Automated detection of follow-up appointments using text mining of discharge records

KARI L. RUUD1, MATTHEW G. JOHNSON1, JULIETTE T. LIESINGER1, CARRIE A. GRAFFT2 AND JAMES M. NAESSENS1

1Division of Health Care Policy and Research, Mayo Clinic College of Medicine, 200 First Street Southwest, Rochester, MN, USA, and 2Division of Nephrology and Hypertension, Mayo Clinic College of Medicine, 200 First Street Southwest, Rochester, MN, USA

Address reprint requests to: Kari Ruud, Division of Healthcare Policy and Research, Mayo Clinic, 200 First Street Southwest, Rochester, MN 55905, USA. Tel: +1-(507)-266-9822; Fax: +1-(507) 284-1731; E-mail: ruud.kari@mayo.edu

Accepted for publication 28 February 2010

Abstract

Objective. To determine whether text mining can accurately detect specific follow-up appointment criteria in free-text hospital discharge records.

Design. Cross-sectional study.

Setting. Mayo Clinic Rochester hospitals.

Participants. Inpatients discharged from general medicine services in 2006 (n = 6481).

Interventions. Textual hospital dismissal summaries were manually reviewed to determine whether the records contained specific follow-up appointment arrangement elements: date, time and either physician or location for an appointment. The data set was evaluated for the same criteria using SAS Text Miner software. The two assessments were compared to determine the accuracy of text mining for detecting records containing follow-up appointment arrangements.

Main Outcome Measures. Agreement of text-mined appointment findings with gold standard (manual abstraction) including sensitivity, specificity, positive predictive and negative predictive values (PPV and NPV).

Results. About 55.2% (3576) of discharge records contained all criteria for follow-up appointment arrangements according to the manual review, 3.2% (113) of which were missed through text mining. Text mining incorrectly identified 3.7% (107) follow-up appointments that were not considered valid through manual review. Therefore, the text mining analysis concurred with the manual review in 96.6% of the appointment findings. Overall sensitivity and specificity were 96.8 and 96.3%, respectively; and PPV and NPV were 97.0 and 96.1%, respectively. Analysis of individual appointment criteria resulted in accuracy rates of 93.5% for date, 97.4% for time, 97.5% for physician and 82.9% for location.

Conclusion. Text mining of unstructured hospital dismissal summaries can accurately detect documentation of follow-up appointment arrangement elements, thus saving considerable resources for performance assessment and quality-related research.

Keywords: patient discharge, medical records, natural language processing, appointments and schedules, quality indicators

Introduction

Recent proposals in the USA by the Centers for Medicare and Medicaid Services have aimed to reduce avoidable hospital readmissions [1]. Clear discharge instructions and timely follow-up visits are among the proposed interventions to decrease the number of readmissions. It is suggested that follow-up appointments be arranged for all patients prior to hospital dismissal to reduce the likelihood of readmission [2]. Therefore, all patients should have in hand a specific appointment for follow-up care when they leave the hospital so they know details about whom to see, when and where the provider is located.

Previous research has demonstrated the importance of arranging and documenting follow-up appointments prior to patient dismissal. In a study that examined hospital readmission rates in patients with heart failure, written documentation of discharge instructions including follow-up
appointment arrangements at hospital dismissal was correlated with reduced readmission rates [3]. Lower readmission rates were also found among inpatient psychiatric patients who attended an outpatient appointment after discharge [4]. Finally, patients whose follow-up appointments were arranged and given to them upon discharge from the emergency department have been shown to have higher rates of follow-up compliance [5, 6].

Given the potential benefits of follow-up appointment scheduling and documentation upon hospital discharge, information contained in the dismissal record is beneficial for performance measurement to support quality improvement activities and quality-related research. However, this information is often contained in free-text format. At Mayo Clinic, each patient on general medicine services is to have a follow-up appointment scheduled prior to dismissal. Appointments within Mayo Clinic can be scheduled by clinical assistants in an electronic scheduling system. However, external appointments and those made by physicians are scheduled by telephone. Mayo Clinic’s in-house electronic scheduling system does not connect to its electronic hospital record. Therefore, follow-up appointment details are typed directly into an unstructured field of the electronic medical record (EMR) by the clinical assistant, attending physician or a trained transcriptionist. Upon discharge from the hospital, each patient receives a copy of his/her dismissal summary, which should include follow-up appointment arrangements.

Our group examined outcomes associated with hospital follow-up arrangements, the findings of which are reported elsewhere [7]. The analysis required manual abstraction of appointment details from a large volume of free-text discharge records, which proved to be a time-consuming and costly process. An alternative to manual review would have been the use of automated processes to extract information from the EMR. Information extraction can be accomplished through many approaches, including natural language processing (NLP) and text mining [8]. Previous studies have demonstrated the effectiveness of NLP in extracting pertinent information from textual fields of the EMR [9–15].

To our knowledge, there have been no studies that examined the accuracy of automatically extracting appointment instructions from the EMR discharge summary. This retrospective study investigated the accuracy of text mining in detecting specific follow-up appointment criteria documented in hospital dismissal summaries. Information retrieval is the simplest form of text mining, although text mining has the capacity to extract patterns through text clustering, automatic summarization and analysis of topic trends [8]. Our analysis was generated using the text parsing capability of SAS® Text Miner 3.1 software of the SAS system for Windows 2003 (Cary, NC). SAS Text Miner software has previously been used successfully to extract patient diagnosis information from the pharmacy order database [17] and to examine patients’ orders, medications and complaints in the Emergency Department EMR [18].

**Methods**

**Site and subjects**

For this study, electronic hospital dismissal summaries were extracted for patients discharged from inpatient hospital services at Mayo Clinic in Rochester, Minnesota in 2006. General medicine service patients were identified, and records of patients discharged from the departments of Primary Care Internal Medicine, General Medicine and Hospital Internal Medicine were used for the analysis. Of the 7153 General Medicine dismissal summaries, 672 (9.4%) records were excluded from analysis due to patients being transferred to another inpatient service, being discharged to hospice care or having refused research authorization.

**Analysis**

A data set consisting of 6481 free-text dismissal records was manually reviewed by a health services analyst to determine whether the instructions contained follow-up appointment arrangements. To be considered complete, documentation needed to provide a specific date, time and either a specific physician name or location for the appointment. Appointments with primary care providers and specialty consultants were considered to be follow-up appointments, whereas appointments for procedures or therapy were not. Furthermore, nurse practitioners and physician assistants were considered as acceptable providers for the purpose of the analysis; and locations could be specific departments, clinics, buildings or desks to which to report. Instructions for setting up appointments were not considered as valid appointments. An example of follow-up arrangement text containing all elements of a follow-up appointment is: ‘The patient will follow up with his primary care provider, Dr. Smith, on Monday, June 5, 2006 at 3:30 pm in the Baldwin Building, 5th Floor, Desk B.’ To assess the reviewer’s reliability, appointment elements were manually extracted by a second reviewer from a random sample of 140 discharge records. The raw agreement between the two reviewers was 0.97, and chance-corrected agreement (kappa) was 0.94 [95% confidence interval (CI): 0.885–0.998]. This high agreement validated the use of the original reviewer’s manual abstraction to use as the gold standard for comparison.

The data set was likewise evaluated for the follow-up appointment criteria using SAS Text Miner software, which extracts words and phrases (labeled as ‘terms’) from large collections of unstructured text documents [19]. Although term parsing is an automated process, the analyst conducting the electronic abstraction had to thoroughly review hundreds of terms that were extracted from the data set of dismissal records and manually select indicators for each appointment element (date, time, physician and location). Figure 1 is a
screen shot of the interactive results term window of the Text Miner tool in which date terms were selected by checking the appropriate boxes in the ‘Keep’ column. The manual selection of relevant terms for this analysis was hindered somewhat by misspellings and inconsistencies in transcription. Date terms, for example, were listed in many different forms. January 20, 2006 could be written as ‘20th of January,’ ‘Jan. 20,’ ‘1/20/2006,’ ‘1-20-06’ or ‘20-jan2006.’ Furthermore, patients were often instructed to report for an appointment on an upcoming day of the week, such as ‘this Thursday’; so days of the week that would likely indicate a specific appointment were also selected as date terms.

The review of parsed terms for the selection process necessitated a fair amount of human-computer interface time, although several features of the Text Miner tool aided in the review. For one, the tool automatically stems or groups terms that it identifies as synonymous, which is indicated in the first column of the term window by a ‘+’ symbol (or a ‘−’ symbol when it is expanded to see all equivalent terms, as seen in Fig. 1). The ‘Freq’ column and ‘#Docs’ columns indicate the number of times the term appears in the data set and the number of documents containing the term, respectively. The Text Miner tool additionally facilitated the identification of relevant terms by assigning ‘entity’ categories based on part of speech including time, date, title, person, location and organization (see ‘Role’ column; the ‘Attribute’ column displays character information determined by the tool, including whether the term is an entity). The ‘Weight’ column assigns importance to terms for clustering analysis and modeling; however, this function was not used for this analysis since our objective was limited to parsing. Finally, the results seen in the term window can be sorted by any column, which facilitated the manual process of discovering relevant terms.

After the selection of relevant terms for each appointment element was considered complete, the data set was filtered for those terms, thereby identifying dismissal records containing follow-up appointment details. The interactive results seen in Fig. 2 depict the filtering process for time terms, with records displayed containing specific times. Finally, the selected terms were saved in ‘start lists.’ Start lists can be used by the Text Miner tool for subsequent analyses on the same or new data sets to automatically identify the selected appointment terms, thus significantly reducing the amount of time required for manual term selection. The start lists created for this project ranged in volume from 441 separate terms (for the time element) to 1655 terms (for location).

**Measures**

The text mining assessment of appointment elements was compared with the manual review findings. Agreement was examined for each of the four appointment elements as well as overall follow-up appointment criteria. Overall accuracy or raw agreement was calculated as the ratio of true positives and true negatives to the total number of patient discharge records in the sample. Lastly, sensitivity, specificity, positive predictive and negative predictive values (PPV and NPV) were computed. Sensitivity was defined as the proportion of records containing follow-up appointment arrangement...
elements that were identified as containing appointment criteria via text mining. Specificity was the percent of records lacking follow-up appointment arrangement elements that were not flagged as containing the criteria through text mining. PPV was the percent of records flagged as containing follow-up appointment arrangement elements using Text Miner software that actually contain the criteria according to manual review. NPV was the proportion of records not identified as containing follow-up appointment arrangement elements via text mining that were truly lacking appointment criteria. All data analysis was completed in SAS Version 9.1.2 (Cary, NC).

Results

Of the 6481 discharge records reviewed, 3576 were identified as containing all criteria (date, time and physician or location) for follow-up appointment arrangements through manual abstraction. A total of 113 (3.2%) of these appointments were missed through text mining. Further review of these records revealed location to be the element missed in over half (65) of the false negative cases. A missed physician explained 20 of the overlooked appointments, whereas time and date were missed in 17 and 11 records, respectively. Text mining incorrectly flagged 107 (3.0%) records as containing follow-up appointments that were not considered valid through manual review, 91 of which were procedure or therapy appointments. Specific directives given to patients to set up their own appointments accounted for 13 of the false positives and 3 were appointments listed in addendums (i.e. the information was added after the patient was dismissed to home, so it was not considered as a true discharge instruction per manual review). Table 1 shows the comparison of results between the manual review and text mining. Overall, the text mining analysis concurred with the manual review in 96.6% (95% CI: 96.1–97.0) of the appointment findings. Sensitivity and specificity were 96.8% (95% CI: 96.2–97.4) and 96.3% (95% CI: 95.6–96.9), respectively; and PPV and NPV were 97.0% (95% CI: 96.4–97.5) and 96.1% (95% CI: 95.4–96.8), respectively.

Analysis of individual appointment criteria resulted in accuracy rates of 93.5% (95% CI: 92.9–94.1) for date, 97.4% (95% CI: 97.0–97.8) for time, 97.5% (95% CI: 97.1–97.9) for physician and 82.9% (95% CI: 82.0–83.8) for location. Sensitivity, specificity, PPV and NPV are shown in Table 2. Discrepancies between manual review and text mining results were also examined for each element. Table 3 summarizes the distribution of misclassifications through text mining.

It is worth noting a comparison of total effort invested to complete each portion of the analysis. Manual abstraction of
the 6481 electronic discharge summaries required 43 h of the reviewer’s time at a rate of approximately 150 records reviewed per hour. The analyst using Text Miner software extracted the appointment information from the same records in a total of 14 h. Specifically, it took approximately 3.5 per appointment element to review parsed terms, select appointment identifiers and filter applicable records. It should be noted that utilizing SAS Text Miner software for the extraction process would have taken roughly the same amount of time whether the data set contained 600, 6000 or 600000 observations.

**Discussion**

In this study, approximately half (55.2%) of the patients were dismissed from the hospital with specific follow-up appointment instructions, which is consistent with national findings [3]. The follow-up appointments were accurately detected (96.6%) in unstructured electronic discharge records using SAS Text Miner software, when compared with a gold standard manual abstraction. Patient records containing specific appointment elements of date, time and physician were similarly distinguished via text mining; however, location was only recognized in the records with relatively fair accuracy (82.9%).

Overall, we were satisfied with the accuracy of appointment detection using the Text Miner tool, especially in consideration of the amount of person time saved as compared with manual abstraction. Although some human–computer interface time was required to review all terms parsed by the Text Miner tool and select relevant appointment elements, this effort necessitated one-third of the time required for manual abstraction for this analysis. Furthermore, the selected terms were saved into start lists, which can be processed through any EMR documents using Text Miner software to search for the same words and phrases; physician and location terms would only be relevant in local records, but date and time terms should be universal. Future endeavors to extract appointment information from a large number of discharge summaries could consequently be completed in a relatively brief amount of time. A validation study would be beneficial to examine the accuracy and time saved by using the established start lists.

Previous studies using automated processes to extract information from free-text hospital discharge summaries have similarly yielded promising results. Adverse events have been identified from discharge summaries with an overall sensitivity of 0.280 and specificity of 0.985 using the NLP system MedLEE [20]. A tool developed to extract key findings regarding airway disease resulted in accuracy rates of 82–90% [16]. Finally, diagnoses were correctly identified from discharge summaries in 80% of cases by extracting index terms related to diseases using the vector space model [21].

Although our assessment produced favorable results, error analysis uncovered important obstacles with text mining.

### Table 1 Comparison of text mining assessment of follow-up appointment with manual review findings

<table>
<thead>
<tr>
<th></th>
<th>Text mining</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No appointment, n (% of total)</td>
<td>Appointment, n (% of total)</td>
<td>Total, n (%)</td>
</tr>
<tr>
<td>Manual review</td>
<td>No appointment</td>
<td>2798 (43.2)</td>
<td>107 (1.7)</td>
</tr>
<tr>
<td></td>
<td>Appointment</td>
<td>113 (1.7)</td>
<td>3463 (53.4)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2911 (44.9)</td>
<td>3570 (55.1)</td>
</tr>
</tbody>
</table>

### Table 2 Accuracy of text mining in identifying individual appointment elements

<table>
<thead>
<tr>
<th>Element</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>PPV, % (95% CI)</th>
<th>NPV, % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
<td>99.6 (99.4–99.8)</td>
<td>84.2 (82.8–85.6)</td>
<td>90.5 (89.6–91.4)</td>
<td>99.3 (98.9–99.6)</td>
</tr>
<tr>
<td>Time</td>
<td>99.5 (99.2–99.7)</td>
<td>94.9 (94.0–95.6)</td>
<td>96.0 (95.4–96.6)</td>
<td>99.4 (99.0–99.6)</td>
</tr>
<tr>
<td>Physician</td>
<td>97.5 (97.0–97.9)</td>
<td>97.7 (94.3–96.1)</td>
<td>98.8 (98.4–99.1)</td>
<td>95.3 (96.1–94.3)</td>
</tr>
<tr>
<td>Location</td>
<td>77.8 (76.5–79.0)</td>
<td>94.2 (93.1–95.1)</td>
<td>96.7 (96.1–97.3)</td>
<td>65.7 (64.0–67.5)</td>
</tr>
</tbody>
</table>

### Table 3 Distribution of text mining errors in 6481 records

<table>
<thead>
<tr>
<th>Element</th>
<th>False positives, n (% of total)</th>
<th>False negatives, n (% of total)</th>
<th>Total errors, n (% of total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
<td>407 (6.3)</td>
<td>15 (0.2)</td>
<td>422 (6.5)</td>
</tr>
<tr>
<td>Time</td>
<td>148 (2.3)</td>
<td>18 (0.3)</td>
<td>166 (2.6)</td>
</tr>
<tr>
<td>Physician</td>
<td>51 (0.8)</td>
<td>108 (1.7)</td>
<td>159 (2.5)</td>
</tr>
<tr>
<td>Location</td>
<td>118 (1.8)</td>
<td>991 (15.3)</td>
<td>1109 (17.1)</td>
</tr>
</tbody>
</table>
For one, 13 records containing detailed instructions to patients for setting up their own appointments were falsely identified as containing follow-up appointment elements. For example, a patient instructed to ‘call Dr. Smith’s office between the hours of 8:00 and 5:00 on 20 January to schedule an appointment’ would be identified through text mining as having been given follow-up appointment arrangements due to the presence of date, time and physician name. Addendums, which were added to the record after the patient was discharged from the hospital, also accounted for three false positives. Nevertheless, the primary challenge with this study was that it was impossible to ascertain the type of appointment using the Text Miner tool, because the tool extracts words and phrases from full sentences. Therefore, records with arrangements for procedure and therapy appointments were flagged erroneously in 65 cases. In the future, we plan to use the Mayo Clinic clinical Text Analysis and Knowledge Extraction System (cTAKES) [22], which can be incorporated into the appointment recognition logic to discover procedures and therapies and therefore exclude non-follow-up appointment elements. cTAKES is available open source at www.ohnlp.org.

Investigation of discrepancies by element showed location to be particularly difficult to identify through text mining due to the inability to view entire sentence context. The analyst was cautious in selecting general clinics, departments and desks as location terms, as it was impossible to know whether they were in reference to appointment locations or simply general references. The element of time was missed through text mining in several cases when it was listed in military format. References to physicians were missed in 38 records when the name was written without a title.

Although it would have been beneficial to use an independent sample for validation, resources were not available for this study. Furthermore, this analysis was conducted at a single institution with a comprehensive EMR system. It is unknown how SAS Text Miner software would perform using similar records at other institutions. Some hospitals may incorporate follow-up arrangements in a structured format in the EMR, thus eliminating the need for a tool to extract free-text details. Further research is needed to determine whether appointment details could be detected accurately from the EMR at other institutions with unstructured fields in the dismissal summary.

Conclusions

Our results suggest that text mining of medical records can accurately detect whether elements of follow-up appointment arrangements are documented in a large volume of hospital discharge summaries, thus saving considerable resources required for manual abstraction for performance assessment and quality-related research. Using an existing software product such as SAS Text Miner may allow generalization and adaptation by institutions with varying formats and could be used for other abstraction tasks.

The Mayo Clinic Institutional Review Board has granted the authors permission to use the patient data used in this study.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.

References


Detection of follow-up appointments