

KATEGORIZACIJA LITERATURE S PODROČJA POD-FREKVENČNEGA RAZBREMENJEVANJA S POMOČJO METODE ZA ROJENJE

Tadej Škrjanc^{1*}, Rafael Mihalič¹, Urban Rudež¹

¹ – Univerza v Ljubljani, Fakulteta za elektrotehniko, Tržaška cesta 25, 1000 Ljubljana
tadej.skrjanc@fe.uni-lj.si rafael.mihalic@fe.uni-lj.si urban.rudez@fe.uni-lj.si

Povzetek

Prvi korak vsake raziskave je pregled področja, ki ga bomo obravnavali. Na ta način se seznanimo s problematiko in obstoječimi rešitvami, skupaj z njihovimi prednostmi in slabostmi. Nova odkritja in rezultate raziskav širimo oziroma delimo z drugimi predvsem s pomočjo člankov in knjig. Vsako delo opisuje določen problem oziroma podaja rešitev le-tega, zato se z leti raziskav na nekem področju nakopiči precej velika količina literature. Posledično kmalu izgubimo pregled in imamo težave z učinkovitim povezovanjem gradiva med seboj. V članku bomo na primeru literature s področja pod-frekvenčnega razbremenjevanja pokazali, kako lahko s pomočjo t-SNE metode rojenja in PCA tehnike reduciranja dimenzij podatkov, zbrano literaturo klasificiramo glede na podobnost. V ta namen smo ustvarili bazo z več kot 380 članki, doktorati, magistrskimi deli in drugimi podobnimi dokumenti s področja zaščite pod-frekvenčnega razbremenjevanja. Pri vsakem izmed njih smo izluščili številne podatke, kot so ključne besede, vsebovane reference, citiranost, leto nastanka itd. ter kakšne so lastnosti opisane sheme PFR (npr. ali se odločitev sprejema lokalno ali globalno, kateri so potrebni podatki za delovanje, ali uporablja kakšne napredne tehnike, ipd.). Na ta način lahko lažje in hitreje poiščemo dokumente, ki opisujejo sheme PFR s podobnimi lastnostmi. Prav tako vidimo, kateri tip je bil najpogosteje obravnavan v preteklosti, kateri tipi potrebujejo dodatne raziskave ter kakšen je trend danes.

Ključne besede:

rudarjenje podatkov, t-SNE, PCA, prepoznavanje vzorcev, pod-frekvenčno razbremenjevanje, pregled literature.

CATEGORIZATION OF THE LITERATURE IN THE FIELD OF UNDER-FREQUENCY LOAD SHEDDING WITH CLUSTERING METHOD

Abstract

The first step in any research is to review the area we will be dealing with. In this way, one becomes familiar with the problems and existing solutions, as well as their advantages and disadvantages. New discoveries and research results are disseminated and shared primarily through articles and books. Since each work describes a specific problem or offers a solution to it, quite a large amount of literature accumulates over the years of research in a particular field. Therefore, it is easy to lose track of the material and have difficulty linking it effectively. In this paper, we use the literature in the field of under-frequency load shedding as an example to show how the collected literature can be classified by similarity using the t-SNE clustering method and the PCA technique to reduce the dimensionality of the data. For this purpose, we have created a database of more than 380 articles, PhD theses, master thesis and other similar documents in the field of under-frequency protection. For each of these documents, we extracted a set of data, such as keywords, references, citations, year of creation, etc. as well as the characteristics of the under-frequency load shedding algorithm described (e.g., whether the decision is made locally or globally, what data is needed for operation, whether it uses advanced techniques, etc.). This way, you can more easily and quickly find documents that describe under-frequency load shedding schemes with similar characteristics. You can also see which types of schemes have been most commonly covered in the past, which types need further research, and what the current trend is.

Keywords:

data mining, t-SNE, PCA, pattern recognition, under-frequency load shedding, literature review.

1 UVOD

Ena od prvih nalog raziskovalca, ki se loti raziskovanja nekega novega področja, je seznanitev s problemi in že obstoječimi rešitvami ne tem področju. Torej, mora poskušati razumeti kaj počnejo drugi raziskovalci in kako se področje razvija. To je kasneje pomembno tudi pri povezovanju raziskovalčevega dela s širšo sliko. Stanje na nekem področju pa se odraža v najnovejših publikacijah, zato je raziskovanje literature nekako prvi potreben korak. Raziskave na nekem področju običajno napredujejo zelo hitro, saj pri njen sodelujejo številni raziskovalci iz različnih držav in organizacij, kar pa pomeni, da obstaja veliko literature, ki jo je predvsem na začetku raziskovanja nekega področja, težko razumeti. Prav tako, ne vidimo povezav med članki, vrzeli in trenda v raziskovanju. Vse to pa nam nekako ponujajo sistematični pregledni članki literature. Torej, to pomeni, da se je že nekdo pred nami soočil z veliko količino literature, jo pregledal in ovrednotil. Faza iskanja, filtriranja in ocenjevanja literature je izredno zamudna. Prav tako imajo različni avtorji pri pisanju različne cilje in različne poglede, pri čemer tudi težko zajamejo vso obstoječo literaturo.

Kadar imamo opravka z velikimi količinami podatkov, ponavadi uporabimo metode podatkovnega rudarjenja, zato se zdi nekako logičen korak, da tovrstne metode uporabimo tudi pri analiziranju literature. V zadnjem času se kaže [1-11], da si pri analiziranju literature lahko pomagamo z različnimi naprednimi tehnikami umetne inteligence kot so obdelava naravnih jezikov (ang. *Natural Language Processing*, NLP), gručenje (ang. *clustering*) in analiza glavnih komponent (ang. *Principal Component Analysis*, PCA). Pri NLP [13] gre za to, da s kombinacijo jezikovne in računalniške znanosti ter umetne inteligence računalnik naučimo prepoznati človeški govor za namene strojnega prevajanja teksta, predvidevanja besedila (tj. predlaganje besed pri vnosih teksta), prepoznave govora in virtualne pomoči uporabnikom spletnih strani in naprav (npr. Siri, Alexa, Bixby), optičnega prepoznavanja besedila (ang. *Optical Character Recognition*, OCR), itd. Najbolj tipični operaciji pri NLP sta tokenizacija, pri kateri razdelimo govor/tekst v stavke in besede ter zarezovanje in lematizacija, pri kateri odstranimo predpone in končnice besed, da dobimo korene besed. Pri metodah gručenja [14,15] združimo objekte v skupine, pri čemer, objekti v isti skupini v nekem smislu izkazujejo enake lastnosti kakor tisti v drugih skupinah. Tovrstne tehnike se uporabljajo pri analizi velike količine nepreglednih podatkov, saj nam omogočajo najti njihove bistvene lastnosti in njihovo vizualizacijo. Tako jih lahko najdemo na številnih področjih kot so razpoznavanje vzorcev, analiza slik, bioinformatika, stiskanje podatkov, detekcija anomalij, itd. PCA [14,15] je statistično orodje, s področja strojnega učenja, ki s pomočjo ortogonalne transformacije podatke preslika v nov manjši linearen prostor, imenovan glavni pod-prostor. Torej PCA uporabimo, ko želimo zmanjšati dimenzije podatkov. PCA se je v preteklosti že izkazala kot primerna metoda za prepoznavanje na številnih področjih kot so obdelava slik, prepoznavanje vzorcev, klasifikacija, odkrivanje anomalij, itd.

Tako v tem članku predstavimo postopek obdelave literature s pomočjo metode gručenja t-SNE (ang. *t-Distributed Stochastic Neighbour Embedding*) in aglomerativnega hierarhičnega združevanja (ang. *Agglomerative Hierarchical Clustering*) ter metode za zmanjšanje dimenzij podatkov PCA. Postopek je sestavljen iz več korakov: iskanje relevantne literature, ekstrakcija različnih podatkov (ključne besede, avtorji, čas nastanka, ipd.), zmanjšanje dimenzij podatkov in združevanje v skupine glede na podobnost. Pri sestavi in urejanju baze z literaturo ter njenem analiziranju, smo si pomagali s programskima orodjema Python in Matlab. Le-ta ponujata številne uporabne funkcije, s katerimi lahko pohitrimo sicer časovno zamuden proces. Relevantno literaturo s smo pridobili iz različnih odprto dostopnih revij, revij s faktorjem vpliva, repozitorijev različnih univerz ter arhivov raznih energetskih podjetij in operaterjev prenosnega omrežja. Za prikaz postopa smo v tem članku uporabili literaturo s področja pod-frekvenčnega razbremenjevanja, vendar je opisan postopek možno uporabiti pri analiziranju literature s kateregakoli področja.

1.1. Zaščita pod-frekvenčnega razbremenjevanja

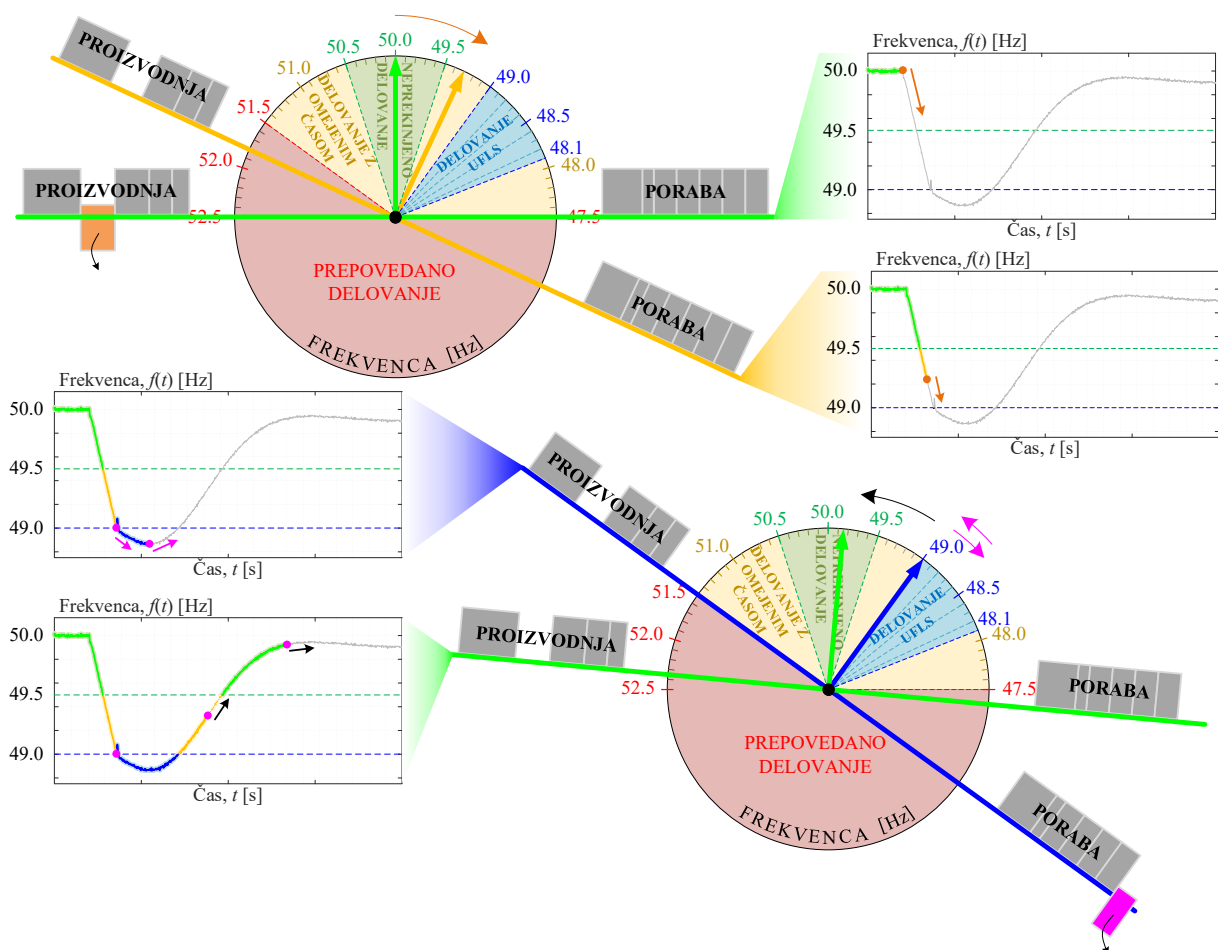
Električne energije ne moremo učinkovito shranjevati v velikih količinah, zato mora proizvodnja le-te ves čas slediti porabi. Indikator, ki nam pove, ali je temu res tako, je frekvenca, ki je zaradi konstrukcijskih omejitev turbin [16], v neprekinjenem obratovanju EES-a omejena na relativno ozek pas okoli nazivne vrednosti 50 Hz v ENTSO-E oziroma 60 Hz v NERC. Večino neravnovesij uspešno odpravi primarna regulacija frekvence, ampak ko pa pride do ekstremnih situacij, pa lahko kontrolni mehanizmi frekvence odreagirajo prepočasi ali pa nimajo zadostne kapacitete. V takih primerih prideta v ospredje pod- in nad-frekvenčna zaščita. Prva zagotovi začasno znižanje porabe v primerih velikih primanjkljajev moči, medtem ko druga, zagotovi znižanje proizvodnje v primerih velikih presežkov moči. V nadaljevanju se bomo osredotočili le na pod-frekvenčno zaščito (PFR).

Slika 1 prikazuje tipičen frekvenčni odziv EES-a ob nepričakovani okvari velike proizvodne enote, kjer barvno kodiranje označuje stanje v katerem se EES trenutno nahaja (tj. neprekinjeno, časovno omejeno in prepovedano

delovanje ter delovanje PFR). Pri normalnem obratovanju je frekvenca blizu nazivne vrednosti (glej zeleno puščico na zgornjem diagramu), ko pa pride do nepričakovane izgube večje količine proizvodnih enot (glej oranžno obarvan blok na zgornjem diagramu) le-ta prične padati (glej zgornji desni graf). Zaradi nezadostne vztrajnosti EES in zakasnjenege odziva primarne regulacije se frekvenca zmanjšuje, dokler se ne aktivira PFR (glej spodnji desni graf). PFR omejuje porabo tako, da odklopi del bremen (glej roza obarvan blok na spodnjem diagramu), in če je količina dovolj velika, frekvenca preneha padati. Sčasoma se zaradi mehanizmov za regulacijo frekvenca le-ta vrne na nazivno vrednost (glej spodnji levi graf).

Sivo obarvani bloki na Sliki 1, ki predstavljajo proizvodnjo in porabo električne energije, kažejo na to, da so proizvodne enote in odjemalci različnih velikosti, kar pomeni, da so tudi izpadi različnih velikosti. Zato, da se pravočasno vzpostavi ravnovesje, mora biti zaščita PFR zasnovana tako, da v različnih obratovalnih pogojih odklopi ravno pravšnje količino bremen. Zaradi kompleksnosti EES-a, velikega števila možnih obratovalnih stanj in operativnih omejitev obratovalne politike (npr. [17]) dizajniranje zaščite PFR ni enostavno opravilo. Tipične odločitve pri načrtovanju PFR so: *i*) najmanjša dovoljena frekvenca, *ii*) frekvenčni pragovi, *iii*) največja količina odklopa, *iv*) število razbremenitev in njihova velikost ter *v*) časovne zakasnitve.

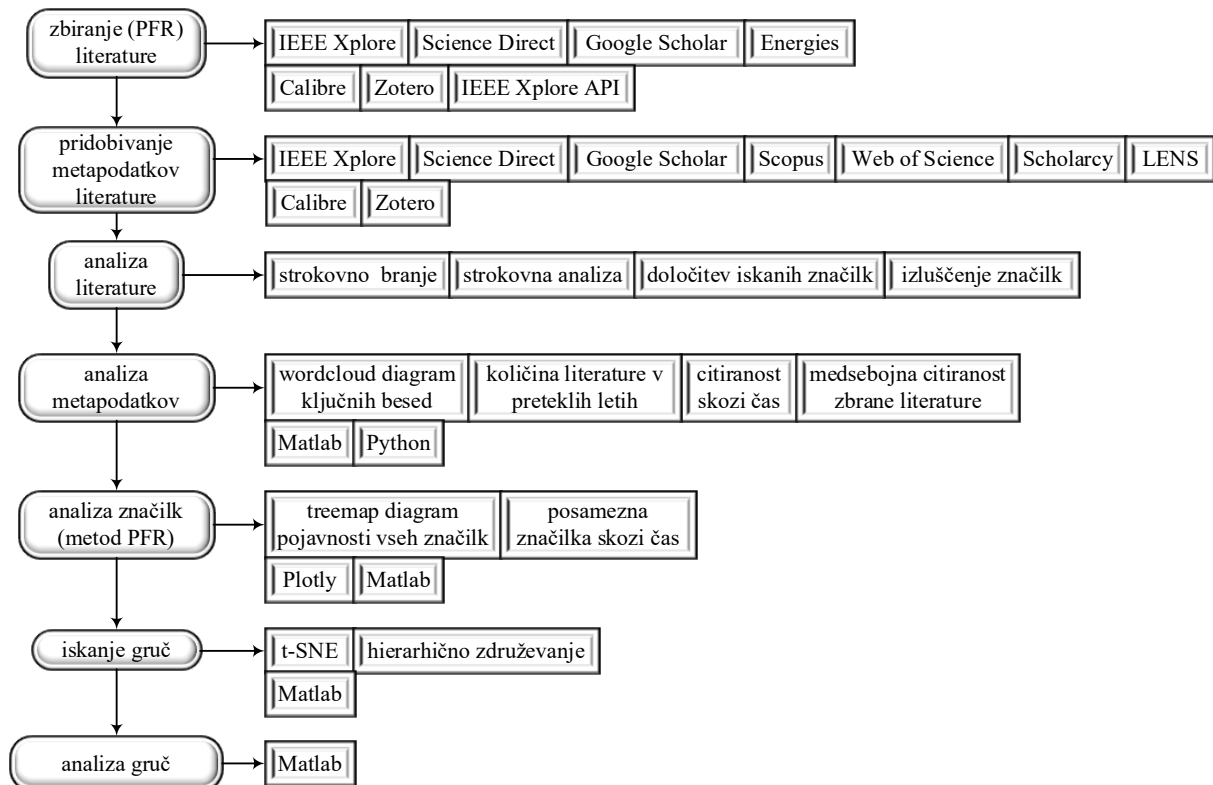
Rezultat več kot 30 let iskanja idealne zaščite PFR (tj. najmanjša možna razbremenitev v katerikoli kritični situaciji, pri čemer frekvenca ostane znotraj predpisanih mej) je veliko število različnih algoritmov PFR. Za sledenje in vzpostavljanje povezav med njimi raziskovalci uporabljajo pregledne članke, ki poskušajo predlagane algoritme PFR razdeliti glede na njihove podobne lastnosti. Obstoječi pregledni članki [18-25] razvrščajo na podlagi strokovnih mnenj in subjektivnih presoj, zato smo se v tem članku odločili, da predlagamo uporabo napredne matematične metode, ki omogoča sistematično in popolnoma objektivno združevanje algoritmov PFR.



Slika 1: Prikaz frekvenčne stabilnosti EES

2 METODOLOGIJA

Količina literature na področju zaščite PFR hitro narašča. To kaže na to, da je tema po eni strani aktualna, po drugi strani pa lahko zelo hitro izgubimo pregled nad vsemi predlaganimi metodami in njihovimi povezavami. Zato, da ohranimo jasnost, se raziskovalci zatekajo k preglednim člankom literature, ki skušajo sistematično kategorizirati algoritme PFR [18-25]. Ker to ni ravno trivialna naloga, je smiselno v ta namen uporabiti napredne matematične tehnike. Zato smo se v tem članku odločili, da bomo podobne sheme PFR poiskali s pomočjo PCA in t-SNE. Seveda, moramo pred njuno uporabo izvest še nekaj drugih korakov povezanih z zbiranjem literature, pridobivanjem informacij in analizo podatkov. Slika 2 na levi strani prikazuje sedem glavnih korakov postopka analiziranja literature opisanega v tem prispevku. Prav tako so na desni strani Slike 2 podani viri, metode in orodja uporabljena v posameznem koraku.



Slika 2: Glavni raziskovalni koraki in uporabljena orodja za analizo literature

2.1. Priprava zbirke z literaturo

Zbiranje literature PFR je potekalo v dveh korakih. V prvem smo literaturo v glavnem iskali preko angleških ključnih besed kot so "under-frequency load shedding", "UFLS", "frequency stability", "adaptive load shedding", itd. na spletnih straneh IEEE Xplore [26], Science Direct [27] in Google učenjaka [28], in sicer za obdobje od 1953 do 2020. Poleg tega smo del člankov z IEEE Xplore zbrali s pomočjo njihovega API-ja [29], implementiranega v Python skripti. Primer uporabe API prikazuje Slika 3. Tako smo v prvem koraku zbrali približno 200 člankov. V naslednjem koraku smo relevantno literaturo iskali med navedenimi referencami v že zbranih člankih. Ker gre tu za dolgotrajen proces, smo si delo nekoliko olajšali s preprosto Python skripto, ki najprej naloži .pdf dokument, izvleče vse besedilo iz tega dokumenta in poišče odsek z navedenimi referencami. Nato gre skozi naslove že zbranih člankov in preveri, ali so prisotni v izvlečenem besedilu. Torej kot rezultat dobimo vse preostale reference, ki so navedene v izbranem članku, ampak se pa ne nahajajo v naši zbirki literature. Ker je ekstrakcija besedila iz .pdf datotek nekaterih člankov, zlasti starejših, težavna zaradi slabe kakovosti .pdf dokumenta in napačne interpretacije nekaterih znakov, smo namesto neposredne primerjave zbranih in izvlečenih naslovov člankov iskali določeno podobnost med njimi. In sicer, zahtevali smo vsaj 90% ujemanje. V ta namen smo uporabili knjižnico mehkega ujemanja nizov TheFuzz.py [30], ki določi ujemanje dveh nizov znakov na podlagi izračunane Levenshteinove razdalje. Vsa na novo pridobljena literatura je bila uvožena, shranjena in organizirana v odprtokodnem upravitelju e-knjig Calibre [31] in orodju za pomoč pri raziskavah Zotero [32].

```

import xplore
import json

# Import API key -----
API_key = open('API_key.txt', 'r')
api_key = API_key.read()
API_key.close()

# Set search query -----
query = xplore.xploreapi.XPLORE(api_key)
query.resultSetMax = 100 # max 200 lahko dobis naenkrat

# Search options
query.abstractText('')
query.affiliationText('')
query.articleNumber('')
query.articleTitle('')
query.authorText('')
query.authorFacetText('')
query.doi('')
query.publisherFacetText('')
query.publicationFacetText('')
query.publicationYear('2020')
query.facetText('')
query.indexTerms('UFLS') # authors, IEEE, mixed terms
query.isbn('')
query.issn('')
query.issueNumber('')
query.metaDataText('')
query.publicationTitle('')
query.thesaurusTerms('') # IEEE terms
query.queryText('')

# Run query -----
data = query.callAPI()
print(data)

# JSON -> Python (dictionary) -----
# data = json.loads(data)

# Save data into .txt -----
file = open('data.txt', 'w')
file.write(data)
file.close()

```

```

PDF_URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9898346
ARTICLE_NUMBER: 9898346
PUBLICATION_NUMBER: 59
ABSTRACT: The reliability of power systems with high penetration of renewables, a fast response mecha
RANK: 1
START_PAGE: 4815
CITING_PAPER_COUNT: 1
TITLE: A Scheme to Mitigate Generation Trip Events by Ancillary Services Considering Minimal Actions
ABSTRACT_URL: https://ieeexplore.ieee.org/document/9898346/
ISSUE: 6
IS_NUMBER: 9239415
INDEX_TERMS
-> author_terms: under-Frequency Load Shedding (UFLS), Ancillary services, frequency nadir predicti
-> ieee_terms: Phasor measurement units, Frequency control, Power system stability, Demand response
PUBLICATION_TITLE: IEEE Transactions on Power Systems
HTML_URL: https://ieeexplore.ieee.org/document/9898346/
VOLUME: 35
ACCESS_TYPE: LOCKED
CONTENT_TYPE: Journals
AUTHORS
-> author_order: 1
-> affiliation: Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan
-> authorUrl: https://ieeexplore.ieee.org/author/37886311249
-> full_name: Chao-Yuan Lai
-> id: 37886311249
-----
-> author_order: 2
-> affiliation: Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan
-> authorUrl: https://ieeexplore.ieee.org/author/37286618800
-> full_name: Chih-Wen Liu
-> id: 37286618800
-----
PUBLICATION_DATE: Nov. 2020
PUBLISHER: IEEE
CITING_PATENT_COUNT: 0
DOI: 10.1109/TPWRS.2020.2993449
ISSN: 1558-0679
PUBLICATION_YEAR: 2020
END_PAGE: 4823

```

Slika 3: Primer uporabe IEEE Xplore API [29] v programskem okolju Python (levo) in rezultat iskanja (desno)

2.2. Pridobivanje metapodatkov

Zato, da lahko smiselno organiziramo zbrano literaturo s pomočjo Calibre, Zotero in drugih podobnih programov moramo pridobiti nekatere informacije (metapodatke) o posamezni publikaciji, kot so naslov, avtorji, založnik, datum objave, revija, citati, ISBN, DOI, itd. Vse te informacije pogostokrat niso na voljo na mestu, kjer smo jih našli, zato jih poskušamo najti s pomočjo drugih spletnih virov kot so Web of Science [33], Scopus [34], Scholarcy [35] in LENS [36]. Včasih je nekatere metapodatke možno pridobiti tudi neposredno iz datotek .pdf z uporabo različnih Python knjižnic kot so pdfssa4met.py [37], Grobid [38], CERMIN [39], itd.

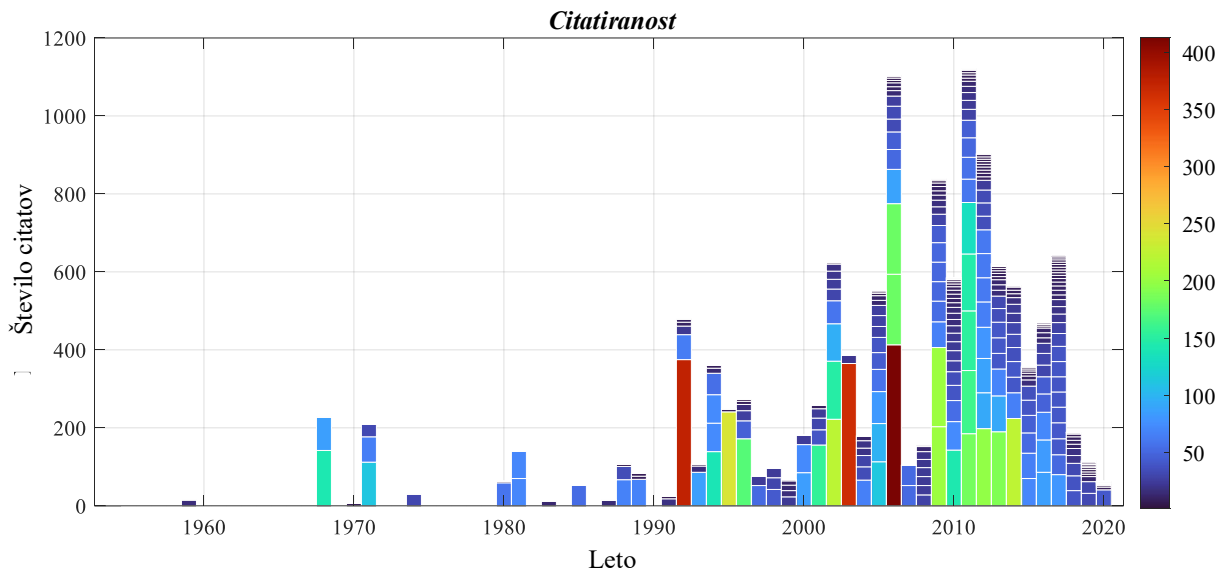
2.3. Analiza literature

Ko zaključimo z zbiranjem in organizacijo literature (s področja PFR), sledi njeno preučevanje in analiza. Raziskave na nekem področju se lahko izvajajo že nekaj desetletij (npr. prve publikacije glede PFR segajo okoli 60 let nazaj), kar pomeni, da je bilo predlaganih nešteto različnih metod. Ker ima vsaka opisana metoda neke svoje značilnosti, njihova primerjava in kategorizacija ni enostavna. V primeru PFR, smo izbrali in kategorizirali metode na podlagi 55 značilik (glej Poglavlje 3). Ti podatki so kasneje uporabljeni kot vhodni podatki v algoritem združevanja v gruče (glej Poglavlje 4).

2.4. Analiza metapodatkov

Eden najpomembnejših sestavnih delov publikacij je razdelek s ključnimi besedami, saj le-te opisujejo bistvo obravnavane tematike, kar olajša iskanje ustrezne literature. Ključne besede običajno določijo avtorji sami na podlagi tehničnih izrazov relevantnih za neko področje. Nekateri založniki in knjižnice pa poleg avtorjevih dodajo še svoje ključne besede, saj za označevanje in urejanje baz uporabljajo neke lastne indekse. Na primer, IEEE plega avtorjevih in lastnih oznak članke dodatno označuje z indeksnimi izrazi Inspec. Inspec je ena izmed največjih zbirk znanstvene in tehnične literature, ki jo je pred več kot 50 leti ustanovil Inštitut za tehniko in tehnologijo (*Institution of Engineering and Technology*, IET) [40]. Večina publikacij v naši zbirki izvira iz knjižnice IEEE Xplore, zato Slika 4 (*wordcloud*) prikazuje štiri različne kategorije ključnih besed, in sicer ključne besede avtorjev, ključne besede IEEE in ključne besede Inspec. Barva in velikost besed označujeta pojavnost ključnih besed, in sicer manjša kot je velikost pisave ter modra barva, manjkraj se neka ključna beseda

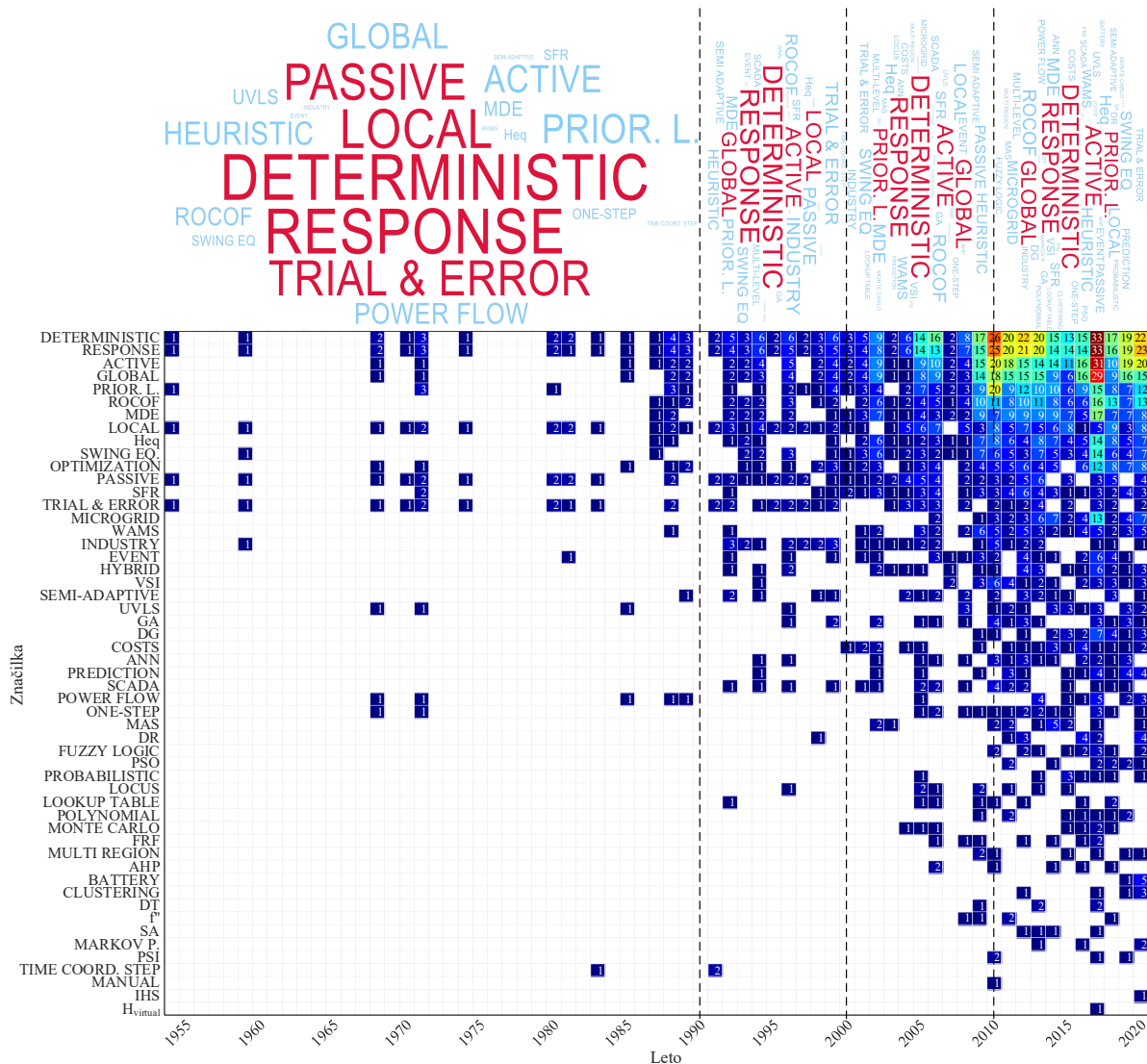
področju. Slika 6 prikazuje skupno število citatov v preteklih letih, kjer barva označuje število citatov posamezne objave (temno modra barva pomeni majhno število citatov in temno rdeča barva pomeni veliko število citatov). Vidimo lahko, da se število citatov z leti spreminja, vendar je mogoče opaziti določen trend z vrhom leta 2011 (skupno 1100 citatov). Najbolj citirana publikacija je članek [41] iz leta 2006, ki mu sledijo članki [42] in [43] iz leta 1992 in 2003.



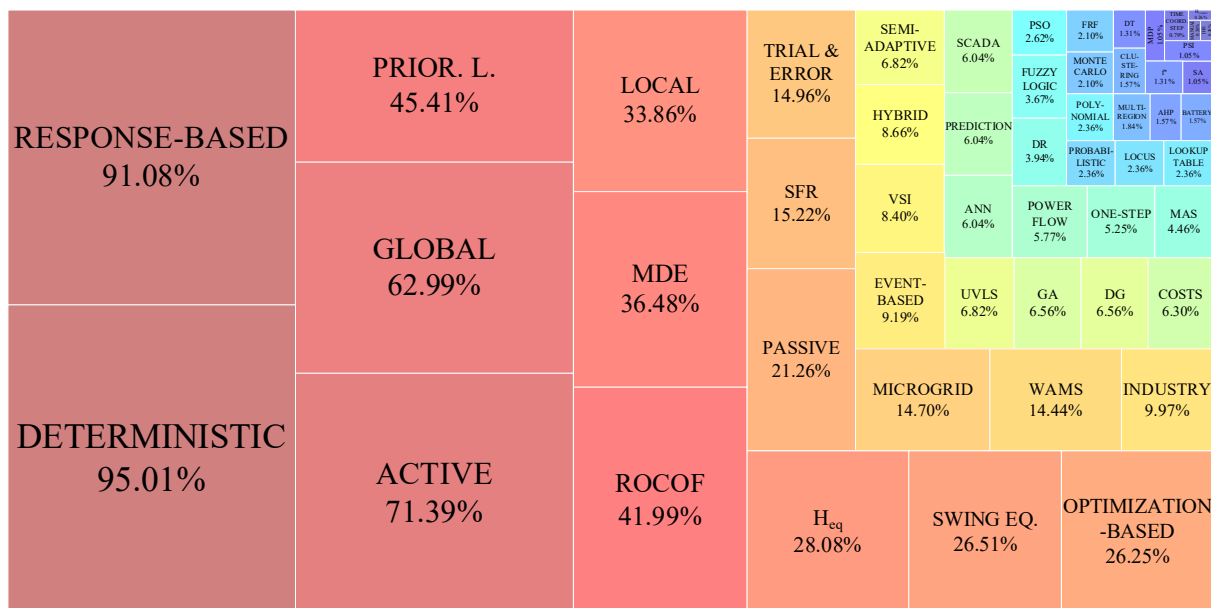
Slika 6: Skupno število navedb v preteklih letih, na podlagi podatkov pridobljenih 20. 1. 2021 na Google Učenjaku [28]

2.5. Analiza značilk metod PFR

Kot smo že omenili, smo za grupiranje algoritmov PFR po podobnosti, najprej poiskali posebne značilnosti (značilke) vsakega algoritma. Slika 7 prikazuje kolikokrat se je vsaka značilka pojavila v vsakem letu od leta 1954 do 2020. Značilke na ordinatni osi so razvrščene glede na skupno število pojavitev v celotnem obdobju opazovanja (glej Sliko 8). Značilka "deterministic" je najpogostejša (pojavlja se v 362/381=95.01% primerov), značilka " H_{virtual} " pa je najmanj pogosta (pojavlja se v 1/381=0.26% primerov). Prazno polje na Sliki 7 pomeni, da v našem naboru literature ni objave z značilko označeno na levi, barva polj pa označuje število objav in se spreminja od temno modre (majhno število publikacij) do temno rdeče (veliko število publikacij). Zgornji del Slike 7 prikazuje najpogostejše značilke v štirih obdobjih. Vidimo lahko, da so bile sheme PFR do leta 1990 opisane z manj značilkami, kar kaže na to, da so bile PFR sheme enostavnejše. In sicer, večina shem PFR je tako imenovanih tradicionalnih/konvencionalnih, kar pomeni, da delujejo lokalno, pri čemer izklopijo predhodno določeno fiksno količino bremen, ko frekvenca doseže predhodno določen frekvenčni prag. Količina bremen in frekvenčni prag sta določena na podlagi obsežnih dinamičnih simulacij oziroma po metodi "*trial & error*". Po letu 1990 se je raznolikost shem PFR začela postopoma povečevati. Sheme postajajo vse bolj prilagodljive, odločitve pa sprejemajo na centralni lokaciji na podlagi lokalno zajetih meritev. Poleg tega je vse več pozornosti usmerjene ključno po katerem se izvaja razbremenjevanje.



Slika 7: Značilke PFR shem skozi čas, razvrščene po skupnem številu nastopov



2.6. Rojenje

V tem prispevku so sheme PFR opisane s 55 značilkami, kar pomeni, da vsaka PFR predstavlja podatek v 55-dimenzionalnem prostoru. Ker je to za človeka nepredstavljivo, moramo uporabiti tehnike za zmanjšanje dimenzij. Sicer obstaja veliko tehnik, ki nam to omogočajo [14], vendar smo na podlagi pozitivnih izkušenj [8,44] za kategorizacijo zbrane literature izbrali metodo t-SNE. t-SNE je nelinearna metoda za zmanjšanje dimenzij podatkov, ki lahko preslika visoko-dimenzionalne podatke v dvo- ali tri-dimenzionalen prostor, pri čemer ohrani lokalno in globalno strukturo podatkov v eni sami preslikavi [44]. Splošni koraki algoritma t-SNE so [44,45]:

- 1) priprava podatkov,
- 2) izračun razdalj med visoko-dimenzionalnimi točkami,
- 3) izračun Gaussovih variacij – generiranje standardnega odklona za vsako visoko-dimenzionalno točko tako, da je efektivno število lokalnih sosedov vsake točke enako predhodno določeni vrednosti,
- 4) izračun matrike podobnosti (tj. skupna verjetnost podatkovnih točk),
- 5) ustvaritev začetnega niza nizko-dimenzionalnih točk,
- 6) optimizacija – iterativno posodabljanje nizko-dimenzionalnih točk zato, da zmanjšamo Kullback-Leiblerjevo divergenco v visoko-dimenzionalnem prostoru in t-porazdelitev v nizko-dimenzionalnem prostoru.

Glavne nastavitve algoritma t-SNE so [44,45]:

- metrika razdalje (*distance metric*) – način izračuna razdalje med podatkovnimi točkami; značilne metrike razdalj so Evklidska, Cityblock, Chebycheva, Minkowska, Kosinusna itd.
- pretiravanje (*exaggeration*) – velikost naravnih skupin v podatkih; večja vrednost povzroči, da se t-SNE nauči večje skupne verjetnosti in ustvari relativno večjo razdaljo med gruči
- število dimenzij (*number of dimensions*) – dimenzija izhoda (dvo- ali tri-dimenzionalen)
- število komponent PCA (*number of PCA components*) – če je dimenzija podatkov večja kot 50 je priporočljivo, da se dimenzijo prvotnih podatkov zmanjša s pomočjo PCA
- zmedenost (*perplexity*) – efektivno število lokalnih sosedov vsake podatkovne točke; večja vrednost pomeni, da t-SNE algoritem vzame več točk pri določanju najbližjih sosedov; za večje količine podatkov je priporočljivo uporabiti veliko vrednost
- stopnja učenja (*learning rate*) – stopnja učenja v procesu optimizacije; če je stopnja učenja premajhna, se lahko t-SNE algoritem konvergira k slabemu lokalnemu minimum, če pa je prevelika, pa lahko optimizacija poveča Kullback-Leiblerjevo divergenco namesto, da bi jo zmanjšala

postopek optimizacije je dolgotrajen, zato t-SNE algoritem ni primeren za uporabo v aplikacijah v realnem času. Običajno se uporablja za vizualizacijo podatkov, obdelavo slik in obdelavo naravnega jezika (ang. *Natural Language Processing*, NLP), kjer čas ni problem. Enako velja za kategorizacijo literature, zato se je izkazalo, da je t-SNE algoritem primeren tudi za tovrstne aplikacije.

Za razvrščanje transformiranih podatkov po podobnosti smo uporabili aglomerativno hierarhično združevanje (ang. *agglomerative hierarchical clustering*) [46]. Ta metoda v bistvu zgradi hierarhijo gruč tako, da začne od spodaj navzgor, pri čemer se vsaka opazovana točka nahaja v svoji gruči, nato pa se združi z najbolj podobnimi točkami. Podobnost med vsakim parom podatkovnih točk se meri z ustrežno metriko razdalje (npr. Evklidska, Manhattan, Mahalanobis, itd.). Podatkovne točke se nato razvrstijo v gruče glede na funkcijo povezovanja (*linkage function* – *Average, Centroid, Ward...*).

Gruče nastale po uporabi t-SNE in rezultati hierarhičnega združevanja v skupine so predstavljeni v Poglavlju 4.

3 MERILA ODLOČANJA

V tem prispevku ločimo dve vrsti značilk: i) splošne značilke, ki jih je mogoče dodeliti katerikoli shemi PFR, ii) specifične značilke, ki jih je mogoče dodeliti samo posameznim shemam PFR.

3.1. Splošne značilke

Splošne značilnosti metod PFR lahko razdelimo glede na:

- * kraj, kjer so izvedeni izračuni in se sprejemajo odločitve, kot so npr. količina razbremenitve in katera bremena bomo odklopili
 - lokalna/decentralizirana/porazdeljena (ang. *local/decentralized/distributed*)
 - globalna/centralizirana (ang. *global/centralized*)
 - kombinirana/hibridna I (ang. *combined/hybrid I*)

- * njihovo fleksibilnost (tj. sposobnost prilagoditve trenutnim obratovalnim razmeram EES-a)
 - pasivna/statična/neprilagodljiva/tradicionalna/konvencionalna/večstopenjska/reaktivna (ang. *passive/static/non-adaptive/traditional/conventional/multi-stage/round/reactive*)
 - pol-prilagodljiva/hibridna (ang. *semi-adaptive/hybrid*)
 - aktivna/prilagodljiva/inteligenta/dinamična/samo-prilagodljiva/proaktivna (ang. *active/adaptive/intelligent/dynamic/self-adapting/proactive*)
 - kombinirana/hibridna II (ang. *combined/hybrid II*)
- * njihov sprožilni mehanizem
 - dogodkovna/vzročna (ang. *event-based/caused-based/contingency-based*)
 - odzivna/efektivna (ang. *response-based/effect-based*)
 - vzročno-efektivna/hibridna III (ang. *event- and response-based/hybrid III*)
- * tip algoritma
 - deterministična (ang. *deterministic*)
 - verjetnostna (ang. *probabilistic*)
 - optimizacijska (ang. *optimization-based*)
- * vrsto delovanja
 - ročna (ang. *manual*)
 - avtomatska (ang. *automatic*)

3.2. Specifične značilke

Preostale značilnosti PFR lahko razdelimo glede na:

- * vrsto omrežja v katerem so implementirane
 - več-regionalno omrežje/prostrano sinhronsko omrežje/interkonekcija/nacionalno omrežje (ang. *multi-regional grid/wide-area synchronous grid/interconnection/national grid*)
 - industrijsko omrežje (ang. *industrial grid*)
 - mikro-omrežje(ang. *microgrid*, MG)
 - konvencionalno omrežje/tradicionalno omrežje/makro-omrežje (ang. *conventional grid/ traditional grid/macro grid*)
- * uporabljeno optimizacijsko metodo
 - genetski algoritem (ang. *genetic algorithm*, GA)
 - optimizacija izboljšano iskanje harmonikov (ang. *improved harmonic search*, IHS)
 - simulirana optimizacija žarjenja (ang. *simulated annealing optimization*, SA)
 - optimizacija roja delcev (ang. *particle swarm optimization*, PSO)
- * metodo za oceno stabilnosti
 - napetostni faktor občutljivosti (ang. *voltage sensitivity indices*, VSI)
 - faktor regulacije frekvence (ang. *frequency regulation factor*, FRF)
 - napoved (ang. *prediction*)
 - polinom (ang. *polynomial*)
 - drugi odvod frekvence (ang. *second frequency time deriver*, f'')
 - gručenje (ang. *clustering*)
- * vzorec razbremenjevanja
 - en korak (ang. *one-step*)
 - po pomembnosti (ang. *priority-based*)
 - odziv povpraševanja (ang. *demand response*, DR)
 - iskalna tabela (ang. *lookup table*)
- * računsko inteligenco
 - umetna nevronska mreža (ang. *artificial neural network*, ANN)
 - mehka logika (ang. *fuzzy logic*, FL)
 - odločitveno drevo (ang. *decision tree*, DT)
 - sistem več agentov (ang. *multi-agent system*, MAS)
 - Markov postopek odločanja (ang. *Markov decision process*, MDP)
 - analitični hierarhični proces (ang. *Analytical hierarchy process*, AHP)
- * metodo za nastavitev parametrov PFR

- poskušanje (ang. *trial & error*)
- Monte Carlo metoda (ang. *Monte Carlo*)
- pretok moči (ang. *power flow (topology-based)*)
- model frekvenčnega odziva EES (ang. *system frequency response (SFR) model*)
- nihajna enačba (ang. *swing equation*)
- velikost primanjkljaja (ang. *magnitude of disturbance estimation, MDE*)
- lokus (ang. *locus (df/dt(f))*)
- hitrost spremembe frekvence (ang. *rate of change of frequency, ROCOF*)
- * sistem za zbiranje podatkov
 - SCADA (ang. *supervisory control and data acquisition*)
 - WAMS (ang. *wide-area measurement system*)
- * na ostale uporabljene informacije
 - napetostna stabilnost (ang. *voltage stability, UVLS*)
 - stroški (ang. *costs*)
 - porazdeljena generacija (ang. *distributed generation, DG*)
 - podpora baterije (ang. *battery support*)
 - virtualna inercija (ang. *virtual inertia, H_{virtual}*)
 - čas (časovno koordinirana PFR stopnja) (ang. *time (time-coordinated UFLS step)*)

4 KATEGORIZACIJA LITERATURE

Slika 9 prikazuje rezultat po uporabi algoritma t-SNE in algoritma za hierarhično združevanje. Nastavitve prvega algoritma so vidna v levem spodnjem kotu Slike 9, nastavitve drugega algoritma pa so kvadratna Evklidska povprečna metrika razdalje in povprečna povezava (*average linkage*).

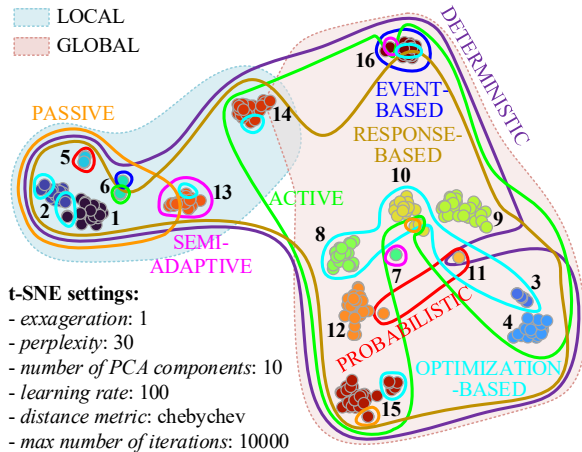
Dobili smo 16 različnih skupin algoritmov PFR (vsaka skupina je na Sliki 9 označena z unikatno okrepljeno številko), ki jih lahko v splošnem razdelimo na "lokalne" (glej modro območje) in "globalne" (glej rdeče območje) sheme PFR.

"Lokalne" sheme PFR lahko nadalje razdelimo na "pasivne" (glej skupine **1, 2, 5, 6** in oranžno krivuljo na), "pol-prilagodljive" (glej skupino **13** in roza krivuljo) in "aktivne" (glej skupino **14** in zeleno krivuljo). Večina shem PFR, ki spadajo v te tri kategorije so "odzivne" (glej skupine **1, 2, 5, 6, del 13 ter 14** in rjavo krivuljo), obstaja pa nekaj shem, ki so "dogodkovne" (glej skupino **6** in temno modro krivuljo). Poleg tega vidimo, da nekatere sheme PFR v vsaki od treh kategorij uporabljajo nekakšno "optimizacijo" (glej skupine **1, 2, 13, 14** in svetlo modro krivuljo). Kot nakazuje rdeča krivulja, lahko "pasivne" sheme PFR med postopkom načrtovanja vključujejo tudi izračun "verjetnosti" (glej skupino **5**).

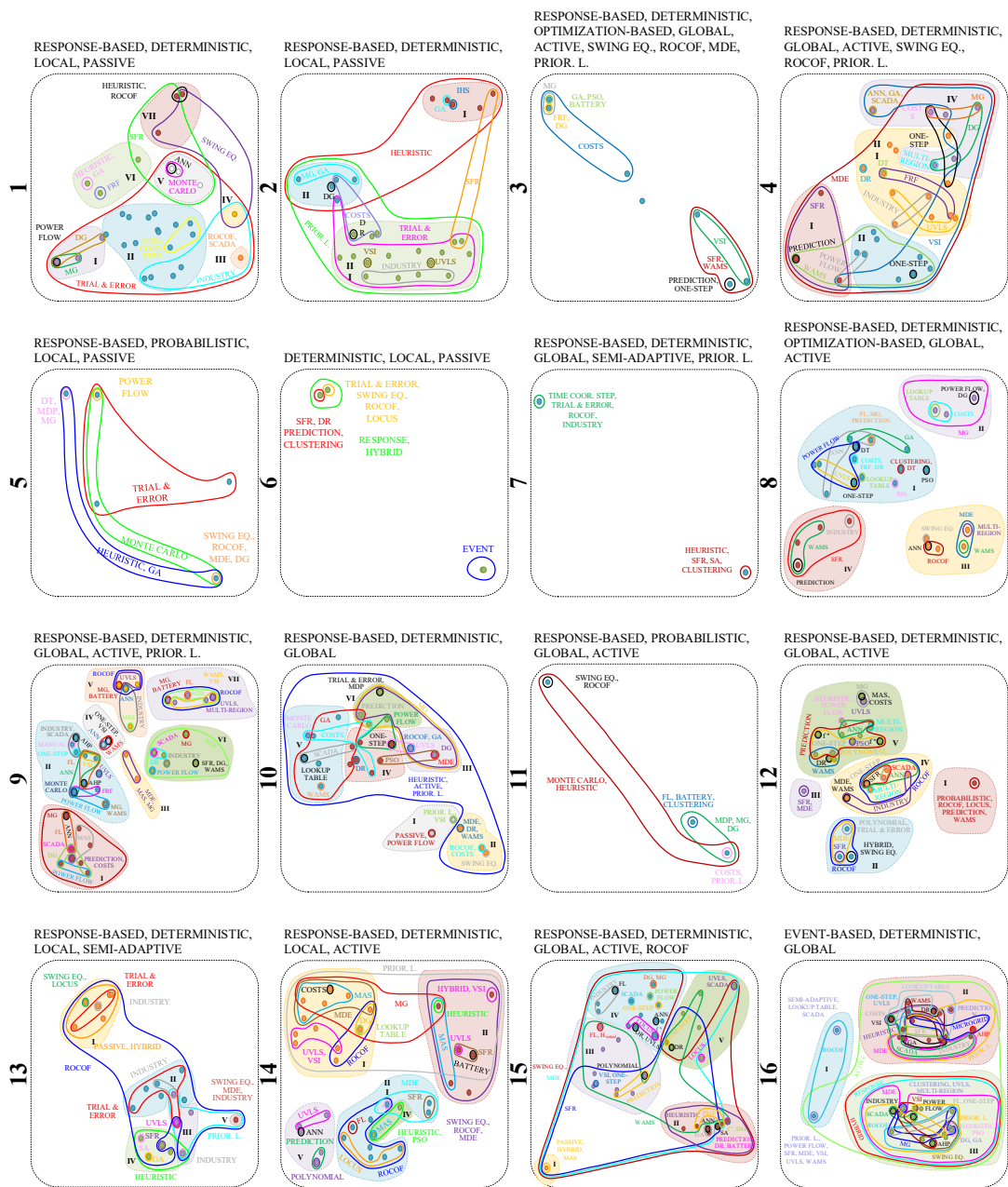
"Globalne" sheme PFR so v glavnem "deterministične", "odzivne" in "aktivne" (glej vijolično, rjavo in zeleno krivuljo na Sliki 9). Nadalje jih lahko razdelimo na "dogodkovne" (glej skupino **16** in temno modro krivuljo), "optimizacijske" (glej skupine **3, 8, 10, 11, 15** in svetlo modro krivuljo) in "verjetnostne" (glej skupini **11 in 12** ter rdečo krivuljo).

Na Sliki 9 lahko prav tako opazimo "hibridne" sheme PFR. In sicer, zelena krivulja pri skupini **6** in oranžna krivulja pri skupini **15** predstavljata sheme PFR, ki združuje lastnosti "aktivne" in "pasivne" sheme, presečišče oranžne in roza krivulje pri skupini **13** predstavlja sheme, ki imajo značilnosti "pol-prilagodljivih" in "pasivnih" shem, presek modrega in rdečega območja pri skupini **14** predstavlja sheme, ki kombinira lastnosti "lokalnih" in "globalnih" shem ter presečišče rjave in temno modre krivulje pri skupini **16** predstavlja sheme, ki imajo lastnosti "dogodkovnih" in "odzivnih" shem PFR.

Slika 9 prikazuje razdelitev shem PFR na podlagi splošnih značilnosti. Če pa nas zanima bolj podrobna razdelitev, pa moramo analizirati posamezno skupino. Tako slika 10 prikazuje bolj podrobno razdelitev posameznih skupin na podlagi specifičnih značilnosti. Pri nadaljnji razdelitvi posamezne skupine smo si ponovno pomagali s tehniko hierarhičnega združevanja, kjer smo ohranili nastavitve iz prejšnjega koraka (torej kvadratna Evklidska razdalja in povprečna povezava). Tako smo večjo skupino razbili na več manjših. Vsaka skupina na Sliki 10 je na levi strani označena s pripadajočo unikatno številko (le-te ustrezajo številkam na Sliki 9), na vrhu pa so podane skupne značilnosti shem PFR v tisti skupini. Hitro lahko opazimo, da so skupine različnih velikosti ter da so nekatere bolj raznovrstne, nekatere pa malo manj.



Slika 9: Splošna razdelitev PFR



Slika 10: Podrobna razdelitev PFR

5 ZAKLJUČKI

V tem prispevku smo predstavili postopek za sistematično analiziranje literature z nekega področja, pri čemer smo za primer uporabili področje zaščite pod-frekvenčnega razbremenjevanja. Z leti je bilo predlaganih veliko različnih shem pod-frekvenčnega razbremenjevanja, kar po eni strani pomeni, da je problem aktualen in kompleksen, po drugi strani pa zlahka zgubimo pregled in povezave med vsemi rešitvami. Prav tako težko zaznamo trend in vrzeli na področju. Zato v tem prispevku predlagamo objektivni in čedalje bolj uveljavljen pristop k analizi literature. Postopek je sestavljen iz sedmih glavnih korakov, in sicer zbiranje relevantne literature, pridobivanje metapodatkov, analiza literature, analiza metapodatkov, analiza značilk, iskanje gruč in analiza gruč. Z analizo metapodatkov lahko ugotovimo kakšno je stanje in dinamika področja ter kolikšen doprinos ima posamezna publikacija na področju. V procesu analize literature smo na podlagi značilnosti opisane metode posamezni publikaciji pripisali določene značilke. Z analizo značilk lahko ugotovimo kako poteka razvoj področja ter kje so možne vrzeli. Značilke so kasneje služile tudi kot vhod v algoritem za gručenje t-SNE in algoritem za hierarhično združevanje. Kot izhod smo dobili 16 skupin shem pod-frekvenčnega razbremenjevanja, pri čemer smo najprej izvedli analizo na podlagi splošnih značilnosti, kasneje pa smo posamezne skupine nadalje razdelili še glede na specifične lastnosti shem.

ZAHVALA

To delo je podprla Slovenska raziskovalna agencija z raziskovalnim programom Elektroenergetski sistemi št. P2-0356, mehanizmom financiranja mladih raziskovalcev in projektom LoLaG, J2-9232. Avtorji se zahvaljujejo Javni agenciji za raziskovalno dejavnost RS za finančno pomoč.

REFERENCE

- [1] W. Akram, S. Noor, A. Manzoor, "The State of the Art in Smart Grid Domain: A Network Modeling Approach," *Broad Research in Artificial Intelligence and Neuroscience (BRAIN)*, Vol. 11, No. 4, pp. 201-230, 2020.
- [2] P. Nieminen, I. Pölönen, T. Sipola, "Research literature clustering using diffusion maps," *Journal of Informatics*, Vol. 7, pp. 874-886, 2013.
- [3] T. Weißer, T. Saßmannshausen, D. Ohrndorf, P. Burggräf, J. Wagner, "A clustering approach for topic filtering within systematic literature review," *MethodsX*, pp. 1000831, 2020.
- [4] S. Consoli, N. I. Stilianakis, "A VNS-based quartet algorithm for biomedical literature clustering," *Electronic Notes in Discrete Mathematics*, Vol. 47, pp. 13-10, 2015.
- [5] Y. Yamamoto, T. Takagi, "Biomedical knowledge navigation by literature clustering," *Journal of Biomedical Informatics*, Vol. 40, pp. 114-130, 2007.
- [6] W.-L. Shiau, Y. K. Dwivedi, H. S. Yang, "Co-citation and cluster analysis of extant literature on social networks," *International Journal of Information Management*, Vol. 37, pp. 390-399, 2017.
- [7] J. P. Minas, N. C. Simpson, Z. Y. Tacheva, "Complete bibliographic data, cluster assignment and combined citation network of emergency response operations research extant literature," *Data in Brief*, Vol. 31, pp. 105908, 2020.
- [8] M. E. Eren, N. Solovyev, E. Raff, C. Nicholas, B. Johnson, "COVID-19 Kaggle Literature Organization," In *Proceedings of the ACM Symposium on Document Engineering*, Nr. 15, CA, USA, 2020.
- [9] P. Li, "Literature Characterization and Similarity Retrieval Based on Hierarchical Clustering," *WRI World Congress on Software Engineering*, Xiamen, China, May 2009.
- [10] T. Tong, D. Dinakarpandian, Y. Lee, "Literature Clustering using Citation Semantics," *Proceedings of the 42nd Hawaii International Conference on System Sciences*, Jan. 2009.
- [11] J. Hong, S.-H. Yoon, W.-S. Hwang, D.-H. Bae, S.-W. Kim, "On Analyzing Scientific Literature Databases," *6th International Conference on Computer Sciences and Convergence Information Technology (ICCIT)*, Seogwipo, South Korea, Dec. 2011.
- [12] J. Yun, S. Ahn, J. Y. Lee, "Return to basics: Clustering of scientific literature using structural information," *Journal of Informetrics*, Vol. 14, pp. 101099, 2020.
- [13] H. Lane, C. Howard, H. M. Hapke, *Natural Language Processing IN ACTION – Understanding, analyzing, and generating text with Python*, Shelter Island, NY, USA: Manning Publications, 2019.
- [14] Z.-H. Zhou, S. Liu, *Machine Learning*, Beach Road, Singapore: Springer Nature Singapore, 2021.
- [15] C. M. Bishop, *Pattern recognition and machine learning*, New York, NY, USA: Springer-Verlag, 2006.
- [16] IEEE Guide for Abnormal Frequency Protection for Power Generating Plants, IEEE Std C37.106-2003, Feb. 2004.
- [17] ENTSO-E, "RG CE OH – Policy 5: Emergency Operations V 3.1," Sep. 2017.

- [18] L. Sigrist, L. Rouco, F. M. Echavarren, "A review of the state of the art of UFLS schemes for isolated power systems," *Electric Power and Energy Systems*, Vol. 99, pp. 525–539, Jul. 2018.
- [19] H. H. Alhelou, M. E. H. Golshan, T.C. Njenda, N. D. Hatziaargyriou, "An Overview of UFLS in Conventional, Modern, and Future Smart Power Systems: Challenges and Opportunities," *Electric Power System Research*, Vol. 179, pp. 106054, Feb. 2020.
- [20] J. A. Laghari, H. Mokhlis, A. H. A. Bakar, H. Mohamad, "Application of computational intelligence techniques for load shedding in power systems: A review," *Energy Conversion and Management*, Vol. 75, pp. 130-140, Nov. 2013.
- [21] P. Lakra, M. Kirar, "Load Shedding Techniques for System with Cogeneration: A Review," *Electrical and Electronics Engineering: An International Journal (ELELIJ)*, Vol. 4, No. 3, Aug. 2015.
- [22] O. Shariati, A. A. Mohd Zin, A. Khairuddin, M. H. A. Pesaran, M. R. Aghamohammadi, "An Integrated Method for Under Frequency Load Shedding Based on Hybrid Intelligent System-Part II: UFLS Design," 2012 Asia-Pacific Power and Energy Engineering Conference, Shanghai, China, 27-29 Mar. 2012.
- [23] Y. Xu, Y. Dai, Z. Y. Dong, Y. Xue, K. P. Wong, "Load Shedding and Its Strategies Against Frequency Instability in Power Systems," 2012 IEEE Power and Energy Society General Meeting, San Diego, CA, USA, 22-26 Jul. 2012.
- [24] N. N. A. Bakar, M. Y. Hassan, M. F. Sulaima, M. N. M. Nasir, A. Khamis, "Microgrid and load shedding scheme during islanded mode: A review," *Renewable and Sustainable Energy Reviews*, Vol. 71, pp. 161-169, May 2017.
- [25] N. M. Sapari, H. Mokhlis, J. A. Laghari, A. H. A. Bakar, M. R. M. Dahalan, "Application of load shedding schemes for distribution network connected with distributed generation: A review," *Renewable and Sustainable Energy Reviews*, Vol. 82, No. 1, pp. 858-867, Feb. 2018.
- [26] Spletna stran IEEE Xplore: <https://ieeexplore.ieee.org/Xplore/home.jsp>
- [27] Spletna stran Science Direct: <https://www.sciencedirect.com/>
- [28] Spletna stran Google Učejak: <https://scholar.google.com.tw/>
- [29] Spletna stran IEEE Xplore API: <https://developer.ieee.org/>
- [30] Python knjižnica TheFuzz: <https://github.com/seatgeek/thefuzz>
- [31] upravitelj e-knjig Calibre: <https://calibre-ebook.com>
- [32] orodje Zotero: <https://www.zotero.org/>
- [33] Spletna stran Web of Science: <https://www.webofknowledge.com/>
- [34] Spletna stran Scopus: <https://www.scopus.com/home.uri>
- [35] Spletna stran Scholarcy: <https://library.scholarcy.com/flashcard-generator.html>
- [36] Spletna stran LENS: <https://www.lens.org/>
- [37] Programsko orodje pdfssa4met: <https://github.com/eliask/pdfssa4met>
- [38] Programsko orodje Grobid: <https://github.com/kermitt2/grobid>
- [39] Programsko orodje CERMINE: <https://github.com/CeON/CERMINE>
- [40] Spletna stran Inspec: <https://www.theiet.org/publishing/inspec/>
- [41] V. Terzija, "Adaptive Underfrequency Load Shedding Based on the Magnitude of the Disturbance Estimation," *IEEE Transactions on Power Systems*, Vol. 21, No. 3, Aug. 2006.
- [42] P. M. Anderson, M. Mirheydar, "An Adaptive Method for Setting Underfrequency Load Shedding Relays," *IEEE Transactions on Power Systems*, Vol. 7, Iss.2, pp. 647-655, May 1992.
- [43] H. You, V. Vittal, Z. Yang, "Self-Healing in Power Systems: An Approach Using Islanding and Rate of Frequency Decline-Based Load Shedding," *IEEE Transactions on Power Systems*, Vol. 18, No. 1, Feb. 2003.
- [44] L. Maaten, G. Hinton, "Visualizing Data using t-SNE," *Journal of Machine Learning Research*, Vol. 9, pp. 2589-2605, Nov. 2008.
- [45] Spletna stran MathWorks, dokumentacija tsne: <https://www.mathworks.com/help/stats/t-sne.html>
- [46] Spletna stran MathWorks, dokumentacija Hierarchical Clustering: https://www.mathworks.com/help/stats/hierarchical-clustering.html#bq_679x-4