შავი ზღვის საერთაშორისო უნივერსიტეტი

კომპიუტერული ტექნოლოგიებისა და ინჟინერიის ფაკულტეტი

Ph.D. პროგრამა

უბნობრივი წრფივი აპროქსიმაციის მეთოდის დამუშავება არასტაციონარული დროითი მწკრივების მსგავსების შეფასებისთვის

დანიარ სატიბალდიევი

კომპიუტერული მეცნიერებების გავრცობილი სადოქტორო სადისერტაციო ნაშრომი

თბილისი-2016

სამეცნიერო ხელმძღვანელი: ალექსანდრე მილნიკოვი

(პროფესორი, დოქტორი, შავი ზღვის საერთაშორისო უნივერსიტეტში)

(ხელმძღვანელის ხელმოწერა)

ექსპერტები (სახელი, გვარი & აკადემიური ხარისხი):

1. პროფესორი ნოდარ მომცელიძე
2. ასოც.პროფ.დოქტორ გიორგი მანდარია

(თუ არსებობს)

1. პროფესორი გურამ ლეჟავა
2. ასოც.პროფ.დოქტ ნიკოლოზ აბზიანიძე
3. ასოც.პროფ.დოქტ. ნურლან ატაბაევი

(თუ არსებობს)

შესავალი

დიდი მონაცემების კომპლექტი ყველაზე სასარგებლოა და ინახავს საყურადღებო ინფორმაციას მნიშვნელოვანი სტატისტიკის საწარმოებლად. საინტერესოა, როგორ შეგვიძლია მათი გამოყენება და ანალიზი. გიგანტური კომპანიები, როგორიცაა Apple და Google არიან დიდი მონაცემების ანალიზის ოსტატები. ამის ერთ-ერთი მიზეზია, რომ მათ ძალიან წარმატებულად შეუძლიათ დაინახონ სასიცოცხლო ცოდნა შეგროვილი მონაცემების ოკეანეში. როგორ უნდა განისაზღვროს დიდი მონაცემები? დიდი მონაცემები შედეგია განსხვავებული დაკვირვებების: საბანკო ოპერაციების, სამეცნიერო-კვლევების, სოციალური ქსელის პოსტების და ა.შ. და ამ ფაქტორების შედეგები ექსპონენციალურად იზრდება, ამიტომ არსებული ანალიზისა და შენახვის სისტემები აღარ არის საკმარისი.

ანალიტიკა დიდ მონაცემებთან დაკავშირებულ სამი V-ს პრობლემას აწყდება: volume(მოცულობა), variability(ცვალებადობა) და velocity(სიჩქარე). ესენი წარმოადგენენ დიდი მონაცემების სამ ძირითად ასპექტს, რომლებიც უნდა იყოს გათვალისწინებული, მასთან მუშაობის დაწყებამდე. მონაცემთა ექსპონენციალური ზრდის მიზეზი მათი წარმოების ევოლუციაა , რომელიც პირველი პროგრესიიდან იწყება. ისტორიულად, მონაცემების გამომუშავებას და დაგროვებას კომპანიების თანამშრომლები აწარმოებდნენ, რომლებსაც ხელით შეჰყავდათ საჭირო მონაცემები კომპიუტერულ სისტემაში. მეორე პროგრესია დაიწყო ინტერნეტის გამოგონებიდან, როდესაც ჩვეულებრივმა მომხმარებლებმა შეძლეს საკუთარი მონაცემების გენერირება. მილიონობით მომხმარებელს შეჰყავს თავისი მონაცემები ბევრ საიტებზე, როგორიცაა Facebook, Twitter, და Instagram. და ბოლოს, მესამე პროგრესიია მოხდა ისეთი მანქანების გამოგონების შემდეგ, რომელთაც შეუძლიათ საკუთარი მონაცემების წარმოება. მაგალითად, არსებობს ბევრი მოწყობილობა, რომელიც ახდენს ტენიანობისა და ტემპერატურის მონიტორინგს. შედეგად ხდება კოლოსალური რაოდენობით მონაცემების გენერირება ყოველ წამს. არსებობს სამი სახის მონაცემები. პირველში, ობიექტის კვლევის შედეგი იწერება, დროის კონკრეტულ წერტილში. მეორე, ობიექტების კვლევის სხვადასხვა დროს წერტილის

შედეგი. ბოლო არის მონაცემები დროის სერიების სახით. მას შემდეგ, რაც მონაცემთა დიდი ნაწილი მჭიდროდ უკავშირდება დროს, ის იწერება დროის სერიების სახით. დროის სერია არის მიმდევრობითი დაკვირვებების შედეგი, რომელიც გროვდება გარკვეულ მიმდევრობით ინტერვალში და იზომება დროის თანაბარ მანძილებში. მაგალითად დროის სერია მოიცავს ადამიანის არტერიული წნევის უწყვეტ მნიტორინგს, ტენიანობის საათობრივ აღწერას, კომპანიის აქციების ყოველდღიურ საბოლოო ღირებულებას,უმუშევრობის ყოველთვიურ დონეს და ა.შ.

ზოგადად, დროის სერიის ანალიზი გამოიყენება, როდესაც არსებობს 50-ზე მეტი მონაცემთა წერტილი და ეს კეთდება იმისათვის, რომ გავიგოთ ფუნდამენტური სტრუქტურა და ფუნქცია რომელიც აწარმოებს დაკვირვებას. დროის სერიის მექანიზმების გააზრება საშუალებას გვაძლევს შევადგინოთ ისეთი მათემატიკური მოდელი, რომელსაც შეუძლია წინასწარი პროგნოზი, მონაცემთა მონიტორინგი და კონტროლი. მაგალითად პროგნოზირება, ფართოდ გამოიყენება ეკონომიკაში და ბიზნესში. გარემო პირობების, შესატანი ან გამომავალი მონაცემების მონიტორინგი კი საჭიროა მეცნიერებისა და ინდუსტრიისთვის.

არსებობს დროის სერიის ანალიზის ორი ძირითადი მეთოდი. პირველი, ანალიზი დროის მიხედვით, როცა სერია გაანალიზებულია დროსთან მიმართებაში და მეორე, ანალიზი სიხშირის მიხედვით, რომელიც ეყრდნობა დროის სერიის ანალიზს სიხშირესთან მიმართებაში. ამ მეთოდებს, როგორც წესი იყენებენ თუ დროის სერია სტაციონარულია.

უმეტესად ბიზნეს და ეკონომიკური დროის სერიები შორსაა სტაციონარულობისგან, როდესაც ისინი აღიწერება ორიგინალ ერთეულებში, და დეფლაციის ან რეგულირების შემდეგაც კი მათ, როგორც წესი, კვლავ აღენიშნებათ ტენდენციები, ციკლები, შემთხვევითი(random) ცვლილებები, და სხვა არასტაციონარული ქცევები. არასტაციონარული დროის სერიებისთვის გამოიყენება სხვა მეთოდები. განსაკუთრებით, არასტაციონარული დროის სერიების მსგავსებათა გაზომვისთვის არსებობს ოთხი კარგად ცნობილია მეთოდი. ესენია:

- ევკლიდური მანძილი: პირდაპირ ადარებს ორ, ტოლი D მანძილის დროს სერიას , და, როგორც წესი, საკმარისია აპლიკაციებისთვის, რომლებიც არ მოითხოვენ გამოკვეთილ თვისებებს შორის კორელაციის ჩვენებას;

- დროის დინამიური დეფორმირება: დროის სერიების მსგავსების საზომი ტექნიკა, დროის გარკვეულ მონაკვეთში, მაშინაც კი, თუ ისინი განსხვავდებიან სიგრძით და სიჩქარით;

- A და B დროის სერიის ყველაზე გრძელი საერთო ქვემონაკვეთ არის ყველაზე გრძელი მონაკვეთები A-დან და B-დან, რომლებიც საერთოა ორ დროის სერიაში.

- უბნობრივი წრფივი წარმოდგენის (PLR): ნებისმიერი სიგრძის დროის სერიის წრფივი ხაზებით წარმოდგენა. PLR არის ყველაზე ხშირად გამოყენებული ტექნიკა. მთავარია, რომ წრფივი ხაზების სიგრძე ბევრად ნაკლები უნდა იყოს დროის სერის სიგრძეზე.

ყველა ზემოთ აღწერილი მეთოდი გამოიყენება ბევრ სფეროში, დროის მონაკვეთების შედარებისთვის. ყველა ზემოთ აღწერილ ტექნიკას გააჩნია თავისი დადებითი და უარყოფითი მხარეები. ყველა მეთოდს აქვს თავისი დაბკოლება მაგალითად მასშტაბირება, დროის სერიის მახსიმალური სიგრძე და დავალების სპეციფიკაცია.

ჩვენ გვინდა შევქმნათ ახალი მეთოდი, რომელიც დაეხმარება გამომთველ მანქანას სიზუსტის და მონაცემთა მოპოვების გაუმჯობესებაში.

კონკრეტულ სწავლებაში ყურადღება გამახვილებულია არასტაციონარული დროის სერიების შედარება უბნობრივი წრფივი აპროკსიმაციის მეთოდით. ასევე ნაჩვენებია სპეციალური ტექნიკები არასაჭირო წერტილების მოსაშორებლად. შედარების სპეციალური კრიტერიუმები იქნა გამომუშავებული დროის სერიების მსგავსების კუთხის განსასაზღვრად.

თეზისის სტრუქტურა

პირველი თავი ეხება ლიტერატურულ მიმოხილვას. დროის სერიების ანალიზის თეორია, მათი წარმოდგენის განსხვავებული ტიპები და მონაცემთა

მსგავსების განსხვავებული მეთოდები არის წარმოდგენილი ამ თავში. ისევე როგორც სხვა ბევრი კომპიუტერული მეცნიერების პრობლემების, მონაცემთა წარმოდგენა ეფექტური გადაწყვეტის გასაღებია. რამოდენიმე მაღალი დონის წარმოდგენის მეთოდებია გამოყენებული: Fourier Transforms, Wavelets, Symbolic Mappings and უბნობრივი წრფივი წარმოდგენის(PLR). PLR ძირითადად გამოიყენება მონაცემთა შენახვაში, მათ გადატანაში და გამოთვლაში, რაც გვაძლევს მათი ეფექტური განხილვის საშუალებას.

მთავარ, მეორე თავში არის უბნობრივი წრფივი აპროკსიმაციის(PLA) მეთოდოლოგიის ფორმალური და თეორიტიკული განსაზღვრება. ამ თავში განსაზღვრულია პრობლემა, მოცემულია წინასწარი და რეალური მოდელის ტერმინოლოგია. PLA მეთოდის მთავარი არსია, ჩამოაყალიბოს რეგრესიის მოდელი, რომელშიც დამატებით ფაქტორად გამოყენებული იქნება სეზონური ბუტაფორიული(dummy) ცვლადი. ბუტაფორიული(dummy) ან ინდიკატორი ცვლადების შეყვანმდე, მათ უნდა მიენიჭოთ ციფრული ეტიკეტი. სხვა სიტყვებით რომ ვთქვათ თვისობრივი ცვლადები გადაყვანილი უნდა იქნას რაოდენობრივში.

მესამე თავში წარმოდგენილია უბნობრივი წრფივი აპროკსიმაციის და ოპტიმიზაციის ეხპერიმენტალური აპლიკაცია. რომელიც ეყრდნობა მეორე თავში ახსნილ მეთოდოლოგიას. სანიმუშო მოდელის და გაანალიზებული მოდელის შედარება, მსგავსებების დადგენა და გამოთვლილი დროის სერიების უბნობრივი წრფივი აპროკსიმაცია საყრდენი წერტილების სიზუსტის დასადგენად. წრფივი სეგმენტების შედეგების ოპტიმიზაცია შემდეგ წარმოებულ ქეისებში გამოსაყენებლად. წარმოდგენილია ყველა ქეისის ემპირიული მაგალითი. წარმოდგენილია სხვადასხვა ქვეყნის ვალუტის კურსის დროის სერიების შედარების მაგალითები, სადაც დროის სერიის ნიმუშად აღებულია ა.შ.შ. დოლარისა და ქართული ლარის გაცვლითი კურსი.

პრობლემის აქტუალურობა:

მსოფლიო მონაცემები ორმაგდება ყოველ 1.2 წელიწადში. დედამიწის მოსახლეობა შეადგენს 7 მილიარდს. 5.1 მილიარდი მათგანი ფლობს მობილურ ტელეფონს. ყოველდღიურად ჩვენ ვაგზავნით 11 მილიარდ ტეხტურ შეტყობინებას, ვუყურებთ 2.8 მილიარდზე მეტ ვიდეოს YouTube-ზე და ვაწარმოებთ თითქმის 5 მილიარდ ძებნას Google-ში. ადამიანები არა მარტო იღებენ ინფორმაციას, არამედ აწარმოებენ მას. ინფორმაციის წარმოების და დაგროვების ექსპონენციალური ზრდის გამო, დიდი ყურადღება ექცევა რეალურად მნიშვნელოვანი ინფორმაციის წამოღებას მონაცემთა ბაზიდან. ობიექტების ვარიაცია შეიძლება იყოს საძებნი ინფორმაცია: ტექსტი, სურათი, სიტყვა, ა.შ. თუმცა მონაცემთა უმეტესობა მოდის დროის სერიების სახით, რადგან ჩვენი ცხოვრება მჭიდრო კავშირშია დროსთან.

დროის სერიები შეიცავენ დიდ და მნიშვნელოვან ინფორმაციას, რომლის ვერბალიზება შეუძლებელია. ინფორმაცია შეიძლება იყოს ჩვენი ცხოვრების განსხვავებული არეალიდან: ფიზიკიდან, ეკონომიკიდან, ფინანსებიდან და ა.შ. ამიტომ დიდი მონაცემთა ბაზებიდან სასიცოცხლო ინფორმაციის ამოღება მონაცემთა მოპოვების ერთერთი მთავარი ამოცანაა. როდესაც ვუყურებთ საფონდო ბირჟაზე აქციების ცვლილების დროის სერიას, გვინდა გავიგოთ მათი ყველაზე მაღალი ფასი . ეს მხოლოდ ერთი მაგალითია და რეალურად მათი რაოდენობა ბევრად მეტია.

დროის სერიათა ანალიზი ახალი ტექნოლოგიების გამოგონებასთან ერთად ვითარდება. მონაცემთა ნაკრებები უფროდაუფრო დიდი ხდება. კომპანიები სჭირდებათ უფროდაუფრო მეტ მონაცემთან გამკვლავება მათ ბაზებში. ამისათვის ისინი ეძებენ მონაცემთა მოპოვების ახალ იდეებს, ალგორითმებს და ამოცანებს. დროის სერიათა მსგავსებათა ანალიზი ასევე გამოიყენება უფრო დიდ ანალიტიკაშიც, როგორიცაა დრო სერიის პროგნოზირება, კლასტერირება ან კანონზომიერების აღმოჩენა. დოის სერიების მსგავსებადობის ანალიზის სისწრაფის და სიზუსტის მოთხოვნილება უბიძგებს ახალი ალგორითმების ძებნისკენ მონაცემთა მოპოვების ინდუსტრიას. დროის სერია T1 და დროის სერია T2-ის შედარების უამრავ მეთოდს

შორის უნდა იყოს ახალი მეთოდი, რომელიც მოახდენს სისწრაფისა და სიზუსტის ოპტიმიზაციას.

ეს ნაშრომი ეხება მსგავსების გაზომვის ოპტიმიზაციასა და აჩქარებას უბნობრივი წრფივი მეთოდის გამოყენებით, განსაკუთრებით არასტაციონარული დროის სერიებისთვის. კვლევა გავრცობილია შედარების კრიტერიუმის გამომუშავებით დოის სეარიათა მსგავსების კუთის განსასაზღვრად.

მეთოდოლოგია

მოცემული მეთოდოლოგია ეყრდნობა სტატისტიკური დროის სერიების ანალიზის მეთოდებს და პრინციპებს, კალკულუსის მეთოდებს, წრფივ ალგებრას და მათემატიკურ სტატისტიკას, მრავალგანზომილებიანი ბუტაფორიული(dummy) ცვლადის რეგრესიის ანალიზში.

უბნობრივი წრფივი წარმოდგენის მეთოდისათვის გამოიყენება MatLab-ის პროგრამული ენა. ის ასევე გამოიყენება არასაჭირო საყრდენი წერტილების გამოსავლენად უბნობრივ წრფივ აპროკსიმაციის მეთოდში და შედარების მთლიანი პროცესის ოპტიმიზაციისთვის. MatLab ასევე გამოიყენება F-ტესტის შედეგი ბუტაფორიული(dummy) საყრდენი წერტილის წარმოსადგენად და მის ცხრილის მნიშვნელობასთან შესადარებლად. ძირითადად გამომუშავებული იქნა 6 განსხვავებული ქეისი მსგავსებათა გაზომვისათვის, როგორებიცაა: სტრუქტურულად იდენტური, სტრუქტურულად იდენტური პროპორციულ ფერდობებზე, სტრუქტურულად იდენტური პატარა შეცდომით, სტრუქტურულად მსგავსი, ნაწილობრივ მსგავსი და სტრუქტურულად განსხვავებული. T-ტესტი გამოყენებულ იქნა სტრუქტრულად იდენტური პატარა შეცდომით ქეისის განსასაზღვრად. რეგრესიის ანალიზის განხორციელებისას და უბნობრივი წრფივი აპროკსიმაციისას სიმარტივისთვის MatLab-ის კოდი იქნა გამოყენებული. კოდის შესრულების შედეგები წარმოდგენილია როგორც საბოლოო სტატისტიკა ქეისებში.

სწავლების მიზანი

ამ სწავლების მთავარი მიზნებია: (I)ახალი მეთოდის შემუშავება, რომელიც ეყრდნობა წრფივი უბნობრივი წარმოდგენის მეთოდს; (II) მონაცემთა ბაზიდან აღებული დროის სერიების მსგავსების კუთხის შედარების ახალი კრიტერიუმის შემუშავება.

სწავლების ამოცანები:

- უბნობრივი წრფივი სეგმენტების ფორმით არაწრფივი გაფანტული წერტილების წარმოდგენა;
- ახალი მეთოდის გამომუშავება, არასაჭირო საყრდენი წერტილების მოსაშორებლად. საყრდენი წერტილები არის წერტილები, რომლებითაც იწყება და მთავრდება წრფივი სეგმენტი;
- ახალი მათემატიკური გზების გამომუშავება, დროის სერიების მსგავსების დონის დასადგენად;
- არასტაცონარული დროის აპროკსიმაციის პროცესის გამარტივება უბნობრივი წრფივი აპროკსიმაციის ოპტიმიზირებით;
- დროის სერიათა უბნობრივი წრფივი აპროკსიმაციის შედეგების შედარების მეთოდის განხორციელება;
- მსგავსების კრიტერიუმის განსაზღვრა მიახლოებითი დროის სერიების დაჯგუფებისთვის;
- MatLab პროგრამირების ენის გამოყენება, ამოცანების განხორციელებისთვის საჭირო კოდის დასაწერად;

კვლევის სიახლე და წვლილი:

1. დროის სერიების მონაცემები წარმოდგენილი იქნა წრფივი სეგმენტების სახით უბნობრივ წრფივ წარმოდგენაზე დაყრდნობით.

2. რამოდენიმე ქეისი იქნა ჩამოყალიბებული ახლად გაცნობილ უბნობრივ წრფივ წარმოდგენაზე დაყრდნობით. ეს ქეისებია: სტრუქტურულად იდენტური, სტრუქტურულად იდენტური პროპორციულ ფერდობებზე, სტრუქტურულად იდენტური პატარა შეცდომით, სტრუქტურულად მსგავსი, ნაწილობრივ მსგავსი და სტრუქტურულად განსხვავებული.
3. ბუტაფორიულ(dummy) მრავალგანზომილებიან წრფივ რეგრესიაზე დაყრდნობით, არასტაციონალური დროის სერიების მსგავსების მეთოდი იქნა შექმნილი. მოდელის შემადგენლობა რამოდენიმე რიცხობრივ მოდელშია დემონსტრირებული;
4. ახალი წარმოდგენის მეთოდის შედეგები შედარებულ იქნა უკვე არსებული მეთოდების შედეგებთან. დაკვირვებებმა აჩვენა,რომ უბნობრივი წრფივი აპროკსიმაციის ინტეგრირებულ მოდელს მარტივად შეუძლია მოიპოვოს დროის სერიის შედეგები და გადათარგმნოს ისინი მანქანური ინტერპრეტაციისთვის.
5. MATLAB-ის პროგრამულ ენაში შექმნილი იქნა საჭირო პროგრამული ინსტრუმენტები.
6. დროის სერიების უბნობრივი წრფივი აპროკსიმაციის მეთოდზე დაყრდნობით მოხდა დროის სერიების მსგავსების კუთხის შედარების ახალი კრიტერიუმის შემუშავება.
7. ფიშერის მეთოდზე დაყრდნობით მოხდა ახალი მეთოდოლოგიის ჩამოყალიბება შემუშავებული მეთოდის ადეკვატურობის შესამოწმებლად.

სწავლების თეორიული ღირებულება და პრაქტიკული მნიშვნელობა:

მოცემული სწავლება ახალი ნაბიჯია დროის სერიათა მონაცემთა მოპოვებაში, განსაკუთრებით მათი მსგავსების ძებნაში; ის ბევრად აასწრაფებს პროცეს რაც ხელს შეუწყობს მონაცემთა ბაზების ანალიზს. არაწრფივი ერთი ცვლადის ფუნქციის

ახალი, უბნობრივი წრფივი აპროკსიმაციის მეთოდი იქნა წარმოდგენილი. წარმოდგენილი აპროკსიმაციის მეთოდის სარგებლები წრფივი ტეხილების აგებაში არის ის, რომ იგი იყენებს n-განზომილებიან წრფივი რეგრესიის ანალიზის პროცედურას და არ მოითხოვს აპროკსიმაციის კვანძების შეზღუდვას. ამჟამინდელი ჩამოყალიბებული მეთოდი საშუალებას გვაძლევს შევადაროთ ორი დროის სერია; ამიტომ ჩვენ შეგვიძლია შევადაროთ მონაცემთა ბაზის შეკითხვის დროის სერია და მასში არსებული ნებისმიერი დროის სერია. მაგალითად, ავიღოთ ორი X და Y დროის სერია. შეკითხვა იქნება მოძრაობენ თუ არა X და Y აქციები თანაბრად? ამ ნაშრომში აღწერილი მსგავსებათა შედარების მეთოდი საშუალებას გვაძლევს დროის სერიების ბუნდოვანი შედარების. შემდეგში, შესაძლებელია მისი გამოყენება ინდექსინგში, თანმიმდევრული მსგავსების დასადგენად, კლასტერინგისთვის და კანონზომიერებათა აღმოჩენის პრობლემაში. რადგან მონაცემთა მოპოვების პროცესის სისწრაფე ჩვენი მეთოდის მთავარი ასპექტია, ნაცადი იქნა ბალანსის დაცვა პრობლემის გადაწყვეტის სიზუსტესა და ეფექტურობას შორის. წარმოდგენილი მსგავსების კრიტერიუმი უფრო მარტივად და ეფექტურად განსაზღვრავს დროის სერიათა მსგავსების დონეს.

სწავლება შეიძლება განსაკუთრებით მნიშვნელოვანი იყოს, როგორც ალტერნატიული ინსტრუმენტი დროის სერიათა ანალიზის, ინდექსინგის და კლასტერინგის კონკრეტულ სფეროებში, როგორც ბიზნეს ანალიზი, სამეცნიერო კვლევები, პრაქტიკოსებისთვის და ა.შ.

კვლევის ძირითადი აღმოჩენები და დასკვნები წარმოდგენლი იყო რამოდენიმე კონფერენციაზე. უფრო მეტიც, კვლევის შედეგები ხაზგასმით იყო წარმოდგენილი სამ სტატიაში, რომლებიც გამოქვეყნდა ეროვნულ და საერთაშორისო ცნობილ აკადემიკურ ჟურნალებში.

ნაშრომის სტრუქტურა და მოცულობა

თეზისის მოცულობაა 134 გვერდი და შედგება 3 თავის, ლიტერატურის სიის, მონაწილეების სიის და ცხრილების სიისაგან.

თავი 1: ლიტერატურული მიმოხილვა

ამჟამად, კომპლექსური სისტემების გამოსაკვლევად, ექსპერიმენტალური სწავლებების ჩათვლით, ფართოდ გამოიყენება მრავალი სისტემის მიერ წარმოებული მონაცემთა ანალიზი. ხანდახან, შეიძლება გვქონდეს დაკვირვების ქვეშ მყოფი მნიშვნელობის რამოდენიმე აღწერილობა, მაგრამ შეუძლებელი იყოს პროცესის მათემატიკური აღწერა. სისტემური ანალიზი, განსაკუთრებით ექსპერიმენტალური კვლევები ხშირად რეალიზებულია სიგნალების აღწერის პროცესით. ჩვეულებრივ, ამ სიგნალებს ქვია დაკვირვებადი, ხოლო კვლევის მეთოდს - დინამიური სისტემების წარმოდგენა. დინამიური სისთემების თეორიის ამ ნაწილს ქვია დროის სერიების ანალიზი. დროის სერიების ანალიზის განვითარებამ, როგორც მეცნიერებამ ბოლო დეკადების განმავლობაში ბევრი მეთოდის, პროცედურის და პროგნოზირების ტექნიკის ჩამოყალიბებას დაუდო საფუძველი. კვლევებზე დაყრდნობით, დროის სერიის ანალიზი უკვე 100-ზე მეტ მეთოდს მოიცავს. კვლევის მაღალი ხარისისათვის და ანალიზის სიზუსტისათვის მასში გაერთიანებული უნდა იყოს მათემატიკა, ეკონომიკა და სტატისტიკა. აშკარაა, რომ შეუძლებელია კომპლექსური, არაწრფივი, მრავალვარიაციული მოდელის ანალიზი მის ჩვეულებრივ, ნორმალურ ფორმაში.

დროის სერიის ეფექტური და ხარისხიანი ანალიზისათვის ძალიან მნიშვნელობანია მონაცემების სწორი წარმოდგენა. მრავალი წარმოდგენის მეთოდი იყო ნაჩვენები: ფურიერული გარდაქმნა, ვეივლეტი, სიმბოლური მეპინგი, უბნობრივი წრფივი წარმოდგენა. დროის სერიათა წარმოდგენა შესამჩნევად აუმჯობესებს დროის უბნების მსგავსებათა ძებნას. მსგავსებების ძებნა ეფექტურია სხვა გამოთვლით პროცესებში როგორებიცაა: ინდექსაცია, ქვეუბენზის მსგავსება, კლასტერინგი, კანონზომიერებათა აღმოჩენა ა.შ.

ამჟამად, კომპლექსური სისტემების გამოსაკვლევად, ექსპერიმენტალური სწავლებების ჩათვლით, ფართოდ გამოიყენება დროის სერიათა ანალიზი. მონაცემთა ანალიზი განსაკუთრებით ექსპერიმენტალურ კვლევებში ხშირად ჩაწერილი დაკვირვებებით არის რეალიზებული. მაგალითად, კარდიოლოგიაში ამ მიზნით ელექტროდიაგრამები გამოიყენება, სეისმოლოგიაში - დედამიწის ქერქის მერყეობა,

მეტეოროლოგიაში - მეტეოროლოგიური დაკვირვებები. ასეთ სიგნალებს დაკვირვებად სიგნალებს უწოდებენ. ამ სიგნალების კვლევის მეთოდს დინამიური სისტემის რეკონსტრუქცია ჰქვია. დინამიური სისტემების ამ თეორიას თავის მხრივ დროის სერიების ანალიზი ჰქვია. ამ კონტექსტში 'დაკვირვებადი' ნიშნავს ცვლადების მნიშვნელობათა თანმიმდევრობას, რომლებიც დროის თანაბარ ინტერვალში აღიწერება. აქედან გამომდინა, 'დაკვირვებადი'-ს ნაცვლად 'დროის სერიები' გამოიყნება. დროის სერია არის რიცხვობრივი მონაცემთა წერტილების მომდევრობა, რომელიც ხშირად ჩნდება ერთგვაროვან ინტერვალებში. სხვა სიტყვებით რომ ვთქვათ, დროის სერია არის გარკვეულ მონაკვეთში დროის თანაბარი ინტერვალებით შეგროვილი რიცხვების მიმდევრობა.

დროის სერიის ანალიზი სხვა შემთხვევითი ნიმუშების ანალიზებისაგან განსხვავებით ეყრდნობა ვარაუდს, რომ მიმდევრობითი მნიშვნელობები მონაცემთა ფაილში წარმოადგენს აღწერას თანაბარ ინტერვალებში. დროის სერიების ანალიზი გამოსადეგია, რომ დავინახოთ, თუ როგორ იცვლება ცვლადი დროის განმავლობაში და ასევე სხვა ცვლადებთან მიმართებაში. მაგალითად, ვთქვათ, თქვენ გინდათ დროის სერიის საშუალებით გააანალიზოთ აქციების ყოველდღიური დახურვითი ფასები მთელი ლის განმავლობაში. თქვენ მიიღებთ ყოველდღიური საბოლოო ფასების ცხრილს, დალაგებულს ქრონოლოგიურად. ასევე თქვენ შეიძლება გაინტერესებდეთ გაჩნია თუ არა აქციის ფასებს სეზონურობა. ეს ნიშნავს, მაგალითად აღწევს თუ არა ფასები პიკს ყოველთვის წლის გარკვეულ მონაკვეთში. ან შეიძლება გვინდოდეს გავიგოთ, როგორ იცვლება აქციის ფასები ეკონომიკურ ცვლადთან დამოკიდებულებაში, მაგალითად უმუშევრობის ცვლილებისას. დროის სერიების ანალიზ გააჩნია ორი მთავარი ამოცანა: პირველი - ფენომენის რაობის აღმოჩენა და მეორე - პროგნოზირება. ეს ორივე ამოცანა მოითხოვს დაკვირვებების ნიმუშების აღწერილობას და განსაზღვრას. მსოფლიოს 15 1974-1989 წლებში გამოქვეყნებული გაზეთების შემთხვევითი 4000 გრაფიკიდან აღმოჩნდა, რომ 75%-ზე მეტი დროის სერია იყო.

თავი 2: თეორიული საფუძვლები

ინტეგრალური აგრეგატული მოდელი რეალიზებული უნდა იქნას ბუტაფორიების(dummies) მეთოდით, რომელიც უკავშირდება დროის სერიაში სეზონური კომპონენტების მოდელირებას. ამ მეთოდის მთავარი არსი არის, რომ ჩამოყალიბებული იქნას რეგრესიული მოდელი, რომელიც დროის ფაქტორის გარდა შეიცავს სეზონურ ბუტაფორიულ(dummy) ცვლადს. ბუტაფორიების(dummies) გამოიყენება მაშინ, როცა დაკვირვება შეიცავს ერთი დამოუკიდებელი მნიშვნელვანი ცვლადის კომპლექტს. მაგრამ რეალურად, ეს კომპლექტი შეიცავს ქვეჯგუფებს რომლებიც იცვლება თვისობრივი მაჩვენებლების მიხედვით. მთავარი, მნიშვნელოვანი ცვლადი განსაზღვრავს რაოდენობრივი ინდექსის დონეს, ხოლო ბუტაფორიული(dummy) ან ინდიკატორი ცვლადი - თვისობრივი მაჩვენებელია. ის შეიძლება იყოს ნებისმიერი ატრიბუტული ნიშანი - პროფესია, სქესი, განათლება, კლიმატი, რეგიონი და ა.შ. ასეთი ცვლადების რეგრესიულ მოდელში შესატანად მათ უნდა მიენიჭოთ ციფრული იარლიყი, ანუ თვისობრივი ცვლადები გადაყვანილი უნდა იყოს რაოდენობრივში. მისაღებია, რომ მათ ვუწოდოთ ბუტაფორიული(dummy), სტრუქტურული, ან ხელოვნური ცვლადები. მაგალითად, 0-კაცი; 1-ქალი; ბუტაფორიული(dummy) ცვლადები საშუალებას გვაძლევენ ჩამოვაყალიბოთ უბნობრივი წრფივი მოდელები, რომლებიც გამოიყენება სტრუქტურული ცვლილებების კვლევაში. პირველი ბუტაფორიული(dummy) ცვლადი არის დამატებითი ცვლადი X_0 ,რეგრესიის მოდელის β_0 ვადის, სადაც ის ყოველთვის 1-ია. არ არის აუცილებელი, რომ გამოვიყენოთ X_0 ცვლადი, რადგან ის უკვე გვაქვს პირველი უბნობრივი წრფივი სეგმენტის საწყის წერტილად. სხვა ცვლადები საჭიროა, რომ შევიტანოთ.

მაგალითად შედეგი y-ის დამოკიდებულობა გამოიკვლევა x_t-ს ზომაზე. საბაზისოდ ვიღებთ t_0 სადაც მოხდა რესტრუქტურიზაცია და დამოკიდებულების ბუნება შეიცვალა. ამ მოდელის შესაფასებლად საჭიროა ორობითი ცვლადის შეტანა. $v_t = \begin{cases} 0 \text{ if } t \le t_0, \\ 1 \text{ if } t \ge t_0 \end{cases}$ გადავწეროთ როგორც: $y = a_0 + a_1 x_t + a_2(x_t - x_{t-t_0})v_t$.

$t \le t_0$ ამ წერტილზე რეგრესიის ხასზს აქვს ფერდობი a_1 , $t > t_0$-ზე ფერდობი ტოლია($a_1 + a_2$) და x_t წერტილზე ჩაღრმავება არ ჩანს. როცა a_2=0, დამტკიცდა, რომ t_0 წერდტილზე სტრუქტურული ცვლილებები არ აღინიშნება. შემდეგ განვიხილავთ ბოლო გარემოებას, რაც ჩაღრმავების არარსებობის მიზეზია, მაგრამ უკავშირდება მრავალცვლადიან რეგრესიის თემას, ბუტაფორიული(dummy) ცვლადების არარსებობა .

ამოცანების სია, სადაც ბუტაფორიული(dummy) ცვლადების მეთოდი ყველაზე ეფექტურია:

1. ცვლადები - ინდიკატორები ეკუთვნის დაკვირვების კონკრეტულ პერიოდს - არათანაბარი სტრუქტურული გადაადგილების მოდელირებისათვის. მაგალითად, ცვლადს მივანიჭოთ 1 თუ დაკვირვება ეკუთვნის 1941-1945 პერიოდს, 0 - სხვამხრივ. ეს არის ბუტაფორიული ცვლადების გამოყენების მაგალითი დროებითი სტრუქტურული ცვლილების მოდელირებისათვის.

2. სეზონური ცვლადები - სეზონურობის მოდელირებისათვის. სეზონური ცვლადები იცვლიან მნიშვნელობას თვეების, კვარტალების ან დღეების მიხედვით. მაგალითად მოხმარების მოდელი, რომელიც დამოკიდებულია სეზონებზე:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3,$$

$$x_1 = \begin{cases} 1 \quad \text{for} \quad \text{winter} \quad \text{months}, \\ \quad\quad \text{else} \quad 0, \end{cases}$$

$$x_2 = \begin{cases} 1 \quad \text{for} \quad \text{spring} \quad \text{months}, \\ \quad\quad \text{else} \quad 0, \end{cases} \tag{1}$$

$$x_3 = \begin{cases} 1 \quad \text{for} \quad \text{summer} \quad \text{months}, \\ \quad\quad \text{else} \quad 0. \end{cases}$$

მოხმარების მოცულობაა:

$y = b_0$ - შემოდგომისათვის;

$y = b_0 + b_1$ - ზამთარში;

$y = b_0 + b_2$ - გაზაფხულზე;

$y = b_0 + b_3$ - ზაფხულში

ამ შემთხვევაში, თუ ანალიზის მიხედვით შედეგია $b_3 = 0$, ნიშნავს, რომ განსხვავება ზაფხულისა და შემოდგომის სეზონებს შორის უმნიშვნელოა. $b_1 = b_2$, ნიშნავს, რომ ზამთარსა და გაზაფხულს შორის განსხვავება უმნიშვნელოა და ა.შ.

დროის წრფივი ტენდენცია – გლუვი სტრუქტურული გადაადგილების მოდელირებისათვის. ბუტაფორიული(dummy) ცვლადი გვიჩვენებს, რამდენი დრო გავიდა დროის ნული წერტილიდან კონკრეტულ წერტილამდე. თუ შედეგი ერთი და იგივეა რამოდენიმე დაკვირვებისათვის მაშინ ტენდენცია შეიძლება შედგეს. განსხვავება ტენდენციასა და ბუტაფორიულ(dummy) ცვლადს შორის გამოიყენება ხარისხად t^2, t^3 ა.შ. ისინი გვეხმარება გლუვი, მაგრამ არაწრფივი ტენდენციის შესადგენად. უფრო მეტიც, ბუტაფორიული(dummy) ცვლადების კომბინაციაცაა შესაძლებელი. ეს საშუალებას გვაძლევს მოვახდინოთ კონკრეტული წერტილის ტრენდის ფერდობის მოდელირება. გარდა ტრენდისა, ამ ცვლადის შეტანაც ხდება რეგრესიაში: შერჩევის დაწყებამდე გარკვეულ დრომდე ის 0-ის ტოლია და შემდეგ წარმოადგენს დროის ტენდენციას(1,2,3... თანაბარი ინტერვალების შემთხვევაში).

ამ მეთოდს რამოდენიმე დადებითი მხარე გააჩნია:

1. არაა საჭირო ინტერვალების ტოლობა და შეგვიძლია ნიმუშში დაკვირვების გამოტოვება;
2. ფიქტიური ცვლადის კოეფიციენტი მარტივად შეიძლება იქნეს ინტერპრეტირებული, ისინი ნათლად წარმოადგენენ დინამიური პროცესის სტრუქტურას ;
3. მოდელის შეფასებისთვის არაა საჭირო გავცდეთ კლასიკურ OLS-ს.

ფიქციური ცვლადების მეთოდის არსის ინტეგრირებულ მოთხოვნის მეთოდთან შედარებით ნათლად შეგვიძლია დავინახოთ, რომ მიუხედავად მსგავსებისა აღებული ამოცანების ამ მეთოდით ფორმალიზება შეუძლებელია. $(1 - \frac{z_i}{x_j})$ გამოსახულება არის x_jცვლადის არაწრფივი r ფუნქცია

$$\left(1-\frac{z_i}{x_j}\right)=\left\{\begin{matrix}\left(1-\frac{z_i}{x_j}\right) if\ z_i \le x_j \\ 0\ if\ z_i \rhd x_j\end{matrix}\right\} \qquad (2)$$

სადაც განისაზღვრება მონაცემთა მატრიცა. ფიქციური ცვლადის მეთოდისაგან განსხვავებით აქ არ არის მხოლოდ ერთი დამოუკიდებელი ცვლადი. ჩვენ გვაქვს r რაოდენობის დამოუკიდებელი ცვლადი, რომელიც განისაზღვრება ფუნქციით(2). რომლებიც მიღებულია საწყისი სერიების z_i ცვლადების გარდაქმნით(2). ჩვენ დავამტკიცეთ შებრუნებული ამოცანის გადაჭრის შესაძლებლობა ინტეგრირებული მოთხოვნის მეთოდის გამოყენებით.

წრფივი განტოლების შემთხვევაში წარმოდგენილი მოდელი

$$D_{A\Sigma}=d_1\left(1-\frac{z}{x_1}\right)+d_2\left(1-\frac{z}{x_2}\right)+d_3\left(1-\frac{z}{x_3}\right)+...+d_r\left(1-\frac{z}{x_r}\right) \qquad (3)$$

ვთვლით $D_{A\Sigma}$-ს z ცვლადის ფუნქციად, მარტივია, რომ დავინახოთ:

1. საყრდენ წერტილშისადაც $z=x_i\left(i=1,2,...,r\right)$ $D_{A\Sigma}$ ფუნქცია განგრძობითია;
2. ამ ფუნქციის წარმოებული უბნობრივი მუდმივაა, სადაც $z=x_i\left(i=1,2,...,r\right)$.

განვიხილოთ $D_{A\Sigma}$ სადაც z=x_i, (26)-ის შედეგად, აშკარაა, რომ:

$$\lim_{z\to -x_i} D_{A\Sigma}=\lim_{z\to +x_i} D_{A\Sigma}=\Sigma_{j=i+1}^{r} d_j\left(1-\frac{x_i}{x_j}\right) \qquad (4)$$

$D_{A\Sigma}(z)$ ფუნქციის ლიმიტი ტოლია მისი მარჯვენა ლიმიტის , რომელიც აჩვენებს წერტილის განგრძობადობას სადაც $z=x_i\left(i=1,2,...,r\right)$

ეს ლიმიტი სადაც $z=x_i$, ტოლია

$$\lim_{z\to -x_i} D_{A\Sigma}=\lim_{z\to +x_i} D_{A\Sigma}=\Sigma_{j=i+1}^{r} d_j\left(1-\frac{x_i}{x_j}\right) \qquad (5)$$

მარჯვენა და მარცხენა ლიმიტების ტოლობა და ბოლო ლიმიტის ტოლობა შედეგია: $\lim\limits_{x\to -x_i}\left(1-\frac{z}{x_i}\right)=0$

მსგავსების დანახვა მარტივია $\left(z_{k,}z_{k+1}\right)\left(k=0,1,2,\ldots,r-1\right)$ ტოლია

$\frac{dD_{A\Sigma}}{d_z}=-\Sigma_{i=1}^{k}\frac{d_i}{x_i}$, რაც მარტივად დასტურდება პირდაპირი დიფერენციაციით (6)

აქედან გამომდინარე, ფუნქცია შედეგია რეგრესიის შედეგი, როგორც დეტერმინანტი r-განზომილებიანი წრფივი განტოლების(6). თუ ჩავთვლით, რომ z ცვლადის ფუნქცია უწყვეტია, მაშინ უბნობრივი წრფივი ფუნქცია, წრფივი ნაწილების რაოდენობით ტოლია საყრდენი წერტილების რაოდენობის.

ჩავთვალოთ, m პოლინომინალი $t \geq 0$ დამოუკიდებელი ცვლადის სისტემად.

$P_i(\alpha_1^i,\ldots,\alpha_{r_i}^i,\tau_i,t)$, (i=1,…,m) (7)

სადაც $\alpha_j^i(\mathrm{j}=1,\ldots,r_i)-$ ith პოლინომის უცნობი კოეფიციენტები; r_i –ith პოლინომის მიმდევრობა .

ჩავთვალოთ, რომ თითოეული პოლინომი ნულის ტოლია სადაც $t>\tau_i$, სადაც τ_i-ნამდვილი რიცხვია, როგორც: $\tau_1 \leq \tau_2 \leq \cdots \leq \tau_{m-1} \leq \tau_m = T$.

ინტერვალებს ვუწოდებთ Ii=$[0,\tau_i]$, როგორც ith *პოლინომს* არგუმენტებს(6), ანუ $P_i\left(\alpha_1^i,\ldots,\alpha_{r_i}^i,\tau_i,t\right)\neq 0$ თუ t∈ $[0,\tau_i]$, და $P_i\left(\alpha_1^i,\ldots,\alpha_{r_i}^i,t_i,t\right)=0$ თუ არა. ყურადღება მივაქციოთ, რომ $I_i \subseteq I_{i+1}$. ინტერვალებს I_i ეწოდებათ პოლინომთა სისტემების ბადე (1), და წერტილებს τ_i – მათი კვანძები. ვამარტივებთ ფუნქციის ნოტაციას (7) და ვწერთ$P_i(t)$, რაც გულისხმობს ამ პოლინომს $P_i(t)$, დამოუკიდებელი t ცვლადის გარდა, ის ასევე დამოკიდებულია r_i-ზე, პარამეტრები τ_i რომლებიც განსაზღვრავს არგუმენტის ზედა ლიმიტს.

გავეცნოთ ახალ ფუნქციას

$F(t)=\sum_{i=1}^{m}P_i(t)$ (8)

ძნელი დასანახი არაა, რომ $F(t)$ფუნქცია სასრულია სადაც მისი არგუმენტი [0,T], და უწყვეტია მასზე, თუმცა მისი წარმოებული წყვეტილია კვანძებზე τ_i (i=1,…,m). t<τ_k $\lim_{t\to\tau_k-}F(t)=\sum_{i=k}^{m}P_i(\tau_k)$ რადგან t>τ_k $\lim_{t\to t_k+}F(t)=\sum_{i=k+1}^{m}P_i(\tau_k)$.

ეს ლიმიტები ტოლია წერტილზე t=t_k, როგორც განსაზღვრებით, $P_k(\tau_k) = 0$. ამ დროს წარმოებული $P_k'(\tau_k) \neq 0$, აქედან გამომდინარე $\lim_{t\to t_k-} F'(t) \neq \lim_{t\to t_k+} F'(t)$ წერტილებზე τ_i (i=1,…,n).

ფუნქციას (8) აპროკსიმაციის აგრეგატი, და პოლინომს (7) - აგრეგატის კომპონენტი. ვთქვათ რომ ბადის კვანძები განსაზღვრული როგორც τ_i (i=1,…,n) ცნობილია, მაშინ ნათელია რომ აგრეგატი განსაზღვრულია პარამეტრით $r = \sum_{i=1}^{m} r_i$, რომელიც შეიძლება შეფასდეს least squares მეთოდით შეზღუდვების დამატებით, რაც უზრუნველყოფს კვანძები უწყვეტობას. უნდა აღვნიშნოთ, რომ აგრეგატის უზნობრივი სტრუქტურა განისაზღვრება პოლინომის გარკვეული კომპონენტების სასრულობით და მათი არგუმენტები Ii=$[0, \tau_i]$ აყალიბებენ არა კლებადი ინტერვალების წყობილ მიმდევრობას $I_i \subseteq I_{i+1}$. აგრეგატის ფორმალურად ჩაწერილი ფუნქცია (8), შეიძლება ჩაწერილი იქნას, როგორც:

$$F(t) = \sum_{i=k}^{m} P_i(t), \text{ for } t \geq \tau_{k-1} \tag{9}$$

ჩვენ შევქმენით m ფუნქციონალი(არგუმენტების რაოდენობაზე დაყრდნობით Ii=$[0, \tau_i]$ $(\mathrm{i} = 1, \dots, \mathrm{m})$)

$$\min_{\alpha_1^k,\dots,\alpha_{r_i}^k} S^2 = \sum_{i=\tau_{k-1}}^{\tau_m} (x(t_i) - \sum_{i=k}^{m} P(t_i))^2 \text{ (k=1,2,…,m)} \tag{10}$$

საწყისი დროის სერიების x(t_i) (i=1,2,…,n) უზნობრივი წრფივი აპროკსიმაციისათვის. აპროკსიმაციის აგრეგატის (8) პოლინომის კომპონენტების ბოლო საჭირო იდენტიფიკაცია (7), მოითხოვს სისტემის კოეფიციენტის იდენტიფიკაციას $\alpha_1^i, \dots, \alpha_{r_i}^i$ (i=1,…,m) პოლინომის აგრეგატის (7). ეს დიდ დროს მოითხოვს, მაგრამ მათემატიკურად არც ისე რთულია ტეხილების სისტემის ჩამოყალიბება.

აღვნიშნოთ პრობლემური განაცხადების ხელოვნური ბუნება (10) როგორც აპროკსიმაციის აგრეგატის სასრული ფუნქციის ჩამოყალიბება, რამაც გამოიწვია თანდართული სასრული არგუმენტების $I_i \subseteq I_{i+1}$ $(\mathrm{i} = 1, \dots, \mathrm{m})$ და სასრული პოლინომების განსაზღვრა მათზე. ნათელია, რომ ამის მაგივრად შეგვიძლია ტრადიციული ტეხილების სისტემის გამოყენება (ალბერგ, ნილსონ, ვოლში, 1967) ინტერვალებზე (τ_i, τ_{i+1}) i=0,1,…,m. ესეთი პრობლემა არ იმსახურებს ცალკე კვლევას,

გარდა მისი ცალკეული ქეისებისა - უბნობრივი წრფივი აპროკსიმაცია, ამ ქეისში ყოველი შემადგენელი პოლინომი წარმოადგენს წრფივ ხაზს.

თავი 3: უბნობრივი წრფივი აპროკსიმაციის ექსპერიმენტარული რეალიზაცია

სეგმენტაციის და აპროკსიმაციის პროცესებისთვის Matlab-ს სპეციალური პროგრამა იქნა შემუშავებული. შესაბამისი Matlab-ის პროგრამულ ენაზე დაწერილი პროგრამა წარმოდგენილია თეზისის პირველ აპენდიქსში. პროგრამის შესატანი პარამეტრებია: შერჩევის მომენტების მასივი *x*, დროის სერიების მნიშვნელობების მასივი *y*. ასევე როგორც შესატანი პარამეტრი გვაქვს მასივი *d* რაც წარმოადგენს დროის სერიების საყრდენი წერტილების მასივს. საყრდენი წერტილების რიცხვი სხვადასხვაა. ასევე მათი მნიშვნელობებიც იცვლება. ინტერვალების მნიშვნელობა შეიძლება იყოს ფიქსირებული ან მომხმარებლის მიერ განსაზღვრული. Matlab პროგრამის შედეგებია –*f* მასივი, რომელშიც უბნობრივი წრფივი აპროკსიმაციის დათვლილი მნიშვნელობებია. ხდება შედეგის შედარება დაკვირვებების მასივის მნიშვნელობებთან. მასივი *y* წარმოდგენილია წერტილების სახით ხოლო უბნობრივი აპროკსიმაციის სეგმენტები როგორც წრფეები *f*.

დროის სერიების მსგავსების კრიტერიუმის შეფასებისათვის რამოდენიმე ქეისი იქნა შემუშავებული. ესენია:

1. სტრუქტურულად იდენტური – ამ ქეისში უბნობრივი წრფივი აპროკსიმაციის საყრდენი წერტილების რაოდენობა იდენტურია, ასევე შესაბამისი წრფივი სეგმენტების ფერდობები.
2. სტრუქტურულად იდენტური (ფერდობებით პროპორციული) – ამ ქეისში უბნობრივი წრფივი აპროკსიმაციის საყრდენი წერტილების რაოდენობა იდენტურია, ასევე შესაბამისი წრფივი სეგმენტების ფერდობები უნდა იყოს იდენტური ან პროპორციული.
3. სტრუქტურულად იდენტური (პატარა შეცდომით) – ვხსნით ახალ ქეისს რომელშიც ვთვლით რომ ფერდობები პროპორციული, მაგრამ აქვს პატარა შეცდომა. ეს პატარა შეცდომა არ ქმნის დიდ განსხვავებას წრფივ აპროკსიმაციაში; გარდა ამისა, ის არ ქმნის დიდ განსხვავებას დროის სერიების

დაკვირვებებში . ეს ყველაფერი სტუდენტური t კრიტერიუმის შედეგია. ეს ყველაფერი დამოკიდებული, თუ რას და როგორ ვაანალიზებთ. შემდეგში კიდევ უფრო გავრცობილად ვისაუბრებთ ამ ქეისზე.

4. სტრუქტურულად მსგავსი – ამ ქეისში უბნობრივი წრფივი აპროკსიმაციის საყრდენი წერტილების რიცხვი იგივეა. შესაბამისი ფერდობები არ არის ტოლი, მაგრამ აქვთ საერთო კოეფიციენტი. სხვა სიტყვებით რომ ვთქვათ წრფეების მიმართულება ერთი და იგივეა.
5. ნაწილობრივ მსგავსი – საყრდენი წერტილების ტოლობა არაა აუცილებელი. ყველა ფერდობს არ აქვს იგივე კოეფიციენტი. ამ ქეისში მსგავსების კუთხე განსაღვრება წრფივი სეგმენტების ტოლი კოეფიციენტების რაოდენობით და თანაბარი ფერდობების სიახლოვის პროცენტულობით.
6. სტრუქტურულად განსხვავებული – საყრდენი წერტილების რაოდენობა და წრფივი თანაბარი ფერდობების კოეფიციენტი სრულიად განსხვავებულია.

ამ 6 ქეისის მსგავსებები შეიძლება იქნეს გავრცობილი. მაგრამ ჩვენი მიდგომა აგებულია ამ მსგავსების ქეისებზე. ქვემოთ მოყვანილია თითოეული ქეისის ემპირიული მაგალითი. ძალიან რთულია ისეთი ეკონომიკური და სამეცნიერო დაკვირვებების დროის სერიების მოძებნა, რომელიც პირველ და მეორე ქეისს შეესაბამება. ამის გამო, ჩვენ მოვიყვანეთ მაგალითები გენერირებულ მონაცემებზე.

ამ კვლევაში წარმოდგენილია ა.შ.შ. დოლარის გაცვლითი კურსის შედარება სხვადასხვა ვალუთასთან ქართული ლარის, ევროს და რუსული რუბლის ჩათვლით. უბნობრივი წრფივი წარმოდგენის გამოყენებით ჩატარებულმა ემპირიულმა სწავლებამ აჩვენა, რომ ბოლო ექვსი წლის განმავლობაში ქართული ლარის ტენდენცია უფრო ახლოს იყო რუსულ რუბლთან ვიდრე ევროპულ ვალუტასთან.

აღწერილი უბნობრივი წრფივი წარმოდგენის მეთდით ასევე შესაძლებელია დამატებითად დროის სერიათა სეზონურობის ჩვენება. დროის სერიების სეზონურობის დასადგენად გამოიყენება იგივე მეთოდები, რაც მათი მსგავსების გაზომვის დროს. განსხვავება მხოლოდ ისაა, რომ მთელი დროის სერია იყოფა პატარა

ნაწილებად და ხდება მათი გაანალიზება სხვადასხვა თანმიმდევრობით. ოპტიმიზაციის ალგორითმს შეუძლია გახადოს გაზომვის მთელი პროცესი უფრო ეფექტური. ამ ქეისში უფრო კარგი იქნება შესაბამისი ალგორითმის გამოყენება. დაყოფილი დროის სერიების კლასიფიცირება შეიძლება მოხდეს ქეისებად, როგორც წინა კვლევაში. ეს დაგვეხმარება სეზონურობის კუთხის უკეთ გასაზომად.

დასკვნა

შემდეგი მიზანი იქნა დასახული: დოის სერიათა უბნობრივი წრფივი წარმოდგენის ოპტიმიზაციის ალგორითმის შექმნა არასაჭირო საყრდენი წერტილების მოშორებით, კომპლექსური არასტაციონარული დროის სერიების გასამარტივებლად და დროის სერიების ნაკრებების მსგავსების დადგენის მთელი პროცესის ასაჩქარებლად. დამატებით, ამ ოპტიმიზაციის პროცესზე დაყრდნობით უნდა იქნეს შემუშავებული დროის სერიების მსგავსების კრიტერიუმი. ასევე, ქვემოთ ჩამოთვლილი მიზნებიც იქნა მიღწეული;

- არასტაციონარული დროის სერიების უბნობრივი წრფივი აპროკსიმაციის ახალი მეთოდი იქნა ჩამოყალიბებული.
- გამომუშავდა დროის სერიების გენერირებული უბნობრივი წრფივი წარმოდგენის ოპტიმიზაციის ალგორითმი.
- ჩამოყალიბდა აპროკსიმაციის პარამეტრების შეფასების კრიტერიუმი.
- ჩამოყალიბდა ნიმუში და მოკვლეული დროის სერიათა შედარების კრიტერიუმი.
- გამომუშავებულ მეთოდებზე დაყრდნობით ჩატარდა გამოთვლები მოცემულ დროის სერიათა მაგალითებზე.
- შეიქმნა შესაბამისი პროგრამული ინსტრუმენტები.
- ჩამოყალიბებულ მეთოდზე და კრიტერიუმზე დაყრდნობით ჩატარდა დროის სერიების სეზონურობის ფაქტორის ანალიზი.

პუბლიკაციები

1. A. Milnikov and D. Satybaldiev. Designing Optimal Integral Aggregate Structure. Journal: Journal of Technical Science and Technologies; ISSN 2298-0032; Volume 3, Issue 2, 2014, pp 21-24.
2. A. Milnikov, D. Satybaldiev and C.Mert. A new method of piecewise linear approximation of non-stationary time series. Journal: Journal of Technical Science and Technologies; ISSN 2298-0032; Volume 4, Issue 1, 2015, pp 5-8.
3. D.Satybaldiev A New Method of Piecewise Linear Approximation of Non-stationary Time Series for Similarity Measurement. ICECCO 2015 September, Almaty Kazakhstan, pp 90-93.
4. D.Satybaldiev Non stationary time series similarity measurement based on Piecewise Linear Approximation: Empirical example, Journal of Science, Innovations and Technologies of Kyrgyzstan N) №2. 2016 ISSN 1694-7649, pp 12-16.

.