Unsupervised Attention-Guided Atom-Mapping

Philippe Schwaller
IBM Research Europe | Reymond group - University of Bern

$u^b$

IBM Research Europe
Exploring the nearly endless chemical space

$10^{60}$ drug-like molecules

Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17

Lars Ruddigkeit, Ruud van Deursen, Lorenz C. Blum, and Jean-Louis Reymond
What molecule to make?
Generative models
Property predictions

How to make it?
Reaction prediction
Synthesis planning

Experimental validation
Chemical reactions data

US Patents

Text-mining
(Lowe 2012/17)

Millions of reactions

BrC(Br)(Br)Br.CC...>>...
CO.Nc1ccc([N+]...>>...
CC(=O)O[BH-]...>>...
(OC(C)=O)OC(C)=O...>>...
...
precursors>>products

Benchmark sets

USPTO_MIT
USPTO_STEREO

Reaction SMILES

CC(C)S.CN(C)C=O.Fc1cccnc1F.O=C([O-])[O-].[K+].[K+]>>CC(C)Sc1ncccc1F

precursors                          >>      products

\[ \text{precursors} \rightarrow \text{products} \]
While typically correct, the atom-maps are wrong in many cases and hence should not be entirely relied on.

https://figshare.com/articles/Chemical_reactions_from_US_patents_1976-Sep2016_/5104873
Chemical reactions data

**US Patents**

Text-mining
(Lowe 2012/17)

**Millions of reactions**

**Benchmark sets**

USPTO_MIT
USPTO_STEREO

**Reaction SMILES**

CC(C)S.CN(C)C=O.Fc1cccnc1F.O=C([O-])[O-].[K+].[K+]>>CC(C)Sc1nccccc1F

precursors >> products
Atoms as letters, molecules as words

Chemical reactions can be represented as text.

\[
\text{HS} + \text{N} \rightarrow \text{F} + \text{O} \rightarrow \text{K}^+ + \text{K}^+ \rightarrow \text{CC(C)S}.
\]

SMILES: \( \text{CC(C)S.CN(C)C=O.Fc1cccnc1F.O=C([O-])[O-].[K+].[K+]>>CC(C)Sc1nccccc1F} \)

Tokens: \( \text{CC(C)S.CN(C)C=O.Fc1cccnc1F.O=C([O-])[O-].[K+].[K+]>>CC(C)Sc1nccccc1F} \)

Natural Language Processing methods
- e.g. Sequence-2-sequence transformer

Attention is all you need
A Vaswani, N Shazeer, N Parmar... - Advances in neural ... 2017 - papers.nips.cc
The dominant sequence transduction models are based on complex recurrent or convolutional neural networks in an encoder and decoder configuration. The best performing such models also connect the encoder and decoder through an attention mechanism...
Chemical reaction prediction

Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction

- Reaching >90% on common reaction prediction benchmark
- Presented in 2018, still state-of-the-art approach

Carbohydrate Transformer: Predicting Regio- and Stereoselective Reactions Using Transfer Learning

- Tackling more challenging reactions with transfer learning
- Experimental validation (14-step synthesis)
Multi-step retrosynthesis planning

Molecular Transformer **Retro** (product>precursors)
- Precursors (reactants & reagents) suggestions

Molecular Transformer **Forward** (precursor>product)
- Path scoring

Predicting retrosynthetic pathways using a combined linguistic model and hyper-graph exploration strategy

Philippe Schwaller, Riccardo Petraglia, Valerio Zullo, Vishnu H Nair, Rico Andreas Haeuselmann, Riccardo Pisoni, Costas Bekas, Anna Iuliano, Teodoro Laino

Chem. Sci., 2020, Advance Article
https://doi.org/10.1039/C9SC05704H
Retrosynthesis feature on https://rxn.res.ibm.com

Not only reactants but also reagents (catalysts, solvents)

→ Also available in interactive mode
Approaches to chemical reaction tasks

Template-based approaches

Planning chemical syntheses with deep neural networks and symbolic AI
Marwin H. S. Segler1,2, Mike Preuss2 & Mark P. Waller3

Prediction of Organic Reaction Outcomes Using Machine Learning
Connor W. Coley, Regina Barzilay, Tommi S. Jaakkola, William H. Green and Klavs F. Jensen

Graph-NN based approaches (bond changes / graph edits predictions)

A graph-convolutional neural network model for the prediction of chemical reactivity
Connor W. Coley, Wengong Jin, Luke Rogers, Timothy F. Jamison, Tommi S. Jaakkola, William H. Green, Regina Barzilay and Klavs F. Jensen

A Generative Model For Electron Paths
John Bradshaw, Matt J. Kusner, Brooks Paige, Marwin H. S. Segler, José Miguel Hernández-Lobato

SMILES-2-SMILES approaches

Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction
Philippe Schwaller*, Teodoro Laino, Théophile Gaudin, Peter Bolgar, Christopher A. Hunter, Costas Bekas, and Alpha A. Lee*

Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy
Philippe Schwaller, Riccardo Petraglia*, Valerio Zullo, Vishnu H. Nair*, Rico Andreas Haeuselmann*, Riccardo Pisoni*, Costas Bekas, Anna Iuliano* and Teodoro Laino Base
Encoder Transformers (BERT)

Revolutionised
Natural Language Processing

Bert: Pre-training of deep bidirectional transformers for language understanding
J Devlin, M W Chang, K Lee, K Toutanova - arXiv preprint arXiv:1810.04805, 2018 - arxiv.org

We introduce a new language representation model called BERT, which stands for Bidirectional Encoder Representations from Transformers. Unlike recent language representation models, BERT is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and …

🌟 ⚾️ Cited by 5292  Related articles  All 20 versions  🔗

https://huggingface.co/transformers/

State-of-the-art Natural Language Processing for PyTorch and TensorFlow 2.0

Yannic Kilcher
35.6K subscribers

https://www.youtube.com/c/yannickilcher

Attention Is All You Need
Yannic Kilcher • 150K views • 2 years ago

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
Yannic Kilcher • 33K views • 1 year ago
How does a reaction BERT look like?
Training

• First stage: **Pretraining**
  **Unsupervised**
  Task with unlimited training data

• Second stage: **Fine-tuning**
  **Supervised**
  • BERT + Classification head
    Reaction classification
  • BERT + Regression head
    Reaction yield prediction
Self-supervised training

Masked Language Modelling
- Predict mask given context
- Unlimited training data
Fine-tuning task - Reaction Classification

Bromo Suzuki-Coupling reaction

**DATA:** Pistachio (NextMove Software)

Table 1: Classification results

| Model                      | Test Accuracy [%] |
|----------------------------|-------------------|
| FastText (autotuned)       | 77.5              |
| MHFP-based $k$-NN classifier | 86.1              |
| transformer enc2-dec1      | 95.2              |
| BERT classifier            | 98.2              |

Mapping the Space of Chemical Reactions using Attention-Based Neural Networks

**Version 3**  Preprint revised on 06.08.2020, 18:28 and posted on 07.08.2020, 09:36 by Philippe Schwaller, Daniel Probst, Alain C. Vaucher, Vishnu H Nair, David Kreutter, Teodoro Laino, Jean-Louis Reymond

COC(=O)CCC(=O)c1ccc(Br)cc1.Cc1ccccccc1.O=C([O][Na])[O][Na].OB(O)c1ccc(Cl)c(Cl)c1.c1ccc([P]c2ccccccc2)[c2cccc cc2][Pd][P][c2ccccccc2]c2ccccccc2)[P][c2ccccccc2]c2ccccccc2)[P][c2ccccccc2]c2ccccccc2)[P][c2ccccccc2]c2ccccccc 2)c1>>COC(=O)CCC(=O)c1ccc(-c2ccc(Cl)c(Cl)c2)ccc1
Visualization of Very Large High-Dimensional Data Sets as Minimum Spanning Trees

Daniel Probst, Jean-Louis Reymond

(Submitted on 16 Aug 2019 v1, last revised 6 Jan 2020 (this version, v3))
Enhanced chemical reaction space exploration

Reaction Atlas tutorial and more on https://github.com/rxn4chemistry/rxnfp
### RXNFP Application: Reaction search

| Query: Bromo Suzuki coupling - 3.1.1 | Query: Nitration - 10.2.1 |
|--------------------------------------|--------------------------|
| ![Reaction SMILES](image1)           | ![Reaction SMILES](image2) |
| Nearest Neighbors (all class 3.1.1): | Nearest Neighbors (all class 10.2.1): |
| ![Reaction SMILES](image3)           | ![Reaction SMILES](image4) |
| ![Reaction SMILES](image5)           | ![Reaction SMILES](image6) |
| ![Reaction SMILES](image7)           | ![Reaction SMILES](image8) |

- No reaction center or atom-mapping! -> just a reaction SMILES
- Link back to similar training reactions
Reaction yield prediction (Regression task)

Brc1cccn1.CCOC(=O)c1cc(C)no1.CN1CCCN2CCCN=C12.COc1ccc(OC)c(P(C)(C)C)C(C)(C)Cc1-c1c(C(C)C)cc(C(C)C)c1C(C)Cc1ccc(N)cc1.0=S(=O)O[O[pd]1c2ccccc2-c2ccccc2[NH2]1]C(F)(F)F>>Cc1ccc(Nc2ccccc2)cc1

Reaction SMILES

|           | R²   | DFT [9] | one-hot [11] | MFF [11] | Yield-BERT |
|-----------|------|---------|--------------|----------|------------|
| random    | 0.92 | 0.89    | 0.927 ± 0.007| 0.944 ± 0.009 |
| test 1    | 0.8  | 0.69    | 0.85         | 0.84     |
| test 2    | 0.77 | 0.67    | 0.71         | 0.90     |
| test 3    | 0.64 | 0.49    | 0.64         | 0.70     |
| test 4    | 0.54 | 0.49    | 0.18         | 0.35     |
| test avg. 1-4 | 0.69 | 0.59    | 0.60         | 0.70     |

• Transformers have better than DFT computed chemical descriptors + RF

[9] Ahneman, D. T., Estrada, J. G., Lin, S., Dreher, S. D. & Doyle, A. G. Predicting reactivity performance in C–N cross-coupling using machine learning. _Science_ **360**, 186–190 (2018).

[11] Sandfort, F., Strieth-Kalthoff, F., Kühnemund, M., Beecks, C. & Glorius, F. A structure-based platform for predicting chemical reactivity. _Chem_ (2020).
But why do Transformers work so well?

Transformer
Self-attention layers

Training set

Contextual Representations

Self-attention layer:
Multiple heads = functions to attend input tokens

Work with Ben Hoover and Hendrik Strobelt

transmized reaction SMILES

BrC(Br)(Br)Br.CC...>>...
CO.Nc1ccc([N+])...>>...
...
(OC(C)=O)OC(C)=O...>>...
precursors>>products

millions of unlabeled reaction SMILES
Visual Analysis of Attention Weights

Discovery: Atom-mapping

trained model

head functionality
Every (AL)BERT model learns atom-mapping
RXNMapper algorithm

Get attention weights for atom-mapping head/layer from model

**Iterate** through product atoms:
- **Map** the most *likely pair*
- **Increase attention** from product neighbour corresponding reactant neighbours

- Pure attention  → ~96% correct maps
- Neighbour multiplier  → >99% correct maps
Focus on mis-matches

99.4% atom-mappings correct

| Category                        | Percentage |
|--------------------------------|------------|
| Matching mapping               | 96.8%      |
| RXNMapper better               | 0.89%      |
| Equivalent mapping             | 0.85%      |
| Questionable reaction          | 0.75%      |
| Unclear / missing reagents     | 0.10%      |
| Data set mapping better        | 0.58%      |

Reactions of 49k test set
RXNMapper

Task

Forward prediction

Chemical reactions

Retrosynthetic prediction

Common approach

Common approach

Available of training data
• Better mapping, better performance

How it is affected

• Interpretability
• Reactant-reagent role assignment
• Reaction classification

• SMILES-based: generate target molecules

• Graph-based: predict bond changes

• Template-based: predict rules to apply

Link template-based/graph-based and SMILES-based approaches
Grammar of chemical reactions

reaction center, bond changes

Synthesis planning
Quantum Mechanical simulations
Reaction accessibility and interpretability

7sec/1k imbalanced reactions
Compared to other tools

Advantage compared to Mappet is speed
- some reactions took >100 seconds to map

Automatic mapping of atoms across both simple and complex chemical reactions

Wojciech Jaworski, Sara Szymkuć, Barbara Mikulak-Klucznik, Krzysztof Piecuch, Tomasz Klucznik, Michał Kaźmierowski, Jan Rydzewski, Anna Gambin & Bartosz A. Grzybowski
How would you map the following reactions?

Epoxidation

Semi-pinacol Rearrangement
How would you map the following reactions?

Epoxidation

Semi-pinacol Rearrangement
Unsupervised Attention-Guided Atom-Mapping

• **Unsupervised** Transformers capture the hidden grammar of chemical reactions.

• Reactions follow **consistent rules**

• **Atom rearrangements** can be extracted from model --> RXNMapper

• **Fast, high-quality** even on strongly imbalanced reactions

• Demo on [http://rxnmapper.ai](http://rxnmapper.ai)
Open-source code

https://github.com/rxn4chemistry/rxnfp | pip install rxnfp

```python
from rxnfp.transformer_fingerprints import (  
    RXNBERTFingerprintGenerator, get_default_model_and_tokenizer, generate_fingerprints
)

model, tokenizer = get_default_model_and_tokenizer()

rxnfp_generator = RXNBERTFingerprintGenerator(model, tokenizer)

example_rxn = "Nc1ccccc2cnccc12.0=C(0)c1cc([N+](=O)[O-])c(Sc2c(Cl)cncc2Cl)s1>>O=C(Nc1ccccc2cnccc12.0)"

fp = rxnfp_generator.convert(example_rxn)
print(len(fp))
print(fp[:5])
```

```
256
[-2.0174953937530518, 1.7602033615112305, -1.3323537111282349, -1.1095019578933716, 1.225454
```
Open-source code

https://github.com/rxn4chemistry/rxnmapper | pip install rxnmapper

```python
from rxnmapper import RXNMapper
rxn_mapper = RXNMapper()
rxns = ['CC(C)S-CN(C)C=0.Fc1cccnc1F.0=C([0-])[O-].[K+].[K+]>>CC(C)S-c1ncccc1F', 'C1C0C0C
results = rxn_mapper.get_attention_guided_atom_maps(rxns)

[{'mapped_rxn': 'CN(C)C=0.F[c:5][n:6][cH:7][cH:8][cH:9][c:10][F:11].0=C([0-])[O-].[CH3:1][CH
'confidence': 0.9565619900376546},
{'mapped_rxn': 'C1C0C0C01.CC(C)(C)[0:3][C:2](=[O:1])[CH2:4][0:5][NH:6][C:7](=[O:8])[NH:9][CH2:
'confidence': 0.9704424331552834}]
```
Conclusion

• Reaction Transformers show excellent performance
  • Yield prediction
  • Reaction classification

• We developed data-driven reaction fingerprints
  • Nearest-Neighbor search

• Unsupervised transformers capture the grammar of chemical reactions
  → Provide missing link between rule-based and data-driven approaches

• Makes ML predictions more interpretable and improves all downstream applications
What’s next?

Combination of ML models, cloud and lab automation
https://rxn.res.ibm.com

#RoboRXN

Retrosynthetic analysis

Automatic conversion to recipe and actions

Synthesis robot performs actions

Analysis with LCMS
Teodoro Laino, Alain Vaucher, Alessandra Toniato, Vishnu H Nair, Joppe Geluykens, Antonio Cardinale, Heiko Wolf, Theophile Gaudin, Costas Bekas, Ben Hoover, Hendrik Strobelt

Jean-Louis Reymond, Daniel Probst, Giorgio Pesciullesi and the Reymond group

Alpha Lee, Peter Bolgar and Ryan-Rhys Griffiths
Many thanks for your attention!

@pschwllr

https://github.com/pschwllr

phs@zurich.ibm.com