Predicting with confidence using conformal prediction

Ulf Norinder
ulf.norinder@swetox.se

Utilizing Information on Uncertainty for In Silico Modeling using Random Forests

Henrik Boström
Dept. of Computer and Systems Sciences
Stockholm University
and
Informatics Research Centre
University of Skövde

Introducing Uncertainty in Predictive Modeling—Friend or Foe?
Ulf Norinder and Henrik Boström

Representing descriptors derived from multiple conformations as uncertain features for machine learning
Ulf Norinder & Henrik Boström
Question of (un)certainty

Utilizing Information on Uncertainty for In Silico Modeling using Random Forests

Henrik Boström
Dept. of Computer and Systems Sciences
Stockholm University
and
Informatics Research Centre
University of Skövde

Ulf Norinder
AstraZeneca R&D Södertälje

Conformal Prediction for Distribution-Independent Anomaly Detection in Streaming Vessel Data

Rikard Laxhammar
University of Skövde
Skövde, Sweden

&
Security and Defense Solutions, Saab AB
Järfalla, Sweden

Göran Falkman
University of Skövde
Skövde, Sweden
goran.falkman@his.se

Introducing Uncertainty in Predictive Modeling—Friend or Foe?

Ulf Norinder*†,‡,§ and Henrik Boström§


Representing descriptors derived from multiple conformations as uncertain features for machine learning

Ulf Norinder • Henrik Boström

Question of (un)certainty

Utilizing Information on Uncertainty for In Silico Modeling using Random Forests

Henrik Boström
Dept. of Computer and Systems Sciences
Stockholm University
and
Informatics Research Centre
University of Skövde

Ulf Norinder
AstraZeneca R&D Södertälje

Proceedings: Stream KDD '10 First International Workshop on Novel Data Stream Pattern Mining Techniques, Washington, D.C. 2010
Conformal Prediction for Distribution-Independent Anomaly Detection in Streaming Vessel Data

Rikard Laxhammar
University of Skövde
Skövde, Sweden
Security and Defense Solutions, Saab AB
Järfälla, Sweden
rikard.laxhammar@his.se

Göran Falkman
University of Skövde
Skövde, Sweden
goran.falkman@his.se

Introducing Uncertainty in Predictive Modeling—Friend or Foe?
Ulf Norinder* and Henrik Boström§

Conformal Prediction
What is it good for ...?

Representing descriptors derived from multiple conformations as uncertain features for machine learning
Ulf Norinder & Henrik Boström
Implementing CP @Lundbeck

- Building ADME(T) & project models
- Learning & understanding CP

Lars Carlsson, AstraZeneca

Martin Eklund, post doc, AZ
Conformal Prediction

Why Conformal Prediction?

• Win situation
• Statistical guarantees (on validity)
Conformal Prediction

If \{\text{Exchangeability}\} then \{\text{conformal predictors are always valid}\}

Mathematical proof

If 20\% \text{ prediction errors on validity acceptable} \implies CP will give, \text{at most}, 20\% \text{ errors}!!
Conformal Prediction

If {Exchangeability} then {conformal predictors are always valid}
Conformal Prediction

If {Exchangeability} then {conformal predictors are always valid}

- Training and test data have the **same distributions**
  the relationship between input and output in training and test data is **the same**.
- Training and test data have the **same distributions**
  the relationship between input and output in training and test data is **different**.
- Training and test data have the **different distributions**
  the relationship between input and output in training and test data is **the same**.
- Training and test data have the **different distributions**
  the relationship between input and output in training and test data is **different**.
Conformal Prediction

Why Conformal Prediction?

- Win situation
- Statistical guarantees (on validity)
- CP is instance-based
- The risk is known up-front for the decision taken
- Applicability domain closely linked to model development
  
  CP strictly defines the level of similarity (conformity) needed
  
  No ambiguity anymore

- Gracefully handles (severely) imbalanced datasets
  
  Ratios of 1:100 – 1:1000
  
  No need for over- or undersampling

- CP is a framework (almost any ML algorithm will work)
Conformal Prediction

• (non-) similarity function $\rightarrow$ (non-) conformity function in CP
• Compares new compounds to old (calibration) compounds
  o Defined by the user
  o Probability from the RF trees
  o Distance to decision plane in SVM
  o (Random numbers)
Conformal Prediction

- (non-) similarity function $\rightarrow$ (non-) conformity function in CP
- Compares new compounds to old (calibration) compounds
  - Defined by the user
  - Probability from the RF trees
  - Distance to decision plane in SVM

**Ranking problem**

CP $p$-value

\[
\left| \{i = 1, \ldots, n : \beta_i \leq \beta_{\text{new}} \} \right| / (n+1) \geq \varepsilon
\]

- $\beta_i = \text{probability for the calibration compound } i$
- $\beta_{\text{new}} = \text{probability for the new test compound}$
- $n = \text{number of calibration set compounds}$
- $\varepsilon = \text{significance level (% acceptable errors)}$

The number of calibration set compounds with probabilities $\leq$ probability for the new compound divided by $(n+1)$ must be $\geq \varepsilon$ to be assigned a class label
Conformal Prediction

How does this work?

Data

Train set

Proper Train set

Model

Calibration set

Calibration set predictions

CP $p$-values (for each class)

classification

Test set


Courtesy Dr. Fredrik Svensson
New compound to predict (is toxic)

Imbalanced dataset (toxic minority class)

A binary RF classifier (100 trees) gives the output:

32 trees: toxic
68 trees: non-toxic
Conformal Prediction

Example: Predicting Toxicity

Calibration set, 7 toxic, 7 non-toxic compounds
N trees predicting correct class

<table>
<thead>
<tr>
<th>Toxic</th>
<th>Non-Toxic</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>91</td>
</tr>
<tr>
<td>42</td>
<td>88</td>
</tr>
<tr>
<td>36</td>
<td>85</td>
</tr>
<tr>
<td>30</td>
<td>82</td>
</tr>
<tr>
<td>28</td>
<td>79</td>
</tr>
<tr>
<td>27</td>
<td>79</td>
</tr>
<tr>
<td>26</td>
<td>77</td>
</tr>
</tbody>
</table>

New compound to predict (is toxic)

32 trees: toxic
68 trees: non-toxic
Conformal Prediction

Example: Predicting Toxicity

Calibration set, 7 toxic, 7 non-toxic compounds
N trees predicting correct class

New compound to predict (is toxic)

<table>
<thead>
<tr>
<th>Toxic</th>
<th>Non-Toxic</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>91</td>
</tr>
<tr>
<td>42</td>
<td>88</td>
</tr>
<tr>
<td>36</td>
<td>85</td>
</tr>
<tr>
<td>30</td>
<td>82</td>
</tr>
<tr>
<td>28</td>
<td>79</td>
</tr>
<tr>
<td>27</td>
<td>79</td>
</tr>
<tr>
<td>26</td>
<td>77</td>
</tr>
</tbody>
</table>

32 trees: toxic
68 trees: non-toxic
Conformal Prediction

Example: Predicting Toxicity

Based on the similarity to the known examples in the calibration set:

Position toxic: 4/8
Position non-toxic: 0/8

<table>
<thead>
<tr>
<th>Toxic</th>
<th>Non-Toxic</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>91</td>
</tr>
<tr>
<td>42</td>
<td>88</td>
</tr>
<tr>
<td>36</td>
<td>85</td>
</tr>
<tr>
<td>30</td>
<td>82</td>
</tr>
<tr>
<td>28</td>
<td>79</td>
</tr>
<tr>
<td>27</td>
<td>79</td>
</tr>
<tr>
<td>26</td>
<td>77</td>
</tr>
</tbody>
</table>

New compound to predict (is toxic)

32 trees: toxic
68 trees: non-toxic

Courtesy Dr. Fredrik Svensson
Conformal Prediction

Example: Predicting Toxicity

Using 80% confidence level (0.2 significance level):

4/8 = 0.5 > 0.2 therefore the compound is assigned to the toxic class

0/8 = 0.0 < 0.2 therefore the compound is not assigned to the non-toxic class

New compound to predict

32 trees: toxic
68 trees: non-toxic

Using 80% confidence level (0.2 significance level):

4/8 = 0.5 > 0.2 therefore the compound is assigned to the toxic class

0/8 = 0.0 < 0.2 therefore the compound is not assigned to the non-toxic class
Mondrian Cross-Conformal Prediction

Mondrian = for computing p-values →
study each class separately

Class 1
\[ \frac{|\{i = 1, \ldots, n : \beta_i \leq \beta_{\text{new}}\}|}{n+1} \geq \varepsilon \]

Class 2
\[ \frac{|\{i = 1, \ldots, n : \beta_i \leq \beta_{\text{new}}\}|}{n+1} \geq \varepsilon \]
Binary Mondrian Conformal Prediction p-values

- Build model using the proper training set
- Use calibration set for calculating p-values

For each class $\left| \{i = 1, \ldots, n : \beta_i \leq \beta_{new} \} \right| / (n+1) \geq \varepsilon$

Sorted decreasing probabilities for the calibration set

Significance level $\varepsilon = 0.20$

Class 1
- New cmpd 1
  - Prob class 1: 0.75
  - Prob class 2: 0.25
- Class 2
  - New cmpd 2
    - Prob class 1: 0.30
    - Prob class 2: 0.70
- New cmpd 3
  - Prob class 1: 0.39
  - Prob class 2: 0.61
- Class both
  - New cmpd 4
    - Prob class 1: 0.37
    - Prob class 2: 0.59

Calibr set class 1

<table>
<thead>
<tr>
<th>8/10 = 0.80</th>
<th>6/10 = 0.60</th>
<th>4/10 = 0.40</th>
<th>2/10 = 0.20</th>
<th>1/10 = 0.10</th>
<th>0/10 = 0.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.82</td>
<td>0.64</td>
<td>0.61</td>
<td>0.57</td>
<td>0.53</td>
<td>0.52</td>
</tr>
<tr>
<td>0.64</td>
<td>0.61</td>
<td>0.57</td>
<td>0.53</td>
<td>0.52</td>
<td>0.49</td>
</tr>
<tr>
<td>0.61</td>
<td>0.57</td>
<td>0.53</td>
<td>0.52</td>
<td>0.49</td>
<td>0.49</td>
</tr>
<tr>
<td>0.57</td>
<td>0.53</td>
<td>0.52</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
</tr>
<tr>
<td>0.53</td>
<td>0.52</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
</tr>
<tr>
<td>0.52</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
</tr>
</tbody>
</table>

Calibr set class 2

<table>
<thead>
<tr>
<th>5/9 ~ 0.56</th>
<th>2/9 ~ 0.22</th>
<th>1/9 ~ 0.11</th>
<th>0/9 ~ 0.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>0.81</td>
<td>0.73</td>
<td>0.69</td>
</tr>
<tr>
<td>0.81</td>
<td>0.73</td>
<td>0.69</td>
<td>0.66</td>
</tr>
<tr>
<td>0.73</td>
<td>0.69</td>
<td>0.66</td>
<td>0.63</td>
</tr>
<tr>
<td>0.69</td>
<td>0.66</td>
<td>0.63</td>
<td>0.60</td>
</tr>
<tr>
<td>0.66</td>
<td>0.63</td>
<td>0.60</td>
<td>0.43</td>
</tr>
</tbody>
</table>
Binary Mondrian Conformal Prediction p-values

In conformal prediction:
If a classification contains the correct class it is correct

*both* = always correct, *empty* = always erroneous

Validity = % of correct classifications (for each class)
Efficiency = % of single label classifications (right or wrong)
Conformal Prediction work at Swetox

PubChem Cytotox Assays

- Results from 16 high throughput cell viability (tox) screens from PubChem
- On average 0.8% toxic compounds

<table>
<thead>
<tr>
<th>AID</th>
<th>Tested compounds</th>
<th>Toxic compounds</th>
<th>%active</th>
<th>ratio non-tox/tox</th>
</tr>
</thead>
<tbody>
<tr>
<td>624418</td>
<td>386 360</td>
<td>524</td>
<td>0.14</td>
<td>736.3</td>
</tr>
<tr>
<td>504648</td>
<td>367 995</td>
<td>600</td>
<td>0.16</td>
<td>612.3</td>
</tr>
<tr>
<td>602141</td>
<td>359 040</td>
<td>1302</td>
<td>0.36</td>
<td>274.8</td>
</tr>
<tr>
<td>620</td>
<td>86 701</td>
<td>364</td>
<td>0.42</td>
<td>237.2</td>
</tr>
<tr>
<td>847</td>
<td>41 152</td>
<td>194</td>
<td>0.47</td>
<td>211.1</td>
</tr>
<tr>
<td>903</td>
<td>52 783</td>
<td>338</td>
<td>0.64</td>
<td>155.2</td>
</tr>
<tr>
<td>2275</td>
<td>29 938</td>
<td>193</td>
<td>0.64</td>
<td>154.1</td>
</tr>
<tr>
<td>588856</td>
<td>404 016</td>
<td>3018</td>
<td>0.75</td>
<td>132.9</td>
</tr>
<tr>
<td>1825</td>
<td>290 605</td>
<td>2259</td>
<td>0.78</td>
<td>127.6</td>
</tr>
<tr>
<td>2717</td>
<td>299 957</td>
<td>3181</td>
<td>1.06</td>
<td>93.3</td>
</tr>
<tr>
<td>648</td>
<td>86 121</td>
<td>924</td>
<td>1.07</td>
<td>92.2</td>
</tr>
<tr>
<td>719</td>
<td>84 841</td>
<td>937</td>
<td>1.10</td>
<td>89.5</td>
</tr>
<tr>
<td>1486</td>
<td>217 851</td>
<td>2408</td>
<td>1.11</td>
<td>89.5</td>
</tr>
<tr>
<td>463</td>
<td>56 465</td>
<td>706</td>
<td>1.25</td>
<td>79.0</td>
</tr>
<tr>
<td>430</td>
<td>62 627</td>
<td>1121</td>
<td>1.79</td>
<td>54.9</td>
</tr>
<tr>
<td>598</td>
<td>85 162</td>
<td>5139</td>
<td>6.03</td>
<td>15.6</td>
</tr>
</tbody>
</table>
Conformal Prediction work at Swetox

PubChem Cytotox Assays

- Results from 16 high throughput cell viability (tox) screens from PubChem
- On average 0.8% toxic compounds
- RDKit descriptors
- RF, 500 trees, ensemble of 100 models
- 80% training set, 20% external test set
Validity of the predictions (test sets) at the 80% confidence level. Models are valid for both classes.
Accuracy of the single label predictions (test sets) at the 80% confidence level. The accuracy is similar for both the active and the inactive class.
PubChem & Hansen Datasets

- Four dataset of different sizes and class imbalances
- 10% randomly selected training sets
- Signature descriptors of heights 0–2 for chemical structure characterization
- Support vectors machines (SVM) C-SVC, RBF kernel, parameters C = 50, gamma = 0.002
- Ensemble of 100 SVM models

Binary classification of imbalanced datasets using conformal prediction

Ulf Norinder*, Scott Boyer

Size and imbalance differs considerably between the datasets.
Fraction predicted active and inactive compounds. Results are similar across the datasets despite the varying imbalance.
#compounds in both class & empty class

@acceptable significance level:

Results from new data ➔

- Many predictions in empty class ➔ outside AD of current model ➔ measure and update model
- Many predictions in both class ➔ inside AD of current model ➔ lack of information ➔
  add new information (descriptors), develop better model (classifier, algorithm)
Applying, validating and trying to understand Conformal Prediction 2015 - 2018

Introducing conformal prediction in predictive modeling for regulatory purposes. A transparent and flexible alternative to applicability domain determination
Ulf Norinder*, Lars Carlsson§, Scott Boyer*, Martin Eklund*§

Conformal prediction to define applicability domain – A case study on predicting ER and AR binding
U. NorinderAb, A. RybackaAb and P.L. AnderssonAb
SAR AND QSAR IN ENVIRONMENTAL RESEARCH, 2016
VOL. 27, NO. 4, 303–316

Conformal Prediction Classification of a Large Data Set of Environmental Chemicals from ToxCast and Tox21 Estrogen Receptor Assays
Ulf Norinder* and Scott Boyer

Modelling compound cytotoxicity using conformal prediction and PubChem HTS data†
Fredrik Svensson,‡ Ulf Norinder*§ and Andreas Bender*‡

Conformal Prediction in Spark: Large-Scale Machine Learning with Confidence
Marco Capucci², Lars Carlsson³, Ulf Norinder and Ola Spjuth³

Binary classification of imbalanced datasets using conformal prediction
Ulf Norinder*, Scott Boyer

Applying Mondrian Cross-Conformal Prediction To Estimate Prediction Confidence on Large Imbalanced Bioactivity Data Sets
Jiangming Sun, Lars Carlsson, Ernst Alldberg, Ulf Norinder, Ola Engkvist, and Hongming Chen

Improving Screening Efficiency through Iterative Screening Using Docking and Conformal Prediction
Fredrik Svensson, Ulf Norinder, and Andreas Bender

Maximizing gain in high-throughput screening using conformal prediction
Fredrik Svensson, A. M. Aftai, Ulf Norinder and Andreas Bender
Journal of Cheminformatics 2018, 10, 7

Computational Toxicology
Volume 6, May 2018, Pages 9-15
Creating an efficient screening model for TRPV1 agonists using conformal prediction
Ulf Norinder*, Scott Boyer, Daniel Muca*, Theodor Pippig*, Anna Forsby*
Conformal Prediction

Why Conformal Prediction?

- Win situation
- Statistical guarantees (on validity)
- CP is instance-based
- The risk is known up-front for the decision taken (signif.level)
- Applicability domain closely linked to model development
  - CP strictly defines the level of similarity (conformity) needed
  - No ambiguity anymore
- Gracefully handles (severely) imbalanced datasets
  - Ratios of 1:100 – 1:1000
  - No need for over- or undersampling
- CP is a framework (almost any ML algorithm will work)
Questions & Comments