

Semantic and Visual Image Clustering

Retrieving Search Term Related Pictures in Structured Clusters

Seminar paper

SEMANTIC MULTIMEDIA

Summer Term 2013

Hasso-Plattner-Institut für Softwaresystemtechnik GmbH

Universität Potsdam

written by

Mandy Roick
Claudia Exeler
Tino Junge
Nicolas Fricke

30. August 2013

Abstract

Abstract goes here.

Write it at the end.

Contents

1	Retrieving Images in Clusters	3
1.1	Problem Statement	3
1.2	Motivation	3
2	Related Work	4
2.1	Semantics	4
2.2	Visuals	4
2.3	Approaches to Combine Semantics and Visuals	4
3	Image Tree Based on Wordnet	5
3.1	Wordnet	5
3.2	Constructing a Searchtree	5
3.3	Synset Detection	5
3.4	Assigning Synsets to Pictures	5
4	Context-related and Visual Clustering	6
4.1	General Approach	6
4.2	Keyword Clusters	6
4.2.1	Keyword Clustering	6
4.2.2	Assigning Images to Keyword Clusters	6
4.3	Visual Clusters	6
4.3.1	Features	6
4.3.2	Clustering	6
5	Evaluation	7
5.1	Method	7
5.2	Results	7
6	Results Discussion	8
6.1	MCL-based Clusters	8
7	Future Work	9
A	Glossary	10
B	Abbreviations and Acronyms	11

1 Retrieving Images in Clusters

1.1 Problem Statement

What do we do?

clustering: creating homogeneous groups of semantically and visually similar pictures

Why do we do that?

seminar challenge: cluster 1M pictures of the MIR1M flickr file set

1.2 Motivation

2 Related Work

Related Subjects: Image Annotation, semantic clustering, content-based image retrieval

2.1 Semantic Clustering and Tags

2.2 Image Annotation and Content-Based Image Retrieval

3 Image Tree Based on Wordnet

3.1 Wordnet

Related to the official web page WordNet is described as a freely and publicly available "large lexical database of english nouns, verbs, adjectives and adverbs, grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept." (ref: <http://wordnet.princeton.edu/>). Synsets are semantically linked with each other e.g. hyponyms and meronyms relations.

network of synsets for discovering semantics between words

3.2 Constructing a Searchtree

two typical types of queries: more or less generic object descriptors, and places
actually construct multiple searchtrees if more than one synset found for a searchterm (i.e. train coach and motorized vehicle for car)

use hyponym relation to span tree of specializations (i.e. apple, banana for fruit; bus, sportscar for car)

if no hyponyms (usually the case for geographic terms), use part-meronyms

3.3 Synset Detection

for each tag and word in title, try to find synset (limiting ourselves to nouns, because they are usually the ones describing the depicted concepts). Other options: description (named entity recognition necessary), comments (noticed little relation to picture), group and album names (for both preprocessing needed to match any to wordnet)

problem: multiple possible synsets for a word, how to find correct meaning?

use best-first search (with limited queue for complexity reasons. idea is that paths at more than position x are unlikely to become best candidate anyways)

still erroneous with words that are meant in a way that is unknown to WordNet, i.e. canon as the camera model is interpreted as [definition of canon.n.01]

therefore preprocessing removes all tags that include numbers. Blacklist could filter even more but would also filter canon in its real sense, and generally not desirable to be flexible with respect to the tag vocabulary.

also removes special characters (more likely to be found on WordNet, and more likely to be identical with other unmatchable tags)

3.4 Assigning Synsets to Pictures

for higher recall: find strongly co-occurring tags that could not be mapped to synset

strong co-occurrence defined on tf-idf (else camera models would be strong co-occurrence with many synsets) observed that it is useful to find translations etc. but of course also introduces noise

take all pictures that are annotated with at least one of them

4 Context-related and Visual Clustering

Why do we do subclusters for nodes?

4.1 General Approach

4.2 Keyword Clusters

4.2.1 Keyword Clustering

MANDY

4.2.2 Assigning Images to Keyword Clusters

4.3 Visual Clusters

4.3.1 Features

color histogram, edge length and angle histogram

4.3.2 Clustering

k-means, late fusion by intersecting

5 Evaluation

5.1 Method

Picture annotation tool, 200 pictures

compare two pictures: for each, does it display food? semantically not similar / same object / same object and same context? visually similar / not similar?

receive testset from users

how are quality measures calculated?

search food: precision / recall of picture inclusion (compare synset detection mechanisms?)

evaluate tree nodes based on same object annotations

evaluate mcl clusters based on same object and on same context annotations (compare both, what does mcl actually do?)

evaluate visuals with large minimal node size

vary parameters given by frontend

5.2 Results

6 Results Discussion

Are our results good? Are they biased by something?

All depends on annotations - inappropriate tagging leads to bad results, as well as limitation to nouns (adjectives, adverbs and verbs are wrongly matched nouns)

6.1 MCL-based Clusters

highly depend on quality of keyword clusters

7 Future Work

How to improve, what other approaches to take

A Glossary

Synset: Ein spezielles Programm, mit dem man über das WWW Zugang zu WWW-Servern erlangen und von diesem angeforderte Dokumente anzeigen kann.

Wordnet: Bezeichnet ein Programm, dass einen Server kontaktiert und von diesem Informationen anfordert. Der im WWW eingesetzte Browser ist in diesem Sinne ein Client. Aber es gibt auch andere Clients im WWW, die WWW-Server kontaktieren und Informationen von diesen herunterladen, wie z.B. Suchmaschinen oder Agenten.

HTML: Hypertext Markup Language; das einheitliche Dokumentenformat für Hypermedia-Dokumente im WWW. Dokumente, die im WWW übertragen und vom Browser dargestellt werden sollen, sind in HTML kodiert.

HTTP: Hypertext Transfer Protocol; das Protokoll, das die Kommunikation von Browsern und WWW-Servern im WWW regelt. Fordert ein Browser ein Dokument vom WWW-Server an oder beantwortet der WWW-Server eine Anfrage, muss diese Anfrage den Konventionen des HTTP-Protokolls gehorchen.

Netzanwendung: Ein Anwendungsprogramm, dessen Ablauf den Zugriff auf Ressourcen einschließt, die nicht lokal auf dem ausführenden Rechner liegen, sondern auf einem entfernten Rechner über das Netzwerk zugegriffen werden.

Server: Bezeichnet einen Prozess, der von Clients kontaktiert wird, um diesen Informationen zurück zu liefern. Oft wird auch der Rechner, auf dem ein Server-Prozess abläuft, als Server bezeichnet.

B Abbreviations and Acronyms

4CIF	4 fach Common Intermediate Format
AAC	Advanced Audio Coding
AAL	ATM Adaption Layer
ABR	Available Bit Rate
AC	Audio Code
ACK	Acknowledgement
ADM	Add Drop Multiplexer
ADSL	Asymmetric Digital Subscriber Line
AH	Authentication Header
AIFF	Audio Interchange File Format
AM	Amplituden-Modulation
ANSI	American National Standards Institute
API	Application Programming Interface
ARP	Address Resolution Protocol
W3C	World Wide Web Community
WWW	World Wide Web

Index

Akronyme, 11

HTML, 10

HTTP, 10

Netzanwendungen, 10

Prozess, 10

Rechner, 10

Server, 10

Synset, 10

Wordnet, 10