Probabilistic Soft Logic for Social Good

Stephen H. Bach¹, Bert Huang¹, and Lise Getoor²

¹ University of Maryland, College Park
² University of California, Santa Cruz

Introduction to PSL

• **Big graph data** has numerous social-good applications

• Probabilistic soft logic (PSL) is a declarative language for defining probabilistic models over big data sets, such as socio-behavioral graphs [Bach et al., UAI 2013; Bröcheler et al., UAI 2010]

• PSL makes important relational tasks in social-good applications easier, such as collective classification, link prediction, and entity resolution

PSL Features

• Intuitive, logic-based syntax, which makes defining relational models easy

• Highly scalable reasoning using convex optimization techniques such as the alternating direction method of multipliers (ADMM)

• Support for learning with latent variables

• New distributed GraphLab inference [Miao et al., Intl. Conf. on Big Data 2013]

Real-World Applications of PSL

• Predict disease outbreaks and civil unrest from news and social media [Ramakrishnan et al., KDD 2014]

• Model student engagement in massive open online courses (MOOCs) as latent variables in order to predict outcomes [Ramesh et al., AAAI 2014]

• Identify latent groups in social media [Bach et al., Inferning Workshop 2013]

• Predict trust in social networks [Huang et al., SBP 2013]

Example PSL Program for Disease-Outbreak Detection Using Social Media

/* Defines predicates */
HasDisease(Location, Disease)
HasLocation(Post, Location)
MentionsEntity(Post, Entity) (closed)
IsLocation(Entity) (closed)
IsDisease(Entity) (closed)
IsGeotagged(Post, Geotag) (closed)
InLocation(Geotag, Location) (closed)
Nearby(Location, Location) (closed)

/* Defines subtype relationships */
Location is a Entity
Disease is a Entity

/* Infers locations from posts from location mentions */
2.0 : MentionsEntity(P, E) & IsLocation(E) -> HasLocation(P, E)
/* Uses any available geotags as strong evidence for determining a post's location */
10.0 : IsGeotagged(P, G) & GeotagInLocation(GT, L) -> HasLocation(P, L)
/* Infers disease presence from disease mentions */
5.0 : MentionsEntity(P, E) & IsDisease(E) -> HasDisease(P, E)
/* Infers disease presence from nearby locations */
1.0 : HasDisease(L1, D) & Nearby(L1, L2) -> HasDisease(L2, D)
/* Requires sufficient evidence in order to infer disease presence */
0.5 : !HasDisease(L, D)

Open-source implementation and video tutorials available at PSL's website!

Learn more: http://psl.cs.umd.edu