Overview of the ImageCLEF 2015 liver CT annotation task

Neda B. Marvasti[1], María del Mar Roldán García[3], Suzan Üsküdarlı [2], José F. Aldana [3], Burak Acar[1]

[1] Boğazici Univ., EE Dept., Istanbul, Turkey
[2] Boğazici Univ., CmpE Dept., Istanbul, Turkey
[3] Univ. of Malaga, Comp. Languages and Comp. Science Dept., Malaga, Spain

Sep 16, 2015

Outline

- CES concept
- CaReRa project
- imageCLEF: Liver CT annotation task
- Task definition: datasets, evaluation methods
- Participants
- Results
- Conclusion

Clinical Experience Sharing (CES)

- Clinical Experience Sharing (CES) refers to a searchable collective medical knowledge database that enables experience sharing among large community of medical professionals, for clinical and educational purposes.
- A CES platform would
  - Empower comparative diagnosis in the clinical use by presenting past cases that are relevant to a query case from a diagnostic point of view.
  - Assist medical students in the educational use by allowing them to browse past cases with similar/dissimilar symptoms and clinical observation but dissimilar/similar diagnoses.
- A CES platform can be implemented in the form of a Content Based Case Retrieval (CBCR) system.

CaReRa[1]: Case Retrieval in Radiology

- CaReRa is a prototype CBCR implementation of the CES concept, which focuses on liver cases.
- Given a query case with incomplete representation, CaReRa searches and retrieves past cases relevant to the query case.
- CaReRa content analysis is context-free and is driven by an underlying ontology, user preferences and user relevance feedback.
- CaReRa (liver) case representation involves:
  - *Demographics*
  - *Clinical history (ICD-10 codes)*
  - *Drugs (ATC codes)*
  - *Laboratory results*
  - *Physical examination*
  - *Semantic radiological (CT) observations (ONLIRA ontology): UsE*
  - *Low-level image (CT) features: CoG*

1. CaReRa-Web is a web-based data application, which can be accessed at https://vavlab.ee.boun.edu.tr:5904/CareraWeb2

imageCLEF: Liver CT Annotation Task

- **Motivation:**
  - The query cases in CaReRa are likely to be incomplete (missing semantic UsE features).
  - It has been shown that using semantic (UsE) features gives a better retrieval performance than low-level computer generated CoG features. [1]
  - Prediction of UsE features from a given CT volume is required to build a query for CaReRa.
  - Besides, an automated semantic annotation using low-level computer generated features would be operational in standardized radiology reporting and CAD systems.

[1] Neda Barzegar Marvasti, Ceyhun Burak Akgül, Burak Acar, Nadin Kökciyan, Suzan Üsküdarl, Pinar Yolum, Rüstü Türkay, and Bars Bakr, Clinical experience sharing by similar case retrieval, in Proceedings of the 1st ACM international workshop on Multimedia indexing and information retrieval for healthcare. ACM, 2013, pp. 6774

Liver CT annotation task:

Task definition and Datasets

- **Task definition:**
  - Given a cropped CT volume enclosing the liver as well as the LiCO[1] ontology, the task is to fill in a standardized radiology report that is composed of UsE features.
- **Datasets:**
  - 50 training and 10 test datasets.
  - Each training dataset is represented as:
    - *A cropped CT volume of the liver.*
    - *A liver mask, which defines liver in the image.*
    - *ROI, which defines lesion area in the image.*
    - *A set of 73 UsE features annotated using ONLIRA.*
  - Test sets has the same format but UsE features are missing, which are asked to be predicted.

[1] N. Kokciyan, R. Turkay, S. Uskudarli, P. Yolum, B. Bakir, and B. Acar, Semantic description of Liver CT images: An ontological approach, 2014.

Liver CT annotation task:
LiCO (Liver Case Ontology)

- Models a patient case with liver observations by describing patient, study and series level of information.
- Patient levels contains information about age, gender, regular drugs, genetic disease, surgeries, and etc.
- Study level, Includes physical examination, lab results, and etc.
- Series level, covers the imaging observations of the liver domain, representing the properties and relationships between liver, hepatic veins and lesions.
- Using CaReRa-Web, LiCO, and by the help of a radiologist, we have gathered a database of liver case annotations (UsE features).
- These annotations are provided in files with RDF format.

Liver CT annotation task:
Evaluation methodology

- Evaluation is based on completeness and accuracy of the predicted annotations with reference to the manual annotations of the test dataset.

$$Completeness = \frac{\text{number of predicted UsE features}}{\text{total number of UsE features}}$$

$$Accuracy = \frac{\text{number of correctly predicted UsE features}}{\text{total number of predicted UsE features}}$$

- For answers, which allow multiple values to a question, the correct prediction of a single value is considered as the correct annotation.
- 7 out of 73 UsE features were excluded from the evaluation due to their unbounded labels (numeric continuous values).

Liver CT annotation task:
Participants in both 2014 and 2015

- 32 groups registered, however 4 of them submitted their results in 2014 and 2015.

Group name	Affiliation	runs
CREDOM	Tlemcen University, Algeria	3
BMET	University of Sydney, Australia	8
CASMIP	The Hebrew University of Jerusalem, Israel	1
piLab	Bogaziçi University, Turkey	1

- The difference between 2015 and 2014:
  - ONLIRA is enlarged to LiCO, which covers whole case of liver patient, instead of representing only imaging observations.
  - No CoG features are provided this year.
  - UsE features are given in RDF format.

Liver CT annotation task: CREDOM group results [1]

- They submitted 3 runs via 2 different methods:
  - Classification by using random forest (RF) classifier
  - Retrieval by considering the specific signature of the liver
- Their best result is achieved by using the retrieval-based method.

Group	Run	Completeness	Accuracy	Score	Method	Feature
CREDOM	1	0.99	0.825	0.904	RF	Feature1
CREDOM	2	0.99	0.822	0.902	RF	Feature2
CREDOM	3	0.99	0.836	0.910	IR	Liver sig.

- Feature 1: 115 liver texture features + 9 lesion geometric features,
- Feature 2: 214 lesion texture features + 9 lesion geometric features.

[1] Nedjar, I., Mahmoudi, S., Chikh, A., Abi-yad, K., Bouafia, Z.: Automatic annotation of liver ct image: Imageclefmed 2015. In: CLEF2015 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org, Toulouse, France (September 8-11 2015)

Liver CT annotation task:

CREDOM group results

- In the retrieval-based method, they have encoded the 2D image extracted from the central slice of the lesion by applying 1D Log-Gabor filter.
- Then break the output of the filter into small blocks and quantize the dominant angular direction of each block to four levels by using Daugman method.
- Afterward, the Hamming distance has been employed as the similarity metric to retrieve the five most similar images to the test image.
- Finally, for each UsE feature, they have used majority voting between retrieved images.

Liver CT annotation task:
BMET group results [1]

- They submitted 8 runs:
  - 4 using classifier-based approach (RBF and linear kernels)
  - 4 using image retrieval algorithm (with feature selection, without feature selection)
- Performed the experiments with 2 different feature sets:
  - Provided CoG features
  - Provided CoG features and bag of visual words (BoVW)
- The best result is achieved in the experiments with image retrieval approach and by using the CoG features only.

1: Ashnil Kumar, Shane Dyer, Changyang Li, Philip H. W. Leong, and Jinman Kim, Automatic annotation of liver ct images: the submission of the bmet group to imageclefmed 2014, in CLEF 2014 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings (CEUR-WS.org), September 2014.

Liver CT annotation task:
BMET group results

- In Classifier-based approach
  - They used 2-stage support vector machine (SVM) classification to annotate every UsE feature.
  - 1st stage is done using 1-vs-all SVM classifier.
  - 2nd stage is done using 1-vs-1 SVM classifier.
- 2nd stage is activated, if the result of 1st step is more than one label, which is applied to the results of the 1st step followed by a majority voting.
- This approach is employed using two different kernels and on two different feature sets.

Liver CT annotation task:
BMET group results

- In image retrieval-based approach
  - The most similar images from the training set to the current image are selected.
  - Then, a weighted voting scheme is applied to assign labels to each of UsE features.
  - Similarity measure is defined as Euclidean distance.
- A sequential feature selection method is applied to use the most distinct features for each question during the similarity calculation.
- This approach is done with and without feature selection on two different feature sets.

Liver CT annotation task:

BMET group results

Group	Run	Completeness	Accuracy	Score	Method	Feature
BMET	1	0.98	0.89	0.935	SVM-linear	CoG
BMET	2	0.98	0.90	0.939	SVM-linear	CoG+
BMET	3	0.98	0.89	0.933	SVM-RBF	CoG
BMET	4	0.98	0.90	0.939	SVM-RBF	CoG+
BMET	5	0.98	0.91	0.947	IR-noFS	CoG
BMET	6	0.98	0.87	0.927	IR-noFS	CoG+
BMET	7	0.98	0.91	0.947	IR-FS	CoG
BMET	8	0.98	0.87	0.926	IR-FS	CoG+

Liver CT annotation task:

CASMIP group results [1]

- They tried 4 different classifiers in the learning phase:
  - Linear discriminant analysis (LDA)
  - Logistic regression (LR)
  - K-nearest neighbors (KNN)
  - SVM
- In learning phase, for every UsE feature, the best classifier and CoG feature sets are learnt via leave-one-out cross validation method.
- CoG features with dimensionality more than one are ignored, which reduces the number of employed CoG features to 39.

1: Assaf B. Spanier and Leo Joskowicz, Towards content-based image retrieval: From computer generated features to semantic descriptions of liver ct scans, in CLEF 2014 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings (CEUR-WS.org), September 2014.

Liver CT annotation task:
CASMIP group results

- Nine additional low-level image features, describing the gray level properties of liver and lesion and also boundary properties of the lesion, are used.
- Three UsE features including: cluster size, segment and lobe, were extracted directly from the image features.
- For most of UsE features, they observed the same performance using any classifier and any combination of CoG features.
- For 6 of them related to density, contrast and location of the lesion, one of the LDA or KNN has been chosen with their selected features.

Group	Run	Completeness	Accuracy	Score	Method	Feature
CASMIP	1	0.95	0.91	0.93	LDA + KNN	CoG+

Liver CT annotation task: piLab group results [1]

- They considered the dataset as a heterogeneous data and applied coupled matrix factorization models using GCTF (generalized coupled tensor factorization) framework.
- Both KL divergence and Euclidean distance based cost functions are applied.
- They considered two groups of UsE features: the 1st group includes UsE features, which have values varying from 0 to 3 and the 2nd group contains UsE features that have binary values.
- Following matrices are provided:
  - X1: UsE features of first group(60*21)
  - X2: UsE features of second group (60*13)
  - Z1: CoG features (60*447)

1: Beyza Ermis and A. Taylan Cemgil, Liver ct annotation via generalized coupled tensor factorization, in CLEF 2014 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings (CEUR-WS.org), September 2014.

Liver CT annotation task: piLab group results

- Then the latent matrices Z2 and Z3 are estimated as:

  X1 = Z1*Z2   X2 = Z1*Z3
- The UsE features of test cases can be predicted using Z2 and Z3 via GCTF.
- Since the predicted values are not discrete values, a binary thresholding has also been applied.
- This group submitted three runs:

Group	Run	Completeness	Accuracy	Score	Method	Feature
piLab	1	0.51	0.39	0.45	GCTF-KL	CoG
piLab	2	0.51	0.89	0.677	GCTF-EUC	CoG
piLab	3	0.51	0.88	0.676	GCTF-KL	CoG

VAVlab

Liver CT annotation task:
Discussion

- The BMET group achieved the highest score of %94.7.
- In terms of accuracy, BMET group has also attained the best performance by using an image retrieval method.
- In terms of classifier-based methods, BMET and CASMIP groups both obtained the total score of %93.

Group	Completeness	Accuracy	Score	Method	Feature
CREDOM	0.99	0.836	0.910	Image Retrieval	Liver Sig.
BMET	0.98	0.91	0.947	Image Retrieval	CoG
CASMIP	0.95	0.91	0.93	LDA + KNN	CoG+
piLab	0.51	0.89	0.677	GCTF	CoG

Liver CT annotation task:

Discussion

- Results of different runs in predicting different groups of UsE features are as:

Group	Liver		Vessel		Lesion-Area		Lesion-Lesion		Lesion-Component	
	Cmp.	Acc	Cmp	Acc	Cmp.	Acc	Cmp.	Acc	Cmp	Acc
CREDOM	1.00	0.925	1.00	1.00	1.00	0.753	1.00	0.48	0.96	0.89
BMET	1.00	0.93	1.00	1.00	0.92	0.79	1.00	0.83	1.00	0.94
CASMIP	1.00	0.93	1.00	1.00	0.85	0.81	0.90	0.82	1.00	0.94
piLab	0.62	0.88	1.00	1.00	0.46	0.77	0.20	1.00	0.12	0.15

Liver CT annotation task:
Discussion

- Results show that all groups have completed the vessel UsE features with high accuracy.
- All groups except piLab, completed liver features in full with accuracy more than %80.
- CASMIP group attains the best accuracy for lesion-area concepts.
- Concepts related to lesion-components are fully completed and annotated with accuracy higher than %90 by both BIMET and CASMIP groups.
- Concepts related to lesion-lesion are annotated completely by both BIMET and CREDOM groups. However, BIMET has got the accuracy of %83, while CREDOM achieved %48.

Liver CT annotation task:
Conclusion

- This was the 2nd time that liver CT annotation task was proposed and organized.
- This year, the liver case ontology (LiCO) was used to generate the annotations in RDF format. No CoG features were provided this year.
- Same as the last year, the challenge was to predict UsE features of patient records, given the CT volume containing the liver.
- The main challenge of the task was due to the unbalanced dataset. Participants tried to overcome this issue with different methods.
- Among all methods, image retrieval scored the best performance.
- It was observed that feature selection is important for the best performance of the prediction method.

Thank you

Any question?