Mixture of metrics optimization for machine learning problems

Magdalena Wiercioch and Marek Śmieja
Faculty of Mathematics and Computer Science, Jagiellonian University

TFML 2015
Będlewo, February 16-21
Goals of this Work

- How to select data representation and metric for a given data set?
Goals of this Work

- How to select data representation and metric for a given data set?
- Combining various data representations and metrics.
Goals of this Work

- How to select data representation and metric for a given data set?
- Combining various data representations and metrics.
- Optimizing a linear combination of selected distance measures.
Mixture of metrics optimization for machine learning problems

Magdalena Wiercioch

Goals of this Work

Outline
Motivation
Background
Approach
Experimental results
Conclusion

References
Motivation

real-life problem of chemoinformatics
Motivation

Mixture of metrics optimization for machine learning problems

Magdalena Wierciuch

Goals of this Work
Outline
Motivation
Background
Approach
Experimental results
Conclusion
References
Representation of molecules

Fingerprints are binary strings where a given bit indicates the absence or presence of particular pattern.
Representation of molecules

Fingerprints are binary strings where a given bit indicates the absence or presence of particular pattern. Problems:

- high dimensionality
Representation of molecules

Fingerprints are binary strings where a given bit indicates the absence or presence of particular pattern. Problems:

- high dimensionality
- they are not unique
Biological activity

- $IC_{50}$, $EC_{50}$, $K_d$
Biological activity

- $IC_{50}$, $EC_{50}$, $K_d$
- a binding constant $K_i$ was used
Biological activity

- $IC_{50}$, $EC_{50}$, $K_d$
- a binding constant $K_i$ was used
- prediction of molecule’s activity is repeated several times
Biological activity

- $IC_{50}$, $EC_{50}$, $K_d$
- a binding constant $K_i$ was used
- prediction of molecule’s activity is repeated several times
- the chemical compound were considered as active if $K_i \leq 100$ while for $K_i \geq 1000$ - inactive
Intuitively: design a measure which gives low values for compounds with similar activities while high values are assigned for compounds with different values of $K_i$. 
Intuitively: design a measure which gives low values for compounds with similar activities while high values are assigned for compounds with different values of $K_i$.

METRIC LEARNING
Background

- Multidimensional Scaling (1994)
- Locally Linear Embedding (Roweis and Saul, 2000)
- Learning a Mahalanobis metric by Xing et al. (2003)
- Kernel regression (Takeda et al., 2006)
- ...
Our aims

- optimize existing metrics and representations
  - use of combination distance measures
  - coefficients
- improve classification and clustering results
Our aims

- optimize existing metrics and representations
  - use of combination distance measures
  - coefficients
- improve classification and clustering results

\[ a(x, y) \approx \omega_1 d_1(x, y) + \ldots + \omega_n d_n(x, y) \]
Optimization

- $X$ - data set
- $a : X \times X \rightarrow [0, \infty)$
Optimization

- $X$ - data set
- $a : X \times X \rightarrow [0, \infty)$
- $d : X \times X \rightarrow \mathbb{R}$
Optimization

- $X$ - data set
- $a : X \times X \rightarrow [0, \infty)$
- $d : X \times X \rightarrow \mathbb{R}$
- $\tilde{d}_\omega(x, y) := \omega_1 d_1(x, y) + \ldots + \omega_n d_n(x, y)$
Optimization

- $X$ - data set
- $a : X \times X \rightarrow [0, \infty)$
- $d : X \times X \rightarrow \mathbb{R}$
- $\tilde{d}_\omega(x, y) := \omega_1 d_1(x, y) + \ldots + \omega_n d_n(x, y)$
- in practice
  $d_\omega(x, y) := \omega_0 + \omega_1 d_1(x, y) + \ldots + \omega_n d_n(x, y)$
Optimization

- $X$ - data set
- $a: X \times X \rightarrow [0, \infty)$
- $d: X \times X \rightarrow \mathbb{R}$
- $d_\omega(x, y) := \omega_1 d_1(x, y) + \ldots + \omega_n d_n(x, y)$
- in practice
  $d_\omega(x, y) := \omega_0 + \omega_1 d_1(x, y) + \ldots + \omega_n d_n(x, y)$ or less
- formally
  $|K_i(x) - K_i(y)| = \omega_0 + \omega_1 d(x, y) + \ldots + \omega_n d(x, y) + \epsilon$,
Optimization

- $X$ - data set
- $a : X \times X \to [0, \infty)$
- $d : X \times X \to \mathbb{R}$
- $\widetilde{d}_\omega(x, y) := \omega_1 d_1(x, y) + \cdots + \omega_n d_n(x, y)$
- in practice
  $d_\omega(x, y) := \omega_0 + \omega_1 d_1(x, y) + \cdots + \omega_n d_n(x, y)$ or less formally
  $|K_i(x) - K_i(y)| = \omega_0 + \omega_1 d(x, y) + \cdots + \omega_n d(x, y) + \epsilon,$
- $\sum_{x,y \in X} (a(x, y) - d_\omega(x, y))^2$
## Data sets

<table>
<thead>
<tr>
<th>receptor name</th>
<th>role</th>
<th>actives</th>
<th>inactives</th>
</tr>
</thead>
<tbody>
<tr>
<td>M₁</td>
<td>modulates few of physiological functions</td>
<td>759</td>
<td>938</td>
</tr>
<tr>
<td>h₁</td>
<td>has an impact on pathophysiological conditions</td>
<td>635</td>
<td>545</td>
</tr>
<tr>
<td>5-HT₇</td>
<td>processes, such as aggression</td>
<td>704</td>
<td>339</td>
</tr>
<tr>
<td>5-HT₂ₐ</td>
<td>has an impact on central nervous system</td>
<td>1835</td>
<td>851</td>
</tr>
<tr>
<td>5-HT₆</td>
<td>mediates both excitatory and inhibitory neurotransmission</td>
<td>1490</td>
<td>341</td>
</tr>
<tr>
<td>5-HT₂₉</td>
<td>has an impact on central nervous system</td>
<td>1210</td>
<td>926</td>
</tr>
</tbody>
</table>
Dissimilarity metrics

- **Buser**: \( \frac{cd+c}{cd+a+b-c} \)
- **Tanimoto**: \( \frac{c}{a+b-c} \)
### Goals of this Work

<table>
<thead>
<tr>
<th>receptor</th>
<th>optimized</th>
<th>B-KR</th>
<th>B-Ext</th>
<th>B-Subs</th>
<th>T-KR</th>
<th>T-Ext</th>
<th>T-Subs</th>
</tr>
</thead>
<tbody>
<tr>
<td>M₁</td>
<td>0.67</td>
<td>0.57</td>
<td>0.55</td>
<td>0.57</td>
<td>0.58</td>
<td>0.54</td>
<td>0.54</td>
</tr>
<tr>
<td>h₁</td>
<td>0.65</td>
<td>0.59</td>
<td>0.56</td>
<td>0.52</td>
<td>0.58</td>
<td>0.6</td>
<td>0.57</td>
</tr>
<tr>
<td>5-HT₇</td>
<td>0.69</td>
<td>0.63</td>
<td>0.61</td>
<td>0.56</td>
<td>0.58</td>
<td>0.59</td>
<td>0.56</td>
</tr>
<tr>
<td>5-HT₆</td>
<td>0.68</td>
<td>0.6</td>
<td>0.62</td>
<td>0.6</td>
<td>0.57</td>
<td>0.57</td>
<td>0.57</td>
</tr>
<tr>
<td>5-HT₂₇</td>
<td>0.66</td>
<td>0.61</td>
<td>0.59</td>
<td>0.49</td>
<td>0.63</td>
<td>0.56</td>
<td>0.5</td>
</tr>
<tr>
<td>5-HT₂₈</td>
<td>0.7</td>
<td>0.64</td>
<td>0.61</td>
<td>0.59</td>
<td>0.64</td>
<td>0.59</td>
<td>0.54</td>
</tr>
</tbody>
</table>

### Outline

- Motivation
- Background
- Approach
- Experimental results
- Conclusion
- References
### Goals of this Work

**Outline**
- Motivation
- Background
- Approach
- Experimental results
- Conclusion

### References

---

### Mixture of metrics optimization for machine learning problems

Magdalena Wiercioch

---

**k-means**

<table>
<thead>
<tr>
<th>receptor name</th>
<th>optimized</th>
<th>B-KR</th>
<th>B-Ext</th>
<th>B-Subs</th>
<th>T-KR</th>
<th>T-Ext</th>
<th>T-Subs</th>
</tr>
</thead>
<tbody>
<tr>
<td>M₁</td>
<td>0.4</td>
<td>0.39</td>
<td>0.36</td>
<td>0.37</td>
<td>0.36</td>
<td>0.37</td>
<td>0.34</td>
</tr>
<tr>
<td>h₁</td>
<td>0.3</td>
<td>0.28</td>
<td>0.27</td>
<td>0.24</td>
<td>0.26</td>
<td>0.26</td>
<td>0.27</td>
</tr>
<tr>
<td>5-HT₇</td>
<td>0.52</td>
<td>0.48</td>
<td>0.49</td>
<td>0.46</td>
<td>0.48</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>5-HT₆</td>
<td>0.33</td>
<td>0.3</td>
<td>0.3</td>
<td>0.31</td>
<td>0.31</td>
<td>0.29</td>
<td>0.27</td>
</tr>
<tr>
<td>5-HT₂₉</td>
<td>0.46</td>
<td>0.44</td>
<td>0.43</td>
<td>0.4</td>
<td>0.42</td>
<td>0.39</td>
<td>0.39</td>
</tr>
<tr>
<td>5-HT₂₆</td>
<td>0.35</td>
<td>0.31</td>
<td>0.3</td>
<td>0.31</td>
<td>0.3</td>
<td>0.31</td>
<td>0.28</td>
</tr>
</tbody>
</table>
hierarchical clustering

<table>
<thead>
<tr>
<th>receptor name</th>
<th>optimized</th>
<th>B-KR</th>
<th>B-Ext</th>
<th>B-Subs</th>
<th>T-KR</th>
<th>T-Ext</th>
<th>T-Subs</th>
</tr>
</thead>
<tbody>
<tr>
<td>M₁</td>
<td>0.45</td>
<td>0.4</td>
<td>0.41</td>
<td>0.35</td>
<td>0.39</td>
<td>0.37</td>
<td>0.36</td>
</tr>
<tr>
<td>h₁</td>
<td>0.23</td>
<td>0.19</td>
<td>0.15</td>
<td>0.17</td>
<td>0.19</td>
<td>0.17</td>
<td>0.16</td>
</tr>
<tr>
<td>5-HT₇</td>
<td>0.41</td>
<td>0.35</td>
<td>0.33</td>
<td>0.35</td>
<td>0.36</td>
<td>0.34</td>
<td>0.33</td>
</tr>
<tr>
<td>5-HT₆</td>
<td>0.4</td>
<td>0.36</td>
<td>0.37</td>
<td>0.35</td>
<td>0.37</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>5-HT₂C</td>
<td>0.52</td>
<td>0.48</td>
<td>0.46</td>
<td>0.45</td>
<td>0.46</td>
<td>0.44</td>
<td>0.45</td>
</tr>
<tr>
<td>5-HT₂A</td>
<td>0.42</td>
<td>0.35</td>
<td>0.33</td>
<td>0.34</td>
<td>0.36</td>
<td>0.34</td>
<td>0.32</td>
</tr>
</tbody>
</table>
Mixture of metrics optimization for machine learning problems

Magdalena Wiercioch

Goals of this Work

Outline

Motivation

Background

Approach

Experimental results

Conclusion

References

after optimization process
after optimization process

5-HT$_7$

ARI
0.65 0.67 0.69 0.71

k
5 10 15 20
more explanatory variables
more explanatory variables
Conclusion

- metric learning problem
- a single function which combines data representation-metric pairs can improve the performance of metric-based algorithms


