Digital Personalized Health and Medicine
L.B. Pape-Haugaard et al. (Eds.)

doi:10.3233/SHTI200287

Ontology-Guided Policy Information Extraction for Healthcare Fraud Detection

Theodora S. BRISIMI [a,1,2], Vanessa LOPEZ [a], Valentina RHO [a], Marco SBODIO [a], Gabriele PICCO [a], Morten KRISTIANSEN [b], John SEGRAVE-DALY [b] and Conor CULLEN [b]

[a] *IBM Research, Ireland*

[b] *IBM Watson Health, GHHS, Ireland*

Abstract. Financial losses in Medicaid, from Fraud, Waste and Abuse (FWA), in the United States are estimated to be in the tens of billions of dollars each year. This results in escalating costs as well as limiting the funding available to worthy recipients of healthcare. The Centers for Medicare & Medicaid Services mandate thorough auditing, in which policy investigators manually research and interpret the policy to validate the integrity of claims submitted by providers for reimbursement, a very time-consuming process. We propose a system that aims to interpret unstructured policy text to semi-automatically audit provider claims. Guided by a domain ontology, our system extracts entities and relations to build benefit rules that can be executed on top of claims to identify improper payments, and often in turn payment policy or claims adjudication system vulnerabilities. We validate the automatic knowledge extraction from policies based on ground truth created by domain experts. Lastly, we discuss how the system can co-reason with human investigators in order to increase thoroughness and consistency in the review of claims and policy, to identify providers that systematically violate policies and to help in prioritising investigations.

Keywords. Healthcare fraud, ontology-based information extraction, claims auditing

1. Introduction

Improper payments for Medicaid in fiscal year 2018 were estimated at 9.8% of Medicaid spending ($36.2 billion) [1]. Examples of improper payments include billing for services exceeding permitted unit limitations, billing separately for services already included in a global fee (unbundling) and billing for medically unnecessary services. Regulations around compliance and accurate billing are described in federal and state policies. Armed with a deep understanding of these policies, FWA investigation units aim at identifying violations in claims submitted for reimbursement by medical providers. However, the sheer volume of claims, benefits, and policy to review, combined with the limited investigation resources and varying skill sets of

[1]Corresponding Author: Theodora S. Brisimi, E-mail: theodora.brisimi@ibm.com

[2]We would like to acknowledge Pierpaolo Tommasi, Carlos Alzate, Spyros Kotoulas, Fabrizio Cucci, Tim Cooper, Mark Gillespie, Mark Gilroy, Jonathon Ryan, Grace Ferguson, John Davis and Jillian Scalvini for their support and insights for this work.

investigators, lead to poor coverage and understanding of the opportunity landscape. The Centers for Medicare & Medicaid Services have developed a fraud detection system that analyses claims to identify providers with suspicious billing patterns [2]. This and similar approaches rely on a combination of applying data analytics techniques to patterns in structured claims data [3, 4] and hand-crafting rules that flag sets of claims for additional investigation. The system presented in this paper focuses on fusing policy and claims together to identify improper payments. It achieves this by semi-automatically converting policy text into *benefit rules*, which are applied directly to claims. This approach is novel to the best of our knowledge and presents a number of research challenges, including:

a) *Domain modelling & knowledge extraction*: Key to our system is building an ontology that is flexible enough to capture the diversity and complexity of the policy *benefit rules* as well as expressive enough to represent the knowledge in a simple, unambiguous and human-readable way to support policy comprehension and human oversight. The ontology is used to integrate relevant terminology and heterogeneous domain sources (e.g., state programs, eligible places of service, coding systems, such as the International Classification of Diseases ICD-9/10 and corresponding updates), as well as to guide the information extraction.
b) *Linking of policy information to the claims data*: Some elements of the *benefit rules* can be easily mapped to claims data columns (e.g., a minimum age constraint will be linked to the birth date of the patient), while others require a more complicated process (e.g., aggregated units of service for a given patient over a given period). The system also needs to map services described in policy to their representation form used in claims, such as the Healthcare Common Procedure Coding System (HCPCS).
c) *Rules polarity*: In policy documents, benefit rules usually express what will be reimbursed (a *positive* rule), e.g. "Adults at high risk for caries may have up to 2 units of dental prophylaxis per year". Sometimes, they express what will not be reimbursed (a *negative* rule), e.g. "Dental prophylaxis is not a covered benefit for children aged 0-4". This provides a challenge for automated identification of improper claims because while the negative rules are generally unambiguous and directly identify violations, the positive rules are often ambiguous and incomplete. In the example, we must infer that more than two units are not allowed for a high-risk adult and that there may be an implied rule for non-high-risk adults and separate rules for children and high-risk children.
d) *Validation of the approach*: Validating the feasibility and impact of our system when used as part of the investigators' workflow requires constructing a set of ground truth rules to evaluate its performance and measuring its generalizability across different policy areas as well as geographic regions.

2. Method

Below we present the components that comprise our system. Fig. 1 runs an example through the end-to-end pipeline.

Policy Ingestion: This component converts the PDF policy document into an HTML document [3], and further enriches the HTML with information identifying sentences, and their hierarchical arrangement (e.g., grouping sentences under a common header or identifying a paragraph introducing a list).

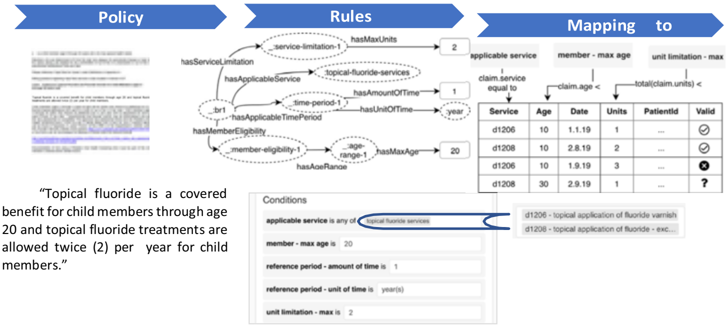

Figure 1. Proposed system's pipeline overview with *benefit rule* extraction and claims validation example.

Rules Extraction: This component is an ontology-based information extraction system based on the enriched HTML policy. Two extractors are implemented on top of two different NLP technologies: Watson X [5] and SystemT [6]. For each textual fragment, the extraction involves the annotation of medical concepts and relations guided by UMLS and the domain ontology. We construct a machine-processable representation of a benefit rule by reasoning over the ontology and translating textual patterns into ontological entities and corresponding relations. Each extracted rule is represented in a knowledge graph (KG). The rule extraction process may generate multiple KGs from the same textual fragment(s), which can be consolidated based on different strategies. Human oversight of this extracted knowledge is essential to correct any errors and establish trust. To that end, we transform each KG into a user-friendly representation, that contains conditions and corresponding values (e.g., *applicable service* with value *topical fluoride* or *maximum age* with value *20*). This empowers the investigators to interact with, adapt and use the extracted knowledge. Rules in policies follow different templates that we captured in the ontology working together with FWA investigators. A detailed discussion of the ontology creation is presented in [7]. The extracted *benefit rules* are then normalised, i.e., properties values are standardised, so that rules execution can unambiguously identify them in claims. The normalisation steps include linking extracted services to specific procedure codes, e.g., *topical fluoride treatments* maps to codes *D1206* and *D1208*, and representing dates in a standard format, e.g., representing *fiscal year* as the appropriate dates range depending on the geographic area.

Mapping to claims: The conditions of each *benefit rule* must be mapped to claims columns in order for the rule to be executed. We assume a claims schema against which the mapping is performed. Examples of mappings are: direct - e.g., a minimum age property in the rule maps directly to an age column that can be calculated from the birth date of the patient and the date of service, aggregations - e.g., a rule that places an upper limit on the number of claims over a period of time. To map this to claims, we need to (i) filter claims, e.g. by service, (ii) group claims, e.g., by patient or date and (iii) compute the aggregation, e.g. an incremental count of claims, value-dependent - e.g., a rule may have a 'requirement' element with values such as 'prior authorisation', 'medical necessity', etc. Each of these values may map to distinct columns in the claim data.

[3]IBM Compare and Comply: www.ibm.com/watson/developercloud/compare-and-comply/

This component applies the *benefit rules* to claims and outputs a status for each claim: improper, correct or unknown, with the latter representing claims that are not represented by any rule. Claims that appear to be improper are linked automatically to both the executed benefit rule and the relevant policy documentation.

3. Results and Discussion

We designed a prototype investigation workbench (Fig. 2) that highlights the impact of bringing policy and claims data together into a unified workflow. Using this workbench, investigators can contrast what providers are authorised to do with what they actually do. This starts by selecting a policy area (e.g., topical fluoride) and the policy sections (rules) to be investigated (Fig. 2: #1, #2). An Investigator can then immediately visualise, inspect and prioritise claims that appear to be out-of-sync with that policy (Fig. 2: #3, #4).

This helps in addressing significant problems in investigation planning and execution: (1) investigators can size potential recovery opportunities before deciding where to allocate scarce resources; (2) maintaining a direct connection between policy and related claims helps build a water-tight case for recovery; (3) automatically showing the relevant claims (and claim fields) for a policy area under investigation saves time by reducing dependencies on other departments for pre-defined data extracts/spreadsheets; and (4) it enables investigators to review and correct the automatically-extracted benefit rules (Fig. 2: #2), thereby building up a shared store of high-quality, executable policy knowledge.

We validate the rules extraction performance via precision and recall. To create Ground Truth (GT) for these measurements, we used a set of documents for Physical Therapy policy from one U.S. state, plus Dental policy from two U.S. states. For each one, a team of three FWA Investigators manually translated the associated benefit rules into a computable form, based on our ontology (**GT rules**). Automatically **extracted rules** were then compared to these, yielding the evaluation results presented in Table 1.

Table 1. We pair a GT rule R with the set of extracted rules from the same text, and we compute a pairing score that takes into account the number of common conditions and values. If the pairing score is always zero, then R is a false negative. Otherwise, the true positive corresponding to R is the extracted rule having the maximum pairing score. Details on the pairing score calculation can be found in [7]. The last column of the table shows the average maximum pairing score across all rules.

Policy	num. GT rules	num. extracted rules	precision	recall	avg. pairing score
Physical Therapy	25	38	0.58	0.88	0.61
Dental (State 1)	50	46	0.65	0.60	0.69
Dental (State 2)	34	42	0.64	0.79	0.56

The ontology used in the experiments comprises of 35 classes, 1151 individuals, 29 properties and 4214 lexicalisations (i.e., ontology entity labels used to annotate textual entities). We are currently extending our work to improve the breadth and coverage of the models and the extraction, in particular across paragraphs, headings and in cases where there are conflicting extracted information. Most of the effort required to generalise across domains and geographical areas is on identifying external (instance) area-specific data to be incorporated into the ontology, such as programs and grouping of codes that are not

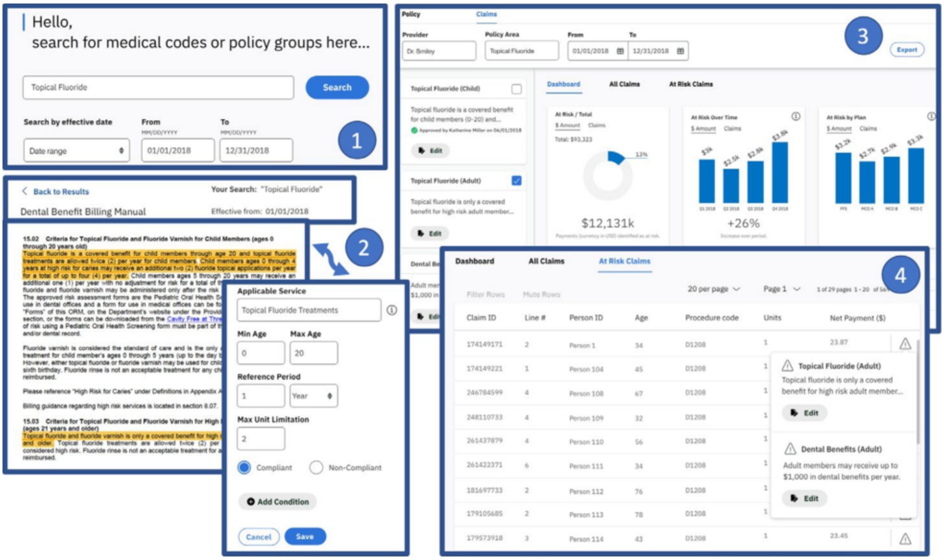

Figure 2. Workflow: The investigators receive a phone tip about a suspicious dental provider *Dr. Smiley* and decide to look for violations related to topical fluoride. Next, they 1) search for related policies, 2) select one or more extracted rules from these policies and modify/correct the automatically extracted rule(s), if necessary, 3) define the scope (e.g., provider and date range), and apply the set of rules to the related claims, seeing the impact of the violations in terms of number of improper claims and potentially recoverable amount of money, 4) carefully inspect each improper claim and the corresponding policy it appears to violate.

part of a global (e.g., federal) code set. Without these groupings, for example, considering the *Dental (State 2)* policy, recall started at around 0.59. When clinical vocabularies such as UMLS are used to find services, treatments and diagnoses, recall increases to 0.76 (with a precision of 0.68). Furthermore, when 135 different groupings of codes were added by a coding expert, recall improved to 0.79, with a small drop in precision to 0.64 (Table 1).

4. References

[1] *U.S. Government Accountability Office*, https://www.gao.gov/key_issues/medicaid_financing_access_integrity, 2018.

[2] *U.S. Government Accountability Office*, https:// www. gao. gov/ assets/690/686849.pdf , 2017.

[3] H. Joudaki, et al., Using data mining to detect health care fraud and abuse: a review of literature, *Glob. J. Health Sci.* **7(1)** (2015), 194.

[4] D. Thornton, et al., Predicting healthcare fraud in medicaid: a multidimensional data model and analysis techniques for fraud detection, *Proc. Technol.* **9** (2013), 1252–1264.

[5] A. Kalyanpur, et al., Structured data and inference in DeepQA, *IBM J. Res. Dev.* **56.3.4** (2012), 10–1.

[6] L. Chiticariu, et al., SystemT: Declarative text understanding for enterprise, *NACCL-HLT.* **3** (2018), 76–83.

[7] V. Lopez, et al., *Benefit graph extraction from healthcare policies*, ISWC, 2019.