

Mining clustering algoritmi u wireless senzorskim mrežama

Miloš Milutinović

Elektrotehnički fakultet – Računarska tehnika i informatika

Beograd, Srbija

milos2906@gmail.com

Sadržaj— Ovaj rad opisuje klaster algoritam zasnovan na protokolu minimalnog obuhvata stabla i maksimalnom izvoru energije na senzorima po imenu MSTME. Takođe, ovaj protokol ograničava grupisanje u wireless senzorskim mrežama, zatim su opisane neke procene metrika. MSTME bolje obavlja grupisanje, kada se procenjuju evaluacije metrika, od već poznatih metoda minimalnih energija adaptivnog klasteringa hijerarhija i baznih stanica kontrolisanih dinamičkim klastering protokolima u wireless senzorskim mrežama.

Ključne reči- data mining; wireless sensor networks; clustering; minimalna obuhvatna stabla i maksimalna energija.

I. UVOD

Data mining je nova disciplina koja ima za cilj da filtrira podatke u bazama podataka, da ih sumira, i pronalazi obrazce. Ogroman broj podataka iz kolekcija može se analizirati i ispitati na neograničen broj načina. Kako vreme odmiče, nove vrste objekata i dezena mogu privući pažnju, i mogu biti vredni u traženju potrebnih podataka.

Tradicionalni data mining otkriva korisne informacije ili znanja iz već poznatih baza podataka. Bežične senzorske mreže obuhvataju stotine ili čak hiljade senzora iz životne sredine. Bežične senzorske mreže proširuju koncepte baza podataka u tradicionalnim data mining sistemima. Kao i tradicionalni data mining podaci, cilj bežične senzorske mreže je da obezbedi korisne informacije ili znanje za ljude iz fizičkog sveta. Bežični senzori su nasumično raspoređeni u fizičkom prostoru da bi prikupljali podatke iz okoline i formirali korisne podatke za nas.

II. CLUSTERING U WIRELESS SENZORSKIM MREŽAMA

Metode grupisanja koje se koriste u bežičnim senzorskim mrežama mogu smanjiti rasipanje energije mreže. Ovo je veoma važno u bežičnim senzorskim mrežama, jer senzori

imaju ograničen energetske resurs a bežične komunikacije troše veliki iznos energije, potrošnja energije koju bežične komunikacije troše je u vezi sa brojem prenesenih podataka i prenosne daljine. Optimalna ruta može biti izabrana da se minimizuje prosečna udaljenost prenosa. Štaviše, iako oni imaju samo ograničene računarske sposobnosti, senzori mogu da smanje podatke kompresijom ili da se oslobode viška podataka i onda da ih pošalju u bazne stanice. Dakle, senzori su grupisani u nekoliko klastera i neki senzori su izabrani kao klaster glave da spajaju ostale senzore i prosleđuju podatke između njih.

Ipak, različite prenosne razdaljine i različiti poslovi “neglavnih” senzora i glavnih klaster senzora, doprinose rasipanju energije u bežičnim senzorskim mrežama. Dakle, svi senzori treba da rade zajedno kao tim i da ravnomerno raspoređuju rasipanje energije da bi produžili životni vek mreže koja se definiše kao vremenski period od raspoređivanja senzora i dužine trajanja senzora. Dakle, životni vek mreže je zavistan od energetske efikasnosti i efektivnosti ravnomerne raspodele energije rasipanja. Od grupisanja u bežičnim senzorskim mrežama se očekuje da će grupa u početku negrupisanih senzora u zavisnosti od blizine lokacija, dati prednosti dužini trajanja cele senzorske mreže tako što će biti smanjen broj prenesenih podataka, smanjena udaljenost prenosa podataka kao i efikasnije rasipanje. Zbog prostornih karakteristika fizičke životne sredine, prirodna fizička udaljenost se koristi kao ključni parametar klastera senzora, jačina signala se povećava za prenos na daljinu u većini slučajeva.

III. PRETPOSTAVKE ZA MODELOVANJE GRUPISANJA U WIRELESS SENZORSKIM MREŽAMA

Postoje neka određena ograničenja kada se algoritmi grupisanja primenjuju u bežičnim senzorskim mrežama. Neke pretpostavke i definicije su date na sledeći način:

- A. *Runda se definiše kao proces kada svi senzori prenose svoje podatke jednom glavnom senzoru. Klasteri formirani u trenutnom krugu bi se razlikovali od onih u poslednjem . Nema fiksnih optimalnih klastera koji se koriste u celom procesu,*
- B. *Senzori visokih energija imaju energetske resurs iznad prosečnog nivoa svih senzora u mrežama. Najmanje jedan senzor visoke energije se distribuira u svakom klasteru u svakom krugu, kao glavni senzor,*
- C. *Pod pretpostavkom da se energija rasipa uglavnom na prenos, prijem i spajanje podataka u bežičnim mrežama senzora. Svi senzori imaju istu početnu energiju resursa. Glavni senzori rasipaju daleko više energije nego ne-glavni senzori, s obzirom da moraju da dobiju više podataka, osiguraju te podatke, i da ih dalje proslede,*
- D. *Senzor se zove "mrtav" kada je istrošio svu svoju energiju. Jedna od prednosti korišćenja clustering algoritma je u tome da ne dozvoli da neki senzor prestane sa radom mnogo pre drugih. To znači da bi svi senzori u mreži trebalo da budu u funkciji istovremeno i da svi istovremeno treba da prestanu da rade,*
- E. *Svi senzori su nasumično i ravnomerno raspoređeni u mreži.*

IV. OGRANIČENJA U KORIŠĆENJU TRADICIONALNIH CLUSTERING ALGORITAMA U WIRELESS SENZORSKIM MREŽAMA

Tradicionalni clustering algoritmi kao što su: DBSCAN, CURE, itd. , nisu dobri metodi za korišćenje u wireless senzorskim mrežama zbog sledećih razloga:

- A. *Mada mogu da pronađu prirodnu distribuciju gustinu senzora, broj senzora u svakom klasteru varira značajno. Takođe, ne može da obezbedi da je barem jedan visoko-energetskih senzor u svakom klasteru. Ovo će prouzrokovati da se energetska rasipanje distribuira u celoj mreži previše neravnomerno,*
- B. *Pravili bi fiksne optimalne klasterne, a time bi neadekvatno bile generisane veze između senzora. To bi uzrokovalo da neki senzori prestanu sa radom mnogo pre drugih senzora,*

- C. *Jedan ili nekoliko obekata kao što je šum će biti zanemareni u tradicionalnim data mining sistemima, ali u bežičnim mrežama senzora, nijedan senzor se ne može smatrati šumom.*

V. KAKO TREBA UNAPREDITI CLUSTERING ALGORITAM U WIRELESS SENZORSKIM MREŽAMA

Dobar klastering algoritam u wireless senzorskim mrežama treba da generiše klasterne u svakoj rundi zadovoljavajući sledećih pet optimalnih pravila:

- A. *Blizina: Što je prosečna razdaljina između dva susedna senzora manja to je sam klaster bolji,*
- B. *Isti broj: Broj senzora u svakom klasteru treba da bude isti,*
- C. *Maksimalna energija: Glavni senzor treba da ima što je moguće veći izvor energije ili makar malo veći od prosečnog izvora,*
- D. *Pozicija senzora: Glavni senzori su ravnomerno raspoređeni,*
- E. *Dinamička promena: Različiti klasteri su dinamički generisani u različitim rundama.*

VI. WIRELESS SENZORSKE MREŽE ZASNOVANE NA KLASTERIMA

Od wireless senzorskih mreža zasnovanih na klasterima se očekuje da, su glavni senzori u mreži oni senzori koji mogu da grupišu otprilike isti broj senzora u svakom klasteru, da imaju visok izvor energije, i da energija u mreži bude što ravnomernije distribuirana.

Takođe one moraju da prevaziđu ograničenja tradicionalnih clustering-algoritama u data mining sistemima, da budu bolje posebno u očuvanju maksimalne energije sistema kao i u dinamičkim promenama.

VII. KLASERING NISKE ENERGETSKI ADAPTIVNE HIJERARHIJE

Niska energetska adaptivna klastering hijerarhija (Leach) daje jednostavan distribuiranu klaster šemu za ravnomernu distribuciju rasipanja energije. Po pravilu, samo senzori koji još nisu glavni senzori, a verovatno imaju više energije na raspolaganju od onih koji su nedavno postali glavni senzori. Glavni senzori mogu postati glavni u MOSME klasteringu za produženje dužine trajanja senzorske mreže u trenutnoj rundi. Kada senzori počinju sa jednakom energijom, jednostavna distribuirana funkcija verovatnoće se koristi za rotiranje pozicije glavnih senzora u svim sensorima. Na taj način, grupisanje u Leach-u glavnog senzora zadovoljava dva optimalna principa maksimalne energije i dinamičnih promena. Međutim, Leach ne uzima u obzir optimalno rasipanje energije u svakoj rundi jer senzori u istom klasteru nisu dovoljno blizu. Štaviše, broj senzora značajno varira i glavni senzori nikada nisu ravnomerno distribuirani u Leach-u.

VIII. KLASITER PROTOKOL DINAMIČKI KONTROLISANIH BAZNIH STANICA

Klaster protokol dinamički kontrolisane bazne stanice (BCDCP) je centralno kontrolisani metod koji treba da poboljša Leach klastering uzimajući u obzir energetske efikasnost kao i efektivnost u distribuciji energetske disperzije. Da bi smanjila rasipanje energije u svakoj rundi BCDCP metod smanjuje razdaljinu između dva senzora, neglavnog senzora i glavnog senzora. Klasteri su formirani pomoću iterativnog klastera-podele algoritma. Senzori koji imaju izvor energije oko prosečnog nivoa biraju se kao mogući kandidati za glavne senzore. Zatim se dva senzora sa najvećom razdaljinom biraju za glavne senzore a ostali senzori se raspoređuju oko najbližeg senzora. Zatim se svi senzori dele u dva podskupa, potom po istom principu i senzori u podskupovima se dele iterativno. Na kraju glavni senzori se povezuju pomoću metode Minimalnog Obuhvatnog Stabla (MST) i jedan glavni senzor se određuje da šalje podatke baznoj stanici. Za ravnomernu distribuciju rasipanja energije, BCDCP metod je usvojio izbalansiranu tehniku grupisanja da bi omogućio da klasteri imaju isti broj senzora .

U poređenju sa tradicionalnim grupisanjem algoritama u Data Mining sistemu, klasične metode grupisanja koje se primenjuju u bežičnim senzorskim mrežama obično mogu da zadovolje pravila maksimalne energije i dinamičkih promena. Algoritmi grupisanja u oba metoda, Leach-u i BCDCP -u, prvo, uzimaju u obzir senzore sa više energetske resursa da budu izabrani kao glavni senzori. Takođe, oni generišu različite klasterne u svakom krugu. Međutim, i Leach i BCDCP ne mogu dobro da odrede pozicije senzora. Dakle, MSTME je dat da bi se poboljšali klasteri dalje u smislu blizine, istog broja, pa čak i lokacije. Takođe, klasteri u MSTME se porede sa onima u Leach-u i BCDCP-u procenjujući ih sa novim klaster-evolucija modelom

IX. METOD GRUPISANJE KORISTEĆI MINIMALNA OBUHVATNA STABLA I MAKSIMALNI IZVOR ENERGIJE NA SENZORU

Minimalno obuhvatno stablo na konačnom skupu tačaka (X_1, X_2, \dots, X_n) u R^2 je povezani graf sa ovim tačkama, kao temenima i sa minimumom ukupne dužine ivice. Aplikacije od MST-a u klasterovanju podataka nazivaju se jedno-vezne klaster analize. U bežičnim senzorskim mrežama, X , označava dve dimenzije koordinate lokacije senzora. Takođe, svaka ivica predstavlja fizičku udaljenost između dva senzora. Isto kao klasterovanje u Leach-u i BCDCP-u, senzori visoke energije smatraju se i glavnim sensorima. Zbog varijacija visokih energija senzora u različitim rundama, pravilo dinamičnih promena je takođe zadovoljeno. Stoga, glavni cilj MSTME algoritma je da izabere one očekivane glavne senzore, koji bi trebalo da budu ravnomerno raspoređeni u senzorskim mrežama, zatim da smanji prenos rasipanja energije glavnih senzora, takođe treba da grupiše približno isti broj bližih senzora u klasterne da bi smanjili prenos energije rasipanja ne-glavnih senzora i da ravnomerno rasporede rasipanje energije glavnih senzora.

X. MSTME ALGORITAM

Kod MSTME algoritma, senzori sa energijom iznad proseka postaju kandidati za glavne senzore. MST je kreiran

da opiše blizinu glavnih kandidata. Ne-glavni senzori postaju pristalice svojih najbližih glavnih senzora. Očekivani broj, N_{ch} , glavnih senzora se bira od strane MSTME algoritma, koji zadovoljavaju svih pet optimalnih principa. Biranje glavnog senzora po pravilima MSTME algoritma je dat u tabeli 1. Koristeći algoritam u tabeli 1, dati broj glavnih senzora je izabran, a zatim ne-glavni senzori se povezuju sa najbližim glavnim sensorima i na taj način se formiraju klasteri u svakom krugu.

TABELA I. IZBOR GLAVNOG SENZORA PO MSTME ALGORITMU

Step 1: Sensors with more energy resource than average level are selected into CH candidates set, S .
Step 2: A MST, T , is used to connect all the items in S .
Step 3: Supporters of a CH candidate x are those non-candidate sensors that are nearest to x among all CH candidates. Compute the number of supporters for each CH candidate including candidate itself.
Step 4: Suppose supporters are just around their candidates and thus the latter can delegate the former to decide which edge would be split.
Step 5 (Initialization): Let the number of already split edge $nSplit = 0$, $T_ = T$ and $S_ = S$.
Step 6 (Loop): Find an edge, which breaks $T_$ into two sub MSTs of T_1 and T_2 and at the same time the nodes in $S_$ are grouped into two subsets of S_1 and S_2 , respectively with the nearest number of supporters in both subsets. Then let $nSplit = nSplit + 1$.
Step 7 (Termination test): If $nSplit \geq N_{CH}$ go to Step 8. Otherwise, go on splitting S_1 and S_2 in turn. If the number of supporters in S_1 (or S_2) is more than N/N_{CH} , then let $S_ = S_1$ and $T_ = T_1$ (or $S_ = S_2$ and $T_ = T_2$). Go to Step 6.
Step 8: The CH candidate with the most energy resource among CH candidates in each subset is chosen as real CH. If more than one CH candidates have maximum energy resource in their subset, then randomly chose one as real CH.

Teorema 1. U Bežičnim senzorskim mrežama, pretpostavimo da svi senzori imaju isti početni energetske resurs i da su senzori raspoređeni ravnomerno i nasumično u mreži, onda su ne-glavni senzori približno isto grupisani oko svojih glavnih senzora u svakom krugu.

Dokaz: Senzori sa istim inicijalnim energetske resursom, u mreži treba da budu povezani sa MST-om, a zatim N_{CH} pod-stabla se formiraju po pravilima MST osnovnih metoda datih u tabeli 1. Zbog toga što se svi senzori distribuiraju u mreži ravnomerno i nasumično, glavni senzori izabrani po MSTME metodi ravnomerno i nasumično se distribuiraju po mreži u prvom krugu. Princip skoro istog broja grupisanja

senzora oko glavnog u MSTME-u obezbeđuje da svako podeljeno pod-stablo pokriva skoro istu površinu u mreži. Isto tako, prema koraku 8 u tabeli, glavni senzori su nasumično izabrani. Zbog toga, glavni senzori se ravnomerno i nasumično distribuiraju po celoj mreži već u prvom krugu. Ovo pruža osnovne uslove da bi se u kasnijim krugovima birali glavni senzori. U drugom, kao i u ostalim krugovima, svi senzori imaju različite energetske resurse. Zato što glavni i neglavni senzori koji su udaljeni od drugih glavnih senzora, troše mnogo više energije prema modelu disperzije energije, koji sam napisao u poglavlju III, oni senzori koji su bliži zadnjem glavnom senzoru mogu da postanu kandidati za glavne senzore u trenutnom krugu. Dakle, u lokalnim mrežama, neglavni senzori se grupišu oko najbližeg kandidata za glavni senzor.

MST je opravdan za grupisanje dok porednako distribuira energiju rasipanja u bežičnim senzorskim mrežama.

Prvo, MST povezuje sve glavne senzor kandidate sa minimalnim ivicama. Kad god je ivica prekinuta, čvorišta u istom podeljenom pod-stablu su bliža. Zbog toga što se neglavni senzori grupišu oko najbližeg glavnog senzora, glavni senzori mogu da delegiraju svoje senzore u odnosu na njihovu lokaciju. Na taj način, senzori u istim klasterima, formiranim od strane MSTME-a, su bliži jedni drugima. Isto tako, otprilike isti broj neglavnih senzora u svakom podskupu kandidata za glavne senzore, obezbeđuje skoro isti broj senzora oko najbližeg glavnog senzora u ovim podskupovima. U isto vreme, jer senzori su ravnomerno i nasumično raspoređeni u mreži, skoro isti broj senzora može da pokrije područja istih veličina. Svaka oblast sadrži po jedan glavni senzor, to jest, glavni senzori su ravnomerno raspoređeni po cijeloj mreži. Konačno, bilo koje čvorište u MST-u su senzori sa visokim resursima energije, većim od ostalih energija senzora u mreži. Dakle MSTME sintetički radi bolje u svih pet optimalnih principa. Kao i BCDCP, MSTME algoritam radi na BS-u. Dakle, MSTME je centralno-kontrolisani klastering metod, pod pretpostavkom da su lokacije svih senzora već poznate.

Uobičajeno je, da optimalan broj glavnih senzora određuje topologija mreže. U mrežama sa dva hopa, kod LEACH i MSTME metoda, koriste se šest glavnih senzora. Ali kod multi-hop topologija, kao što je BCDCP, optimalan broj u stablu-povezanih glavnih senzora nije isti kao kod LEACH-a. U mrežama sa multi-hopovima, kod BCDCP-a i MSTME-a, koriste se devet glavnih senzora.

U multi-hop mrežama, broj glavnih senzora utiče na prosek prenosa na daljinu svakog senzora i broja podataka koji se spajaju. Prvo, što je veći broj glavnih senzora, manji je u proseku kvadrat prenosa na daljinu ne-glavnih senzora, ali je veći prosek kvadratnih prenosa između udaljenih glavnih senzora. Štaviše, MDT se koristi za povezivanje glavnih senzora, da bi proizveli što više podataka koji se spajaju. Dakle, što je veći broj glavnih senzora, više energije se rasipa na podacima koji se spajaju. Optimalni kompromis, na prenosu energije rasipanja ne-glavnih senzora i energije rasipanje oba glavna senzora, prenos podataka i fuzija se određuje po broju glavnih senzora.

XII. POREĐENJE UZORKA KLASTERA U MSTME-U SA LEACH I BCDCP

Primer bežične senzorske mreže sa $N = 100$ čvorova nasumično raspoređenih u $M \times M$ ($M = 100m$) oblasti je usvojen da se procene klasteri koje generišu MSTME, LEACH, i BCDCP. U stvari, tri grupisana metoda ne zavise od apsolutne udaljenosti od bežične senzorske mreže i ne zahtevaju primenu parametara vezanih za to. To su njihove prednosti u poređenju sa metodama grupisanja u tradicionalnim data mining sistemima. Štaviše, tri metode klastera u bežičnim senzorskim mrežama zadovoljavaju optimalna pravila klastera, maksimalne energije i dinamičnih promena. Prvi princip izbora glavnog senzora je baziran na visoko-energetskim sensorima u svakom metodu grupisanja i to je takođe obezbedilo da najnoviji glavni senzori ne mogu da budu izabrani kao glavni senzori ponovo u tekućem kolu.

XIII. ZAKLJUČAK

MSTME predlaže dobro i optimalno rešenje za klaster senzora u bežičnoj senzorskoj mreži. Na osnovu MST metode, članovi u bilo kom od podskupova glavnih senzora kandidata postavljeni su bliže. I zbog toga što glavnih senzora kandidati imaju energije iznad prosečnog nivoa, senzori kod većine energetske resursa u svakom od podskupova imaju više energetske resursa između senzora u čitavoj mreži. Najbliži broj pristalica u svakom od podskupova čine konačnim klasterima one koji imaju što je više moguće približan broj senzora, a time i energije rasipanja kod glavnih senzora su približno iste. Glavni senzori sa više energetske resursa takođe produžavaju život senzorske mreže. Pored toga i bliži senzori u klasterima smanjuju celine energetske rasipanja mreža. Rezultati simulacija pokazuju da MSTME metode daju bolje rezultate od LEACH i BCDCP metoda u mrežama sa dva ili više hopena.

LITERATURA

- [1] Data Mining: Concepts and Techniques, Second Edition Jiawei Han and Micheline Kamber. 2006 by Elsevier Inc J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [2] W. B. Heinzelman, A. P. Chandrakasan and H. Balakrishnan, Energy efficient communication protocol for wireless microsensor networks, in Proc. 33rd Hawaii Int. Conf. Sys. Sci. (2000).
- [3] M. Ester et al., A density-based algorithm for discovering clusters in large spatial databases with noise, in Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining (AAAI Press, Menlo Park, Calif., 1996), pp. 226–231.
- [4] G. Karypis, E.-H. Han and V. Kumar, Chameleon: Hierarchical clustering using dynamic modeling, IEEE Computer 32 (1999) 68–75.
- [5] S. Guha, R. Rastogi and K. Shim, CURE: An efficient clustering algorithm for large databases, in Proc. ACM SIGMOD Int. Conf. Management of Data (ACM Press, New York, 1998), pp. 73–84
- [6] W. B. Heinzelman, A. P. Chandrakasan and H. Balakrishnan, An application-specific protocol architecture for wireless microsensor networks. IEEE Transactions on Wireless Communications 1 (2002) 660–670
- [7] W. B. Heinzelman, An application-specific protocol architectures for wireless networks, Ph.D. thesis, Massachusetts Institute of Technology (MIT) (2000).
- [8] S. D. Muruganathan, D. Ma, R. I. Bhasin and A. O. Fapojuwo, A centralized energy-efficient routing protocol for wireless sensor networks, IEEE Communications Magazine (2005) 8-13
- [9] S. Ghiasi et al., Optimal energy aware clustering in sensor networks, in MDPI Sensors (2002) 40-50.
- [10] J. C. Gower and G. J. S. Ross, Minimum spanning trees and single linkage cluster analysis, Applied Statistics 18 (1969) 54–64.

