Graph-based Semi-supervised Learning: Realizing Pointwise Smoothness Probabilistically

Yuan Fang, University of Illinois at Urbana-Champaign
Kevin Chang, University of Illinois at Urbana-Champaign
Hady Lauw, Singapore Management University

06/24/2014 @ ICML
Graph-based semi-supervised learning

Data points: $X \in \mathcal{X} = \{x_1, ..., x_n\}$
Labels: $Y \in \mathcal{Y}$, the label of $X$

Labeled data + Structures in unlabeled data

$W_{ij} = \begin{cases} \exp\left(-\frac{||x_i - x_j||^2}{2\sigma^2}\right) & i \neq j \\ 0 & i = j \end{cases}$

Affinity: pairwise geodesic distance
Central notion in SSL: smoothness

Two points $x_i, x_j$ are close

Their respective labels $y_i, y_j$ are likely to be the same
Current graph-based smoothness does not precisely realize the nature of smoothness.

- Realizing Pointwise Nature
- through Probabilistic Modeling
Goal 1: **Pointwise** nature of smoothness

Inherently a property occurring "everywhere" on every point

Relating the behaviour of *each* point to that of its close points
Precisely expressing pointwise smoothness

If two points are close, their labels are likely to be the same.

P1. How to judge if two points are close?
   (Data Closeness)

Vague!
   (Rigollet, 2007; Lafferty+, 2007)

P2. Meaning of labels “likely to be the same”?
   (Label Coupling)
Existing graph-based methods only express aggregate, not pointwise, smoothness

$$\min \sum_{ij} W_{ij} (y_i - y_j)^2$$

aggregate pairwise differences

\{ "Average" smooth across the graph
Not pointwise formulation (P1 & P2) \}

Not an explicit realization of smoothness!
**Goal 2:** Probabilistic model of smoothness

**P1.** How to judge if two points are close?

How close is *sufficiently* close? → Deterministic binary decision

Probabilistic formulation

**P2.** Meaning of labels “likely to be the same”?

Deterministic  ×  Probabilistic  ✓

Probabilistic Modelling: $P(Y|X)$, or $P(X|Y)$ and $P(Y)$
Our Proposal: **Probabilistic Graph-based Pointwise Smoothness (PGP)**

- **P1. Data Closeness**
  - Capture closeness distribution of two points $X, X'$ on graph

- **P2. Label Coupling**
  - Couple the label distribution of two close points $X, X'$
P1. Data Closeness: Random walk on graph

- Consider a random walk on the graph \( \{V_t: t = 0,1, \ldots \} \)
  - A traversal from \( x_i \) to \( x_j \) (\( V_t = x_i, V_{t+1} = x_j \))

A traversal from \( X \) to \( X' \) "connects" \( X, X' \) as close points

- The event that \( x_i, x_j \) close: \( (X = x_i, X' = x_j) \)
  - Distribution of traversal from \( x_i \) to \( x_j \) in the long run (\( t \to \infty \))
  - \( (X, X') \) is a pair of limiting random variables:
    \[
    (V_t, V_{t+1}) \xrightarrow{d} (X, X')
    \]
P2. Label Coupling:
Statistical indistinguishability

- \( X \) is \( x_i \) itself
  \( (X = x_i) \)

- or

- \( X \) is close to \( x_i \)
  \( (X' = x_i) \)

\[
\begin{align*}
\text{or} & \quad \text{Y of } X \text{ distributes similarly} \\
\text{or} & \quad P(Y|X = x_i) \quad \text{are alike} \\
\text{or} & \quad P(Y|X, X' = x_i) \quad \text{are alike}
\end{align*}
\]

Statistical Indistinguishability
P2. Label Coupling:
Statistical indistinguishability

\[ p(Y | X = x_i) \rightarrow p(Y | X, X' = x_i) \]

\( \alpha \)-statistical indistinguishable

Intuition: indistinguishability will slowly fade along a “chain” of close points
Constraint-based solution for $P(X|Y = y)$

Labelled data:

\[
p(X = x_i | Y = y) = p(Y = y | X = x_i) p(X = x_i) / p(Y = y)
\approx p(Y = y | X = x_i) Z_i
\]

Unlabelled data (smoothness P1 & P2):

\[
\text{unique solution to } p(X|Y = y)
\]
Experiment 1: PGP is smooth “everywhere”

L: labeled point
Size: magnitude of decision function
Color: predicted class

Non-smooth area
Misclassification
Experiment 2: PGP is **consistently** the best

<table>
<thead>
<tr>
<th></th>
<th>Digit1</th>
<th>Text</th>
<th>ISOLET</th>
<th>Cancer</th>
<th>USPS</th>
<th>Yeast</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRF</td>
<td>.894</td>
<td>.451</td>
<td>.627</td>
<td>.871</td>
<td>.638</td>
<td>.510</td>
</tr>
<tr>
<td>LSVM</td>
<td>.833</td>
<td>.428</td>
<td>.719</td>
<td>.886</td>
<td><strong>.698</strong></td>
<td>.562</td>
</tr>
<tr>
<td>GGS</td>
<td>.855</td>
<td>.567</td>
<td>.677</td>
<td>.867</td>
<td>.666</td>
<td>.540</td>
</tr>
<tr>
<td>MP</td>
<td><strong>.901</strong></td>
<td>.558</td>
<td>.692</td>
<td>.898</td>
<td><strong>.713</strong></td>
<td>.574</td>
</tr>
<tr>
<td>PARW</td>
<td>.881</td>
<td><strong>.587</strong></td>
<td>.721</td>
<td>.893</td>
<td><strong>.706</strong></td>
<td>.575</td>
</tr>
<tr>
<td>PGP</td>
<td><strong>.910</strong></td>
<td><strong>.592</strong></td>
<td><strong>.734</strong></td>
<td><strong>.910</strong></td>
<td><strong>.704</strong></td>
<td><strong>.593</strong></td>
</tr>
</tbody>
</table>
Conclusion

- Smoothness is pointwise in nature
- Probabilistic modelling of smoothness