

Folksonomyによる 階層構造画像データベースの 構築

CVIM研究会

電気通信大学 情報工学専攻

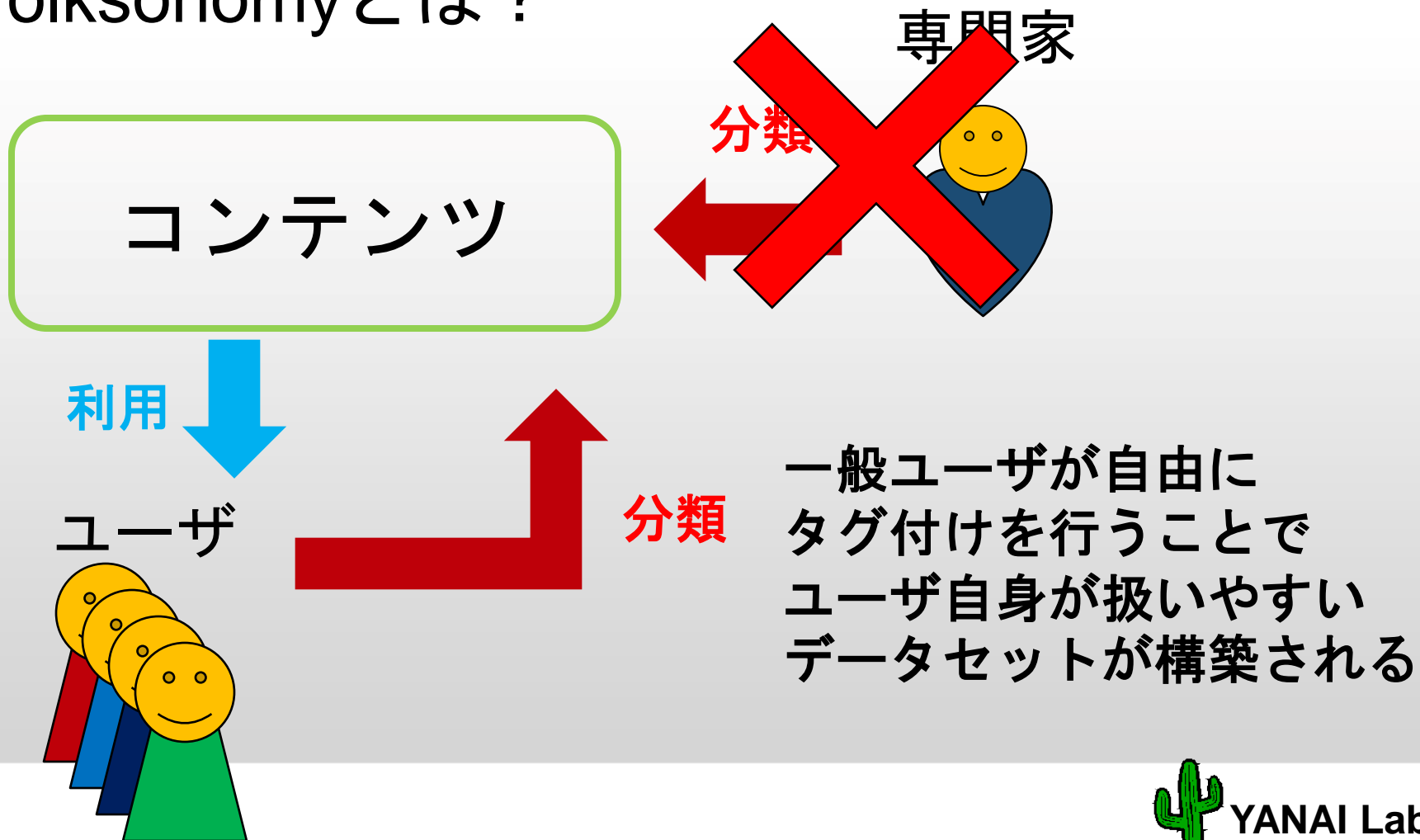
柳井研究室

秋間 雄太

2010年3月19日

背景

■Folksonomyとは？




背景

- Folksonomyの欠点



語彙が統一されていない！

メン



語彙間の関係を
構築することで解決



背景

Folksonomyに基づいた画像データセット Flickr

- タグなどによって意味的な価値を画像に付与



しかし、

- 基本的に付与されているタグは単なる単語の羅列
- 概念間の関係を加味したものではない



視覚情報を含んだオントロジーが作成できれば、
より詳細な画像検索などが可能になる

オントロジーとは



検索エンジン
機による
関連付け
メニュー

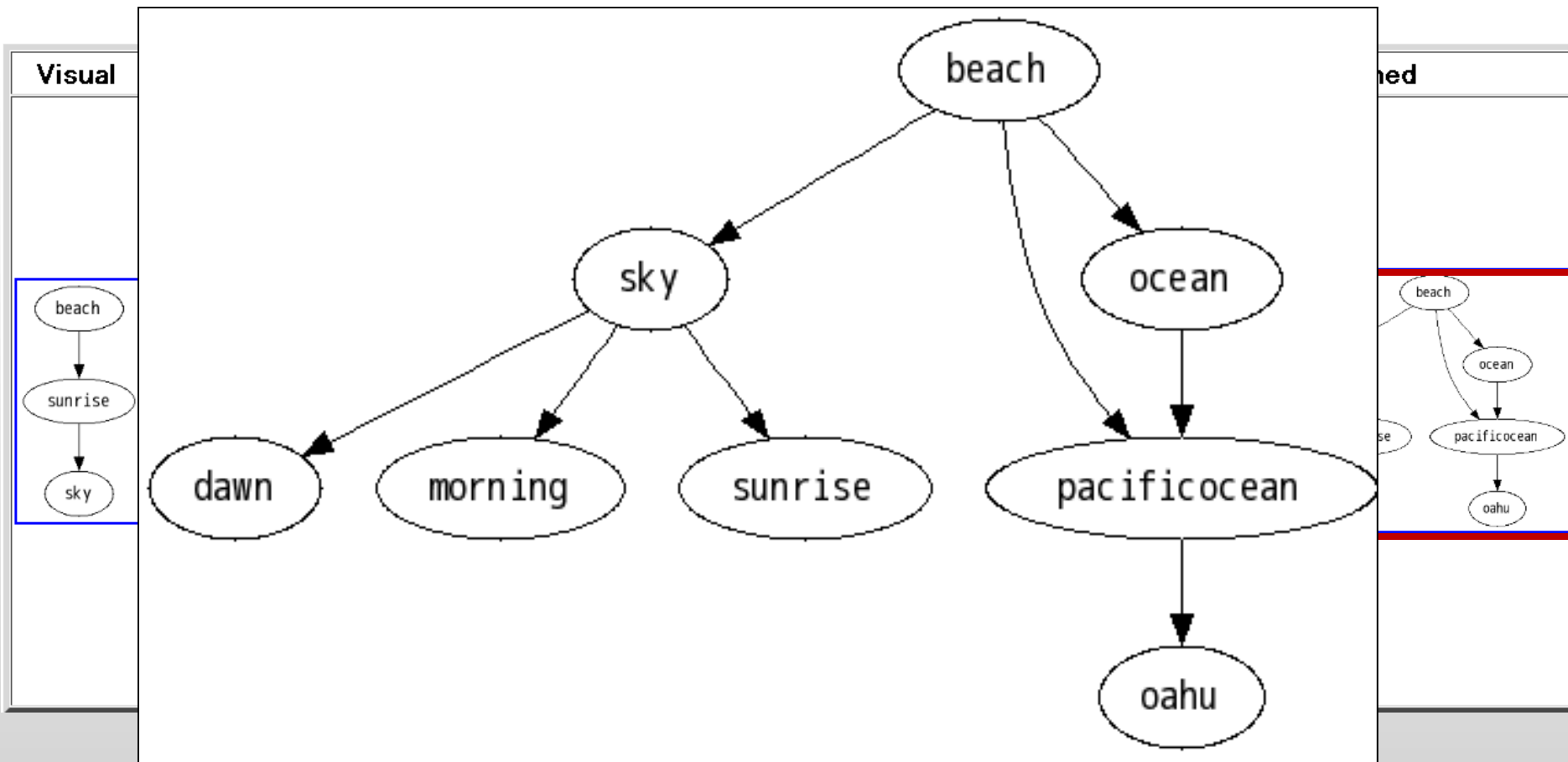
- 単なる画像検索から意味情報を考慮した検索へ
 - sandやbeachとしかタグが付いていない画像をseaやskyでも検索可能に

目的

■本研究の目的

- 階層構造を持った画像データベースの構築
 - 200万枚の画像からなる大規模画像データセットを用いる
 - 視覚による表現, タグによる表現, 視覚とタグを統合した表現の3種類で構築
 - 意味的にあるいは視覚的に分布の広い概念ほど上位の概念であると仮定して上位下位関係を推定

得られる階層構造の例



階層構造自動構築の利点

- 画像検索の補助として利用可能
 - 作成された構造から検索の興味を広げる
 - 検索による画像の絞り込み
 - 画像検索ワードの対応しない画像にもヒット
 - ヒット数の少ない検索ワードでも豊富な画像を表示
 - タグの付いていない画像の検索
- 画像認識や画像アノテーションへの利用
 - 構築する画像オントロジーは各概念に対応する画像の平均的な画像特徴量の分布情報を保持

関連研究

■ WordNet

- 概念間の意味関係を構造化したデータベース
- 上位下位関係、同義関係、反対関係などあり、検索が可能
- 画像検索システムに活用
- 固有名称の処理が難しい

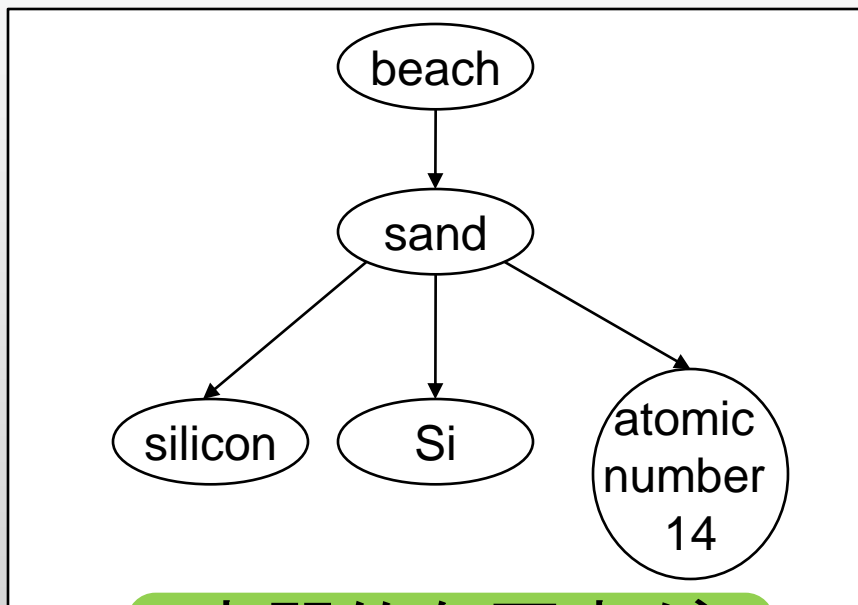
■ 視覚情報

- Tang et al. IJCAI 2007: 視覚情報による自動的に作成された階層構造をオントロジーとするような研究は存在しない
- Plangprasitthichok et al. ACM WWW Conf 2009: Flickrの既存の関係を利用

視覚情報による
自動的に作成された
階層構造を
オントロジーとするような
研究は存在しない

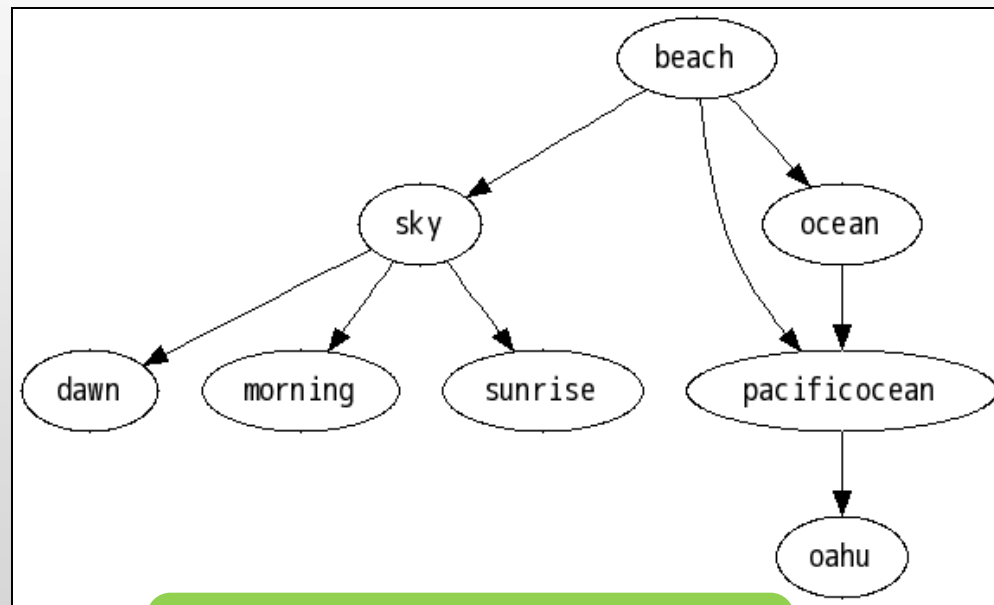
WordNetとの比較例

WordNetの階層の例



専門的な要素が強い

本実験で作成される階層の例



直観的な関係

階層構造構築の手順

大量画像取得と特徴抽出

- Flickrからおよそ200万枚のタグ付き画像を取得
- 画像特徴とタグ共起を用いた3種類の方法で画像を表現

ノイズ画像除去

- pLSAを用いてノイズ画像を除去

概念ベクトルの作成

- 各概念を表現するトピックベクトルを生成

概念間距離の測定

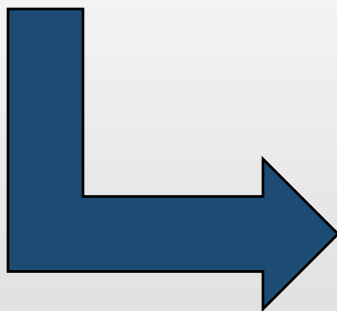
- 概念ベクトルから各概念間の距離を測定
- 距離によって関係する概念を取得

概念間階層構造の構築

- 概念ベクトルからエントロピーを算出
- エントロピーの大小で上下関係を決定

Flickr APIでの画像取得について

ランダムに
200万枚分の
画像情報を取得



タグ

- bird
- goose
- swanny
- Bali Bird
Park

取得できる画像例

タグ付き画像表現ベクトルの種類

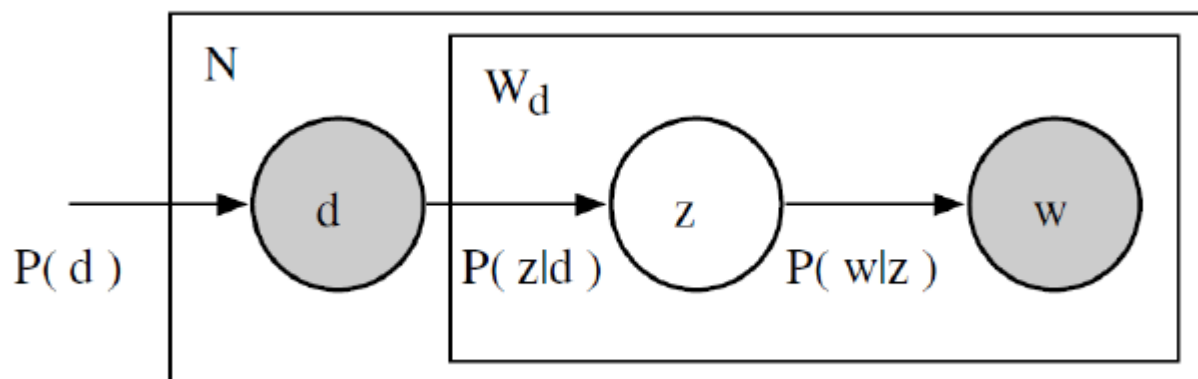
- 視覚特徴による表現
 - Bag-of-Keypoints表現
 - ベクトル次元：1000
- タグ共起による表現
 - Bag-of-Words表現(Bag-of-Tags表現)
 - ベクトル次元：4345
- 視覚特徴とタグ特徴を統合した表現
 - それぞれの表現のpLSAトピックベクトルの結合
 - ベクトル次元：200

probabilistic Latent Semantic Analysis

■ 確率的なクラスタリング手法

- Wordベクトル, Bag-of-Keypointsどちらも適応可能
- 確率的分類が最適になるような,
画像 d , word w , トピック z 間の関係を算出

$$P(d_i, w_j) = \sum_{k=1}^K P(d_i|z_k)P(w_j|z_k)P(z_k)$$



Fold-in heuristics

200万枚全てでpLSAは困難

大規模なデータセット

小分けにしたデータセット
データセット

特定のパラメータ
(全体パラメータ
と同一と仮定)

推定パラメータ

推定パラメータを
用いてpLSAの計算

タグ特徴と視覚特徴の統合

視覚特徴

タグ共起特徴



pLSA



pLSA

$$P_{visual}(z | d)$$

$$P_{tag}(z | d)$$

結合

統合特徴



ノイズ画像除去

$$P_{Concept}^{z_k} = \sum_{d \in Concept} P(d|z_k) / |Concept|$$

Concept : moon

PLCAICによる
トピック作成

$$P(Concept|z_k) = \frac{P_{Concept}^{z_k}}{P_{Concept}^{z_k} + P_{NotConcept}^{z_k}}$$

トピックのConceptらしさの
割合

Not moon dataset

$$P_{NotConcept}^{z_k} = \sum_{d \in NotConcept} P(d|z_k) / |NotConcept|$$

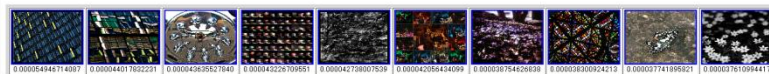
Topic No.31 0.0115582596



Topic No.32 0.0063095228



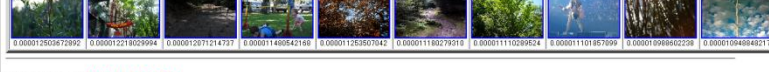
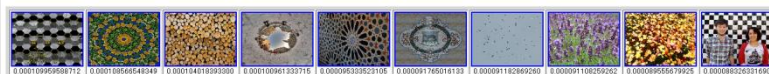
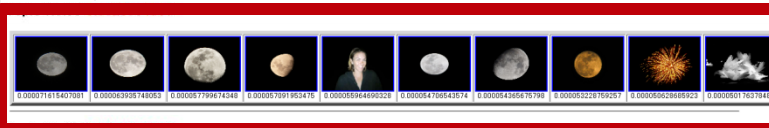
Topic No.33 0.0239038816



Topic No.34 0.0581119124



重みが大きくなる
ことが考えられる



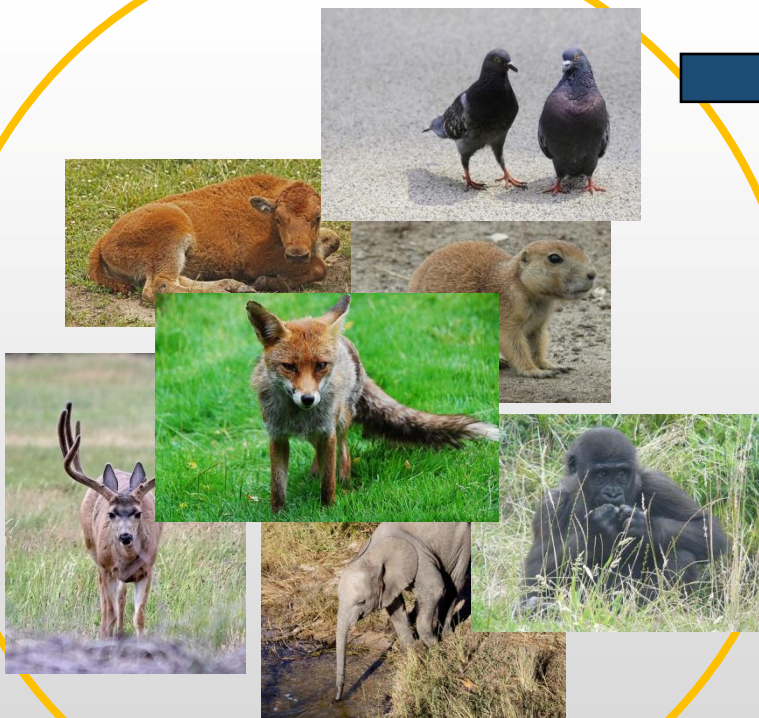
Topic No.39 0.0046782810

概念に所属しやすいトピック
で画像に重みづけ

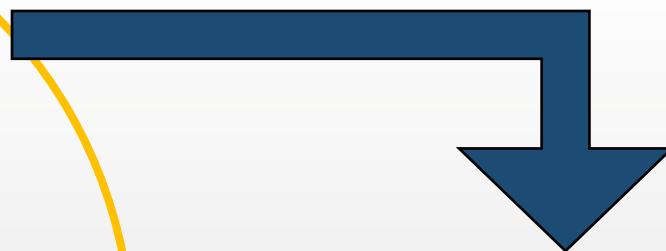
$$\sum_{k=1}^K P(Concept|z_k) P(z_k|d_i)$$

概念ベクトルの生成

$$P(z|Concept) = \left(\sum_{d \in Concept} p(z|d) \right) / (Concept \text{ の画像枚数})$$



Animal



Animal
概念ベクトル

概念間関係

■ 概念間距離

- JSダイバージェンスを利用

- P, Q はそれぞれの概念の確率分布(ベクトル表現)

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

$$D_{JS}(P||Q) = \frac{D_{KL}(P||(P/2 + Q/2))}{2} + \frac{D_{KL}(Q||(Q/2 + P/2))}{2}$$

- 距離が近い : 概念同士が密接に関係
- 距離が遠い : 概念同士の関係性がうすい

概念間関係

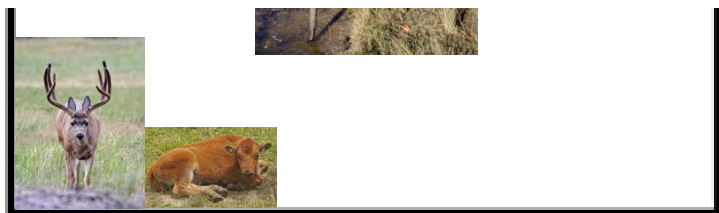
■概念間の上下関係

$$H(\text{Concept}') = - \sum_{z \in Z} P(z|\text{Concept}) \log(P(z|\text{Concept}))$$

上位概念(ばらつき大)

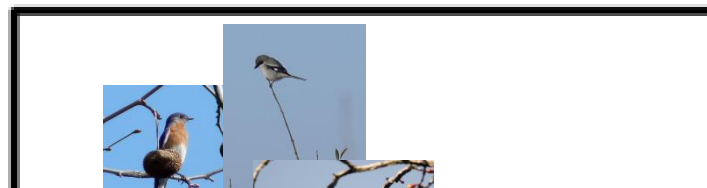


エントロピー大



Animal

下位概念(ばらつき小)



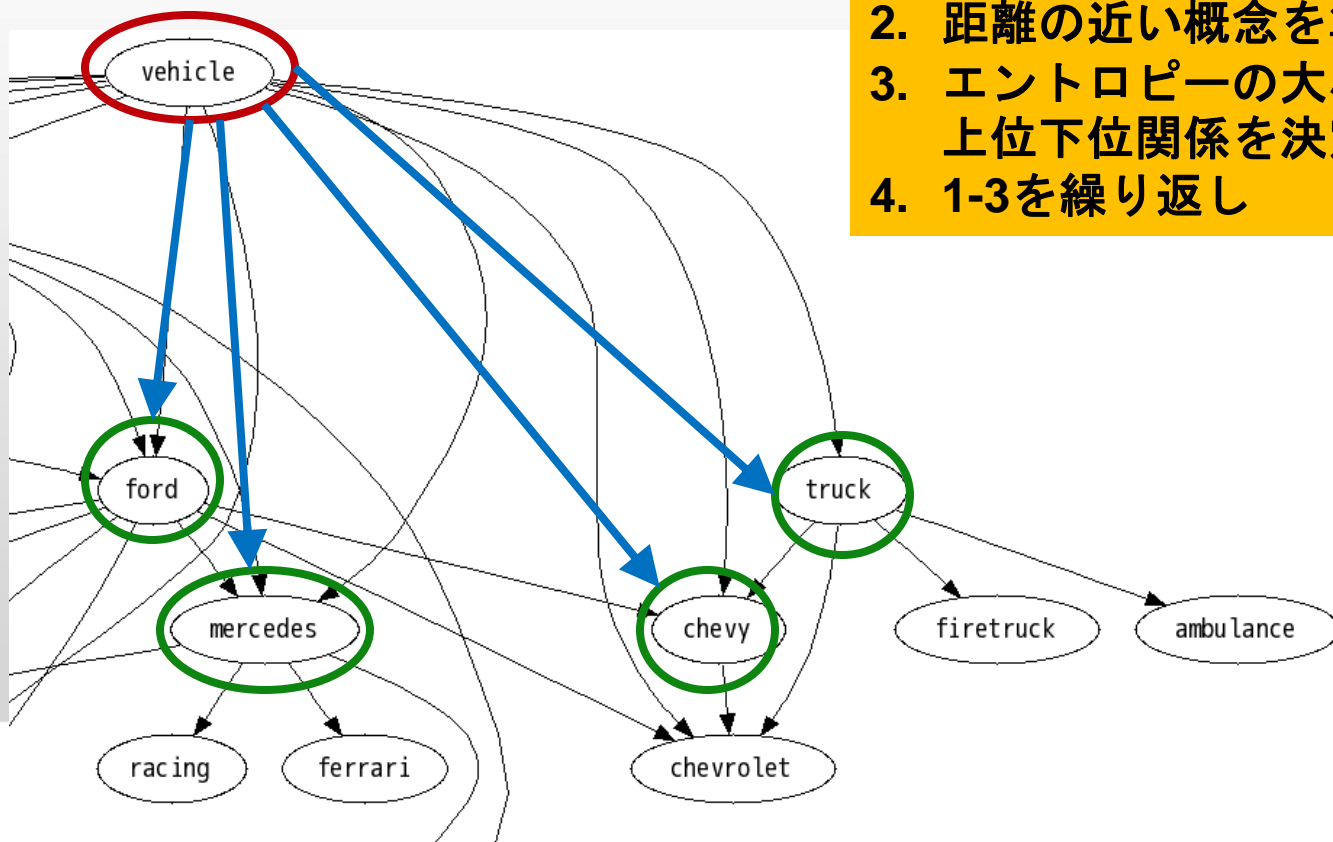
エントロピー小



Bird

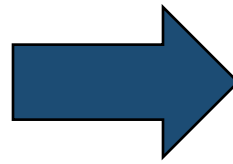
階層構造の構築方法

- 非巡回有向グラフ(Directed Acyclic Graph, DAG)による表現



1. 起点となる概念を決める
2. 距離の近い概念を算出する
3. エントロピーの大小によって上位下位関係を決定
4. 1-3を繰り返す

関係とは



road

本実験では
これらのそれぞれの区別は
行っていない
(今後の課題)

red

修飾詞



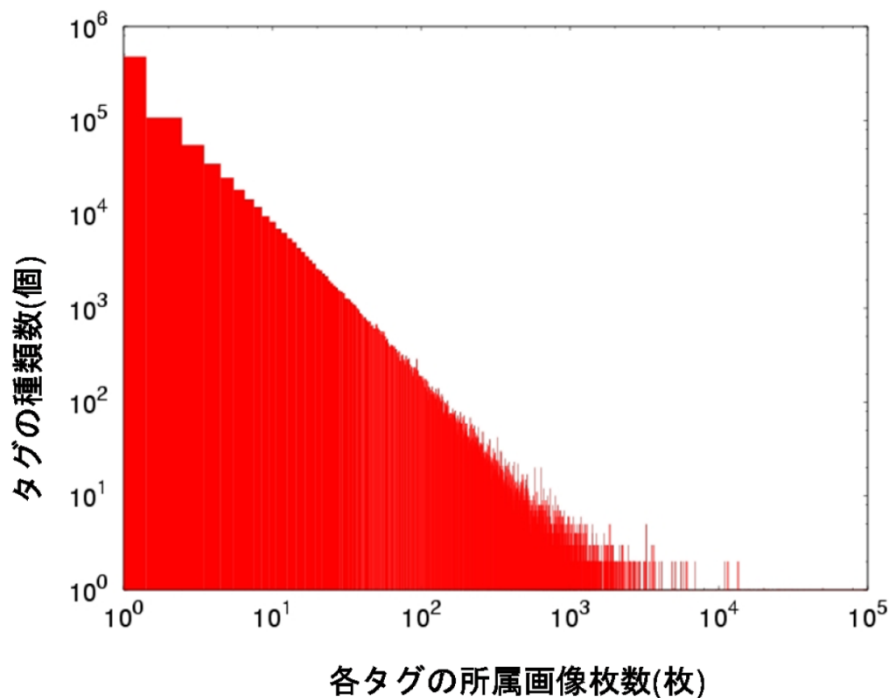
YANAI Lab.

実験について

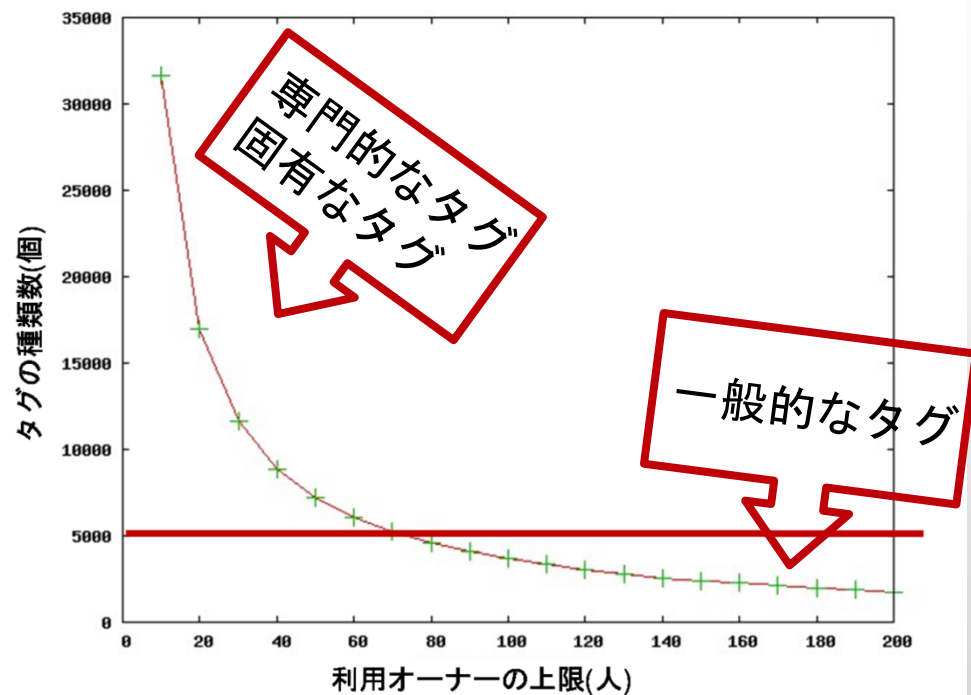
- Flickrによる大規模画像データセット
 - 合計およそ**200万枚**のタグ付き画像
 - 非巡回有向グラフを作成する概念は計2657個
- 各設定
 - ノイズ除去によって、各概念の上位n枚は
 $5 \times \sqrt{(\text{各概念の取得枚数})}$ を用いて計算する
 - これは画像枚数が少ない概念(200枚程度)でもおよそ100枚の画像を利用できるように設定

実験データセットの傾向

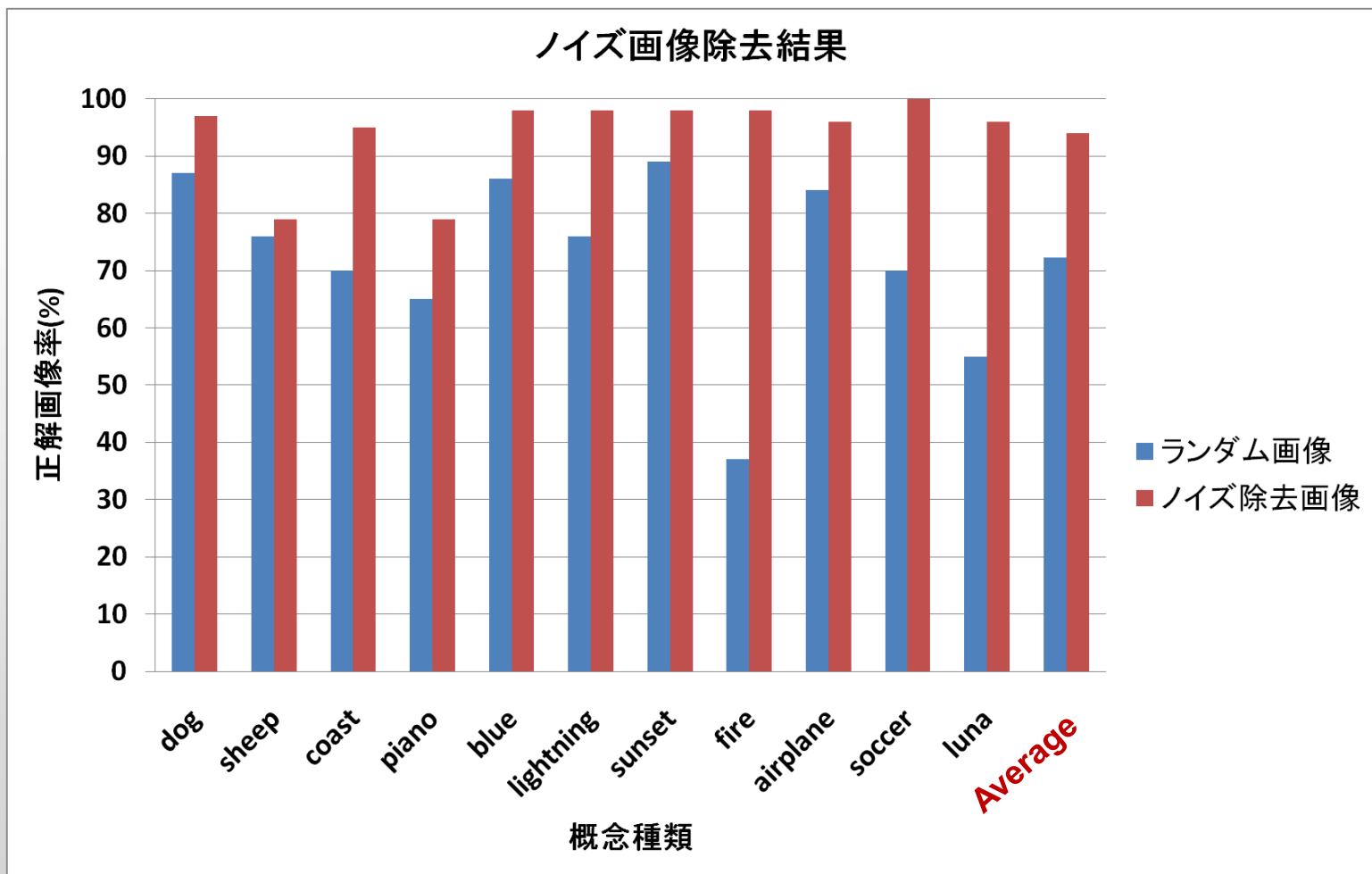
所属枚数ごとのタグ種類数



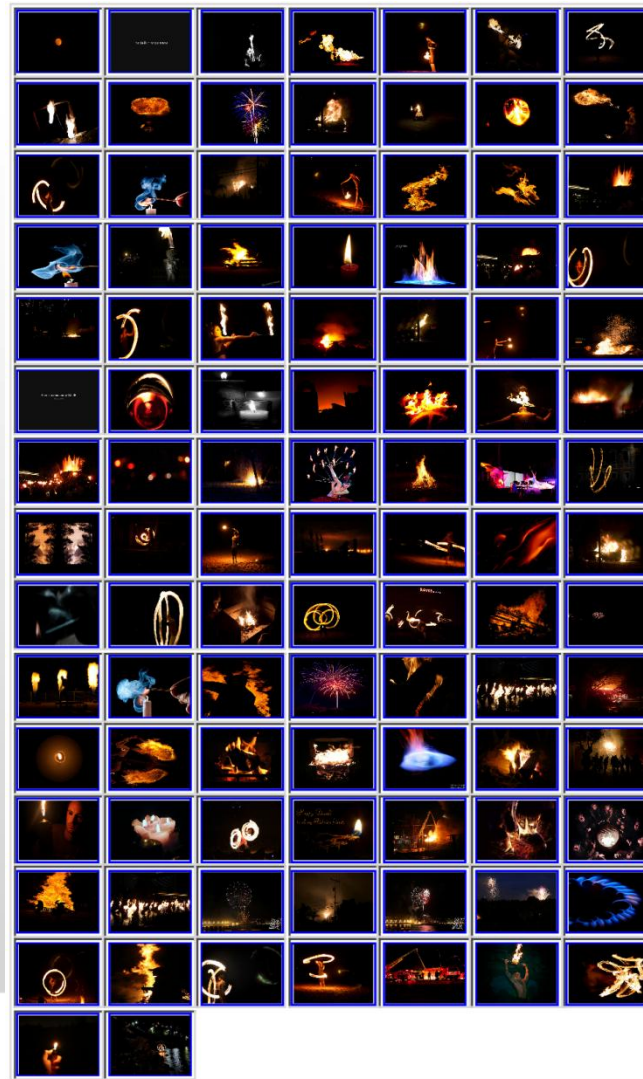
利用オーナー数とタグの種類数の関係



ノイズ画像除去の評価



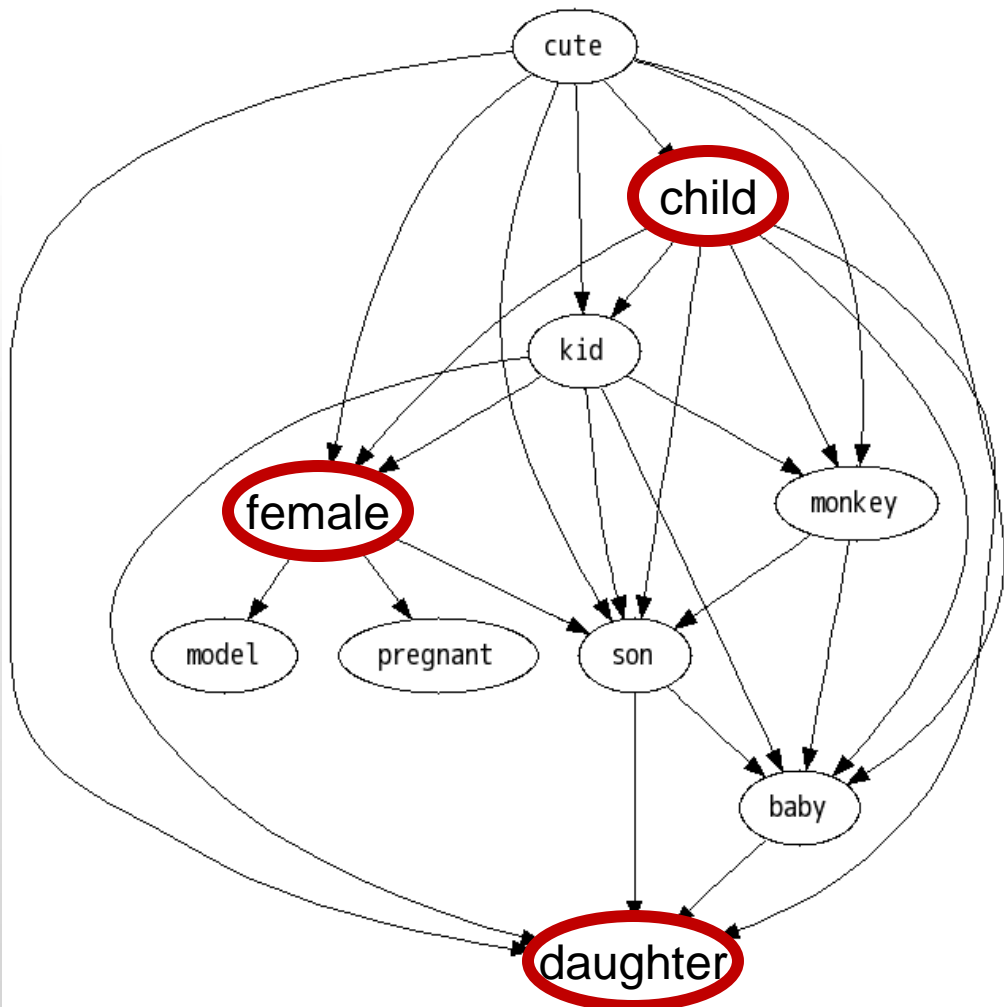
ノイズ画像除去例 (fire)



階層構造の例

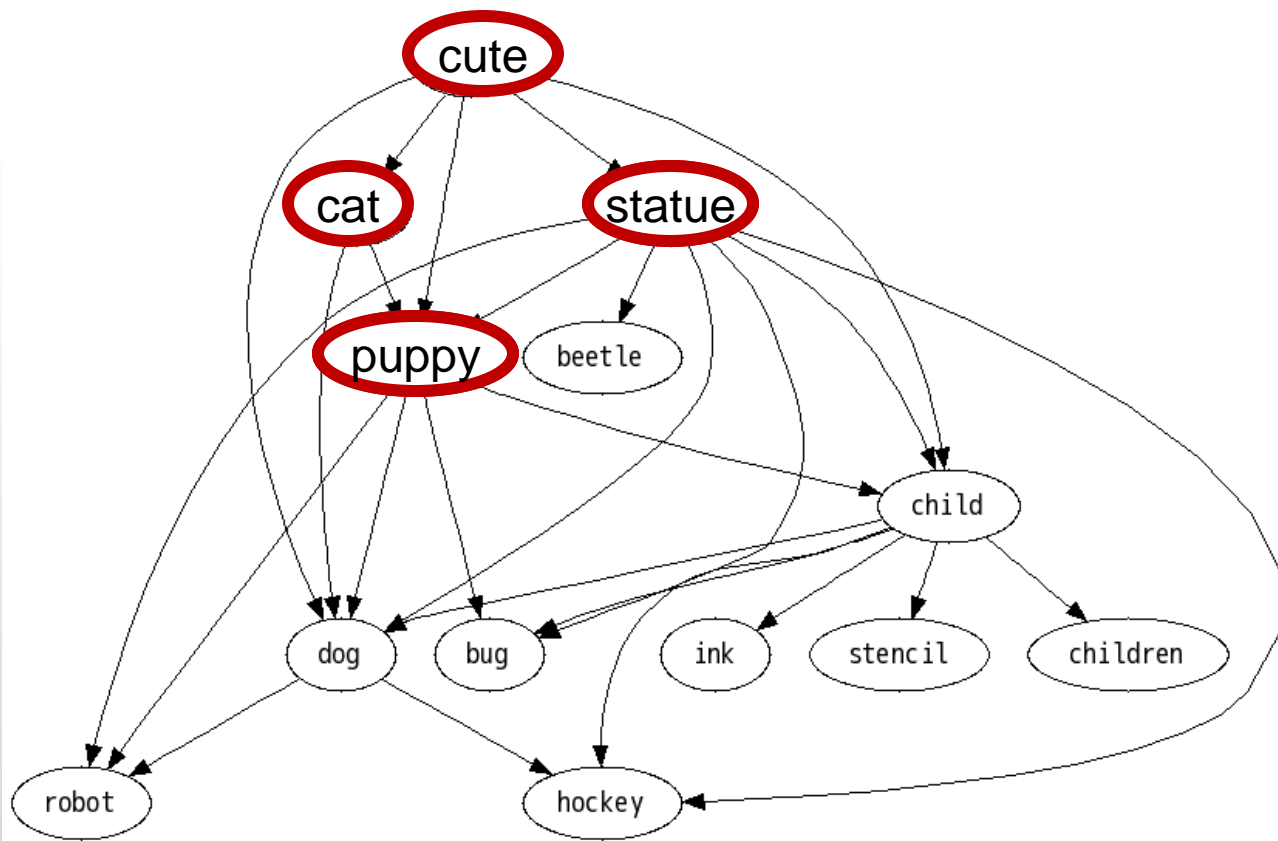
- 実際に作成した階層構造
 - http://mm.cs.uec.ac.jp/akima/flickr_dag/

cuteの下位構造比較



タグ表現による
階層構造

cuteの下位構造比較

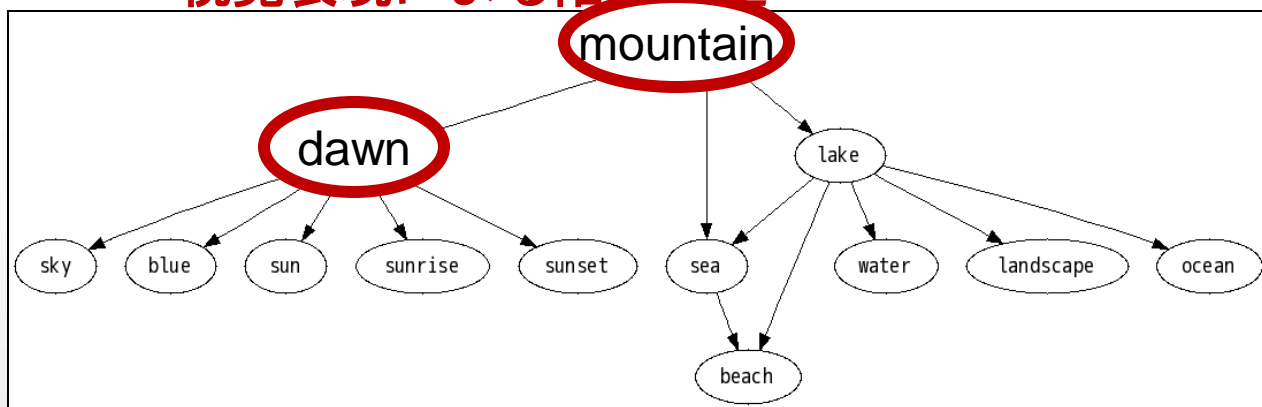


視覚表現による
階層構造

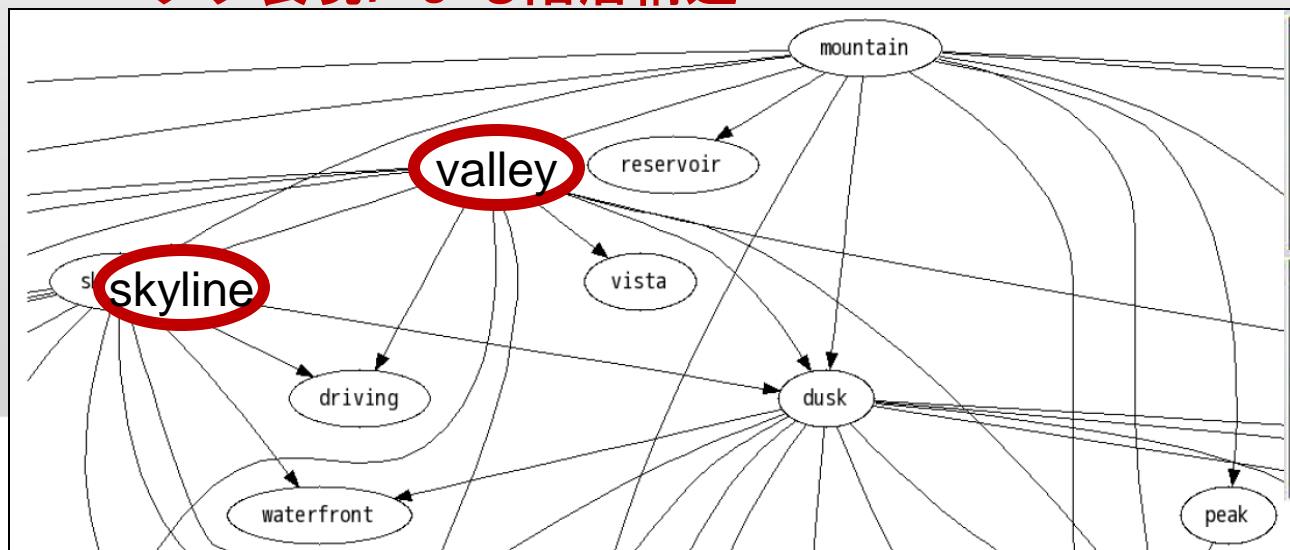


mountainの下位構造の比較

視覚表現による階層構造

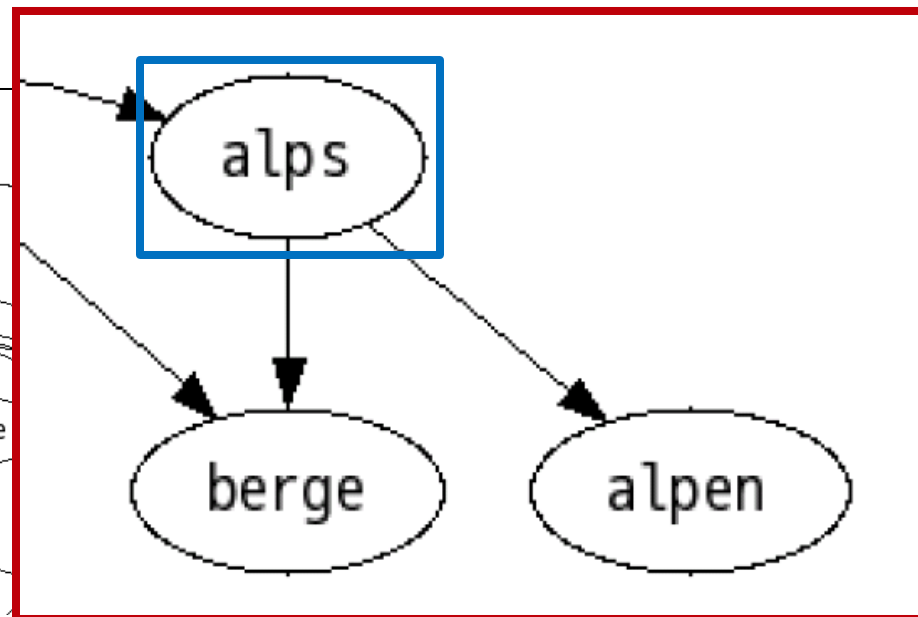
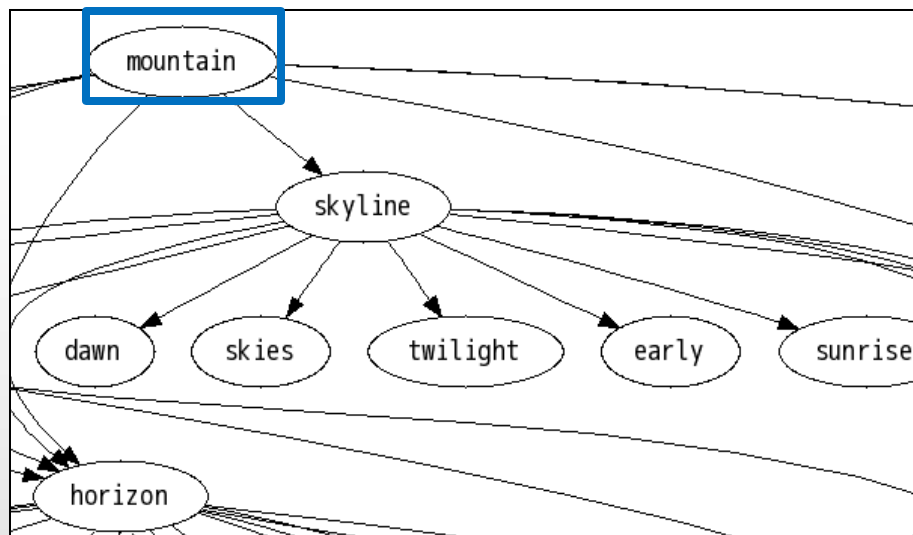


タグ表現による階層構造



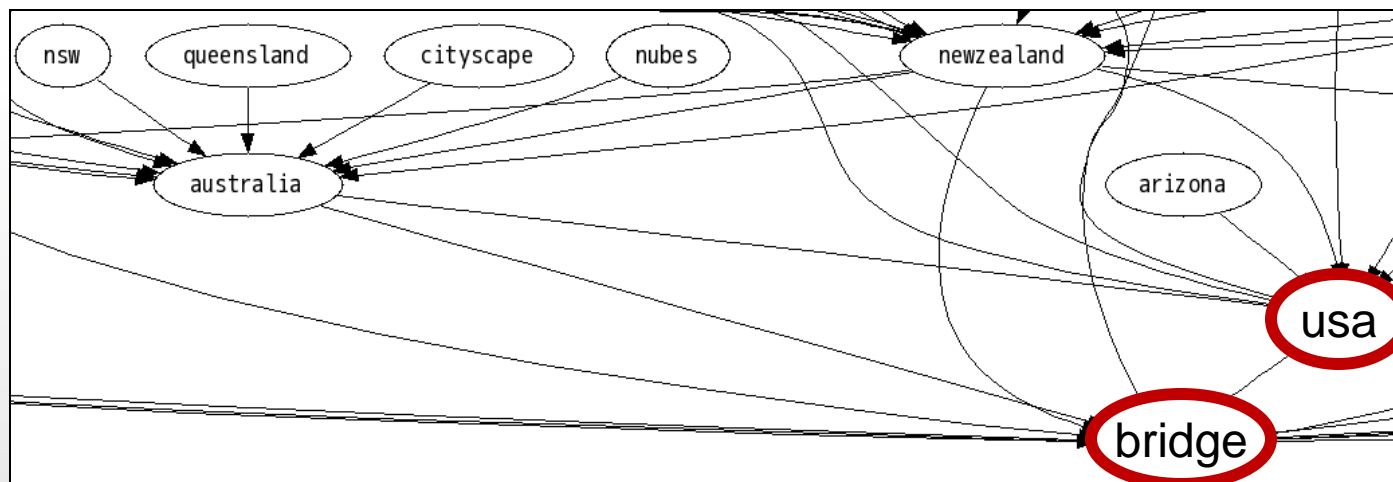
mountainの下位構造の比較

表現統合による階層構造



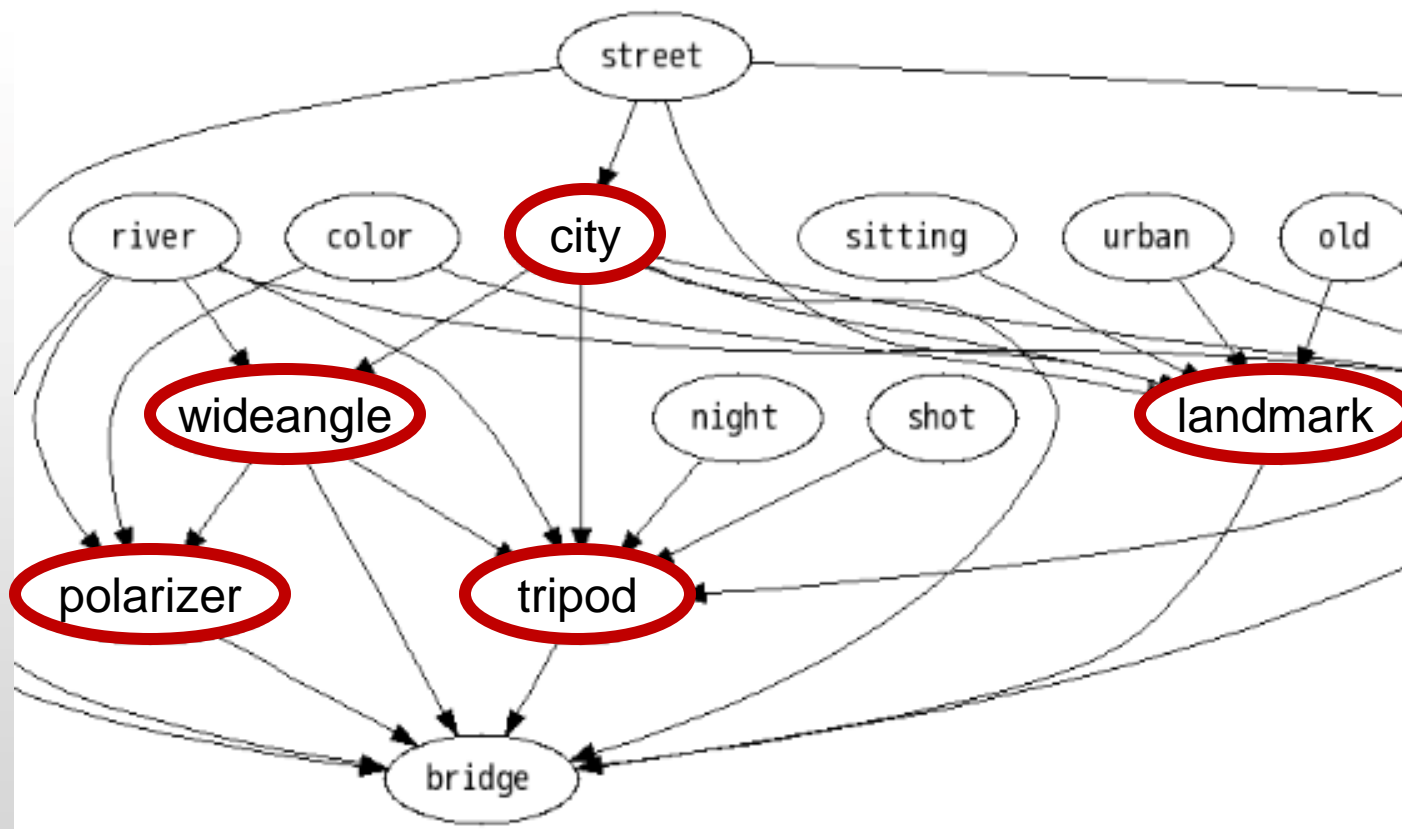
bridgeの上位構造の比較

視覚表現による階層構造



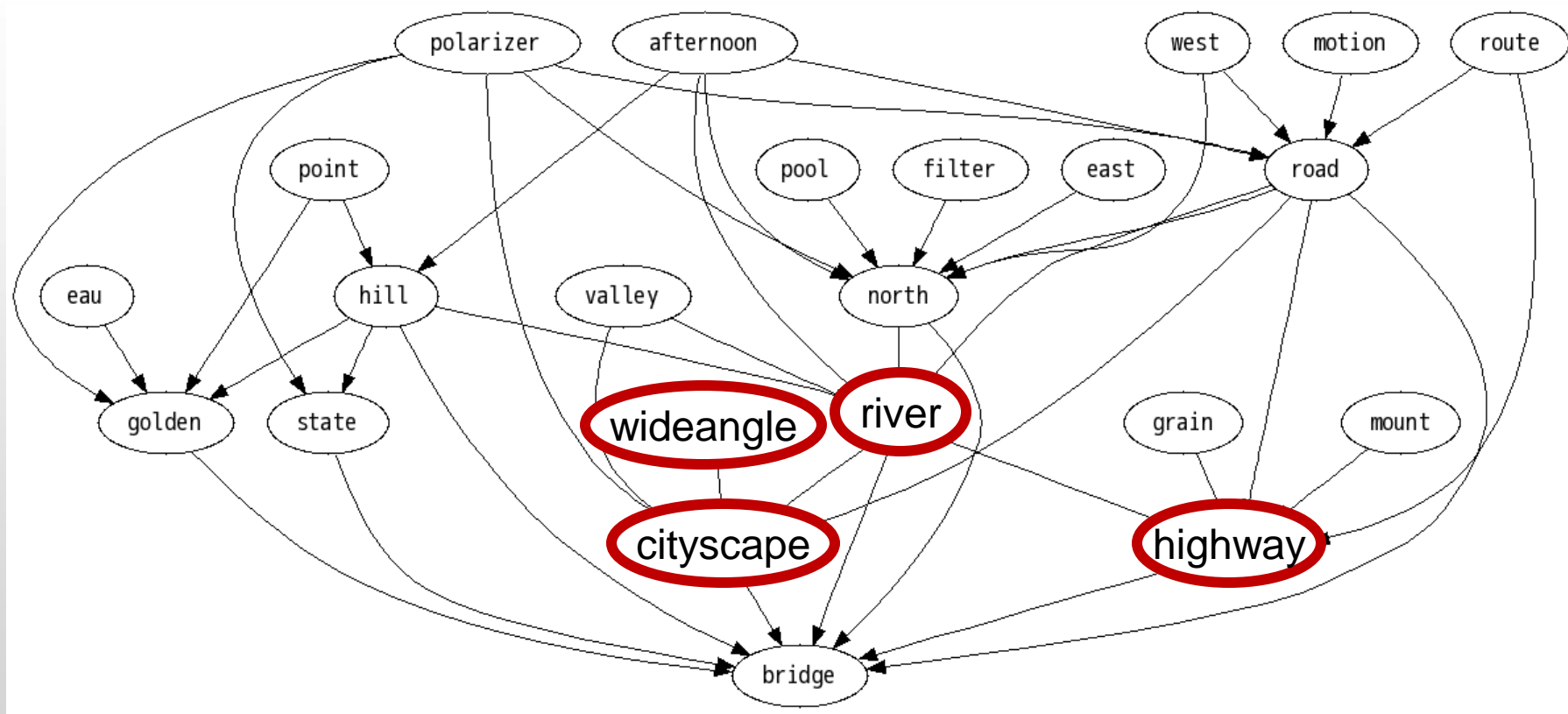
bridgeの上位構造の比較

タグ表現による階層構造



bridgeの上位構造の比較

統合表現による階層構造



階層構造比較の考察

■ 概念の表現方法ごとの違い

- **視覚表現**：視覚的に類似性のある関係性
 - データベース特有あるいは人間の知り得ない新しい関係性を抽出できる可能性がある
- **タグ表現**：意味的に関連する言葉としての関係に近い
 - 人間の認識に近い関係を抽出できる可能性がある
- **統合表現**：両方を足し合わせたような関係で画像検索に利用できる可能性が高い

まとめ

- 作成した階層構造に関して
 - 言葉の意味と視覚的類似性を考慮した関係を抽出
 - データセットの傾向の考察に有効
 - 表現方法の違いで多様な形を表す
 - 検索システムなどの応用の可能性を発見

今後の課題

- 現在は形状のみなので、色などその他の特徴も組み込む
- 概念のばらつきを維持したままのノイズ除去
- 関係性の種類の認識を行えるようにする
- 各概念の画像枚数をより豊富に
 - 現在はランダムに200万枚取得したため、枚数の少ない概念が存在
- 数値的な評価をおこなう
 - 画像検索の補助などに利用できれば可能



ご清聴ありがとうございました