Universität Rostock

Traditio et Innovatio

visual & analytic computing

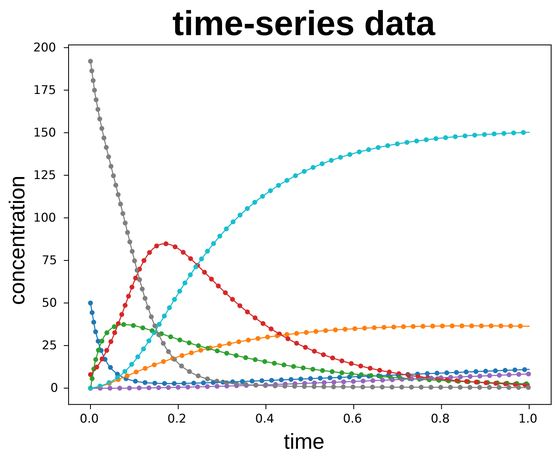
Combining Natural Language and Time Series to Infer Reaction Networks

The inference of *chemical reaction networks (CRN)* is concerned with determining lists of reactions whose dynamics follow a specification, e.g., data or constraints. Combining our previous work, we here consider inferring CRN using two kinds of specifications, time-series data and (partially) known reactions expressed in natural language.

CRN Inference from Time-series Data

Paper

We use a combination of the *Sparse Identification of Non-linear Dynamics (SINDy)* and an *evolutionary algorithm* to infer a CRN capable of reproducing given time-series data. The algorithm also accepts additional knowledge, such as known reactions or certain constraints, to further steer the inference process.



Inference

purge

Knowledge extraction with LLM

Paper

We use a fine-tuned *Large Language Model (LLM)* to extract formal CRN specifications from natural language expert input. Unspecified information is substituted with new variables, which are determined during inference (a). Completely known reactions are used to relax the problem by "subtracting" them from the time series (b).

User
E and S bind. ES produces EP at rate 3.0. S can be infected by I.

Assistant

b

$E + S \xrightarrow{k_0} ES$

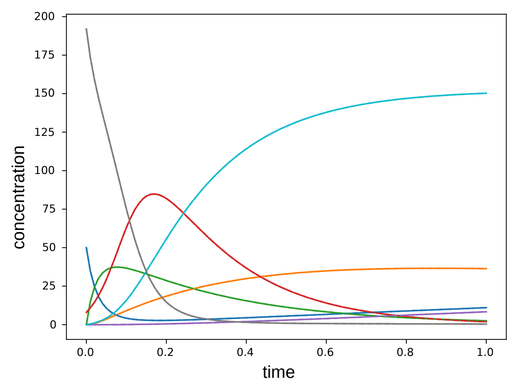
$ES \xrightarrow{3.0} EP$

a $S + I \xrightarrow{k_1} 2I$

Results

First results indicate the helpfulness of the supplied knowledge in exactly recovering the ground truth (center). The inclusion of (partially) known reactions constrains the search space in general, excluding unwanted reactions.

with knowledge



$S + I \xrightarrow{0.2} 2I$

$I \xrightarrow{5.0} R$

$E + S \xrightarrow{0.2} ES$

$ES \xrightarrow{0.2} E + S$

$ES \xrightarrow{3.0} EP$

$EP \xrightarrow{0.3} E + P$

$S + I \xrightarrow{0.2} 2I$

$I \xrightarrow{5.0} R$

$E + S \xrightarrow{0.2} ES$

$ES \xrightarrow{0.2} E + S$

$ES \xrightarrow{3.0} EP$

$EP \xrightarrow{0.3} E + P$

time series only

$S + I \xrightarrow{0.2} 2I$

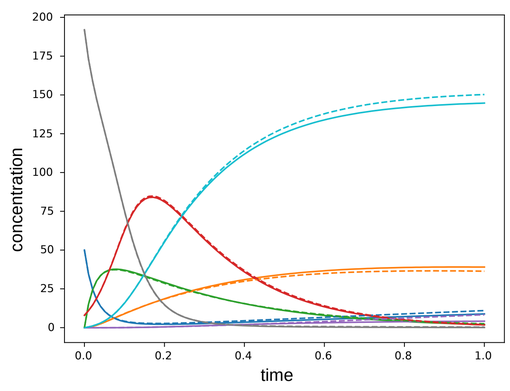
$I \xrightarrow{5.05} R$

$E + S \xrightarrow{0.2} ES$

$ES + R \xrightarrow{0.004} E + P$

$ES \xrightarrow{2.91} EP$

$EP + P \xrightarrow{0.05} E + P$



prior knowledge/ground truth grayed out

Conclusions

→ We demonstrated an **integration of LLM with CRN inference** methods
→ LLM can be used to **obtain prior knowledge** about reaction networks from natural language
→ Prior knowledge can effectively **steer and constrain the inference** process

Future Work:

- determine confidence scores for reactions or complete translation
- close the loop between inference and user input (chatbot): iterate on results using natural language
- retrieve knowledge directly from publications (LLM as reading engine)

Justin N. Kreikemeyer, Adelinde M. Uhrmacher

Modeling and Simulation Group | Institute for Visual and Analytic Computing | University of Rostock

This poster is funded by DFG Grant No. 320435134