

Machine Learning from Health Insurance Administrative Data: Opioids, Obamacare, and Other Applications

Kush R. Varshney
krvarshn@us.ibm.com
<http://krvarshney.github.io>

IBM Research AI / November 14, 2017 / © 2017 IBM Corporation



Machine Learning on Administrative Data: Lots of Examples I've Worked On

- HR and compensation data to predict voluntary resignation of IBM employees
- Work products data to estimate skills and expertise of IBM employees
- Enterprise data to predict successful team compositions of IBM employees
- Roll call, bill co-sponsorship, and bill text data to predict voting patterns of members of Congress
- Application form and repayment data to predict behaviors of pay-as-you-go solar power customers in rural India
- Application form data and evaluation history to predict recipients of a prestigious social entrepreneurship fellowship
- Grand slam tennis match statistics data to predict the winner of each point

IBM Research AI / November 14, 2017 / © 2017 IBM Corporation

2

Medical Claims

Used by payers (insurance companies) to reimburse health providers (physicians, hospitals, etc.)

Date of service

Diagnosis codes (ICD-10), procedure codes (CPT), drug codes (NDC)

Billed amount and paid amount

Provider and patient information (demographics)

Useful for many different machine learning tasks

- Cost prediction, e.g. Obamacare
- Understanding health patterns, e.g. opioid addiction

Medical Claim Form
 This form can be used with all medical plans. It's not intended for Dental or Pharmacy claims.
 You only need to fill out this form if your health care professional isn't filing the claim for you.
 Even if not part of the Cigna network (out-of-network), your health care professional still can file the claim for you.
 We've added instructions on the back of this form to make it easier for you to complete.
 You can find Dental and Pharmacy claim forms on mycigna.com. Go to: Review My Coverage/Dental or Pharmacy-Related Claims.

PRIMARY CUSTOMER INFORMATION: Primary Customer: complete this section

A. PRIMARY CUSTOMER'S NAME AND ADDRESS: FIRST NAME, LAST NAME, STREET ADDRESS, CITY, STATE, ZIP CODE, PHONE NUMBER, FAX NUMBER, ACCOUNT NO. (if the basis of relationship is)

B. DATE OF BIRTH: MM, DD, YYYY

C. PRIMARY CUSTOMER'S MAILING ADDRESS (if different from A): STREET ADDRESS, CITY, STATE, ZIP CODE, PHONE NUMBER, FAX NUMBER, ACCOUNT NO. (if the basis of relationship is)

D. SOCIAL SECURITY NUMBER (if applicable): SOCIAL SECURITY NUMBER, ACCOUNT NO. (if the basis of relationship is)

E. EMPLOYER INFORMATION: EMPLOYER NAME, EMPLOYED, RETIRED, RETIREMENT DATE, RETIREMENT TYPE

PATIENT INFORMATION: Complete this section only if the patient is not the primary customer

F. PATIENT'S NAME: FIRST NAME, LAST NAME, MIDDLE INITIAL, PREFIX, SUFFIX, ACCOUNT NO. (if the basis of relationship is)

G. DATE OF BIRTH: MM, DD, YYYY

H. PATIENT'S ADDRESS: STREET ADDRESS, CITY, STATE, ZIP CODE, PHONE NUMBER, FAX NUMBER, ACCOUNT NO. (if the basis of relationship is)

I. IS THE PATIENT COVERED UNDER ANOTHER HEALTH PLAN? YES/NO

J. IS THE PATIENT COVERED UNDER ANOTHER LIFE INSURANCE POLICY? YES/NO

ACCIDENT/OCCUPATIONAL CLAIM INFORMATION: Complete this section only if you are filing the claim because of an accident or occupational (work-related) illness or injury

A. OCCURRENCE: OCCURRED, OCCURRED

B. DATE OF ACCIDENT OR ONSET OF ILLNESS: MM, DD, YYYY

C. DESCRIPTION OF ILLNESS OR OCCURRENCE: ILLNESS OR OCCURRENCE

D. DATE OF ACCEPTANCE OF SETTLEMENT OF CLAIM: MM, DD, YYYY

E. ARE YOU OR YOUR REPRESENTATIVE FILING A CLAIM OR LITIGATION AGAINST A THIRD PARTY INVOLVED IN THE ACCIDENT OR OCCURRENCE? YES/NO

FAMILY/OTHER COVERAGE INFORMATION: Complete only if claim is for a dependent and/or other coverage is in effect

A. MEMBER EMPLOYER: EMPLOYER NAME, EMPLOYED, RETIRED, RETIREMENT DATE, RETIREMENT TYPE

B. MEMBER'S DATE OF BIRTH: MM, DD, YYYY

C. MEMBER'S SOCIAL SECURITY NUMBER: SOCIAL SECURITY NUMBER, ACCOUNT NO. (if the basis of relationship is)

D. MEMBER'S ADDRESS: STREET ADDRESS, CITY, STATE, ZIP CODE, PHONE NUMBER, FAX NUMBER, ACCOUNT NO. (if the basis of relationship is)

E. IS THE PATIENT COVERED UNDER ANOTHER HEALTH PLAN? YES/NO

F. IS THE PATIENT COVERED UNDER ANOTHER LIFE INSURANCE POLICY? YES/NO

G. MEMBER'S POLICY NUMBER: POLICY NUMBER, ACCOUNT NO. (if the basis of relationship is)

H. MEMBER'S EMPLOYER INFORMATION: EMPLOYER NAME, EMPLOYED, RETIRED, RETIREMENT DATE, RETIREMENT TYPE

I. MEMBER'S SOCIAL SECURITY NUMBER: SOCIAL SECURITY NUMBER, ACCOUNT NO. (if the basis of relationship is)

J. MEMBER'S ADDRESS: STREET ADDRESS, CITY, STATE, ZIP CODE, PHONE NUMBER, FAX NUMBER, ACCOUNT NO. (if the basis of relationship is)

K. IS THE PATIENT COVERED UNDER ANOTHER HEALTH PLAN? YES/NO

L. IS THE PATIENT COVERED UNDER ANOTHER LIFE INSURANCE POLICY? YES/NO

M. MEMBER'S POLICY NUMBER: POLICY NUMBER, ACCOUNT NO. (if the basis of relationship is)

N. MEMBER'S EMPLOYER INFORMATION: EMPLOYER NAME, EMPLOYED, RETIRED, RETIREMENT DATE, RETIREMENT TYPE

O. MEMBER'S SOCIAL SECURITY NUMBER: SOCIAL SECURITY NUMBER, ACCOUNT NO. (if the basis of relationship is)

P. MEMBER'S ADDRESS: STREET ADDRESS, CITY, STATE, ZIP CODE, PHONE NUMBER, FAX NUMBER, ACCOUNT NO. (if the basis of relationship is)

IF YOU ANSWER THE "OR" OR "AND" QUESTIONS, YOU MUST PROVIDE THE COMPANY'S ADDRESS. PLEASE PRINT OR TYPE ON THIS FORM AND FOR A COPY OF THE MEMBER'S BENEFIT GUIDE AND THE MEMBER ID CARD FOR YOUR CLAIM.

Any person who knowingly and with intent to defraud any insurance company or other person (1) files an application for insurance or statement of claim containing any material false information or (2) conceals for the purpose of obtaining information concerning any material fact (bribe, concealment, fraud, insurance act which is a crime). For residents in the following states, please see the last page of this form: Alaska, Arizona, California, Colorado, Connecticut, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming.

A Tale of Two Laws

PATIENT PROTECTION AND AFFORDABLE CARE ACT

Changed the landscape of the health insurance market in the United States

Health insurance companies had to decide which new markets to enter

- New markets defined by geography, age group, and other prospect base criteria

HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT

Required all releases of health-related information about individuals to protect their privacy

- Even for health insurance companies' internal planning uses

k-Anonymity is a common mathematical interpretation of the privacy condition

Desiderata of Health Insurance Companies

Desire low-cost (healthy) people enroll in their plans

Not allowed to accept or deny enrollment on an individual basis

Allowed to offer or not offer plans in well-defined markets

(Allowed to have marketing strategies)

Use data-driven decision making for determining whether or not to offer plans in new markets

Desiderata of Health Insurance Companies

Desire cost data on people who will enroll in new markets

Only have cost data on people who have enrolled in existing markets

Have demographic data on existing market

Have demographic data on new market

Regression problem with covariate shift

– Also need to consider enrollment: three-population shift

Demographic and Cost Data Availability

	Existing Market	New Market
Enrolled	insurance company has demographic data and cost data	insurance company has no demographic data or cost data
Everyone (enrolled and not enrolled)	insurance company can get demographic data from public sources	insurance company can get demographic data from public sources

IBM Research AI / November 14, 2017 / © 2017 IBM Corporation

The Market Risk Assessment Regression Problem

Use cost and demographic data for enrolled members in the existing market, demographic data for the existing market, and demographic data for the new market to estimate cost for enrolled members in the new market

- Can use regression technique of choice
- Ordinary least-squares with and without log-transformed data
 - Two-part models
 - Generalized linear models
 - Multiplicative regression

Need a type of covariate shift to account for the differences between the features of the existing and new market

For the three population shift and regression workload, need a privacy transformation that preserves the probability distribution of the data

IBM Research AI / November 14, 2017 / © 2017 IBM Corporation

8

Results

Developed such an approach, which required a new privacy-preservation method

Excellent empirical performance

Successfully used by a large health insurance company

D. Wei, K. N. Ramamurthy, and K. R. Varshney, "Health Insurance Market Risk Assessment: Covariate Shift and k-Anonymity" SIAM International Conference on Data Mining, pp. 226-234, April-May 2015.

Best Research Paper Honorable Mention

Opioid Epidemic

175 deaths per day due to opioid overdose in the United States

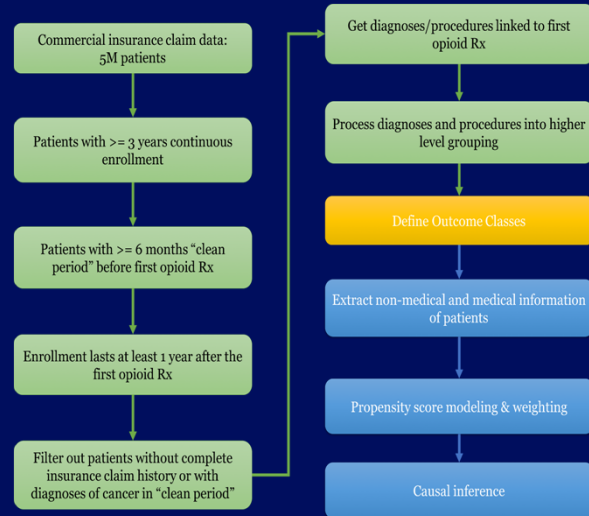
Physicians could be indirectly contributing to the epidemic by overprescription

One reason is because they have generally lacked data to guide opioid prescribing decisions

Use medical claims data to explore the causal relationships between the characteristics of the initial opioid prescriptions and outcomes



Approach



IBM Research AI / November 14, 2017 / © 2017 IBM Corporation

11

Results

In patients who are given synthetic opioids for no more than 7 days, using natural or semi-synthetic opioids instead could potentially reduce the risk of long-term use or addiction by 36.5%.

For natural or semi-synthetic opioids, a shorter days of supply could potentially reduce the risk by 65.3% in patients with longer days of supply.

Longer days of supply on the initial opioid prescription is a driving force of long-term use

A significant difference in average treatment effect between synthetic opioids and natural or semi-synthetic ones

J. Zhang, V. S. Iyengar, D. Wei, B. Vinzamuri, H. Bastani, A. R. Macalalad, A. E. Fischer, G. Yuen-Reed, A. Mojsilović, and K. R. Varshney, "Exploring the Causal Relationships between Initial Opioid Prescriptions and Outcomes," AMIA Workshop on Data Mining for Medical Informatics, November 2017.

IBM Research AI / November 14, 2017 / © 2017 IBM Corporation

12

