

Problēma

Darbā tiek pētīta klasifikācijas problēma. Klasifikācijas problēmā ir dati, kuri ir jāsagrupē, piemēram, klasēs. Tas ir veids, kā analizēt datus. Iespējams klasifikācijas problēmas risinājums ir lēmumu koks. Šādus risinājumus var izmantot GA (ģenētiskie algoritmi), veicot to kombinēšanu, lai iegūtu labākus risinājumus.

Mērķis

Maģistra kursa darba mērķis ir izpētīt GA izmantošanas iespējas klasifikācijā.

Uzdevumi

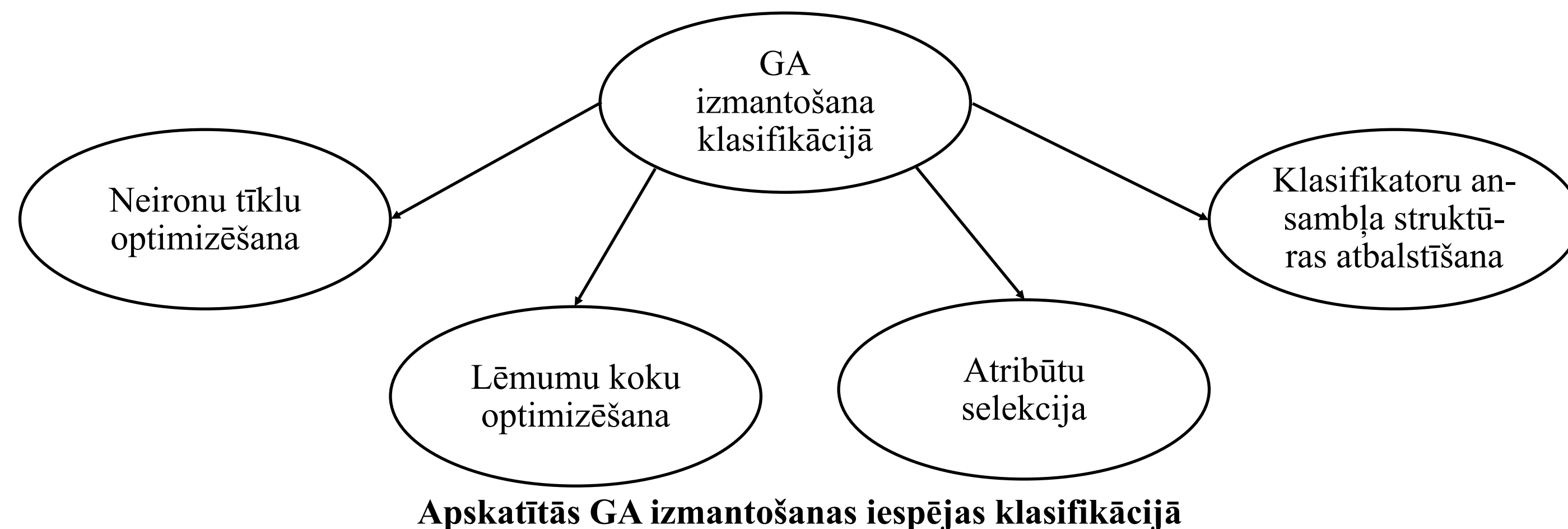
1. Izpētīt veiktos pētījumus par klasifikāciju, kuri izmanto GA;
2. Noteikt galvenās idejas priekš autora risinājuma izstrādes;
3. Iesākt izstrādāt autora risinājumu klasifikācijas problēmai, izmantojot GA.

Rezultāti

Tika apskatīti citi veiktie pētījumi par klasifikāciju, kuri izmanto GA. GA izmantošanas iespējas ir apkopotas zemāk esošajā diagrammā.

Tika noteiktas galvenās idejas priekš autora risinājuma izstrādes. Piemēram, eksperimenti tika veikti ar publiski pieejamām UCI (University of California, Irvine) datu kopām [2]. Tika iesākta autora risinājuma izstrāde klasifikācijas problēmai valodā JAVA, izveidojot GA atribūtu selekcijas risinājumu. Atribūtu selekcija tika veikta WEKA k tu-vāko kaimiņu klasifikācijas algoritmam [1]. Tiek piedāvāta unikāla risinājuma struktūra un jaunas idejas krustmijai un mutācijai. Izstrādāto struktūru var apskatīt pa labi.

Veiktajos eksperimentos apstiprinājās tas, ka GA atribūtu selekcija var izmest lielu daļu atribūtu, kā arī iegūt dažu procentu precizitātes uzlabojumu. Ievērojams rezultāts ir izmesti 70% atribūtu un vairāk daļā gadījumu, bet uzlabojumi precizitātē ne vienmēr tika iegūti. Vislielākais uzlabojums eksperimentos bija par 5% UCI Viskonsinas krūts vēža prognozēšanas datu kopas [2] testēšanas rezultātam.

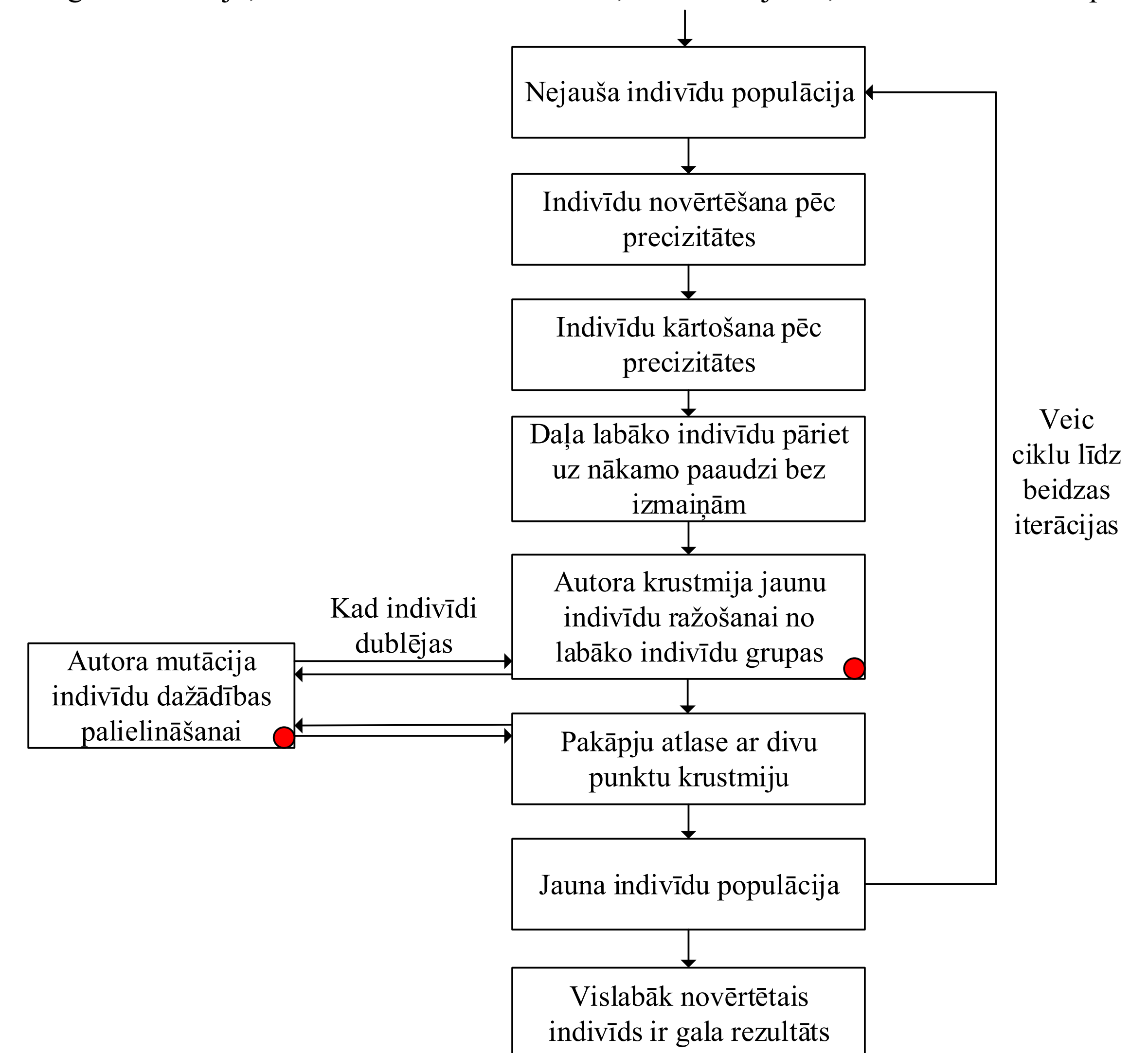


Secinājumi

Problēmas risinājums ir iesākts, piedāvājot GA atribūtu selekcijas risinājumu priekš klasifikācijas. Autora jaunā ideja priekš krustmijas ļauj no mazākas labāko indivīdu grupas izveidot tiem līdzīgu lielāku indivīdu grupu. Izstrādātā autora mutācija ļauj palielināt indivīdu dažādību, likvidējot vienādus indivīdus.

Turpmākie darbi

Izstrādāto atribūtu selekciju var izmantot, kā pirmo soli klasifikācijā. Varētu veidot ansambļa tipa struktūru, kura izmanto klasifikācijas noteikumus, iesaistīt citus klasifikatorus šajā struktūrā, izmēģināt dažādus struktūras variantus, izstrādāt citu derīguma funkciju, veikt indivīdu inicializēšanu, kas nav nejauša, un veikt tālākus eksperimentus.



Sākotnējais izstrādātais autora risinājums ar jaunajām autora idejām (apzīmētas ar sarkaniem punktiem)

Atsauces

1. F. Eibe, H.A. Mark, W.H. Ian, "The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufman, Fourth Edition, 2016," *The University of Waikato*, 2016. Pieejams: https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf.
2. M. Lichman, "UCI Machine Learning Repository," *Irvine, CA: University of California, School of Information and Computer Science*, 2013. Pieejams: <http://archive.ics.uci.edu/ml>.