Query Formulation Use Cases

Daniel Rubin, MD, MS

The Query Formulation Phase II project will deliver functionality that will enable researchers to discover image datasets pertinent to their research questions on the caGrid and to analyze the semi-automatically by accessing AIM-based image annotations. What follows are some use cases for this capability obtained from discussions with clinical researchers in those groups.

RTOG

The RTOG cooperative group has built a rich database that captures a diversity of rich distinctions in their clinical trial data:

- The full imaging study for RT planning (CT image volume data set)
- Target volume contours: (1) gross tumor volume (GTV); (2) Clinical target volume (CTV), (3) Planning target volume (PTV)--an envelope in which encloses the CTV.
- Critical organs at risk (OAR), such as bladder, rectum, spinal cord (dose usually governed by OAR).
- Full 3-d dose distribution (volumetric dose).
- Beam arrangements for delivering dose.
- Quality assurance (QA) information—external judge who scores the contours
- Score of the dose coverage of target volume, and heterogeneity
- Verification images (port films)
- Clinical outcomes: complications, mortality, treatment failures

Currently in RTOG, all these data are stored in a variety of databases: tissue samples, image database, outcomes database, and proteomics databases. Queries that cross these databases are very challenging. Here are two queries based on the need to answer clinical questions that could be directly addressed if query formulation were available. Both of these were actual queries that were at the core of two supported grants.

1) **Biological treatment planning in prostate cancer**: A researcher wanted to quantify the normal tissue complication probabilities (NTCP) as a function of prostate tumor volume and dose (in order to determine quantitative metrics for more robust treatment planning that would minimize the risk to injuring normal tissues in RT planning). The research was carried out by querying the outcomes database to get clinical outcomes on a cohort of patients, with corresponding images in those patients. The images analyzed to calculate NTCP, correlating that with tumor volume and dose received by the patient. If the images had been annotated with AIM, they could have been queried via Query Formulation to obtain the patient cohort based on image metadata query, the spatial regions in the images needed to calculate NTCP could have been accessed, tumor volume accessed via image query, and finally, the results could have been linked to the clinical outcomes. Most of this analysis would have been semi-automated through Query Formulation and AIM.

2) **Understanding spatial dependence on treatment failures**. In another grant-funded study, a researcher explored the effect of spatial location of treatment on efficacy; specifically, how close the tumor was to the spinal cord as a function of treatment success. They found that more treatment failures occurred with disease close to the spinal cord.

3) **Calculating tumor control probabilities (TCP)**: Goal is to determine at what dose level would patients have a particular threshold probability of cure. If images were annotated with AIM, the Query Formulation could find images having particular diagnoses, the image metadata in AIM could be queried to pull out the dose and geometric size of lesions to get the dove-volume histogram (DVH), and these results could be linked to outcomes to calculate probabilities of cure.

4) **Effect of treatment on incidence of impotence**: A researcher reviewed image data on a cohort of patients and re-countoured images to demarcate the penile bulb. RTOG had calculated dose-volume histograms (DVH) across all images. The investigator could conclude that particular doses beyond a certain threshold would correlate with high probability of impotence. If Query Formulation were available, the patient cohort could have been identified by querying image metadata for diagnoses and lesion size characteristics, and the DVH could have been calculated semi-automatically by querying the AIM annotations for lesion size and combining that with dose information. Finally, by querying AIM annotations to locate the geometric region corresponding to penile bulb and combining that with the DVH info and linking it to the outcome data, the researcher could semi-automatically establish the relation between dose to the bulb and clinical outcome.

5) **Improved image segmentation and deformable registration methods**. Modern radiation treatment uses volumetric CT images that are contoured (segmented). If these annotations were in AIM, then Query Formulation could be used to query the image annotation information to get the coordinates and measurements on the contours to drive adaptive radiation therapy (custom treatment based on the patient contours).