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

Todo list

1 Retrieving Images in Clusters

1.1 Problem Statement & Motivation

training data for image categorization and content detection
flickr and other online photo communities are good sources for annotated images
problems: low annotation quality, only search for specific term (with different meanings and visual characteristics)
for example, want to test the quality of my algorithm and identifying different foods: would have to think of all kinds of food and search and filter images manually

What do we do?
clustering: creating homogeneous groups of semantically and visually similar pictures

Why do we do that?
seminar challenge: cluster 1 million pictures of the MIR1M flickr file set
improving the complex task of searching for pictures according to a given keyword
facing different challenges like: multiple meanings of the keyword, bad picture annotations, taking semantic and visual information of a picture into account

1.2 Clustered Tree Nodes Approach

idea: provide ready-to-use semantically and visually homogeneous image clusters for a given topic. Span tree of subordinate pictures, retrieve related images and cluster them to distinguish different settings of the pictures and to identify outliers.

After giving an overview of Related Work in chapter 2, we will present our methods for retrieving (chapter 3) and clustering (chapter 4) appropriate images. Chapter 5 explains how we evaluate our approach, while the evaluation results will be discussed in chapter 6. At last, chapter 7 gives ideas for improvement and possible future work.

2 Related Work

Much research been done recently in image clustering and semantic clustering, with application areas in image segmentation, compact representation of large image sets, search space reduction and avoiding the semantic gap in content based image retrieval ($http://www.cscjournals.org/csc/manuscript/Journals/IJIP/Finalversion/Camera_ready_IJIP-304.pdf$)
However, most of this work presents new algorithms for one of the above use cases, not methods to retrieve training data
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

Ideas exist to use visual features to semantically classify

2.3 Approaches to Combine Semantics and Visuals

3 Image Tree Based on Wordnet

3.1 Wordnet

Related to the offical 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). Further source 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) problem: multiple possible synsets for a word, how to find correct meaning?
use best-first search with limited queue, distances are based on Leacock and Chodorows Normalized Path Length (lch similarities, which is perceived as closer to human understanding than regular path similarity [BH01])

3.4 Assigning Pictures to Tree Nodes

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 the related tags or the synset itself.

4 Context-related and Visual Clustering

nodes are generally large, somewhat semantically homogeneous but very visually diverse. Therefore: create finer clusters within each node.

4.1 General Approach

clustering all images visually is expensive
semantics more important for humans
idea: create clusters with semantically similar pictures and build subclusters within with visually similar pictures

4.2 Keyword Clusters

good for context(?), outlier identification, basic clustering for part-meronym spanned trees (Africa example)

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
easy to calculate, obvious / humanly comprehensible, no need for complex features since only refinement of semantics (not trying to distinguish concepts by visual features)
visual clustering as a refinement for the semantic clustering, therefore basic visual features seem to be sufficient

Can we explain or prove that somehow?

4.3.2 Clustering

k-means separately for colors and edges with k chosen by rule of thumb
feature vectors are pyramidal slices (up to 5) 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

References

[BH01] Budanitsky, Alexander ; Hirst, Graeme: Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures. In: *IN WORKSHOP ON WORDNET AND OTHER LEXICAL RESOURCES, SECOND MEETING OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS*, 2001