketaan seuraavasti. Jos tietyllä kohdalla oleva numero on oikein, niin se lisää hyvyysfunktion arvoa +1 pisteellä. Väärin olevat merkit eivät vaikuta hyvyysfunktion arvoon. Jos kaikki merkit ovat väärin saa hyvyysfunktio arvon nolla. (9 pistettä)