



QUANTIFYING RISK, ENABLING OPPORTUNITY

Subject:
Natural Language Processing in
Healthcare

Author: Matt Zylstra
Organization: Discovery Health
Date: 20 August 2019

Content

Natural Language Processing for Dummies

Secret language of doctors

Tell me how you really feel

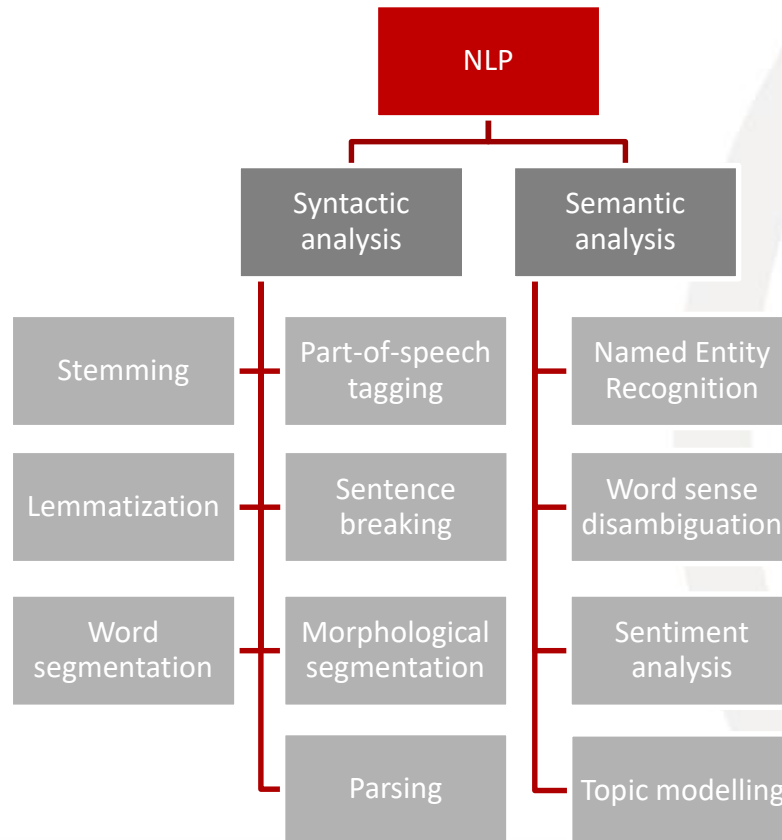
Going off-topic

Chatty robots

Questions

Natural Language Processing for Dummies

- Computers understand human's natural language
- Not new
- NoAI



Secret Language of Doctors

- “Natural language” of doctors
- Sources of valuable clinical information
 - Medical histories
 - Doctors notes
 - Radiology reports
 - Referral notes
 - Authorisation details
 - Discharge summaries

Emergency Department Triage Notes

States started last night, upper adb, took alka seltzer approx. 0500, no relief. Nausea no vomiting

Since yeaterday 10/10, contant Tylenol 1 hr ago. +nausea diaphoretic. Mid abd radiates to back.

Hospital Discharge Note

The patient is a frail 88 year old male was admitted to hospital for complaints of nausea and vomiting and suspected urinary tract infection

He has a past medical history of hypertension, atrial fibrillation and chronic right hip pain after total hip replacement in 2012.

The patient was started on Antibiotics. Urine culture confirmed and E. coli urinary tract infection sensitive to trimethoprim.

During admission an episode of possible coffee ground vomiting coupled with his non-steroidal inflammatory drug use prompted an upper GI endoscopy at which no abnormality was detected. Fecal occult blood was negative.

The patient was also provided with physiotherapy and fully remobilized

Secret Language of Doctors

Doctor's Note

Patient is nauseous and vomiting and has pain in her stomach on the left side. There is no history of gastric disorders. Suspected gastric ulcer. Admitting for a gastro-scope on 15 November 2016.

Symptoms

History

Diagnosis

Treatment

Secret Language of Doctors

Symptoms

History

Diagnosis

Treatment

Doctor's Note

Patient is nauseous and vomiting and has pain in her stomach on the **left**

side. There is **no** history of gastric disorders. **Suspected** gastric ulcer.

Admitting for a gastro-scope on **15 November 2016**.

Laterality

Negation

Certainty

Time

Secret Language of Doctors



Boundary Detection	... The patient underwent a CT scan in April which did not reveal lesions in his liver. ...																			
Tokenization	The	patient	underwent	a	CT	scan	in	April	which	did	not	reveal	lesions	in	his	liver	.			
Normalization	-	-	undergo	-	-	-	-	-	-	do	-	-	lesion	-	-	-	.			
Part-of-speech Tagger	DT	NN	VBD	DT	NN	NN	IN	NNP	WDT	VBD	RB	VB	NNS	IN	PRP\$	NN	.			
Entity Recognition	CT scan Procedure UMLS ID: C0040405					Lesion Disease/ Disorder UMLS ID: C0022198					Liver Anatomy UMLS ID: C0023884					Biomedical End-Use				
Chunking	NP		VP		NP		PP		NP		VP		NP							
Constituency Parsing	S		NP		DT		NN				VP ...		...							
Dependency Parsing	...																			
SRL	undergo.01 (A1.patient; A2.scan; AM-TEMP.in) reveal.01 (A0.scan; R-A0.which; AM-NEG.not; A1.lesions; AM-LOC.in)																			
Entity Properties	CT scan Negated: no Subject: patient					Lesion Negated: yes Subject: patient					Liver Negated: no --					Biomedical End-Use				
Coreference	identity (the patient, his)																			
UMLS Relation	locationOf (lesions, liver)																			
Event, Temp. Expr.	CT scan					April					Reveal								Lesions	
Temporal Relation	April					CONTAINS CT scan					CT scan								CONTAINS lesions	

Secret Language of Doctors

Hospital discharge note

The patient is a frail 88 year old male was admitted to hospital for complaints of nausea and vomiting and suspected urinary tract infection

He has a past medical history of hypertension, atrial fibrillation and chronic right hip pain after total hip replacement in 2012.

The patient was started on Antibiotics. Urine culture confirmed and E. coli urinary tract infection sensitive to trimethoprim.

During admission an episode of possible coffee ground vomiting coupled with his non-steroidal inflammatory drug use prompted an upper GI endoscopy at which no abnormality was detected. Fecal occult blood was negative.

The patient was also provided with physiotherapy and fully remobilized

SNOMED-CT

16932000	Nausea and vomiting
68566005	Urinary tract infectious disease
38341003	Hypertensive disorder (Historic)
49436004	Atrial fibrillation (Historic)
49218002	Hip Pain (Historic)
301011002	Escherichia coli urinary infection
40835002	Coffee ground vomiting
167667006	Fecal occult blood: negative
52734007	Total hip replacement (Historic)
11701004	Urine culture
76009000	Esophagogastroduodenoscopy
91251008	Physical therapy procedure

N39.0 - Urinary tract infection B96.2 - Escherichia coli 43235 – Esophagogastroduodenoscopy

Secret Language of Doctors

CONCEPT ID	SNCT DESCRIPTION	MAP GROUP	MAP PRIORITY	MAP CATEGORY SNOMED CT IDENTIFIER	MAP RULE	MAP ADVICE	TARGET CODE	ICD DESCRIPTION	Workbench Tooling
8619003	Infertile (finding)	1	1	CD	IFA 21071002 Infertility due to germinal cell aplasia (finding)	IF INFERTILITY DUE TO GERMINAL CELL APLASIA CHOOSE N46	N46	Male infertility	
8619003	Infertile (finding)	1	2	CD	IFA 237145004 Unexplained infertility (finding)	IF UNEXPLAINED INFERTILITY CHOOSE N97.9	N97.9	Female infertility, unspecified	
8619003	Infertile (finding)	1	3	CD	IFA 248153007 Male (finding)	IF PATIENT IS MALE, CHOOSE N46	N46	Male infertility	IF GENDER 248153007 THEN N46
8619003	Infertile (finding)	1	4	CD	IFA 248152002 Female (finding)	IF PATIENT IS FEMALE, CHOOSE N97.9	N97.9	Female infertility, unspecified	IF GENDER 248152002 THEN N97.9
8619003	Infertile (finding)	1	5	NC	ELSE	MAP IS CONTEXT DEPENDENT FOR GENDER MAP SOURCE CONCEPT CANNOT BE CLASSIFIED WITH AVAILABLE DATA	NULL		ELSE ALWAYS NOT CLASSIFIABLE

Secret Language of Doctors

- More granular and **coded** clinical data enables
 - Better quality of care and patient experience
 - Clinical Decision Support
 - Prevention of medical errors
 - Better risk management and pricing
 - Better disease management
 - Better managed care and care coordination
 - Drug Utilisation Review in real-time
 - Automated authorisation process
 - Operational efficiencies from interaction data mining
 - Comprehensive EHRs
 - Computer Assisted Coding (ICD-10 and CCSA coding suggestions)

Tell me how you really feel

- Sentiment analysis used to determine customers emotional state
 - Positive
 - Negative
 - Neutral
- Customer surveys
 - Time consuming
 - Annoying
 - Reach less then 5%
 - Performance score not sentiment
- Interaction free-text data
 - Call transcripts
 - WhatsApp messages
 - Virtual agent interactions
 - Emails
 - Social media comments
 - Survey responses
 - Reach over 40%

Tell me how you really feel

Step 1

- Clean data
 - Clear email chain
 - Remove greetings and signatures
 - Handle misspellings and abbreviations
- Adjust for emoticons

Step 2

- Sentiment lexicon
 - 9,000 word dictionary with sentiment scores
 - Additional idiomatic expressions
 - Slang
 - “Lekker”
 - Remove product words
 - “Benefit”

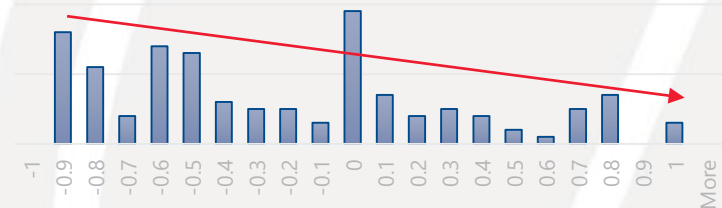
Step 3

- Rules to allow for context
 - Negators
 - “Never”
 - Intensifiers
 - “Very”
 - Punctuation
 - “!!!?”
 - Capitalisation
 - “HORRIFIED”
 - Acronyms
 - “FUBAR”
- Model generates sentiment
 - Polarity
 - Intensity

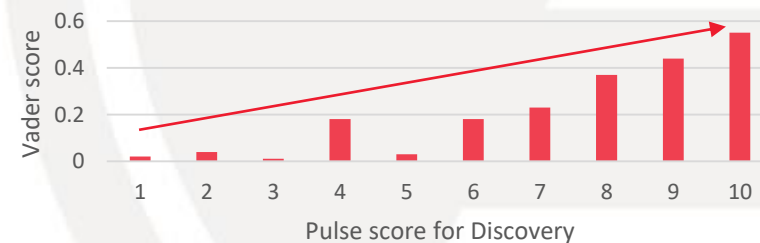
Step 4

- Validate
 - Complaints
 - Survey scores

Validation of NLP model against complaints:



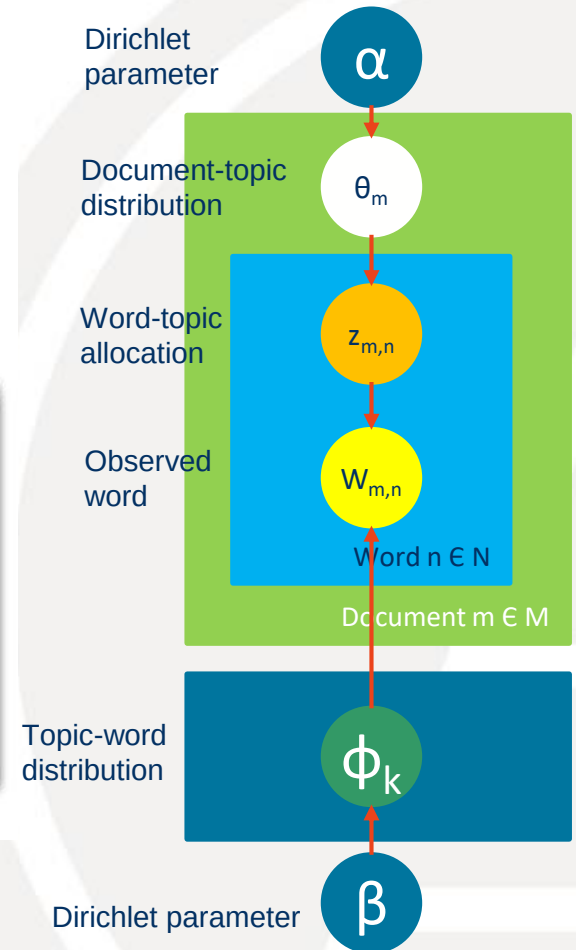
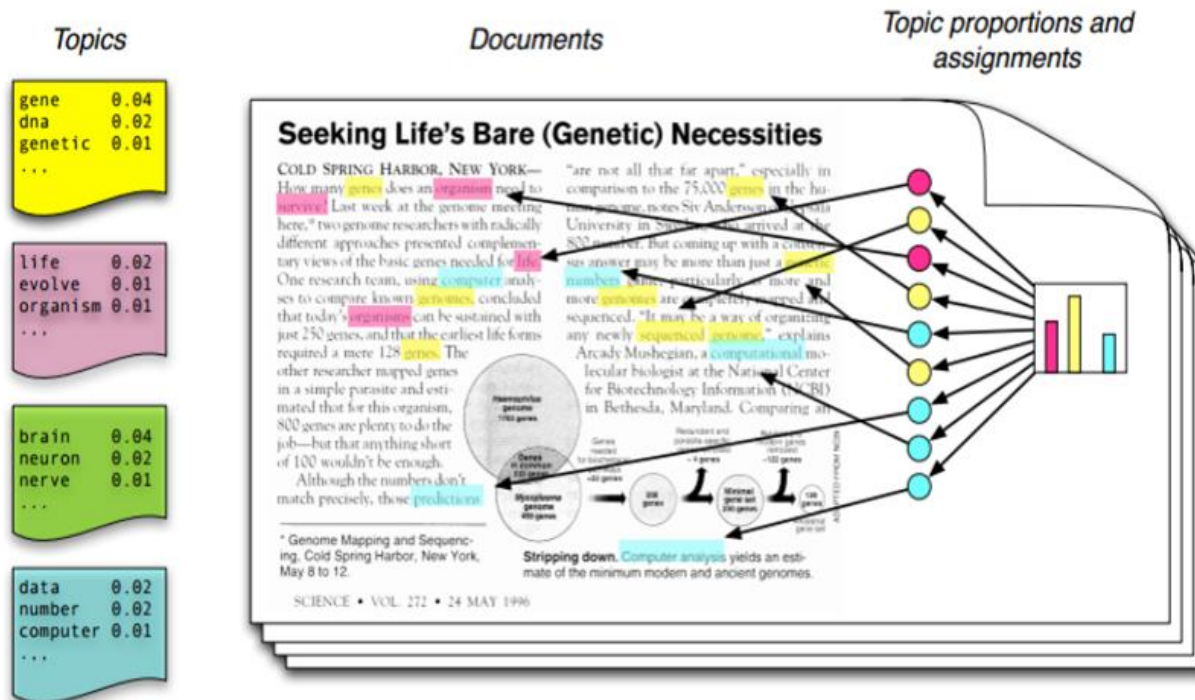
Validation of NLP against survey scores



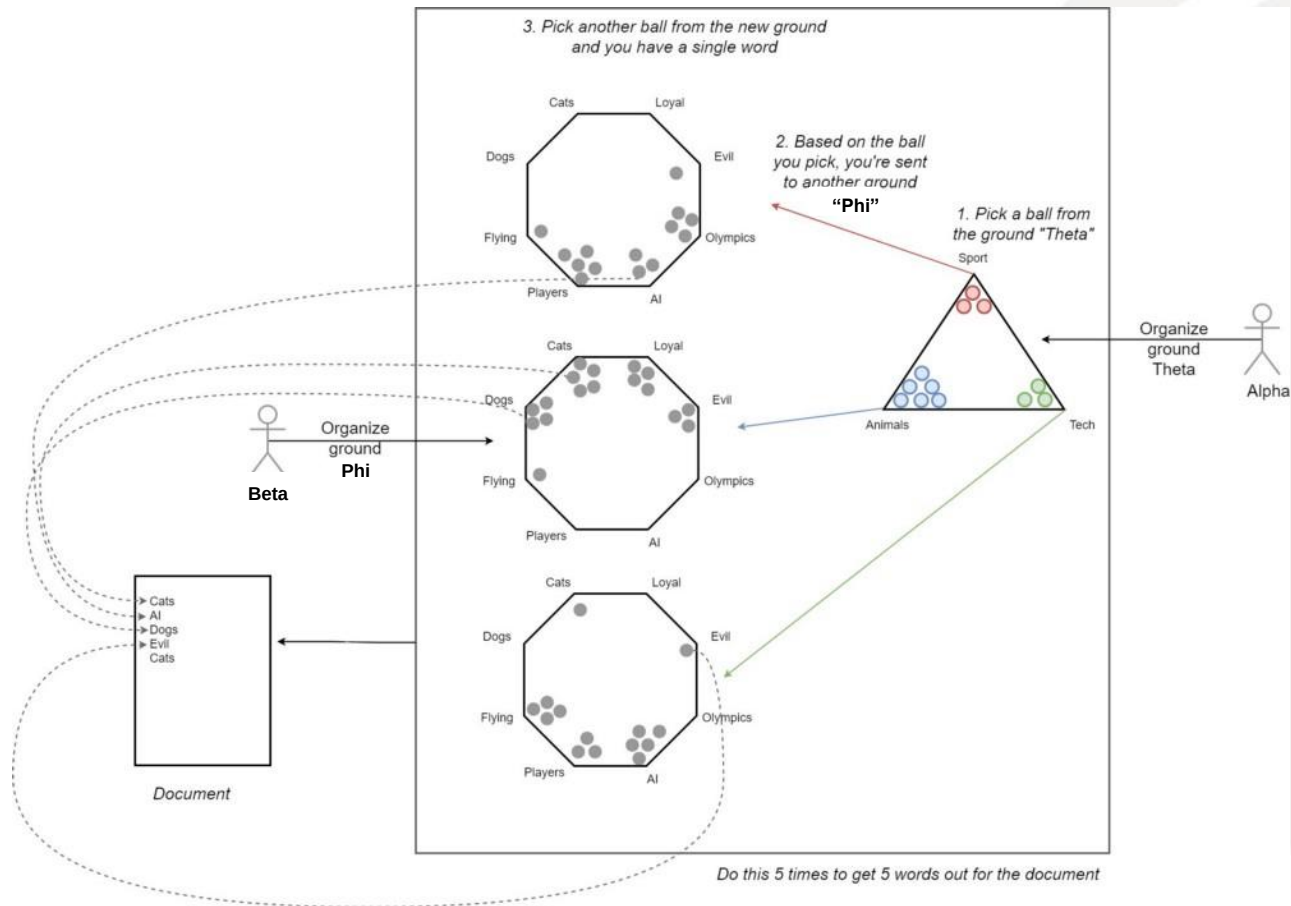
Tell me how you really feel

- Sentiment analysis allows an organisation to
 - Improve customer service
 - Improve product messaging
 - Improve timing and content of communications
 - Reduce customer service workload
 - Positive sentiment relayed to service team to boost morale
 - Prevent lapses
- Remaining challenges
 - Other 10 official languages
 - Sarcasm
 - Medical and insurance jargon
 - Lack of well-formed sentences
 - Combining sentiment analysis with topic modelling

Going off-topic



Going off-topic



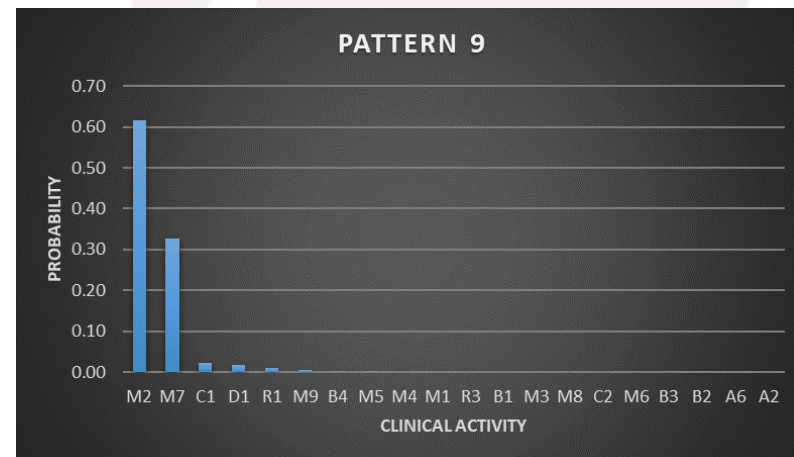
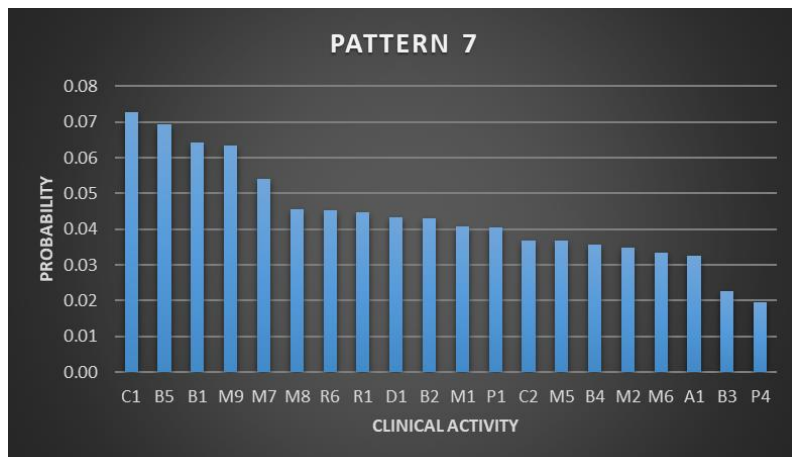
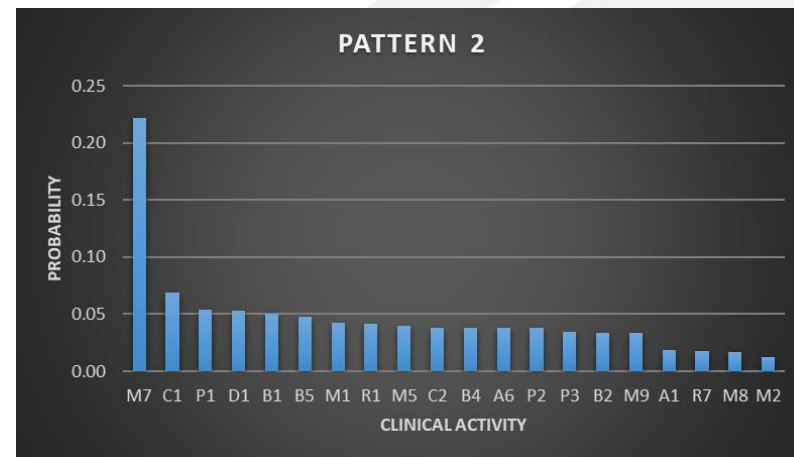
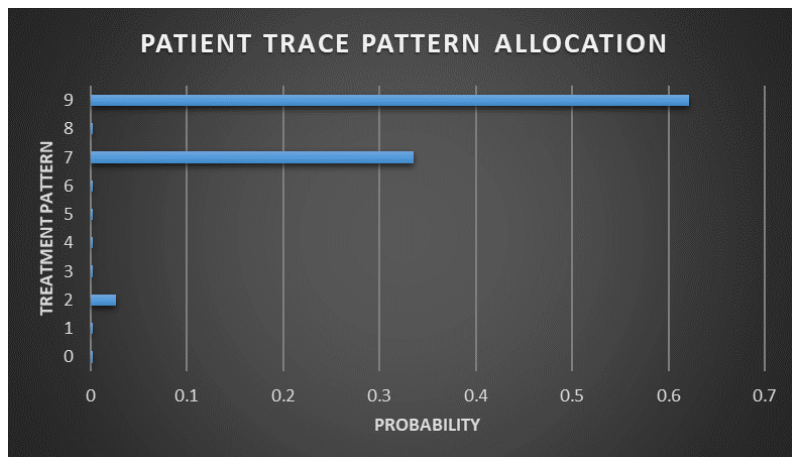
Going off-topic

Measure	Value
Episode of care	Ischaemic Heart Disease
Years included	2008 - 2018
Minimum months on scheme before episode	3
Excluded lives also registered for	Diabetes
Number of patient traces	69,998
Number of clinical activities	1,673,534
Number of activity types	44
Number of treatment patterns	10
LDA prior alpha	0.1
LDA prior beta	0.01

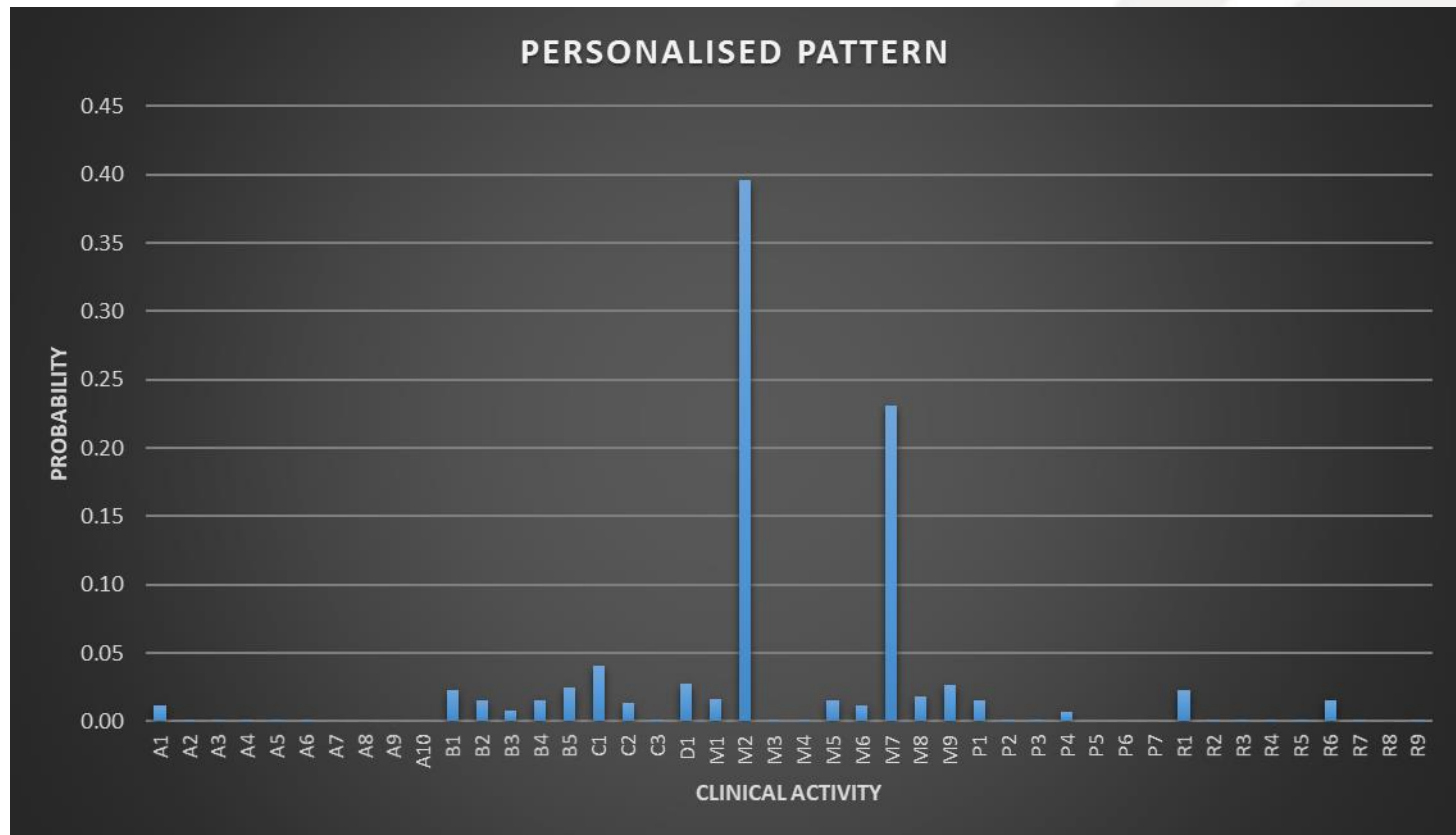
**ACTUARIAL
SOCIETY**
OF SOUTH AFRICA

QUANTIFYING RISK, ENABLING OPPORTUNITY

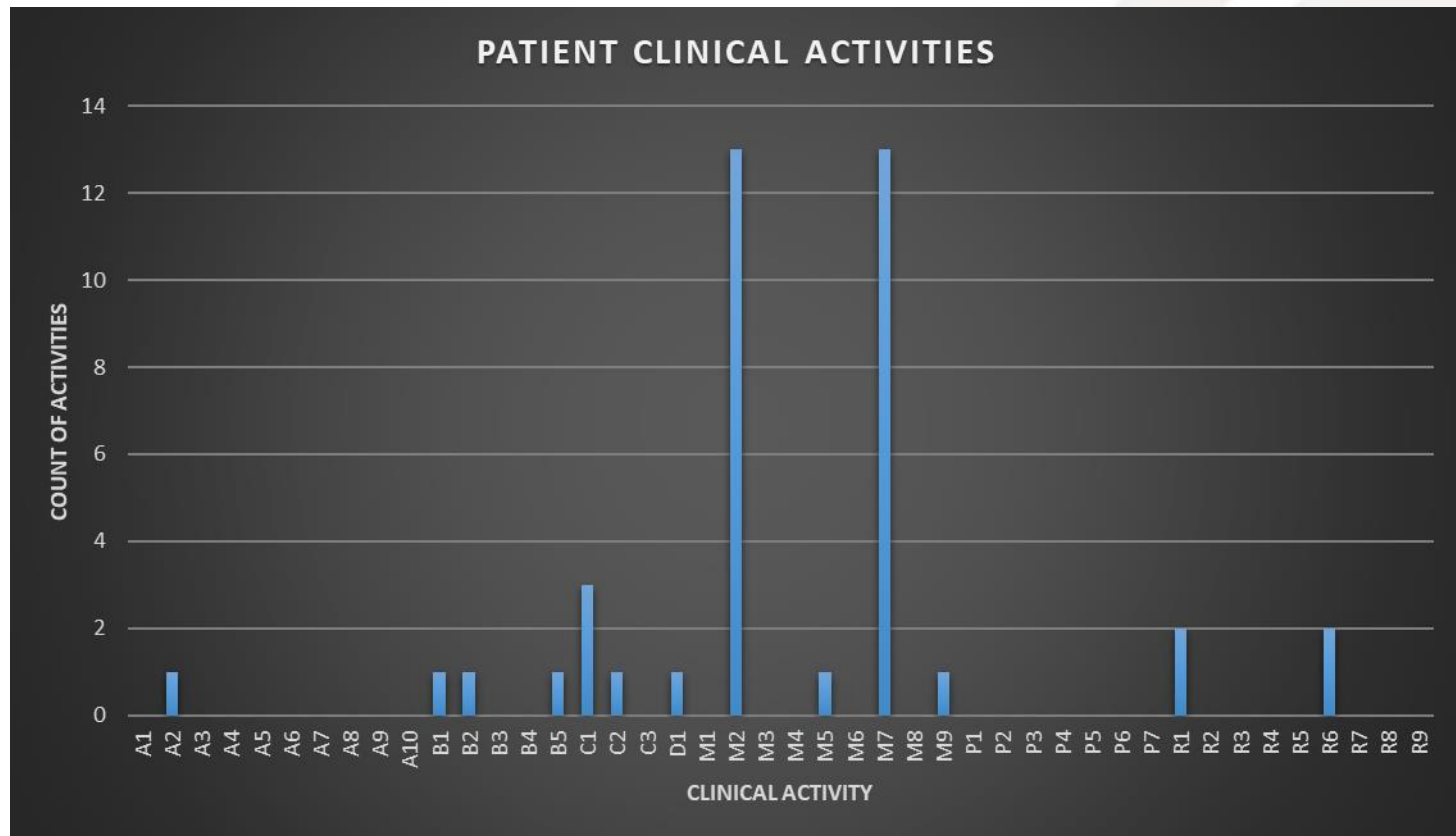
Going off-topic



Going off-topic



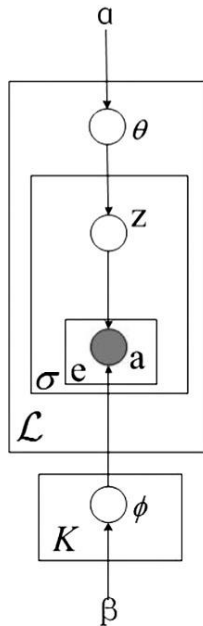
Going off-topic



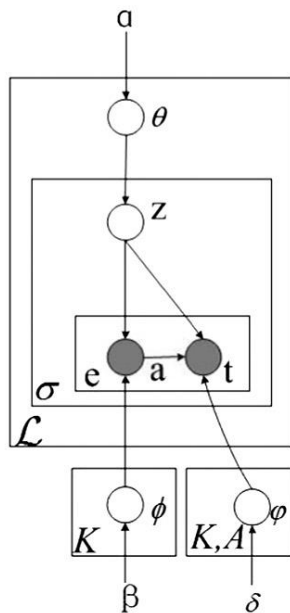
Going off-topic

Standard LDA extended to include timing of clinical activities

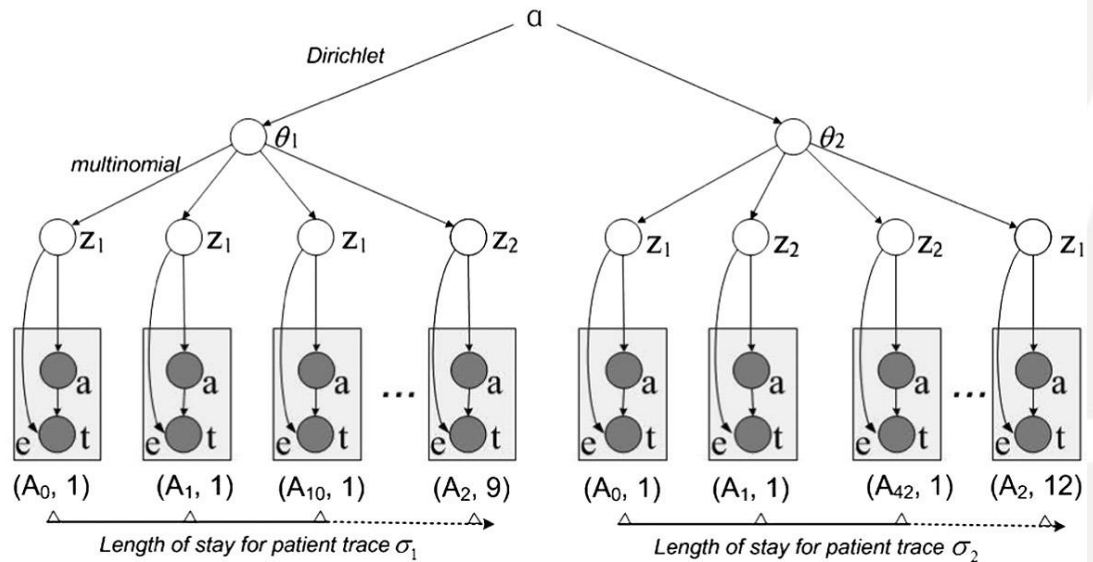
(A) LDA



(B) Clinical Pathway Model

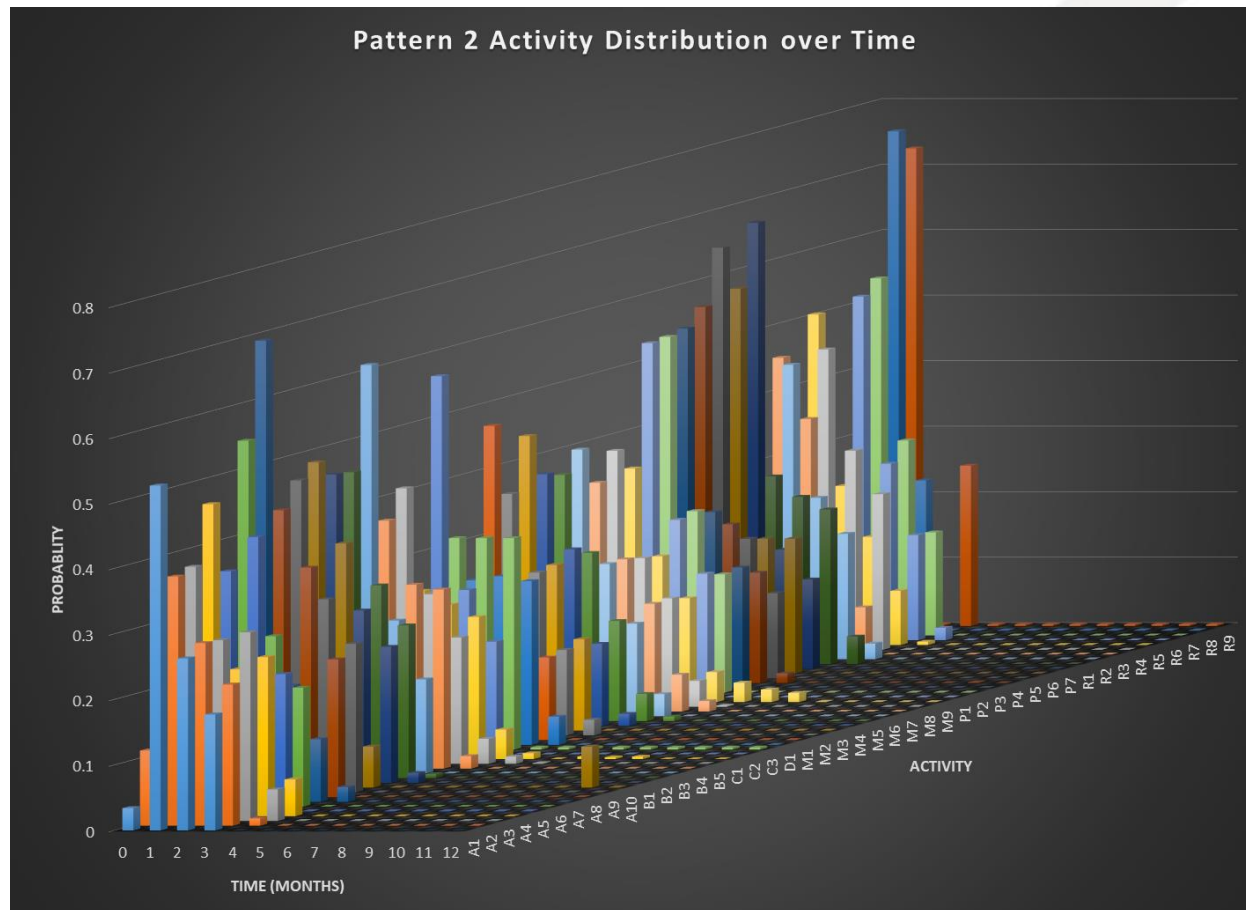


(C) Example of CP pattern generation with Clinical Pathway Model

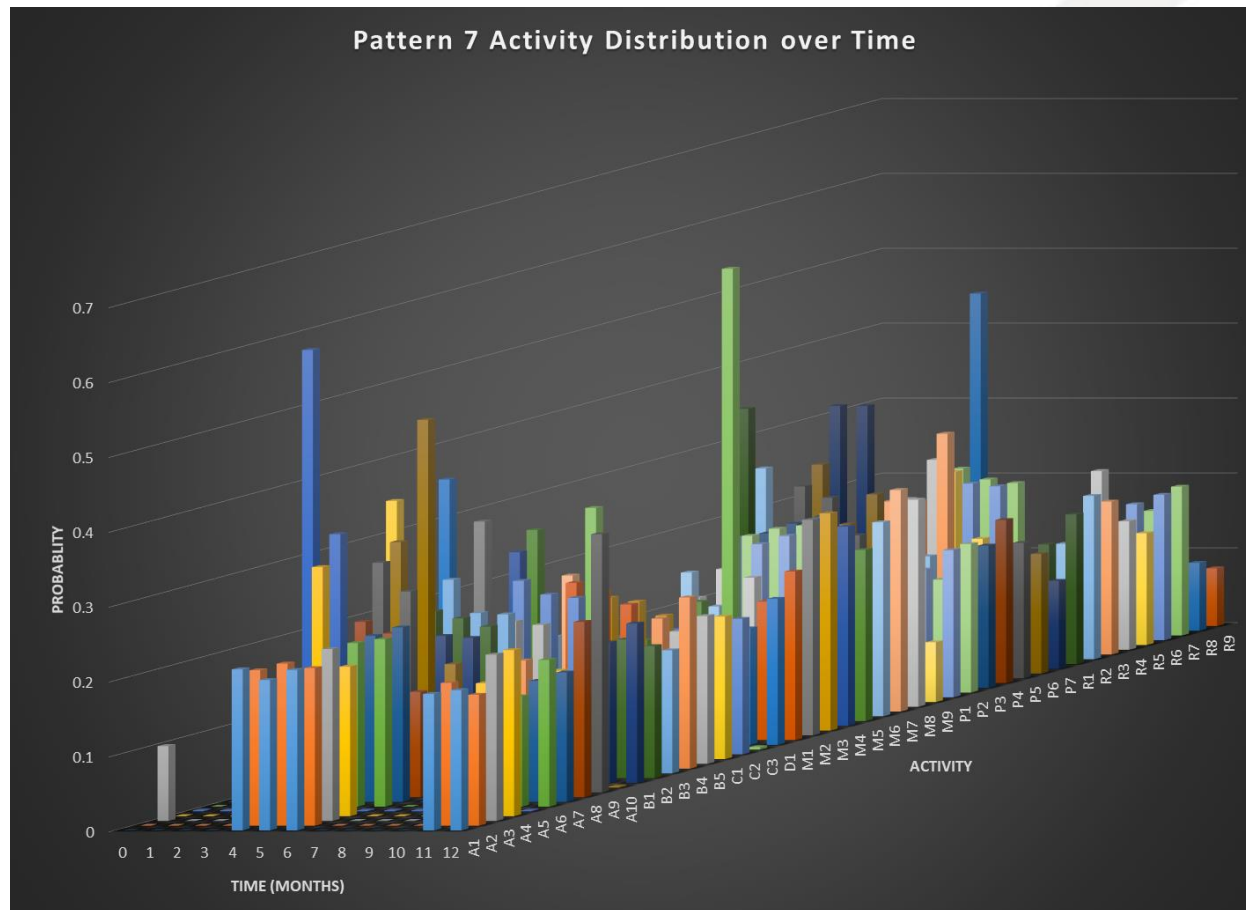


$$P(z_{\sigma,i} = z | z_{\sigma,-i}, \mathbf{e}, \alpha, \beta, \gamma) \propto \frac{n_{\sigma,z} + \alpha}{\sum_Z n_{\sigma,z} + K\alpha} \times \frac{n_{z,a} + \beta}{\sum_a n_{z,a} + |A|\beta} \times \frac{n_{z,a,t} + \delta}{\sum_t n_{z,a,t} + |T|\delta}$$

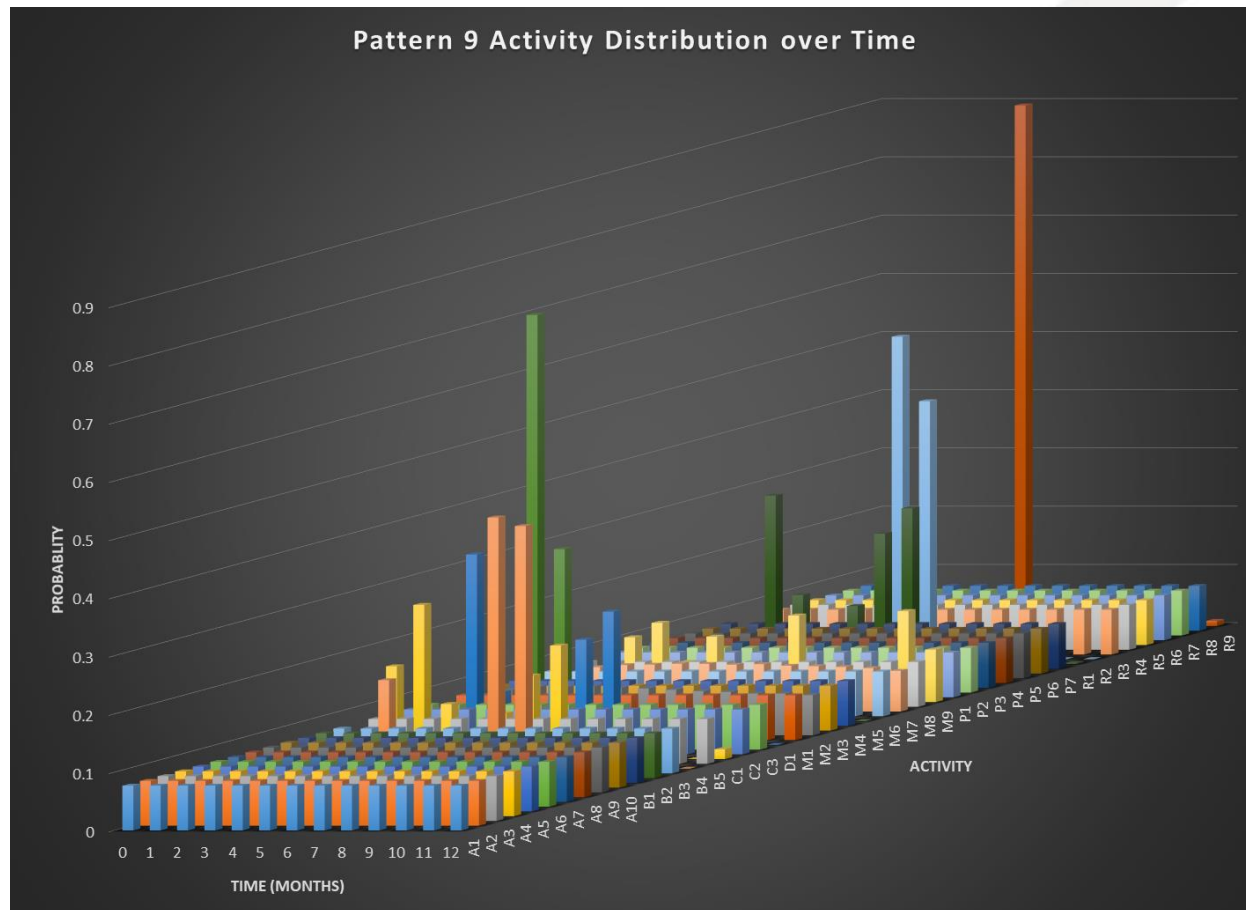
Going off-topic



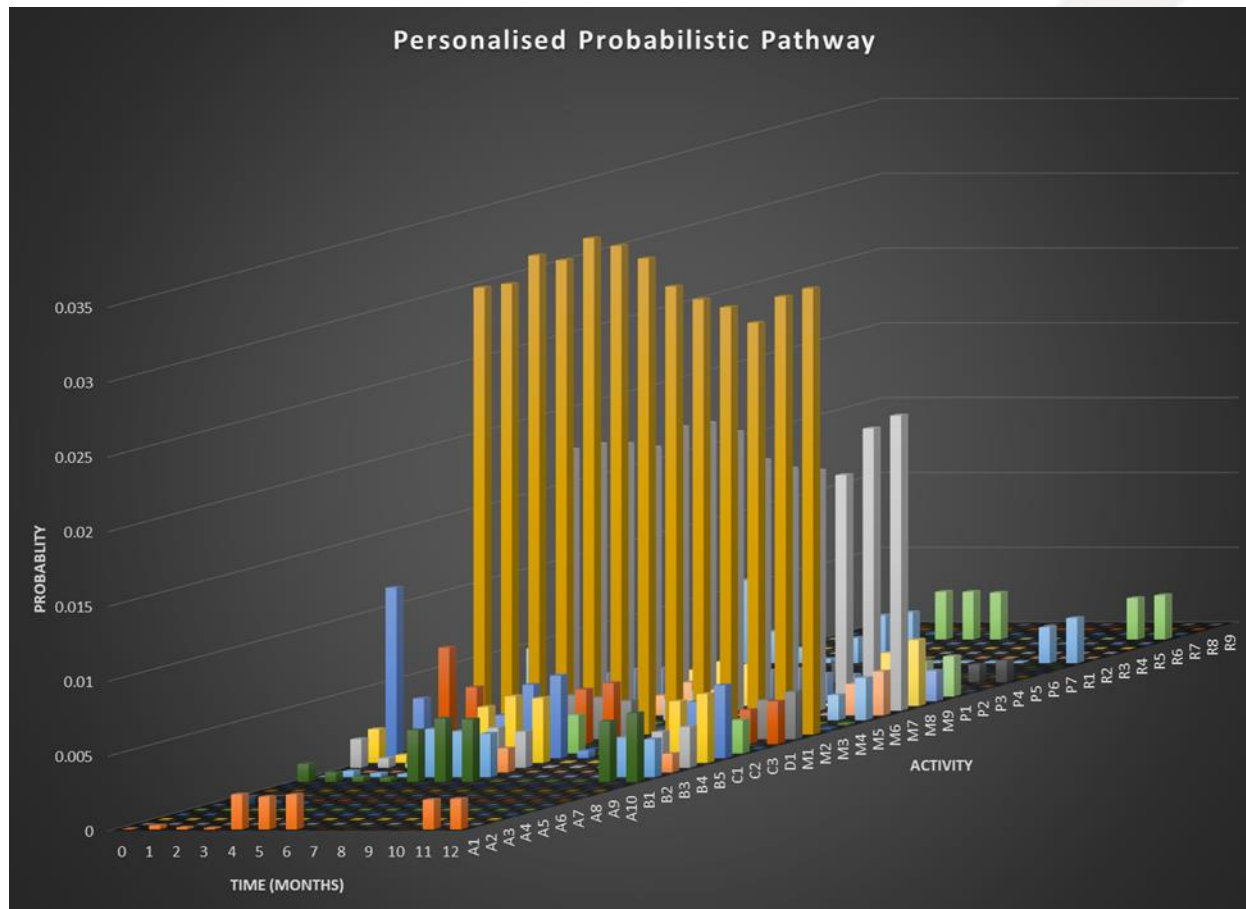
Going off-topic



Going off-topic

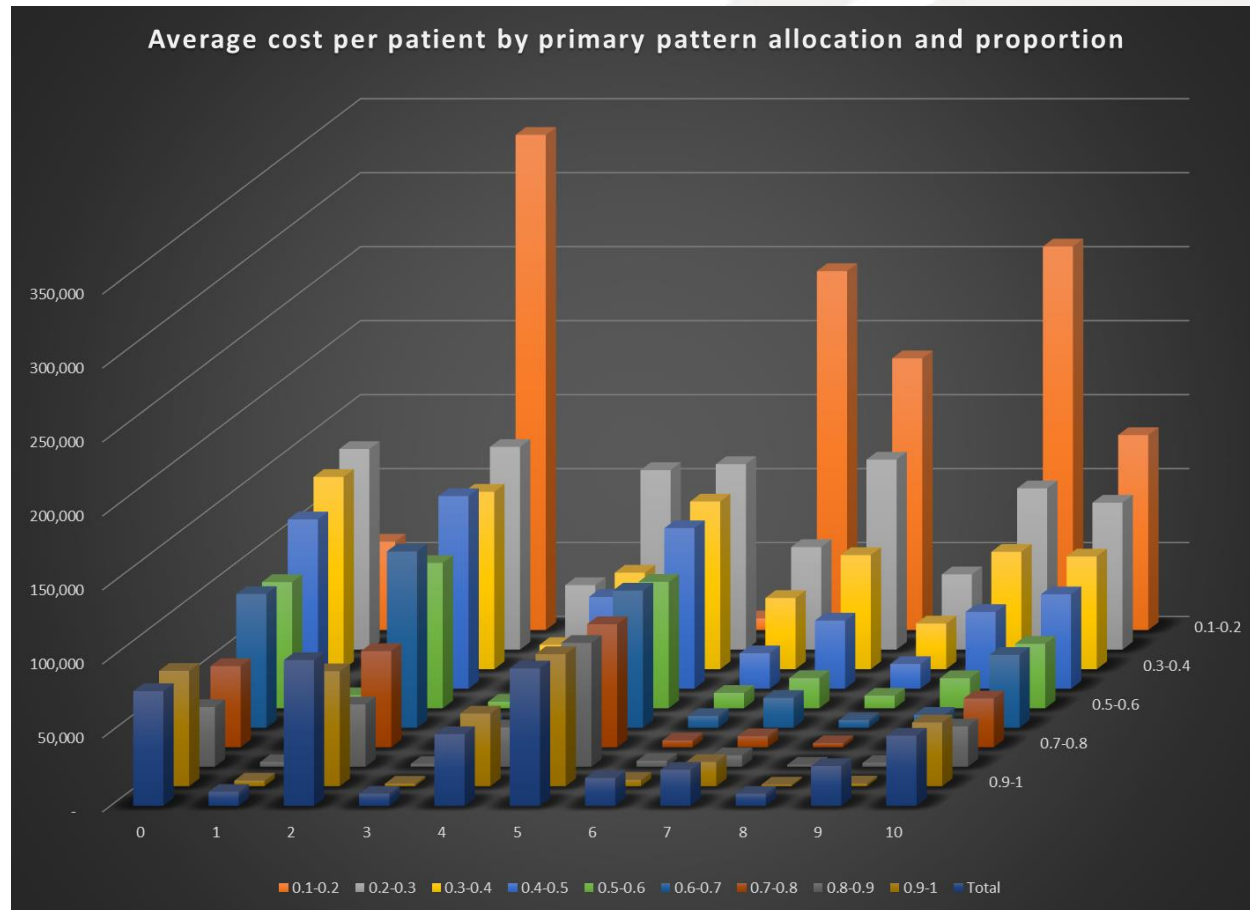


Going off-topic



Going off-topic

- Primary allocated treatment patterns **0, 2 and 5** have the **highest average cost per IHD claimant (R80,000)**
- Costs increase as proportion of primary allocated pattern decrease
- Primary allocated pattern 2 **with less than 0.2** proportion has **average cost of R330,000**
 - 3 claimants cost scheme R1 million
- Top 20 IHD claimants have combined costs of R55.7 million
 - 17 have primary pattern 2
 - Average allocation of 52%
 - 10 died
- Use primary pattern allocation together with proportion to segment members and identify high risk



Going off-topic

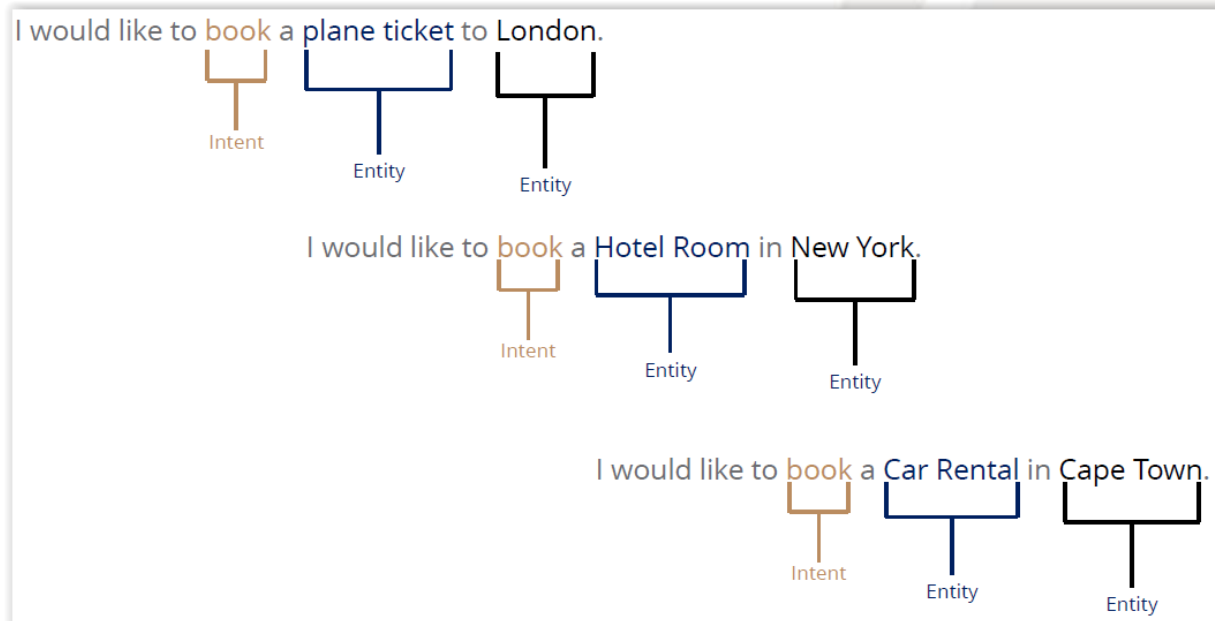
Clinical Pathway	Description	CP with highest prob	Average IHD cost per claimant	% IHD cost of total cost	% Deaths
2	Complex cardiac disease	6%	98,402	59%	5.86%
5	Complex cardiac disease with primary intervention single/multiple stent(s)	13%	92,751	69%	1.84%
0	Complex cardiac disease with primary intervention CABG	25%	77,419	62%	2.66%
4	Progressive cardiac disease with later interventions	4%	48,305	48%	1.61%
9	Asymptomatic observation	9%	26,798	41%	0.76%
7	Complex cardiac disease with cardiomyopathy	9%	24,224	33%	1.99%
6	Observational with signs of cardiac dysfunction	7%	18,637	35%	0.78%
1	Observational with progressive cardiac dysfunction	10%	9,368	13%	2.43%
3	Significant non cardio-vascular disease	3%	8,088	13%	3.68%
8	Observational with late onset cardiac dysfunction	13%	8,037	17%	1.19%

Chatty Robots

- Chatbot computer program acts on behalf of an organisation using textual conversation
 - Customer service
 - Information acquisition
- Simple chatbots scan for keywords in an input and match to database
 - Most matching keywords
 - Most similar wording pattern
- More advanced chatbots have multiple dialogue pathways
 - Conditions to choose a pathway
 - Question and answer conversations
- Accessed by
 - Virtual assistants (Google Assistant, Amazon Alexa)
 - Messaging Apps (Facebook Messenger, WeChat)
 - Organisation's apps and websites

Chatty Robots

- Chatbots built, tested and deployed in cloud platforms
 - Oracle Cloud Platform
 - IBM Watson
- Provide NLP and AI for chatbot development



Chatty Robots

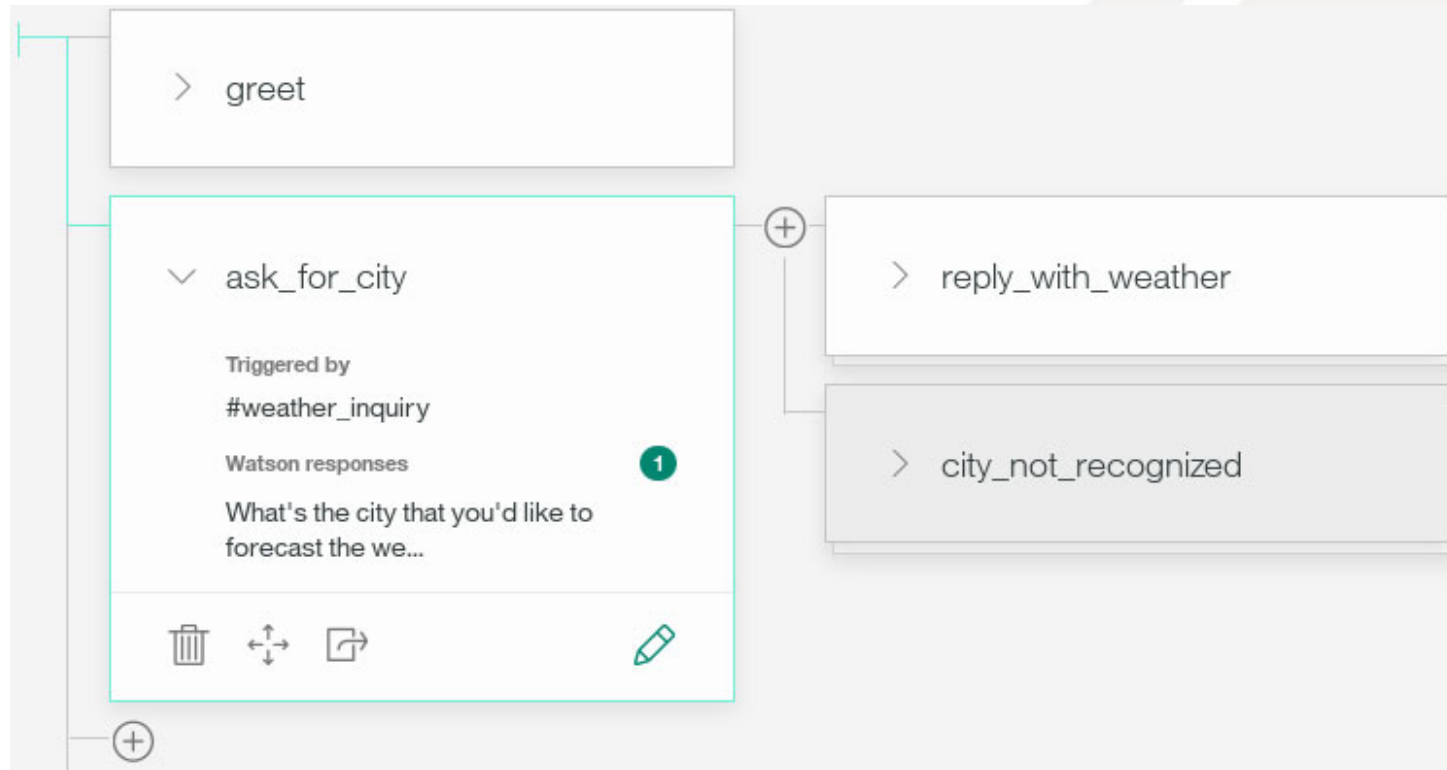
- These entities need to allow for a wide range of synonyms and patterns

Entities	@BookingType
Value	Synonyms
Plane Ticket	Air ticket, flight
Car Rental	Transport, Car,
Hotel Room	Place to stay, accommodation

Entities	@Email
Value	Patterns
Email	\b[A-Za-z0-9._%+-]+@[A-Za-z0-9.-]+\.[A-Za-z]{2,}\b

Chatty Robots

- Users unlikely to provide all required information in one pass



Chatty Robots

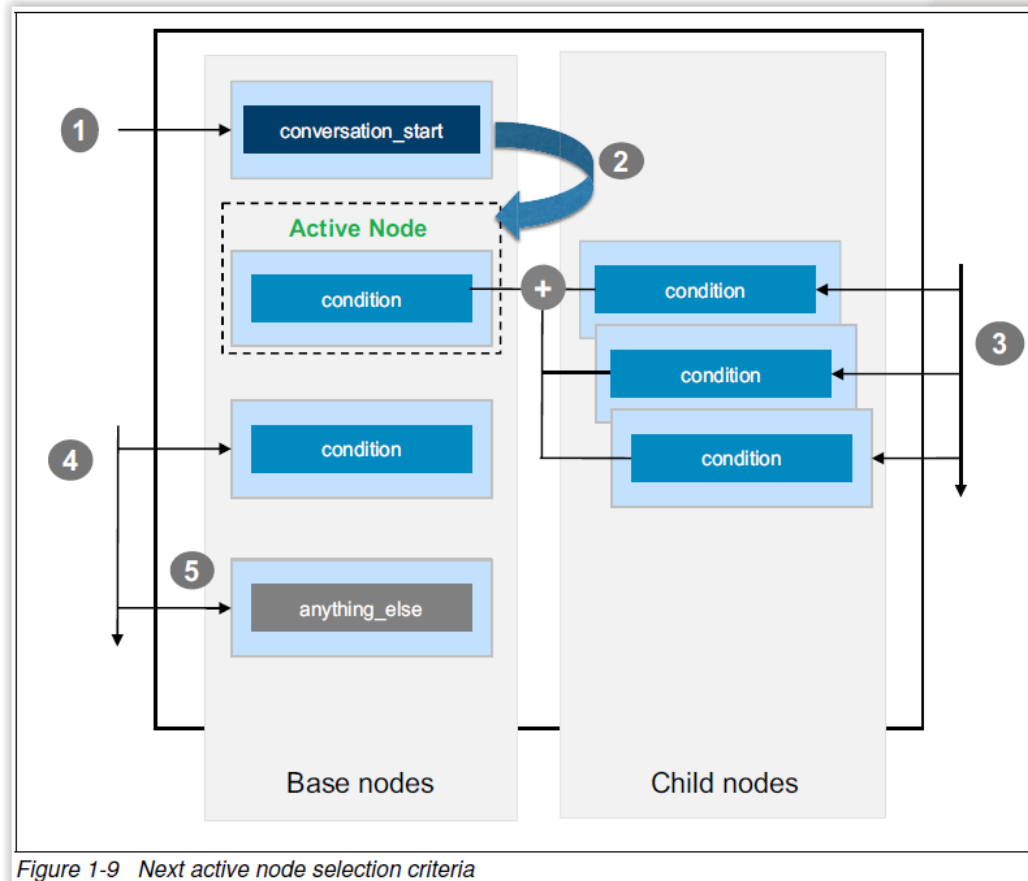
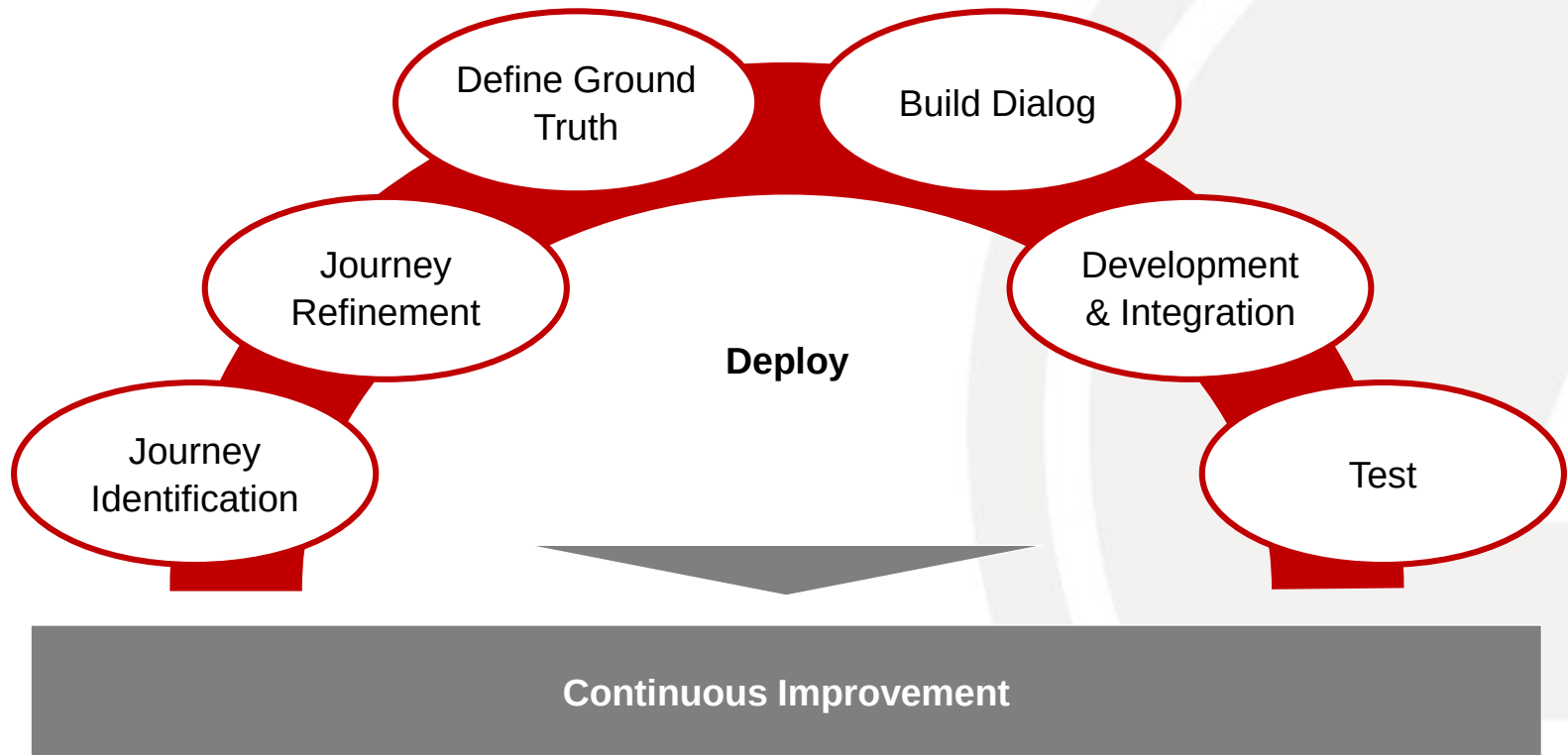


Figure 1-9 Next active node selection criteria

Chatty Robots

Lite	Standard	Premium
\$0 USD	\$0.0025 USD per API call	Custom
• Up to 10,000 API calls • Up to 5 workspaces • Up to 100 intents • Up to 25 entities	• Unlimited API queries / month • Up to 20 workspaces • Up to 2,000 intents • Up to 1,000 entities • Shared public cloud	• Higher level of security and isolation for sensitive data

Chatty Robots



Questions

- Where can I find great data science resources for NLP?
 - Python Natural Language Toolkit (NLTK)
 - Python Scikit-learn to add machine learning models to NLTK
- What courses should I go on?
 - Udemy Python for Data Science and Machine Learning Bootcamp includes chapters on NLP
 - Udemy NLP – Natural Language Processing with Python
 - Udemy Natural Language Processing with Deep Learning in Python
 - Udemy Data Science: Data Mining and Natural Language Processing in R
- What do I need to use cTakes?
 - A UMLS account
 - A license to use SNOMED-CT
- My favourite mis-translation of a sentence?
 - “The vodka is good, but the meat is rotten”
- Recent webinar that I would recommend?
 - <https://www.brighttalk.com/webinar/state-of-the-art-natural-language-understanding-in-healthcare>