A Distributional Approach to Summarization of Radiology Reports

Eamon Johnson*
ebj8@case.edu

W. Christopher Baughman†
wbaughman@metrohealth.org

Gultekin Ozsoyoglu*
tekin@case.edu

*Department of Electrical Engineering and Computer Science
Case Western Reserve University
Cleveland, OH

†Department of Radiology
MetroHealth Medical Center
Cleveland, OH

***Abstract*—Diagnostic radiology reports contain a summary written by a radiologist for communication with primary care providers. Information not included in the summary may be lost in the communication, resulting in substandard patient care. We introduce a notion of salience based on distributional characteristics of term usage in a corpus of radiology reports and present an algorithm for generation of suggestions for inclusion of additional information in report summaries. We evaluate our method on a corpus of 98,913 reports and show our method suggests additions to 11-28% of reports, depending on body location and imaging modality.**

I. Introduction

Efficient human-to-human communication of information is critical in clinical medicine, where time pressure on physicians is high. Automated information extraction and document summarization techniques offer an avenue to improve efficiency of communication, yet difficulties remain in producing reliable information required by the ethical and legal standards of medical practice. In this work we examine state-of-the-art computational natural language processing approaches to the challenges of clinical document summarization and propose a novel method for filtering for salient information. We evaluate our method on three data sets: chest x-rays, computed tomography (CT) of the head, and pelvic ultrasound.

Diagnostic radiology reports may vary in structure, but generally contain a section of *findings* that describe the radiologist's detailed interpretation of the image, and an *impression* section that summarizes the findings for the primary care physician. The question is how well the impression captures the important findings. From the radiology literature, we know that primary care physicians often only read the impression section, though in 36% of instances some critical diseases are not repeated in the impression section, so information loss is a concern for delivery of care [1]. Our task is to improve the summary in a radiology report by assessing the salience of information from the body of the report and then suggesting inclusion of salient details in the summary.

We posit that in general, salient information from the findings section of the report is being included in the summary section, and further that we may use the rate at which terms appear in both the body and summary of a report as a proxy for salience. Based on our salience scoring of terms, we are able to algorithmically generate suggestions for additional details to include in the summary. We evaluate our suggestion algorithm on a corpus of 98,913 reports and suggest additions to 11-28% of reports.

II. Filtering for Salience

Through an IRB-approved request, we received three report sets: 66,888 chest x-rays, 24,683 head CTs, and 7,342 pelvic ultrasounds. We use the cTAKES pipeline [3] to perform named-entity recognition on these records, and we refer to the text of each recognized entity as a *term*.

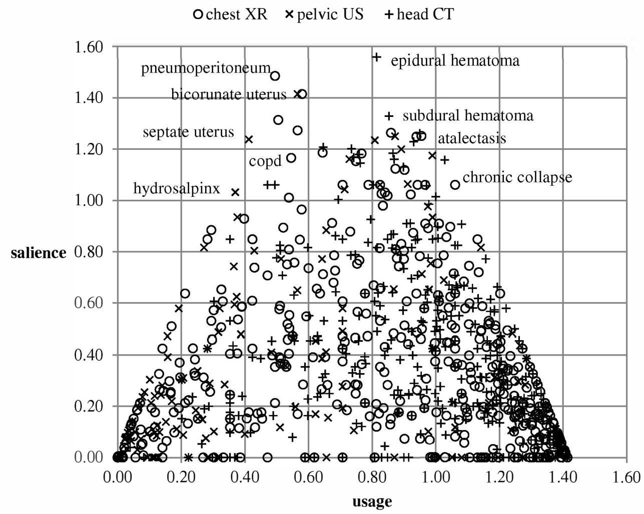
A. Analysis and Projection of Term Usage

For each report in a given set, we start by creating a set of terms that appear in the findings section, T_f, and a set of terms that appear in the impression section, T_i. From these sets, we define three sets for comparison of term usage: $T_F = T_f \setminus T_i$, $T_I = T_i \setminus T_f$, and $T_B = T_f \cap T_i$.

Next, we create a master set of all terms, T, and compute usage statistics for each term: number of times used in findings, number of times used in impression, number of times used in both, and total usage count. We calculate rate of usage by dividing the count of each usage type by the total usage count. To visualize the relative rates of usage, we perform a dimensionality reduction by projecting each term into a number space defined by $x_0 = R_B - R_F$ and $y_0 = R_B - R_I$ where the rates R_F, R_I, and R_B are calculated as$R_z = |T_z|/|T|$. In this projection, terms that are frequently mentioned in both sections have relatively high values in both the x and y dimensions–these are the terms of high salience in the context of the current body location and imaging modality.

To align our projection with the customary x-y axes, we use a simple line-to-point distance transformation which expresses the x-dimension as distance from the line $x - y - 1 = 0$ and the y-dimension as distance from the line $-x - y - 1 = 0$. As shown in Figure 1, transforming the data in this manner allows us to refer to the x-axis as a dimension of usage:

Figure 1. Term Salience and Usage



Algorithm 1 Impression Term Suggestions

```
generate_suggestions(Report R, threshold Q)
    suggestions = []
    candidates = R.findings_terms - R.impression_terms
    for each term Tf in candidates:
        if salience(Tf) > Q:
            for each term Ti in impression:
                if salience(Tf) > salience(Ti):
                    suggestions.append(Tf)
    return suggestions
```

lower values represent terms most frequently used in only the findings, and higher values represent terms most used in only the impression. The y-axis is now our dimension of salience. Higher values reflect more frequent usage of a term in both the findings and impression sections.

B. Generating Suggestions

Our proposed algorithm compares the relative salience metric for terms in the findings with the terms in the impression and suggests terms of higher salience. Algorithm 1 expresses a simple formalization of our proposal.

The threshold value Q is used to express a minimum salience value for suggestion of terms. In our implementation of the suggestion generation we use the mean salience across all terms occurring more than $k = 10$ times as the value for Q.

Table I
STATISTICS OF GENERATED SUGGESTIONS

	chest XR	head CT	pelvic US
mean salience (Q)	0.324	0.369	0.318
reports	66,888	24,683	7,342
suggested terms	12,754	7,383	3,785
reports with suggestions	7,971	3,866	2,103
report suggestion rate	0.119	0.156	0.286
avg terms suggested	1.60	1.90	1.79
avg suggestion distance	4.37	2.84	3.33

Table II
GENERATED SUGGESTIONS

location/modality	rank	term	count
chest XR	1	+effusion	2552
	2	+atelectasis	2348
	3	+infiltrate	2299
	4	+scarring	1913
	5	+opacity	1324
head CT	1	+hematoma	1846
	2	+hemorrhage	873
	3	+atrophy	870
	4	+fracture	812
	5	+infarct	695
pelvic US	1	+cm	2398
	2	+cyst	733
	3	+free fluid	619
	4	+cystic	397
	5	+simple cyst	341

Table I shows the results of execution of our term suggestion algorithm on our three data sets. The details of terms most suggested according to each combination of body area and modality are shown in Table II.

The row marked "avg suggestion distance" reflects the average minimum path distance between the suggested terms and the terms found in the impression section, as calculated over the graph formed by hierarchical "is-a" relationships of the SNOMED-CT ontology [2]. The ontology graph contains 406,099 vertices and 552,761 edges, and path distance between two disease concepts is rarely larger than 10 hops. We use path distance as a conservative estimate of average semantic difference between our suggestions and the existing summary.

III. DISCUSSION

Our algorithm uses two criteria to make recommendations: (1) candidate terms drawn from the findings section must have salience greater than threshold value Q, which we set at average salience, and (2) candidate terms must have higher salience than the lowest salience term in the impression section. As shown in Table I, our algorithm generates suggestions for improvements to the impression section of reports in 11% of chest x-rays, 15% of head CTs, and 28% of pelvic ultrasounds.

The chest x-ray sample displays the lowest percentage of reports with suggestions (11%), yet the suggestions generated are the highest quality (average path distance of 4.37 hops), which we attribute as an effect of the diverse information available in this relatively large sample, as compared with the smaller head CT and pelvic ultrasound samples.

REFERENCES

[1] E. F. Gershanik, R. Lacson, and Ramin Khorasani, and R. Khorasani, "Critical Finding Capture in the Impression Section of Radiology Reports," in *AMIA Annu Symp Proc. 2011*, vol. 2011, Jan. 2011, pp. 465–469.

[2] IHTSDO. (2007) SNOMED CT. Copenhagen. [Online]. Available: http://www.ihtsdo.org/

[3] G. Savova, J. Masanz, P. Ogren, J. Zheng, S. Sohn, K. Kipper-Schuler, and C. Chute, "Mayo Clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications," *JAMIA*, vol. 17, pp. 507–513, 2010.