

Python Pakket en GUI voor FlowSOM

Elias Nijs

Computationele Biologie
Universiteit Gent

ABSTRACT

Dit verslag presenteert de ontwikkeling en implementatie van het FlowSOM package voor python en de ontwikkeling van een grafische gebruikersinterface (GUI) ervoor. Op deze manier maken we het FlowSOM algoritme toegankelijk voor zowel mensen met als zonder technische achtergrond.

11 June 2023

Python Pakket en GUI voor FlowSOM

Elias Nijs

Computationale Biologie
Universiteit Gent

1. Introductie

Flow cytometry is een techniek die in toenemende mate wordt gebruikt in biomedisch onderzoek. Deze techniek maakt het mogelijk om de fysische en chemische kenmerken van individuele cellen in een stroom van vloeistof te meten, wat tot inzichten in het functioneren van cellen kan leiden. Een van de grootste uitdagingen van flow cytometry is echter de analyse van de geproduceerde data, die vaak uit duizenden tot miljoenen cellen bestaat, elk met meerdere gemeten parameters. Het is een hoogdimensionale dataset, en traditionele analysetechnieken zijn vaak ontoereikend.

In deze context is het FlowSOM-algoritme¹ ontwikkeld. FlowSOM maakt gebruik van Self-Organizing Maps (SOM) en meta clustering om de data te visualiseren en interpretatie mogelijk te maken. Het heeft zich bewezen als een hulpmiddel voor flow cytometry data-analyse. Het blijft echter een uitdaging voor onderzoekers met een minder technische achtergrond om dit algoritme te gebruiken. Daarnaast gebeurt de computationele biologie steeds vaker in python. Het huidige flowsom pakket is echter geschreven voor de taal R.

Het doel van dit werk is om zowel een python pakket als grafische gebruikersinterface (GUI) voor het FlowSOM-algoritme te implementeren. Op deze manier kan de toegankelijkheid en de gebruikerservaring van dit algoritme aanzienlijk verbeteren. Daarnaast kan het de drempel verlagen voor onderzoekers om het FlowSOM-algoritme te gebruiken.

In dit verslag bespreken we onze aanpak voor het implementeren van de GUI, de technische uitdagingen die we hebben overwonnen, en de snelheid en efficiëntie die we hebben bereikt. Tot

slot bespreken we ook mogelijke uitbreidingen voor het pakket en de GUI.

2. Methoden

De methodologie van deze studie valt uiteen in twee componenten: de implementatie van het FlowSOM python pakket en de implementatie van de bijbehorende grafische gebruikersinterface (GUI).

2.1. FlowSOM Implementatie

De FlowSOM implementatie wordt beschikbaar gesteld in een klasse genaamd FlowSOM, een uitbreiding van de *BaseEstimator* en *ClusterMixin* klassen. Deze klasse bevat een tal aan functies voor het trainen, voorspellen en visualiseren van de data, alsook een functie om een verslag te genereren. Daarnaast houdt het object ook de relevante data zoals de parameters en de tussenresultaten van het trainen bij.

De hoofd functie van het FlowSOM object is de *fit* methode. Dit is de methode die een model zal trainen. Dit trainen bestaat uit drie delen: trainen van de som, bouwen van de minimum spanning, construeren van de metaclustering. Het resultaat van elk van deze delen wordt ook intern opgeslaan in het object en kan de gebruiker dus later nog verder gebruiken.

We raden aan het Flowsom object op volgende manier te gebruiken:

```
model = FlowSOM(  
    som_param,  
    mst_param,  
    hcc_param).fit(data)
```

We zien hierbij twee zaken: de initialisatie van het object en het oproepen van de *fit* methode. De argumenten van de initialisatie zijn dataklassen die de parameters van elk deel van het algoritme

bevatten. Er werd gekozen om deze in dataklassen te steken om de parameters overzichtelijk te structureren. De parameters van de fit functie zijn: een pandas dataframe die de cytometrie data bevat, een optie om logging aan te zetten met `verbose=True`.

De functionaliteit van de FlowSOM-klasse strekt zich verder uit tot data- en visualisatiefuncties waaronder functies om de minimum spanning tree met of zonder metacusters te plotten, een functie om het som grid met of zonder metaclusters te plotten, ... Voor een globaal overzicht van de functies en hun mogelijke parameters verwijzen we naar de meegeleverde documentatie.

Het FlowSOM-pakket omvat ook een uitgebreide rapport functie, deze is in staat is een verslag van de analyse te genereren in markdown syntax. Dit rapport bevat een samenvatting van de FlowSOM-analyse waaronder de model parameters, visualisaties en numerisch resultaten voor elke node en elke metacluster. Er werd gekozen dit in markdown te voorzien zodat de gebruiker kan kiezen in welke stijl hij het verslag wilt renderen.

De implementatie voor het FlowSOM pakket is terug te vinden onder:

1. `src/flowsom/flowsom.py`,
2. `src/flowsom/fs_dataclasses.py`,
3. `src/flowsom/fs_plotting.py`,
4. `src/flowsom/fs_reporting.py`

Hierbij zijn tal van functies geschreven in de `fs_...` bestanden. De functies in deze geschreven, zijn ofwel hulpfuncties, ofwel functies die in het flowsom object, dat zich in `flowsom.py` bevindt, omwikkeld worden. De zaken die in het pakket publiek gesteld worden zijn het FlowSOM object uit `flowsom.py` en de parameter klassen uit `fs_dataclasses.py`.

2.2. FlowSOMTool: De GUI

De tweede component van onze methodologie betreft de ontwikkeling van FlowSOMTool, een GUI voor FlowSOM. De GUI streeft ernaar de interactie met het FlowSOM-algoritme zo eenvoudig mogelijk te maken zodat mensen met minder technische achtergrond ook het algoritme kunnen gebruiken. Deze GUI is een prototype en is bedoeld als platform waar meer geavanceerdere interactie mogelijkheden kunnen op verder gebouwd worden.

FlowSOMTool biedt gebruikers een aantal aan mogelijkheden. Ten eerste geeft het de gebruiker de optie om FlowSOM-bestanden te selecteren op interactieve wijze. Daarnaast stelt het de gebruikers in staat om de parameters voor het algoritme aan te passen, het algoritme uit te voeren, en de resultaten te exporteren voor verdere analyse. Een belangrijk onderdeel van de GUI is het vermogen om de gegenereerde plots te visualiseren. Hierdoor krijgen gebruikers de mogelijkheid om de FlowSOM-analyse op een visueel aantrekkelijke manier te inspecteren.

Het programma is gebouwd aan de hand van de volgende technologieën:

1. ImGui
2. SDL2
3. OpenGL

Het bestaat uit initialisatie, programmalus en opkuis. Tijdens de initialisatie worden de contexten voor ImGui, SDL, OpenGL en Flowsom opgezet. Tijdens de programma lus worden telkens de contexten geupdate, de ImGui windows opgezet en tot slot alles weergegeven.

De implementatie voor de FlowSOMTool is terug te vinden onder:

1. `src/flowsom/flowsomtool.py`,
2. `src/flowsom/fst_dataclasses.py`,
3. `src/flowsom/fst_utils.py`,
4. `src/flowsom/fs_windows.py`

Het lanceren van FlowSOMTool, kan met het script in de `scripts/` folder genaamd `flowsom-tool.sh`.

3. Resultaten

In deze sectie bespreken we de prestaties van onze implementatie van het FlowSOM algoritme.

We doen eerst een tijdsmeting. We houden hierbij het aantal iteraties op 10 keer de grootte van de dataset. We bekomen dan het volgende resultaat:

4. Conclusie

5. Toekomstig werk

Hoewel de huidige versie van onze FlowSOM-tool al krachtige visualisatie- en analysefuncties biedt, zetten we ons in voor een voortdurende

evolutie van de tool, gericht op het verhogen van de bruikbaarheid en het uitbreiden van de functionaliteit. We identificeren een aantal mogelijke verbeteringen die voor toekomstige implementatie worden overwogen:

5.1. Highlighting en Tooltip-functie

Door een functie toe te voegen die knooppunten markeert wanneer een gebruiker met de muis over het knooppunt beweegt, wordt de interactie tussen de gebruiker en de visualisaties verder versterkt. Dit verbetert niet alleen de leesbaarheid, maar zorgt ook voor een uitgebreid begrip van de onderliggende gegevens. Het combineren van deze functie met een tooltip kan meer gedetailleerde informatie over het geselecteerde knooppunt weergeven, wat de gebruiker toegang geeft tot alle individuele dimensies van de gegevens.

5.2. Filtermogelijkheden

Een andere geplande verbetering is het geven van meer controle aan de gebruikers over welke knooppunten worden weergegeven. Dit kan worden bereikt door filterregelaars toe te voegen die gebaseerd zijn op verschillende criteria zoals clusterlidmaatschap, waardebereik, enz.

5.3. Kleurenkaart aanpassing

De mogelijkheid voor gebruikers om de kleurenkaart van de gegevens te wijzigen kan een waardevolle toevoeging zijn. Deze functionaliteit kan helpen om de knooppunten visueel te differentiëren op basis van verschillende dimensies of metriecken.

5.4. 3D-visualisatie

Voor bijzonder complexe, hoog-dimensionale gegevens kan de toevoeging van 3D-visualisatie, inclusief de mogelijkheid om de 3D-weergave te roteren, bijzonder waardevol zijn.

5.5. Details op aanvraag

We plannen om een optie toe te voegen waarbij gebruikers op een knooppunt kunnen klikken om een meer gedetailleerd beeld te krijgen in een apart paneel. Dit paneel kan alle dimensies en andere details van de gegevens bevatten.

5.6. Vergelijkingsmodus

Een functie die gebruikers in staat stelt om meerdere knooppunten of datapunten te selecteren en een vergelijking van hun gegevens weer te geven, kan bijzonder nuttig zijn bij het

identificeren van overeenkomsten of verschillen in hoog-dimensionale gegevens.

5.7. Zoekfunctie

Tot slot overwegen we de toevoeging van een zoekfunctie waarmee gebruikers specifieke knooppunten of datapunten kunnen zoeken op basis van verschillende parameters.

Deze voorgenomen verbeteringen weerspiegelen onze inzet om de FlowSOM-tool nog gebruiksvriendelijker en krachtiger te maken. We zijn enthousiast over de voortzetting van dit project en de potentiële waarde die deze toekomstige verbeteringen zullen bieden voor de gebruikersgemeenschap.

6. Conclusies

Onze FlowSOM-tool is een veelzijdig en krachtig hulpmiddel voor het visualiseren en analyseren van complexe, hoogdimensionale gegevens. Door gebruik te maken van zelforganiserende kaarten en minimum spannende bomen, biedt FlowSOM een effectieve methode voor clustering en visualisatie, waardoor gebruikers bruikbare inzichten kunnen halen uit hun gegevens.

Hoewel de huidige versie van de tool al aanzienlijke waarde biedt, erkennen we dat er altijd ruimte is voor verbetering en uitbreiding. De geplande verbeteringen die we hebben geïdentificeerd - zoals het toevoegen van functies voor markering, tooltip, filtering, aanpassing van kleurenkaarten, 3D-visualisatie, gedetailleerde weergave op aanvraag, tijdreeksanimatie, vergelijkingsmodus en zoeken - beloven de bruikbaarheid en functionaliteit van de tool aanzienlijk te verhogen.

References

1. Van Gassen, Sofie, Callebaut, Britt, Van Helden, Mary J., Lambrecht, Bart N., De-meester, Piet, Dhaene, Tom, and Saeys, Yvan, "FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data," *Cytometry Part A* **87**(7), pp. 636-645 (2015).