Machine Learning in Health Informatics: Making Better use of Domain Experts

Byron C Wallace

joint work with Kevin Small, Chris Schmid, Issa Dahabreh, Joseph Lau, Thomas Trikalinos, and Carla Brodley
Advancing Machine Learning Through its Application

• Work on *actual problems* with meaningful implications
• Use off-the-shelf methods when possible; *develop new tech when necessary*
This Work, More Specifically

- New machine learning methods for evidence-based medicine & clinical informatics tasks (and similar domains)

- Citation screening, a crucial part of evidence-based medicine, as a motivating (and exemplary) task
  - Machine learning research motivated by this work
Evidence-Based Medicine

• What you kind of would have hoped all medicine is
EBM: Systematic Reviews

1. Formulate question, protocol & query
2. Search database
3. Screen retrieved citations
4. Extract data
5. Synthesize extracted data

Diagram:
- Question cloud connected to PubMed
- PubMed connected to grid with outcomes a, b, c, d
- Grid with outcomes a, b, c, d connected to extracted data
- Extracted data connected to synthesized data
EBM: Systematic Reviews

1. Formulate question, protocol & query
2. Search database
3. Screen retrieved citations
4. Extract data
5. Synthesize extracted data
Abstract Screening
Abstract Screening
In 2010 eleven systematic reviews were published every single day.

[Bastian et al, PLoS Medicine 2010]
Semi-Automating Abstract Screening via Supervised Machine Learning

manually screened abstracts

labeled data \( \mathcal{L} \)

e.g. SVM

expert

doctor conducting review

unlabeled data \( \mathcal{U} \)

citations from PubMed search

predictive model
Open Problems in Machine Learning
(or, the inadequacy of ‘off-the-shelf’ tech)

- **Class imbalance** – far fewer relevant than irrelevant abstracts
  - *asymmetric costs* sensitivity more important than specificity

- Doctor time is scarce and expensive
  - better models, fewer labels: *active learning* and *dual supervision*
Open Problems in Machine Learning
(or, the inadequacy of ‘off-the-shelf’ tech)

• Class imbalance – far fewer relevant than irrelevant abstracts
  – asymmetric costs

• Doctor time is scarce and expensive
  – better models, fewer labels: active learning and dual supervision

Similar concerns in many real world learning tasks (in clinical informatics and beyond!)
Contributions

- New perspectives on *class imbalance*
- Exploiting *alternative forms of supervision*
- Methods for *real-world active learning*
- *And* a deployed system that integrates the above for the task of citation screening
Class Imbalance, Redux

Wallace, Small, Dahabreh, Brodley and Trikalinos

ICDM 2011, ICDM 2012 and the Journal of Knowledge And Information Synthesis (KAIS) in 2013
The Problem of Class Imbalance

- ML methods fare poorly on imbalanced data

- Accuracy is not a good metric (usually): prefer sensitivity/specificity

- Simply *undersampling* works *frustratingly* well [Van Hulse et al, ICML 07]. *Why??*
The Problem of Class Imbalance

minority class

majority class

P

G
Bias

the one we get

the separator we want
Undersampling

P

G
Undersampling
Undersampling
classifiers induced over undersampled datasets tend to be less biased
When is Bias Likely?

\[(1 - \pi)C_{fp} \int_{\mathcal{R}_+}^{} G(x)dx > \pi C_{fn} \int_{\mathcal{R}_-}^{} P(x)dx\]

prevalence

density of majority points on the ‘wrong’ side of \(w^*\)

ditto, minority points

the empirical cost will probably be less for a biased (w.r.t. \(w^*\)) separator
When is Bias Likely?

\[(1 \times \pi) C_{fp} \int_{\mathbb{R}^{w*}_{+}} G(x)dx \geq C_{fn} \int_{\mathbb{R}^{w*}_{-}} P(x)dx\]

density of majority points on the ‘wrong’ side of \(w^*\)

ditto, minority points

the empirical cost will probably be less for a biased (w.r.t. \(w^*\)) separator
“Towards Modernizing the Systematic Review Pipeline: Efficient Updating via Data Mining”

*Genetics in Medicine 2012*
Results (Updating Reviews)

We can achieve 100% sensitivity while substantially reducing workload.

“Towards Modernizing the Systematic Review Pipeline: Efficient Updating via Data Mining”
Genetics in Medicine 2012
The Trouble with Probability Estimates for Imbalanced Data

- Probability estimates for minority estimates are unreliable
Class Probability Estimates for Imbalanced Data

$$\sum_{i=0}^{N} (y_i - \hat{P}\{y_i | x_i\})^2$$

$$\frac{N}{N}$$
Class Probability Estimates for Imbalanced Data

\[
BS^+ = \frac{\sum_{i=0}^{N} (y_i - \hat{P}\{y_i|x_i\})^2}{N} \\
BS^- = \frac{P_{y_i=0} (y_i - \hat{P}\{y_i|x_i\})^2}{N_{neg}} \\
P_{y_i=1} (y_i - \hat{P}\{y_i|x_i\})^2 \\
N_{pos} \\
\]
Better Probability Estimates for Imbalanced Data

Want this to be low!
Contributions

- New perspectives on *class imbalance*
- Exploiting *alternative forms of supervision*
- Methods for *real-world active learning*
- Practical application of the above to the task of citation screening via a deployed system
The Constrained Weight Space Support Vector Machines (CW-SVM)

Small, Wallace, Brodley and Trikalinos

ICML 2011
Dual Supervision

Experts offer:  

- β-blockers, humans

- aspirin, rats
Dual Supervision

Experts offer: $\beta$-blockers, humans $\checkmark$  aspirin, rats $\times$

How to exploit labeled features during classifier induction?
**Ranked Labeled Features**

- terrific
- witty, solid
- lively
- stumbles
- obvious
- terrible, awful

- Incorporates *ranked* labeled features and instance labels directly into the SVM

- Allows additional expressivity for annotator
  - natural interaction for citation screening
Support Vector Machines

- support vectors
- margin
Support Vector Machines

\[
\text{argmin}_{w,b,\xi} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{m} \xi_i \\
\text{s.t.} \quad y_i (w \cdot x_i + b) \geq 1 - \xi_i \quad \forall i = 1 \ldots m \\
\xi_i \geq 0 \quad \forall i = 1 \ldots m
\]
Support Vector Machines

\[
\text{argmin}_{w,b,\xi} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{m} \xi_i \\
\text{s.t.} \quad y_i (w \cdot x_i + b) \geq 1 - \xi_i \quad \forall i = 1 \ldots m \\
\xi_i \geq 0 \\n\forall i = 1 \ldots m
\]

We want to bias the weight space to reflect expert knowledge
The Constrained Weight Space SVM

$$\arg\min_{w, b, \xi, \rho} \frac{1}{2} \|w\|^2 + C_1 \sum_{i=1}^{m} \xi_i - C_2 \sum_{\alpha, \beta} \rho_{\alpha, \beta}$$

s.t. $$y_i (w \cdot x_i + b) \geq 1 - \xi_i \quad \forall i = 1 \ldots m$$
$$w_\alpha - w_\beta \geq \rho_{\alpha, \beta} \quad \forall \alpha, \beta$$
$$\tau_- \leq w_\alpha, w_\beta \leq \tau_+ \quad \forall \alpha, \beta$$
$$\xi_i \geq 0 \quad \forall i = 1 \ldots m$$
The Constrained Weight Space SVM

$$\text{argmin}_{w, b, \xi, \rho} \quad \frac{1}{2} \|w\|^2 + C_1 \sum_{i=1}^{m} \xi_i - C_2 \sum_{\alpha, \beta} \rho_{\alpha, \beta}$$

s.t. $$y_i (w \cdot x_i + b) \geq 1 - \xi_i \quad \forall i = 1 \ldots m$$

$$w_\alpha - w_\beta \geq \rho_{\alpha, \beta} \quad \forall \alpha, \beta$$

$$\tau_- \leq w_\alpha, w_\beta \leq \tau_+ \quad \forall \alpha, \beta$$

$$\xi_i \geq 0 \quad \forall i = 1 \ldots m$$

add a term to reward agreement with terms
Abstract Screening: Proton Beam

70 (+) features in 6 ranks, 11 (-) features in 3 ranks
Contributions

- New perspectives on *class imbalance*
- Exploiting *alternative forms of supervision*
- Methods for *real-world active learning*
- Practical application of the above to the task of citation screening via a deployed system
Active Learning for Biomedical Citation Screening

Wallace, Small, Brodley and Trikalinos

KDD 2010
Active Learning

- **Expert Annotator**: Selects $x^*$ from $U$ for labeling.
- **Labeled Data**: $\mathcal{L}$
- **Learned Classifier**: Rebuild model and evaluate classifier.
- **Unlabeled Pool**: $U$
- **Test Data**: Evaluates classifier.
AL with Imbalanced Data

- random sampling
- active (uncertainty)

sensitivity

accuracy
Co-testing

- Select $x^*$ from $U$ for labeling by disagreement.
- Label $x^*$.
- Labeled data.
- Rebuild models.
- Learn classifiers.
- Test data.
- Evaluate classifier.

[Muslea et al.; AAAI 00]
Co-testing

If model 1 disagrees with model 2 about $x$, then $x$ is a good point to label.
Co-Testing with Labeled Terms

Query strategy

1. Find documents where models based on two views disagree
2. Select $x^*$ with maximum score for view 2
Results (Micronutrients Review)
Who Should Label What?
Instance Allocation in Multiple Expert Active Learning

Wallace, Small, Brodley and Trikalinos
SDM 2010
Multiple Expert Active Learning

- e* labels x*
- select x* from U for labeling and specify expert e* to label it
- labeled data
- rebuild model
- evaluate classifier
MEAL: Key Questions

• How to select \((x^*, e^*)\), given:
  – budgetary constraints
  – workload balance constraints
  – some experts lousy, others pricey

• Want to align queries with appropriate experts (maximum accuracy for minimum cost)
Meta-Cognitive MEAL

**Ideal strategy**: novice (cheap) labelers annotate easy instances; experienced (expensive) labelers reserved for difficult instances

(rare) difficult / borderline examples

(common) easy examples
Meta-Cognitive MEAL

**Key idea** Give most instances to novice labelers; let them designate some as ‘too hard’, and pass those on to the experienced experts.
Results (Micronutrients Review)

Two labelers; one expert, one novice
Putting it All Together in a Deployed System: abstrackr

http://abstrackr.cebm.brown.edu
Monstrous infants and vampyric mothers in Bram Stoker's "Dracula".

Almond BR

Bram Stoker's "Dracula" continues to fascinate and horrify audiences, inviting a psychoanalytic explanation. While previous interpretations have emphasized oedipal dynamics and perverse sexuality, this paper proposes that early developmental issues are central. Vampires and the state of being "undead" are representations of intense oral needs, experienced in a context of passivity and helplessness. Aggressive invasion and possession of the other, with a colonization of body and soul, offer a solution to this dilemma but one devoid of true object-relatedness. The imaginative source of the Dracula figure is posited as Stoker's early invalidism and his later idealization of a powerful and charismatic actor. Implicit in the Dracula story are ideas of intrusively experienced "monstrous" babies and intrusively controlling "vampyric mothers". The author offers studies of key passages from "Dracula" in support of this reading, followed by comparative material to illustrate the spectrum of vampyric mothering: a clinical example and excerpts from a modern novel. The horror of the vampire myth is located in the unending internal attachment to a deeply needed but problematic object.

keywords: Affect,Female,History, 19th Century, Humans, Infant, Medicine in Literature, Mother-Child Relations, Mothers/psychology, Parenting, Psychoanalytic Interpretation, United States

ID: 403973

label instances (documents)

label features (terms)

auxiliary supervision (tags)
Clopidogrel 2

9079 citations have not been screened yet.
1790 citations are probably relevant.

Probabilities of remaining studies being relevant
Conclusions

Motivated by the citation screening problem, we have:

- shed light on problem of imbalance; motivated undersampling + bagging, both for classification and class probability

- Exploited labeled terms to improve classifier performance
  - and to inform active learning!

- Better active learning!
  - Proposed a MEAL method that takes advantage of meta-cognition
  - and exploited predicted labeling time (see thesis)

- Put these into a working tool
Questions?
byron_wallace@brown.edu
www.cebm.brown.edu/byron

Thanks to Carla Brodley, Kevin Small, Foster Provost, Anselm Blumer, Roni Khardon, Joseph Lau, Chris Schmid, Jaime Carbonell, Issa Dahabreh and Tom Trikalinos