ML for readmission reduction, DRG classification and resource allocation

Natalia Serna
Quantil
Alvaro J. Riascos Villegas
Universidad de los Andes y Quantil

November, 2016

Contenido

Introduction

- Data driven decisions for reducing readmissions for heart failure: General methodology and case study. PLOS ONE. 2014.
- ML approaches for early DRG classification and resource allocation. Informs. 2015.

Introduction

- Data driven decisions for reducing readmissions for heart failure: General methodology and case study. PLOS ONE. 2014.
- ML approaches for early DRG classification and resource allocation. Informs. 2015.

Contenido

Data driven decisions for reducing readmissions for heart failure

- Motivation: More than 12 billions USD spent in preventable readmissions.
- Data: 793 hospital visits for heart failure.
- Objective I: Construct a classifier to predict readmissions within 30 days of discharge.
- Objective II: Introduce a decision problem, post discharge intervention costs vrs. readmission, and evaluate cost effectiveness.
- Results: Using out of sample 379 cases they report: Readmission mean cost is $\$13,000$ USD. A post discharge plan reduces 30-day hospitalizations by $35\,\%$. If the post discharge plan costs $\$1,214$ then a this ML guided decision problem would reduce readmissions by $18,2\,\%$ and costs by $3,8\,\%$

Data driven decisions for reducing readmissions for heart failure

- Motivation: More than 12 billions USD spent in preventable readmissions.
- Data: 793 hospital visits for heart failure.
- Objective I: Construct a classifier to predict readmissions within 30 days of discharge.
- Objective II: Introduce a decision problem, post discharge intervention costs vrs. readmission, and evaluate cost effectiveness.
- Results: Using out of sample 379 cases they report: Readmission mean cost is $\$13,000$ USD. A post discharge plan reduces 30-day hospitalizations by $35\,\%$. If the post discharge plan costs $\$1,214$ then a this ML guided decision problem would reduce readmissions by $18,2\,\%$ and costs by $3,8\,\%$

Data driven decisions for reducing readmissions for heart failure

- Motivation: More than 12 billions USD spent in preventable readmissions.
- Data: 793 hospital visits for heart failure.
- Objective I: Construct a classifier to predict readmissions within 30 days of discharge.
- Objective II: Introduce a decision problem, post discharge intervention costs vrs. readmission, and evaluate cost effectiveness.
- Results: Using out of sample 379 cases they report: Readmission mean cost is $13,000 USD. A post discharge plan reduces 30-day hospitalizations by 35 %. If the post discharge plan costs $1,214 then a this ML guided decision problem would reduce readmissions by 18,2 % and costs by 3,8 %

Data driven decisions for reducing readmissions for heart failure

- Motivation: More than 12 billions USD spent in preventable readmissions.
- Data: 793 hospital visits for heart failure.
- Objective I: Construct a classifier to predict readmissions within 30 days of discharge.
- Objective II: Introduce a decision problem, post discharge intervention costs vrs. readmission, and evaluate cost effectiveness.
- Results: Using out of sample 379 cases they report: Readmission mean cost is $\$13,000$ USD. A post discharge plan reduces 30-day hospitalizations by 35 %. If the post discharge plan costs $\$1,214$ then a this ML guided decision problem would reduce readmissions by $18,2$ % and costs by $3,8$ %

Data driven decisions for reducing readmissions for heart failure

- Motivation: More than 12 billions USD spent in preventable readmissions.
- Data: 793 hospital visits for heart failure.
- Objective I: Construct a classifier to predict readmissions within 30 days of discharge.
- Objective II: Introduce a decision problem, post discharge intervention costs vrs. readmission, and evaluate cost effectiveness.
- Results: Using out of sample 379 cases they report: Readmission mean cost is $\$13,000$ USD. A post discharge plan reduces 30-day hospitalizations by 35 %. If the post discharge plan costs $\$1,214$ then a this ML guided decision problem would reduce readmissions by 18,2 % and costs by 3,8 %

Data driven decisions for reducing readmissions for heart failure

- ML methodology:
- LASSO type logistic regression was used to selct the most important variables using cross validation (Paco y Felipe! esta es la forma de racionalizar que se eliminen ciertas variables y/o se fuercen a cero si su intepretacion es contraintuitiva)
- Compared with LACE (index of readmissions) using ROC and reclassification analysis(?). See supplementary information.

Data driven decisions for reducing readmissions for heart failure

- ML methodology:
- LASSO type logistic regression was used to selct the most important variables using cross validation (Paco y Felipe! esta es la forma de racionalizar que se eliminen ciertas variables y/o se fuercen a cero si su intepretacion es contraintuitiva)
- Compared with LACE (index of readmissions) using ROC and reclassification analysis(?). See supplementary information.

Data driven decisions for reducing readmissions for heart failure

- ML methodology:
- LASSO type logistic regression was used to selct the most important variables using cross validation (Paco y Felipe! esta es la forma de racionalizar que se eliminen ciertas variables y/o se fuercen a cero si su intepretacion es contraintuitiva)
- Compared with LACE (index of readmissions) using ROC and reclassification analysis(?). See supplementary information.

Data driven decisions for reducing readmissions for heart failure

- Decision methodology:
- Cost of intervention and readmission the same for all patients: $C_{intervene}$, $C_{readmit}$.
- Efficacy of intervention is a priori the same $P_{success}$.
- Without intervention expected cost of readmission is $C_0(p) = p \times C_{readmit}$.
- With intervention is: $C_1(p) = C_{intervene} + p(1 - p_{success}) \times C_{readmit}$.
- For $p \geq p^* = \frac{C_{intervene}}{p_{success} C_{readmit}}$, $C_0(p) > C_1(p)$ so the agent should be intervened.

Data driven decisions for reducing readmissions for heart failure

- Decision methodology:
- Cost of intervention and readmission the same for all patients: $C_{intervene}$, $C_{readmit}$.
- Efficacy of intervention is a priori the same $P_{success}$.
- Without intervention expected cost of readmission is $C_0(p) = p \times C_{readmit}$.
- With intervention is: $C_1(p) = C_{intervene} + p(1 - p_{success}) \times C_{readmit}$.
- For $p \geq p^* = \frac{C_{intervene}}{p_{success} C_{readmit}}$, $C_0(p) > C_1(p)$ so the agent should be intervened.

Data driven decisions for reducing readmissions for heart failure

- Decision methodology:
- Cost of intervention and readmission the same for all patients: $C_{intervene}$, $C_{readmit}$.
- Efficacy of intervention is a priori the same $P_{success}$.
- Without intervention expected cost of readmission is $C_0(p) = p \times C_{readmit}$.
- With intervention is: $C_1(p) = C_{intervene} + p(1 - p_{success}) \times C_{readmit}$.
- For $p \geq p^* = \frac{C_{intervene}}{p_{success} C_{readmit}}$, $C_0(p) > C_1(p)$ so the agent should be intervened.

Data driven decisions for reducing readmissions for heart failure

- Decision methodology:
- Cost of intervention and readmission the same for all patients: $C_{intervene}$, $C_{readmit}$.
- Efficacy of intervention is a priori the same $P_{success}$.
- Without intervention expected cost of readmission is $C_0(p) = p \times C_{readmit}$.
- With intervention is: $C_1(p) = C_{intervene} + p(1 - p_{success}) \times C_{readmit}$.
- For $p \geq p^* = \frac{C_{intervene}}{p_{success} C_{readmit}}$, $C_0(p) > C_1(p)$ so the agent should be intervened.

Data driven decisions for reducing readmissions for heart failure

- Decision methodology:
- Cost of intervention and readmission the same for all patients: $C_{intervene}$, $C_{readmit}$.
- Efficacy of intervention is a priori the same $P_{success}$.
- Without intervention expected cost of readmission is $C_0(p) = p \times C_{readmit}$.
- With intervention is: $C_1(p) = C_{intervene} + p(1 - p_{success}) \times C_{readmit}$.
- For $p \geq p^* = \frac{C_{intervene}}{p_{success} C_{readmit}}$, $C_0(p) > C_1(p)$ so the agent should be intervened.

Data driven decisions for reducing readmissions for heart failure

- Decision methodology:
- Cost of intervention and readmission the same for all patients: $C_{intervene}$, $C_{readmit}$.
- Efficacy of intervention is a priori the same $P_{success}$.
- Without intervention expected cost of readmission is $C_0(p) = p \times C_{readmit}$.
- With intervention is: $C_1(p) = C_{intervene} + p(1 - p_{success}) \times C_{readmit}$.
- For $p \geq p^* = \frac{C_{intervene}}{p_{success} C_{readmit}}$, $C_0(p) > C_1(p)$ so the agent should be intervened.

Data driven decisions for reducing readmissions for heart failure

- Results:
- LACE: AUC 0,11 %
- Logistic LASSO: 0,66 % They use more information than what we use (i.e. Serna (2016) et.al). Cross validation training AUC mean is 0,69 %
- Significant readmissions to other hospitals. Removing this patients improves AUC 0,71 %.
- Best model selects 253 out of $3,300$.

Data driven decisions for reducing readmissions for heart failure

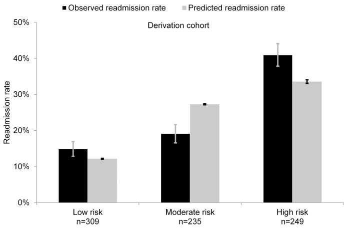
- Results:
- LACE: AUC 0,11 %
- Logistic LASSO: 0,66 % They use more information than what we use (i.e. Serna (2016) et.al). Cross validation training AUC mean is 0,69 %
- Significant readmissions to other hospitals. Removing this patients improves AUC 0,71 %.
- Best model selects 253 out of 3, 300.

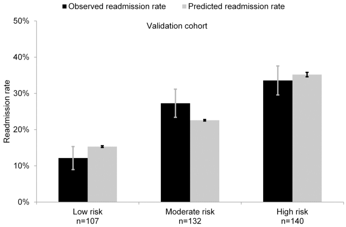
Data driven decisions for reducing readmissions for heart failure

- Results:
- LACE: AUC 0,11 %
- Logistic LASSO: 0,66 % They use more information than what we use (i.e. Serna (2016) et.al). Cross validation training AUC mean is 0,69 %
- Significant readmissions to other hospitals. Removing this patients improves AUC 0,71 %.
- Best model selects 253 out of 3, 300.

Data driven decisions for reducing readmissions for heart failure

- Calibración:





Data driven decisions for reducing readmissions for heart failure

- Variables are clustered and classified according to their evidential support.

Reclassification matrix

		Classifier			
		Low risk	Moderate risk	High risk	Total reclassified (%)
LACE	**Low risk (%)**	35.8	36.2	28.0	**64.2**
	Moderate risk (%)	16.8	33.6	49.6	**66.4**
	High risk (%)	0.0	23.5	76.5	**23.5**

doi:10.1371/journal.pone.0109264.t001

Data driven decisions for reducing readmissions for heart failure

- Reclassification of LACE according to new learner:

Top supportive evidence

Variable class	Variable description	Log Odds Ratio	Log Odds Ratio Standard Error[1]
Lab Results	Lymphocyte % is low	0.0128	0.0027
Patterns of Engagement	Patient was admitted in past 6 months	0.0112	0.0031
Lab Results	BUN is high	0.0038	0.0012
Lab Results	Glucose level random is elevated	0.003	0.0012
Lab Results	Monocyte absolute is low	0.0028	0.0012
Other Diagnoses	History of nondependent abuse of drugs (ICD9 305.x)	0.0018	0.001
Other Diagnoses	History of chronic airway obstruction, not elsewhere classified (ICD9 496.x)	0.0017	0.0008
Other Diagnoses	History of gastrointestinal hemorrhage (ICD9 578.x)	0.0014	0.0007
Lab Results	AST is elevated	0.0013	0.0006
Other Diagnoses	History of cardiomyopathy (ICD9 425.x)	0.001	0.0006
Lab Results	Magnesium is low	0.001	0.0006
Lab Results	INR is elevated	0.0009	0.0004
Patterns of Engagement	Patient has been in isolated room in hospital	0.0009	0.0006
Lab Results	BNP is high	0.0007	0.0005

These are variables that receive positive log-odds ratio with the largest magnitude.
1. Obtained from sample standard error for cross-validation odds ratios
doi:10.1371/journal.pone.0109264.t002

Data driven decisions for reducing readmissions for heart failure

- Odds: Top supportive:

Top supportive evidence

Variable class	Variable description	Log Odds Ratio	Log Odds Ratio Standard Error[1]
Lab Results	Lymphocyte % is low	0.0128	0.0027
Patterns of Engagement	Patient was admitted in past 6 months	0.0112	0.0031
Lab Results	BUN is high	0.0038	0.0012
Lab Results	Glucose level random is elevated	0.003	0.0012
Lab Results	Monocyte absolute is low	0.0028	0.0012
Other Diagnoses	History of nondependent abuse of drugs (ICD9 305.x)	0.0018	0.001
Other Diagnoses	History of chronic airway obstruction, not elsewhere classified (ICD9 496.x)	0.0017	0.0008
Other Diagnoses	History of gastrointestinal hemorrhage (ICD9 578.x)	0.0014	0.0007
Lab Results	AST is elevated	0.0013	0.0006
Other Diagnoses	History of cardiomyopathy (ICD9 425.x)	0.001	0.0006
Lab Results	Magnesium is low	0.001	0.0006
Lab Results	INR is elevated	0.0009	0.0004
Patterns of Engagement	Patient has been in isolated room in hospital	0.0009	0.0006
Lab Results	BNP is high	0.0007	0.0005

These are variables that receive positive log-odds ratio with the largest magnitude.
1. Obtained from sample standard error for cross-validation odds ratios
doi:10.1371/journal.pone.0109264.t002

Data driven decisions for reducing readmissions for heart failure

- Odds: Top discomforming evidence:

Top supportive evidence

Variable class	Variable description	Log Odds Ratio	Log Odds Ratio Standard Error[1]
Lab Results	Lymphocyte % is low	0.0128	0.0027
Patterns of Engagement	Patient was admitted in past 6 months	0.0112	0.0031
Lab Results	BUN is high	0.0038	0.0012
Lab Results	Glucose level random is elevated	0.003	0.0012
Lab Results	Monocyte absolute is low	0.0028	0.0012
Other Diagnoses	History of nondependent abuse of drugs (ICD9 305.x)	0.0018	0.001
Other Diagnoses	History of chronic airway obstruction, not elsewhere classified (ICD9 496.x)	0.0017	0.0008
Other Diagnoses	History of gastrointestinal hemorrhage (ICD9 578.x)	0.0014	0.0007
Lab Results	AST is elevated	0.0013	0.0006
Other Diagnoses	History of cardiomyopathy (ICD9 425.x)	0.001	0.0006
Lab Results	Magnesium is low	0.001	0.0006
Lab Results	INR is elevated	0.0009	0.0004
Patterns of Engagement	Patient has been in isolated room in hospital	0.0009	0.0006
Lab Results	BNP is high	0.0007	0.0005

These are variables that receive positive log-odds ratio with the largest magnitude.
1. Obtained from sample standard error for cross-validation odds ratios
doi:10.1371/journal.pone.0109264.t002