Subject: Re: First few research items
From: Yang Shen <yshen@tamu.edu>
To: Arghamitra Talukder <arghamitra.talukder@tamu.edu>
Date Sent: Wednesday, August 26, 2020 12:00:19 AM GMT-05:00
Date Received: Wednesday, August 26, 2020 12:00:48 AM GMT-05:00

Sorry about the delay.

Best,
Yang

On Tue, Aug 25, 2020 at 11:16 PM Arghamitra Talukder <arghamitra.talukder@tamu.edu>
wrote:
> Great, thank you Professor. I thought to remind you by tomorrow morning.
>
> Sincerely,
> Arghamitra Talukder
>
> On Tue, Aug 25, 2020, 11:14 PM Yang Shen <yshen@tamu.edu> wrote:
>> Hi Arghamitra,
>>
>> Here's the breakdowns of the initial efforts. Points 1-5 are for replication, i.e. for the
>> understanding of the problem, the state of the art, the data, and the assessment. Please feel
>> free to let me know if you have any questions.
>>
>>
>> Goal 1: Inter-protein residue-residue contact map (can extend to distance map)
>>
>> Read https://www.biorxiv.org/content/10.1101/2019.12.24.887877v1
>> 1--5 replicate 6- design new method
>> Objectives: 1. Understand the direct coupling analysis (DCA) approach and access the tool
>> plmDCA (a representative traditional DCA)
>> 2. Understand the rationale of filterDCA (why would it improve against DCA) and access the
>> tool
>> 3. Understand the concept of the competing deep learning approach and its limitations
>> 4. Access the dataset and understand the data (including multiple sequence alignment),
>> compare the performances between plmDCA and filterDCA, replicate some figures and
>> analysis of the paper
>> 5. Potential comparison with some deep learning approach such as
>> https://academic.oup.com/nar/article/46/W1/W432/5001161
>> ------
>> 6. Augment the dataset with available structure data for individual / unpaired proteins
>> (converted to the form of intra-protein contact or distance maps)
>> 7. Sanity check to establish the benefit of the additional data toward inter-protein contact
>> map prediction
>> 8. Design models to incorporate both sequence and structure information of individual
>> proteins for inter-protein contact prediction (to be expanded and implemented over 1-2
>> months)
>>
>>
>> Goal 2: 3D structure of the protein-protein complex
>> See a review https://www.sciencedirect.com/science/article/abs/pii/S0959440X18300691
>> We do not emphasize Goal 2 and we will demonstrate the power of our Goal 1
>> accomplishment here: in other words, start with individual proteins, dock them with existing

tools, using no, other's, or our predicted inter-protein contacts, and demonstrate better docking performances

Best,
Yang

Great, thank you Professor. I thought to remind you by tomorrow morning.

Sincerely,
Arghamitra Talukder

On Tue, Aug 25, 2020, 11:14 PM Yang Shen <yshen@tamu.edu> wrote:
> Hi Arghamitra,
>
> Here's the breakdowns of the initial efforts. Points 1-5 are for replication, i.e. for the understanding of the problem, the state of the art, the data, and the assessment. Please feel free to let me know if you have any questions.
>
>
> Goal 1: Inter-protein residue-residue contact map (can extend to distance map)
>
> Read https://www.biorxiv.org/content/10.1101/2019.12.24.887877v1
> 1--5 replicate 6- design new method
> Objectives: 1. Understand the direct coupling analysis (DCA) approach and access the tool plmDCA (a representative traditional DCA)
> 2. Understand the rationale of filterDCA (why would it improve against DCA) and access the tool
> 3. Understand the concept of the competing deep learning approach and its limitations
> 4. Access the dataset and understand the data (including multiple sequence alignment), compare the performances between plmDCA and filterDCA, replicate some figures and analysis of the paper
> 5. Potential comparison with some deep learning approach such as
> https://academic.oup.com/nar/article/46/W1/W432/5001161
> ------
> 6. Augment the dataset with available structure data for individual / unpaired proteins (converted to the form of intra-protein contact or distance maps)
> 7. Sanity check to establish the benefit of the additional data toward inter-protein contact map prediction
> 8. Design models to incorporate both sequence and structure information of individual proteins for inter-protein contact prediction (to be expanded and implemented over 1-2 months)
>
>
> Goal 2: 3D structure of the protein-protein complex
> See a review https://www.sciencedirect.com/science/article/abs/pii/S0959440X18300691
> We do not emphasize Goal 2 and we will demonstrate the power of our Goal 1 accomplishment here: in other words, start with individual proteins, dock them with existing tools, using no, other's, or our predicted inter-protein contacts, and demonstrate better docking performances
>
> Best,
> Yang

Hi Arghamitra,

Here's the breakdowns of the initial efforts. Points 1-5 are for replication, i.e. for the understanding of the problem, the state of the art, the data, and the assessment. Please feel free to let me know if you have any questions.

Goal 1: Inter-protein residue-residue contact map (can extend to distance map)

Read https://www.biorxiv.org/content/10.1101/2019.12.24.887877v1
1--5 replicate 6- design new method
Objectives: 1. Understand the direct coupling analysis (DCA) approach and access the tool plmDCA (a representative traditional DCA)
2. Understand the rationale of filterDCA (why would it improve against DCA) and access the tool
3. Understand the concept of the competing deep learning approach and its limitations
4. Access the dataset and understand the data (including multiple sequence alignment), compare the performances between plmDCA and filterDCA, replicate some figures and analysis of the paper
5. Potential comparison with some deep learning approach such as
https://academic.oup.com/nar/article/46/W1/W432/5001161
------
6. Augment the dataset with available structure data for individual / unpaired proteins (converted to the form of intra-protein contact or distance maps)
7. Sanity check to establish the benefit of the additional data toward inter-protein contact map prediction
8. Design models to incorporate both sequence and structure information of individual proteins for inter-protein contact prediction (to be expanded and implemented over 1-2 months)

Goal 2: 3D structure of the protein-protein complex
See a review https://www.sciencedirect.com/science/article/abs/pii/S0959440X18300691
We do not emphasize Goal 2 and we will demonstrate the power of our Goal 1 accomplishment here: in other words, start with individual proteins, dock them with existing tools, using no, other's, or our predicted inter-protein contacts, and demonstrate better docking performances

Best,
Yang