Tutorial on searching text and images in the medical domain

Henning Müller
MIE 2012, Pisa, Italy

Overview

• Improving search in the medical domain
  • Allan and Henning
• Searching for medical images
• Who searches for medical images and how?
• Combining text and visual search
• Challenges for search
  • Allan and Henning
Improving search in the medical domain

Scenario

- A close friend of yours has her 5-year-old daughter diagnosed with Leukemia
- As you are regarded as a medical professional in the largest sense they ask you to help find more information on the disease
  - How do you search for this information, what are the strategies?
  - What type of information are you targeting?
  - How can you assure that trustful information is being transferred?
  - How do you explain it to the 5-year-old?
  - What are difficulties and disadvantages of this approach?
Search target?

- Is a document searched for?
- Or an answer to a specific question?
- Maybe an expert in the domain?
- Educational material?
  - Maybe videos, or nice pictures also for children?
- Description of the disease vs. symptoms, treatments, chances of survival?
  - With the goal of making a choice of treatment, for example, comparing various options for the choice

Trustfulness

- How can we make sure that information is correct, or at least not totally wrong?
  - Sometimes differing opinions exist, knowledge changes over time
  - Cross check several sources, but this can also lead to wrong ways
  - Quality labels or certificates of trust can help
    - HONcode
- Classification of web pages into several classes
  - What are the characteristics
  - Success needs to be monitored closely (spam filtering)
Level of understanding

• Different target groups have a totally different level of information that is required
  • Also physicians between a specialist and a GP
• What about children and the elderly?
  • Level of understanding changes the more we read about a specific topic, we become experts
  • Not a fixed thing …
• Users can give more information to get personalized results
• Cyberchondria

Document and page ranking

• Should wikis and blogs be rated highest?
  • Google does this but for medical information several studies show that this might not be the best strategy
• Most queries are short (1-2 terms) and thus not specific at all, how to deal with this?
• Is the search target known?
  • Users want advanced search options but then these are rarely used in practice
• Go beyond ranking, explore the content space
Searching for medical images

Motivation

- Medical imaging is estimated to occupy 30% of world storage capacity in 2010!
- Mammography data in the US in 2009 amounts to 2.5 Petabytes
### Retrieval of images

- **Text retrieval of images**
  - Is there any text attached to the images?
  - Doing this manually is expensive, subjective, **language** dependent, …
  - Take text close to the images (such as captions)
  - Semantic **concepts** could help in some cases

- **Visual retrieval of images**
  - Using automatically extracted visual features
  - Content-based image retrieval (**CBIR**)
  - Query by Image Example(s) (**QBE**)
  - **Multimodal retrieval** (**text+images**)
Content vs. context

- Most often text around images does not describe the image content itself
  - Unless specifically annotated for retrieval
  - Text often gives the context in which the images were taken (private, also medical)
- Image content is rarely described precisely and completely with text
  - Visual features describe the pure content
  - Low level of semantics
- Content and context are complementary for search

Age matters
Types of information needs/searches

- Known-item search (i.e. telephone number, …)
  - Question answering
- Exploration, exploitation, informational
  - Topic search or open ended search
- Comparison search (between things)
- Expert search, person search, entity search
- Geographical search
- Literature search
- Multimedia search (increasingly given as results)
- Browsing, no specific goal

Types of image searches

- Recent (time) pictures (journalists)
  - Date, given with camera
- Pictures of specific places, monuments
  - GPS in many cameras
- Pictures with particular persons (private search)
  - Face detection, recognition (Facebook, Picasa)
- Pictures with particular objects, types of images
- Pictures evoking specific feelings
  - Fear, joy, happiness, …
- Similarity search/browsing (Medicine, journalism)
Visual information for retrieval

- **Object detection**
  - Then potentially mapping the objects to an ontology
  - Usually works well for a small number of objects

- **Image classification**
  - Training data, limited set of classes
  - *Global* classification of images vs. *local* classification of pixels, regions

- **Similarity retrieval of images**
  - Global image information, regions of interest (ROIs), small
  - No training data, relevance as criterion for quality
Visual classification and retrieval

http://www.youtube.com/watch?v=cMoONCqI+2c

Components of image retrieval systems
Model for CBIR

<table>
<thead>
<tr>
<th>Feature</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature 1</td>
<td>0.5</td>
</tr>
<tr>
<td>Feature 2</td>
<td>0.7</td>
</tr>
<tr>
<td>Feature 15</td>
<td>0.1</td>
</tr>
<tr>
<td>Feature 25</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Steps of a retrieval system

- Image pre-processing
  - Normalization, background removal, ...
- *Salient region* or point detection
- Visual feature extraction
  - Feature selection
  - Then feature modeling in the case of visual words
- Distance calculation or similarity measurement
- Results fusion (text, visual, ...) for ranked list
- Results filtering (i.e. by modality)
- …
A typical (old) interface

Classifying visual features (Eakins)

- **Level 1**: primitive features
  - Color, texture, shape, spatial organization

- **Level 2**: derived features
  - Individual objects or persons (Eiffel tower, Britney Spears)
  - Objects of a specific type (Volkswagen car)

- **Level 3**: abstract attributes
  - High level reasoning about meaning and purpose
  - Emotional or religious significance
  - Find images of “suffering”
Visual words

Salient regions
All pixels, grid, high gradient
Original feature space
• N-dimensional feature vector for each salient region

Quantization of the feature space
• Division of the feature space into M groups: visual words
• Clustering (k-means)
• Cluster centers are words

Visual words space
• Optimal number of words needs to be found
• For each image a histogram can be created
• Analogy to text words

Bag of Visual Words
• Histogram of words for an image or an image region based on the salient points
• M-Dimensional feature vector

www.springerimages.com (3.3 mio)
Goldminer.arrs.org (249,000 images)

medgift.hevs.ch/, demos (300,000 images)
Who searches for medical images and how?
Background

- Health professionals’ image search behavior has been subject to several surveys and log file analyses (MedLine, HONmedia, Goldminer)
- Goldminer log files of over 200'000 searches most comprehensive so far
- Khresmoi project performed a survey among radiologists to develop first prototypes
- Results change over time as user requirements change with the generation of physicians
- Current medical students grow up with Google, Facebook and iPhones

Methods of Khresmoi survey

- Survey among radiologists, mainly in Geneva and Vienna University hospitals
  - Paper and electronic version
  - Survey filled in with a person explaining the goals
  - Survey took approximately 1 hour, research/clinical /teaching work separated with same questions
- 26 radiologists answered, 13 from Austria, 9 from Switzerland, form only sent on invitation
- 17 males, 9 females, mainly 26-35 years, few years of experience
Reasons for image search

- Clinicians: Information on unclear cases, illustrate presentations, differential diagnosis were most frequent
- Teaching: Find similar cases, for example an easy, a medium and a tricky case for the same disease, problem-based learning
- Problem-based learning requires increased search skills for the students

Where people search

- The Internet replaces books and colleagues
- Personal files are not always optimal
Determination of relevance

- Experience determines relevance
- Often additional proof such as biopsies or other clinical tests are requested

Success rates

- For teaching success rates are higher
- Clinical work might have less well defined tasks, average of success at 60%
Search time

- 70% of successful searches less than ten minutes
- Failure often after over 15 minutes
- Less time available for clinical search than research/teaching
- Could a **faster and more targeted** image search system help?

Useful additions

- Search by
  - pathology (13 times)
  - modality (10 times)
  - patient demography (6 times)
  - similar images (8 times)
- Other comments:
  - Multilingual retrieval
  - Pathology classification (using **ontologies**)
  - Search in 3D data
  - Confidence in the diagnosis
A perfect search system

- Many free comments
  - Like Google but DICOM and text
  - Structure information and confidence in diagnosis
- Search by regions of interest
- Social networking, comments of other physicians
- Search for similar cases
- Quantification of structures (size, volume)

Goldminer log files

- Monday 12h15, talk on the analysis
- 25'000 consecutive queries
Combining text and visual search

Background

- ImageCLEF image retrieval benchmark has been run each year since 2004
- 12-20 research groups compare their tools and approaches on the same tasks and DBs
  - Visual, textual and combined approaches are used
  - Multilingual approaches are also possible
- Sometimes visual, sometimes textual and sometimes mixed approaches perform best
  - No clear outcome
  - Combination of results can be delicate, unstable
Example search topic

- Show me x-ray images of a tibia with a fracture.
- Zeige mir Röntgenbilder einer gebrochenen Tibia.
- Montre-moi des radiographies du tibia avec fracture.

Visual vs. semantic vs. mixed searches

- Experts can predict what type of technique will most likely perform best
  - Can this prediction be modeled automatically?
  - If results were visually relatively homogeneous visual search can work
    - Same anatomic region, same view, same modality
  - If results are expected to be very different (no modality given), text would work best
  - **Combinations** often work best when some common aspects but some variability as same modality
- Visual and text retrieval are complementary
Early vs. late fusion

- **Early fusion**
  - Feature spaces are directly combined, so visual features and textual words treated in the same way
  - Number of features needs to be similar to avoid bias

- **Late fusion**
  - Results of systems are combined, not features
  - Each system can have a varying number of features
  - For text/visual combinations late fusion is often simpler to employ and works better
  - When using visual words both could be used

Score based vs. rank-based fusion

- **Score-based fusion**
  - Score of the single systems is used for the combination of the results sets
  - Score needs to be normalized, potentially to have similar characteristics

- **Rank-based fusion**
  - The rank of an element is used to calculate fusion
  - Can be linear or logarithmic or in another form
  - Avoids the bias that very differing results sets of system can have
  - Often have better results when visual and textual systems are combined
Types of fusion techniques

- Many types of fusion techniques exist
  - \( \text{combSUM} : V_{\text{combSUM}}(i) = \sum_k V_k(i) \)
  - \( \text{combMAX} : V_{\text{combMAX}}(i) = \text{argmax}(V_k(i)) \)
  - \( \text{combMNZ} : V_{\text{combMNZ}}(i) = f(i) \cdot V_{\text{combSUM}}(i) \)
    - Where \( f(i) \) is the frequency of image \( i \) in the results

- At the ICPR 2010 conference a competition on fusion techniques was organized using the best ImageCLEF runs
  - Rank-based techniques using logarithmic decrease performed best in a variety of different approaches

Distribution of relevant documents

- Visual: \( y = -24.011 \ln(x) + 107.93 \)
  - \( R^2 = 0.9978 \)
- Textual: \( y = -6.6563 \ln(x) + 29.508 \)
  - \( R^2 = 0.7295 \)
### Other combinations

- Modality detection (using visual techniques or text of the captions) can work very well (80+%)
- Allowing to select the **target modality** can improve image search
  - Tests with all runs of ImageCLEF 2009
  - Many search engines allow for this such as Goldminer
    - This can be used for tabbed browsing as well
- **Exclusion criteria** for images can be chosen based on the text
  - Age group, gender, …

### Combinations for *case-based* retrieval

- **Mix** of free text, structured data, images, and many other forms
- **Interactions** of the data can vary strongly between patients and diseases, also over time
- More **complex combinations** for images need to be found
  - Match images between case to match for similarity
- Currently text is better than most fusions
- Case description including images without diagnosis, find images for differential diagnosis
Some ImageCLEF lessons learned

- **Text** retrieval techniques are **stable** and deliver **good** results (i.e. Lucene is above average)
- Visual has had less evolution than text retrieval
  - GIF (old!) has still relatively good results
    - Semantic gap is very present
  - **Visual words**-based approaches can be much better when using high quality training data
- **Interactive** retrieval can improve visual retrieval
- Many features combined deliver best results
- Mapping of images and text to ontologies can help
  - Improve **semantic** retrieval

Attention: advertisement

- ImageCLEF book
- All on **image retrieval**
  - Methods of evaluation
  - Task overviews
  - Participant reports
    - The best techniques
    - Industrial requirements
  - Industry perspectives
  - Specific techniques such as **fusion**
Challenges for search?

BIG data

- How can scalability be assured when treating extremely large amounts data?
  - 250’000 images per day in Geneva …
  - 150 TB of images in Catalonia archive
- Extremely large scales allow solving many new problems
  - Rare diseases
  - Sufficiently large training data sets
- Hadoop/Mapreduce as is also used by Google, Yahoo or Facebook
- Use of cloud computing
  - Costs, confidentiality, also bandwidth …
Confidentiality

• Can patient records be made available?
  • Maybe partly, anonymized, only internally?
  • Could the data warehouse be used for this
  • Secondary use of data

• Availability for data mining not a specific very limited scenario (as ethics committees request)

• Can interoperability be assured using the same semantic standards

• How to link the literature with specific cases
  • Images not in the same quality, much more than just case descriptions, …

Search for medical cases not images

• Combine several data sources
  • Importance of each source is not fully clear
    • Interaction between content importance is complex
  • Different media from free text, structured data to semantics, signals or images including 3D, 4D

• Some data sources might be missing
  • Questions not asked, not documented, errors
  • How to deal with missing data

• What is a case exactly?
  • Limited period of time? Or on a patient basis including old data?
Diversity in the results

- Having a list of almost identical texts as result is not useful
  - Google filters out near duplicates
- **Consistency vs. diversity** have limits
- Some search systems cluster results and the present each cluster in a first step
- Diversity can favor **data exploration** and user relevance feedback
- Understand links between documents and content in the results

Retrieval from social networks

- People share data in social networks
  - Blogs, facebook, …
  - Sometimes more than any physician would share
    - People with rare diseases are sometimes desperate …
- People can **comment** on data
- Goals of comments are not always clear
  - **Spam** can create problems
- Some metadata is available, so free text, semi-structured metadata, images
  - Can semantics help with this?
A medical blog on movement disorder

MOONDAY, MARCH 5, 2011
A pleasant genetic appointment!

Brendan’s 3-day-old ring bearer smiling wide seriously!

Brendan’s “good egg” mom attended to today’s genetic/methanol follow-up. Brendan was found to be “improved”, much to the glee and pleasure of his grandparents.

WEIGHT CHALLENGE
03/04/11 - 2435 lbs. 15oz.
03/05/11 - 2438 lbs. 13oz.
03/05/11 - 2438 lbs. 15oz.
03/07/11 - 2438 lbs. 18oz.

STAMINA CHALLENGE
Sleep Apnea: 1-2 hrs. daily
03/04/11 - 3 hrs.
03/05/11 - 6 hrs.
03/07/11 - 13 hrs.
03/07/11 - 10 hrs.
03/08/11 - 8 hrs.
03/08/11 - 8 hrs.

ABOUT BRENDAH
Born in December 2007, Brendan is a charming, curious young man. He lives in Salt Lake City, UT and has global developmental delay (GDD), microcephaly, leukodystrophy, intractable seizures, epilepsy (Dorner, 2005), peripheral neuropathy, liver fibrosis, gastroesophageal reflux.

Patientslikeme
Visceral project

- VIsual Concept Extraction challenge in RAdioLogy
  - Most likely as a MICCAI workshop in 2013 and 2014
- Two challenges
  - Identify organs in 3D data sets
  - Find similar cases using 3D data and radiology reports
- Very large amounts of data (~10 TB)
  - Data distribution via the cloud
    - Participants will get a virtual machine for free
  - Creation of a gold and silver standard

LinkedIn Khresmoi group

Most Popular Discussions

- Why is medicine often not evidence based? (Ben Goldacre)
  - Many reasons are given, but the following paragraph applies to what KHRESMI is aiming to do.

- 30% of world storage is estimated to be medical imaging
  - According to an EU report by now 30% of worldwide data storage is estimated to be medical imaging, more to be read in http://www.saraal.com...

- Big data - and the access to them for scientists
  - The fact that companies often keep big data sets private but publish with them cause some problems for science.
A few references


Conclusions

- Medical information search and access is an important technique in medicine
- And this includes images!!
- Image information is most often complementary to text
- Visual information such as regions of interest can be used to formulate queries
- Radiologists request this increasingly instead of searching in books and discussing with colleagues
- There is still much to be learning for combining visual and textual techniques
Questions?

- More information available from
  - http://www.imageclef.org/
  - http://khresmoi.eu/
  - http://medgift.hevs.ch/
- Contact henning.mueller@hevs.ch