

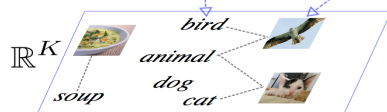
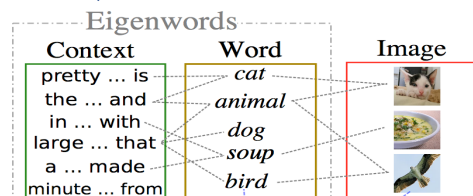
数理統計学チーム (1)

マルチモーダルデータの多変量解析と応用

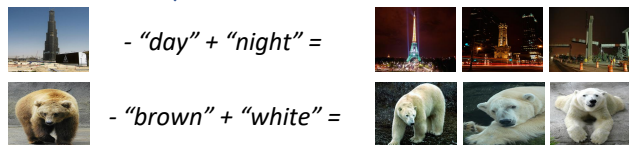
グラフ（ネットワーク）を低次元に埋め込むデータ解析手法を考える

Shimodaira, Neural Networks 2016

↓
文書データとソーシャルメディアのデータを埋め込んでみる



↓
画像と単語を引いたり足したりして検索してみた

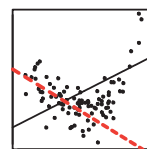


Fukui, Oshikiri and Shimodaira, Textgraphs 2017

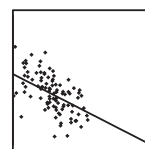
共変量シフト（転移学習）の理論

学習とテストでデータの分布が変わるとき
の統計理論をひっそりと考える

training



test



確率密度比(density ratio)

$$w(x) = \frac{f_{\text{test}}(x)}{f_{\text{train}}(x)}$$

Shimodaira, Journal of Statistical Inference and Planning 2000



共変量シフトと命名！(covariate shift)

機械学習の分野でよく使われるようになる

Shimodaira (2000)の論文被引用数は1000くらい



知らないうちに...

ディープラーニングを加速する手法
(Batch Normalization)に組み込まれる

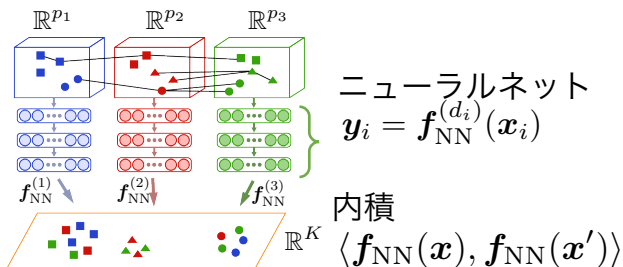
Batch normalization: Accelerating deep network training by reducing internal covariate shift (Ioffe and Szegedy, ICML 2015)

深層学習の標準手法として広く利用されている

数理統計学チーム (2)

内積を拡張する類似度関数の理論

マルチモーダルデータのニューラルネットによるグラフ埋め込みを考える



↓
機械学習では内積を類似度につかうことが多いけど、どうしてかな？

内積でも任意の正定値カーネルを学習できる！
Okuno, Hada, Shimodaira, ICML 2018

↓
双曲空間や疑ユークリッド空間に拡張したらもっとよくなる？

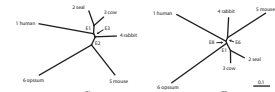
内積にバイアス項つけたり負の重みを入れたりするだけで実装できる！

Okuno, Kim, Shimodaira, AISTATS 2019
Kim, Okuno, Fukui, Shimodaira, IJCAI 2019

信頼度の計算 (仮説検定) の理論と応用

系統樹や階層型クラスタリングで木やクラスタの信頼度を計算したい

↓
多重比較法を対数尤度に適用した



系統樹の検定に使われるようになる

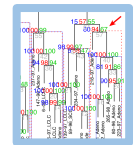
↓
Shimodaira and Hasegawa (1999) の論文被引用数は3800くらい

多重比較は本当の発見を見逃しやすい

↓
ランダムネスのスケール則から全く新しいリサンプリング法 (マルチスケール・ブートストラップ) を考案. Shimodaira (2002) など3本の論文被引用数は5500くらい

データをみてクラスタ (仮説) を決めてる.
選択バイアスを修正しないと. . .

(出版バイアスとかp-hackingなどと同じ問題)



↓
選択的推測 (selective inference) の理論と手法を研究
Terada and Shimodaira (2017, 2019), Shimodaira and Terada (2019).

<http://stat.sys.i.kyoto-u.ac.jp> 「最近の論文 (2019/02/28)」 など参照してください