Computer-aided diagnosis from weak supervision: A benchmarking study

Melih Kandemir
melih.kandemir@iwr.uni-heidelberg.de

June 20, 2014
Background

- Cancer diagnosis from H & E stained tissue images
- Barrett’s cancer
- Oesophagus/Eosophagus/Gullet
- Diabetic retinopathy screening
- Weakly supervised learning = multiple instance learning
Supervised Learning (SL)
Supervised Learning (SL)
Multiple Instance Learning (MIL)
Multiple Instance Learning (MIL)
Motivation 1: Annotation cost

1http://biometrics.cse.msu.edu/projects/auto_gleason.html
Motivation 1: Annotation cost \(^2\)

\(^2\)http://biometrics.cse.msu.edu/projects/auto_gleason.html
Motivation 2: Annotation noise
Motivation 3: The slide google

- **Bag**: Whole slide(s), **Instance**: Patch
- Any solution better than raster scanning saves time
Diagnosis pipeline

- Bag
  - Feature Vector (Instance)
  - MIL Classifier

-1: Normal
+1: Diseased
mi-SVM$^3$: Problem

\[
\begin{align*}
\min_{y, w, b, \xi} & \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i, \\
\text{s.t.} & \quad y_i(w^T \phi(x_i)) \geq 1 - \xi_i, \quad \forall i, \\
& \quad \xi_i \geq 0, \quad \forall i, \\
& \quad \max(y_b) = Y_b, \quad \forall b.
\end{align*}
\]

$^3$Andrews et al., NIPS, 2003
mi-SVM\textsuperscript{4}: Solution

- Start from the configuration that all positive bag instances are positive
- While not converged
  - Train SVM
  - Assign positive bag instances to the class that the SVM predicts

\textsuperscript{4}Andrews et al., NIPS, 2003
MI-SVM: Problem

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \| \mathbf{w} \|^2 + C \sum_{i=1}^{N} \xi_i,$$

s.t. \( \forall b : Y_b \max_{\mathbf{x}_{bn} \in \mathbf{X}_b} (\mathbf{w}^T \phi(\mathbf{x}_{bn})) \geq 1 - \xi_b, \quad \xi_b \geq 0, \)

---

\(^5\)Andrews et al., NIPS, 2003
MI-SVM \textsuperscript{6}: Solution

Initialize: i) Random selection, ii) Mean instance of each bag.
While not converged

- Train SVM on the chosen instance set
- Choose the most representative instance (with maximum predicted margin) from each bag

\textsuperscript{6}Andrews et al., NIPS, 2003
1. Choose an instance-level kernel function $k_{\text{inst}}(\cdot, \cdot)$ and a threshold $\tau$

2. Construct one graph for each bag $b$:
   - Add one node per instance
   - Add an edge between two instances if $k_{\text{inst}}(x, x') > \tau$

3. Compute the adjacency matrix of the graph
   \[ [W_b]_{mn} = w_{mn}^b = \mathbb{I}(k_{\text{inst}}(x_{bm}, x_{bn}) > \tau) \]

4. Compute the bag-level kernel for all bag pairs
   \[
   k_{\text{bag}}(X_b, X_c) = \frac{\sum_{n=1}^{N_b} \sum_{m=1}^{N_c} v_{bn} v_{cm} k_{\text{inst}}(x_{bn}, x_{cm})}{\sum_{n=1}^{N_b} v_{bn} \sum_{m=1}^{N_c} v_{cm}}
   \]
   where $v_{bn} = 1/\sum_{u=1}^{N_b} w_{nu}^b$, $v_{cm} = 1/\sum_{u=1}^{N_c} w_{mu}^c$.

5. Run any kernel learner (e.g. SVM) on the bag-level Gram matrix

---

Zhou et al., ICML, 2009
### Features

<table>
<thead>
<tr>
<th>Color features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Color histograms of the entire patch, normal, cell pixels, cancer cell pixels, lymphocyte cell pixels, and stroma pixels</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Texture features</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 Mean of local binary pattern histograms of 20x20-pixel grids</td>
</tr>
<tr>
<td>3 Mean of SIFT descriptors</td>
</tr>
<tr>
<td>4 Box count for grid sizes 2,3,...,8</td>
</tr>
</tbody>
</table>
Case Study 1: Barrett’s cancer

Normal example 1
Case Study 1: Barrett’s cancer

Normal example 2
Case Study 1: Barrett’s cancer

Cancer example 1
Case Study 1: Barrett’s cancer

Cancer example 2
Case Study 1: Barrett’s cancer

Cancer example 3
Case Study 1: Barrett’s cancer

Cancer example 4
### Corel-level supervision/Core-level detection (Barrett’s cancer)

<table>
<thead>
<tr>
<th>Method</th>
<th>Acc (%)</th>
<th>F1 score</th>
<th>AUC-ROC</th>
<th>AUC-PR</th>
</tr>
</thead>
<tbody>
<tr>
<td>mi-Graph</td>
<td>86.4</td>
<td>0.90</td>
<td>0.93</td>
<td>0.97</td>
</tr>
<tr>
<td>MILBoost</td>
<td>83.0</td>
<td>0.88</td>
<td>0.91</td>
<td>0.96</td>
</tr>
<tr>
<td>B-KI-SVM</td>
<td>82.6</td>
<td>0.88</td>
<td>0.91</td>
<td>0.95</td>
</tr>
<tr>
<td>GPMIL</td>
<td>81.2</td>
<td>0.88</td>
<td>0.90</td>
<td>0.93</td>
</tr>
<tr>
<td>I-KI-SVM</td>
<td>80.3</td>
<td>0.86</td>
<td>0.89</td>
<td>0.93</td>
</tr>
<tr>
<td>iAPR</td>
<td>79.4</td>
<td>0.87</td>
<td>0.88</td>
<td>0.94</td>
</tr>
<tr>
<td>Citation k-NN</td>
<td>74.5</td>
<td>0.83</td>
<td>0.72</td>
<td>0.82</td>
</tr>
<tr>
<td>EMDD</td>
<td>72.2</td>
<td>0.83</td>
<td>0.72</td>
<td>0.82</td>
</tr>
<tr>
<td>mi-SVM</td>
<td>68.4</td>
<td>0.81</td>
<td>0.86</td>
<td>0.76</td>
</tr>
<tr>
<td>MI-SVM</td>
<td>68.1</td>
<td>0.81</td>
<td>0.89</td>
<td>0.94</td>
</tr>
<tr>
<td>SIL-SVM</td>
<td>68.1</td>
<td>0.81</td>
<td>0.92</td>
<td>0.95</td>
</tr>
<tr>
<td>Fully S.SVM</td>
<td>85.0</td>
<td>0.90</td>
<td>0.92</td>
<td>0.96</td>
</tr>
</tbody>
</table>
Core-level supervision/Patch-level detection (Barrett’s cancer)

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy (%)</th>
<th>F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPMIL</td>
<td>71.8</td>
<td>0.74</td>
</tr>
<tr>
<td>GPMIL</td>
<td>65.8</td>
<td>0.54</td>
</tr>
<tr>
<td>I-KISVM</td>
<td>65.4</td>
<td>0.45</td>
</tr>
<tr>
<td>B-KISVM</td>
<td>64.7</td>
<td>0.48</td>
</tr>
<tr>
<td>mi-SVM</td>
<td>62.7</td>
<td>0.71</td>
</tr>
<tr>
<td>MISVM</td>
<td>46.9</td>
<td>0.64</td>
</tr>
<tr>
<td>Fully S.SVM</td>
<td>83.5</td>
<td>0.82</td>
</tr>
</tbody>
</table>

DPMIL: Dirichlet Process Multiple Instance Learning

M. Kandemir and F. Hamprecht, Instance Label Prediction by Dirichlet Process Multiple Instance Learning, UAI, 2014
DPMIL: Dirichlet Process Multiple Instance Learning

Normal cores  Local tumors  Core-wide tumors

M. Kandemir and F. Hamprecht, Instance Label Prediction by Dirichlet Process Multiple Instance Learning, UAI, 2014
Diabetic Retinopathy Screening
Diabetic Retinopathy Screening

Lesions:

- Microaneurysms (Blue)
- Hemorrhages (Green)
- Hard exudates (Red)
# Image-level supervision/Image-level diagnosis (Diabetic Retinopathy Screening)

<table>
<thead>
<tr>
<th>Method</th>
<th>Acc.</th>
<th>F1 score</th>
<th>AUC-ROC</th>
<th>AUC-PR</th>
</tr>
</thead>
<tbody>
<tr>
<td>mi-Graph</td>
<td>72.5</td>
<td>0.75</td>
<td>0.81</td>
<td>0.85</td>
</tr>
<tr>
<td>MILBoost</td>
<td>64.1</td>
<td>0.66</td>
<td>0.70</td>
<td>0.73</td>
</tr>
<tr>
<td>Citation k-NN</td>
<td>62.8</td>
<td>0.68</td>
<td>0.65</td>
<td>0.69</td>
</tr>
<tr>
<td>GPMIL</td>
<td>59.2</td>
<td>0.43</td>
<td>0.76</td>
<td>0.80</td>
</tr>
<tr>
<td>SIL-SVM</td>
<td>58.4</td>
<td>0.72</td>
<td>0.78</td>
<td>0.82</td>
</tr>
<tr>
<td>B-KI-SVM</td>
<td>55.9</td>
<td>0.68</td>
<td>0.60</td>
<td>0.64</td>
</tr>
<tr>
<td>I-KI-SVM</td>
<td>55.5</td>
<td>0.44</td>
<td>0.61</td>
<td>0.65</td>
</tr>
<tr>
<td>EMDD</td>
<td>55.1</td>
<td>0.69</td>
<td>0.58</td>
<td>0.61</td>
</tr>
<tr>
<td>MI-SVM</td>
<td>54.5</td>
<td>0.70</td>
<td>0.68</td>
<td>0.73</td>
</tr>
<tr>
<td>mi-SVM</td>
<td>54.5</td>
<td>0.71</td>
<td>0.58</td>
<td>0.62</td>
</tr>
<tr>
<td>iAPR</td>
<td>54.4</td>
<td>0.70</td>
<td>0.53</td>
<td>0.60</td>
</tr>
<tr>
<td>Fully S.SVM</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Implications/Ideas

- MIL > SL when annotations are noisy
- mi-Graph generalizes over different core-level diagnosis problems well.
- DPMIL is good at locating tumors given core-level supervision.
- Relation of within-bag instances is discriminative! → Relational learning \(^{11}\)

\(^{11}\)M. Kandemir et al., *Empowering multiple instance histopathology cancer diagnosis by cell graphs*, MICCAI, 2014
Collaborators

Annette Feuchtinger
Helmholtz Zentrum München

Axel Walch
Helmholtz Zentrum München

Fred A. Hamprecht
Heidelberg University, HCI/IWR