Prototype implementation on 3cixty Nice KB

Rifat

Tuesday, October 25, 2016

In this implementation we presents, 3cixty Nice 2 class quality analysis report based on four quality characteristics: (i) Persistency (ii) Historical Persistency (iii) Consistency (iv) Completeness. We divided the report based on each quality characteristics. Also we presented 3cixty Nice 2 class growth analysis measure.

Quality Problem Report Overview

Input

In this implementation the datasets is presented in folder "~/ExperimentalData/3cixtyNice" where each class property count over 3cixtyNice 8 releases present in the CSV files. Also, the entity count of 2 classes presented in "3cixty2ClassEntityCount.csv" file.

```
location="~/ExperimentalData/3cixtyNice/"
```

The dataset presented in the folder is already processed and build in a intermediate data structure for the quality assessment approach.

To run the rmd file only need to specify the folder location for input data.

Basic Measure Definition

(i) Persistency of a Classe is 1 if, on the KB releases (i=1....n) , En > (En-1) else Persistency = 0

Where En is the distnct entity of a class.

(ii) Historical Persistency of a Classe is 1 if, on the KB releases (i=1....n) , Persistency = 0 else historical persistency= 0

Where En is the distnct entity of a class.

(iii) Consistency of a predicate = 1 if frequency of a predicate $f_i > 100$, where KB releases, i=1...n

else Consistency = 0 ,if $f_i < 100$.

(iv) Completeness of a predicate = 1 if normalized frequency of a predicate, $f_i > (f_{i-1})$; where Time Series, i=1....n.

else Completeness = 0

(v) For KB growth we applied a Linear regression over the KB releases (i=1....n).

From the linear regression, We calculate the normalize distance.

The normalized distance(ND) = (abs(Last TimeSeries Entity Count - Predicted Value)/mean(abs(Residuals))

So, KB growth is 1 if ND<1 or KB growth is 0 if ND>=1

Results Structure

(i) Persistency: A line graph presented to visualize variation on last two 3cixty Nice Release.

(ii) Historical Persistency: A table with this classes with persistency issues over the 8 Releases.

(iii) Consistency: Result of property values presented in a table. 1st table with incorrect property values and 2nd table with filtered property values

(iv) Completeness: Result of property values presented in a table with normalized property values and difference between two version of normalized property values.

(v) KB growth: A graph visualizing KB growth.

Experiment Results

Input: Set Entity file location such as -

```
location<- "~/ExperimentalData/DBpedia/DBpedia10ClassEntityCount.csv"
```

Class: lode:Event[http://linkedevents.org/ontology/Event]

Input: set file location such as -

```
location<- "~/ExperimentalData/3cixtyNice/lode-Event.csv"
```
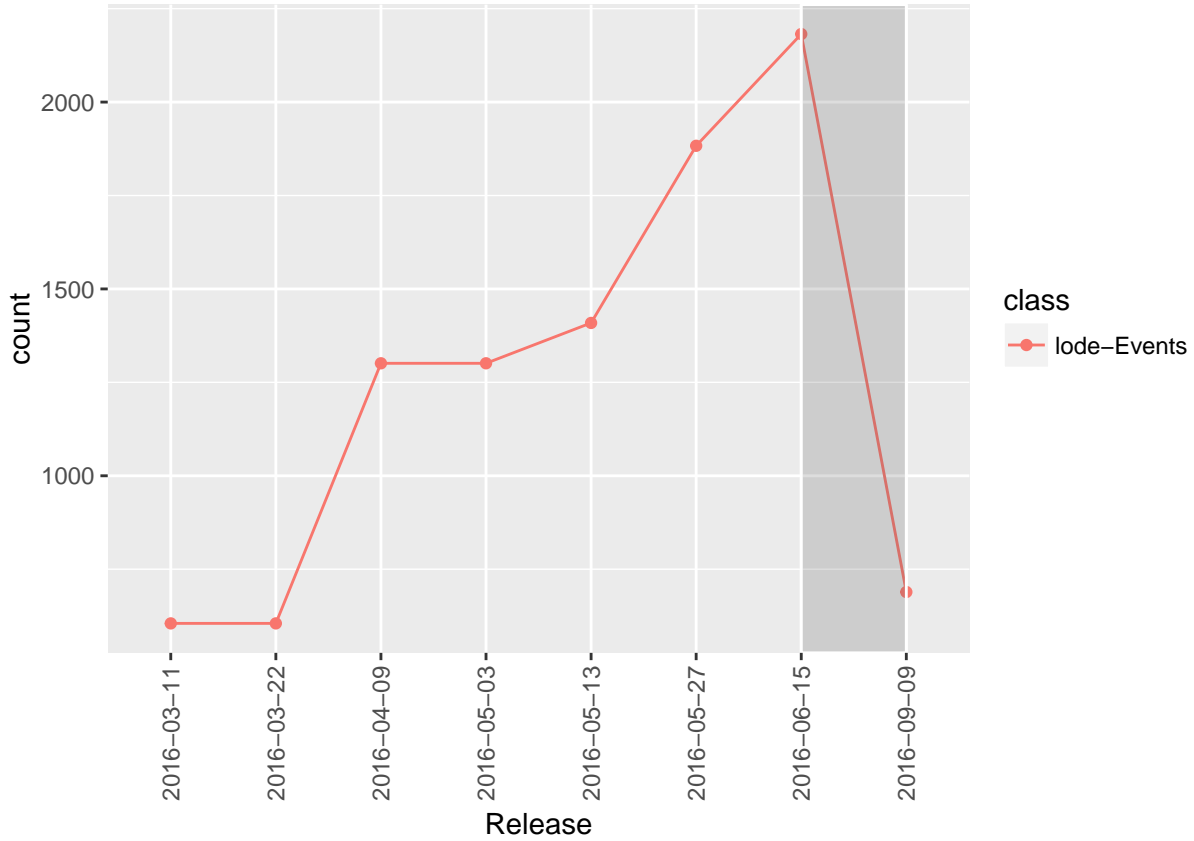
Entity Count for 11 Release:

Release	class	count
2016-03-11	lode-Events	605
2016-03-22	lode-Events	605
2016-04-09	lode-Events	1301
2016-05-03	lode-Events	1301
2016-05-13	lode-Events	1409
2016-05-27	lode-Events	1883
2016-06-15	lode-Events	2182
2016-09-09	lode-Events	689

Persistency

Persistency Measure Result:

(i) 689 value

(ii) 2182 value

(iii) Persistency= 0

Historical Persitency

Releases with persistency values:

Release	class	count	Persistency
2016-03-11	lode-Events	605	1
2016-03-22	lode-Events	605	1
2016-04-09	lode-Events	1301	1
2016-05-03	lode-Events	1301	1
2016-05-13	lode-Events	1409	1
2016-05-27	lode-Events	1883	1
2016-06-15	lode-Events	2182	1
2016-09-09	lode-Events	689	0

Historical Persistency Measure Result:

(i) Total no. of class = 2
(ii) No. of persistency=1
(iii) HistPersistencyMeasure= 85.7142857 %

Consistency

|	Release	Property	freq	count	class	NormalizedFreq
112	2016-06-15	http://linkedevents.org/ontology/atPlace	1632	2182	lode-Events	0.7479377
113	2016-06-15	http://linkedevents.org/ontology/atTime	2014	2182	lode-Events	0.9230064
114	2016-06-15	http://linkedevents.org/ontology/hasCategory	1698	2182	lode-Events	0.7781852
116	2016-06-15	http://linkedevents.org/ontology/involvedAgent	266	2182	lode-Events	0.1219065
118	2016-06-15	http://purl.org/dc/elements/1.1/description	2024	2182	lode-Events	0.9275894
126	2016-06-15	http://www.w3.org/2000/01/rdf-schema#seeAlso	851	2182	lode-Events	0.3900092
127	2016-06-15	http://www.w3.org/2002/07/owl#sameAs	1777	2182	lode-Events	0.8143905
128	2016-06-15	http://purl.org/dc/elements/1.1/identifier	770	2182	lode-Events	0.3528873
131	2016-09-09	http://linkedevents.org/ontology/atPlace	424	689	lode-Events	0.6153846
132	2016-09-09	http://linkedevents.org/ontology/atTime	490	689	lode-Events	0.7111756
133	2016-09-09	http://linkedevents.org/ontology/hasCategory	584	689	lode-Events	0.8476052
135	2016-09-09	http://linkedevents.org/ontology/involvedAgent	42	689	lode-Events	0.0609579
136	2016-09-09	http://linkedevents.org/ontology/poster	380	689	lode-Events	0.5515239
137	2016-09-09	http://purl.org/dc/elements/1.1/description	650	689	lode-Events	0.9433962
138	2016-09-09	http://purl.org/dc/elements/1.1/issued	684	689	lode-Events	0.9927431
144	2016-09-09	http://www.w3.org/2000/01/rdf-schema#label	686	689	lode-Events	0.9956459
145	2016-09-09	http://www.w3.org/2000/01/rdf-schema#seeAlso	334	689	lode-Events	0.4847605
146	2016-09-09	http://www.w3.org/2002/07/owl#sameAs	439	689	lode-Events	0.6371553
147	2016-09-09	http://purl.org/dc/elements/1.1/identifier	453	689	lode-Events	0.6574746

Completeness

(i) No. of incomplete properties= 8
(ii) List of Incomplete properties:

Property	Release20160909	Release20160615	NormFreqDiff
http://linkedevents.org/ontology/atPlace	0.6153846	0.7479377	-0.1325531
http://linkedevents.org/ontology/atTime	0.7111756	0.9230064	-0.2118308
http://linkedevents.org/ontology/involvedAgent	0.0609579	0.1219065	-0.0609486
http://linkedevents.org/ontology/poster	0.5515239	1.0022915	-0.4507675
http://purl.org/dc/elements/1.1/issued	0.9927431	1.1342805	-0.1415374

Property	Release20160909	Release20160615	NormFreqDiff
http://purl.org/dc/elements/1.1/subject	1.4804064	1.7241063	-0.2436999
http://www.w3.org/2000/01/rdf-schema#label	0.9956459	1.0032081	-0.0075622
http://www.w3.org/2002/07/owl#sameAs	0.6371553	0.8143905	-0.1772352

KB growth
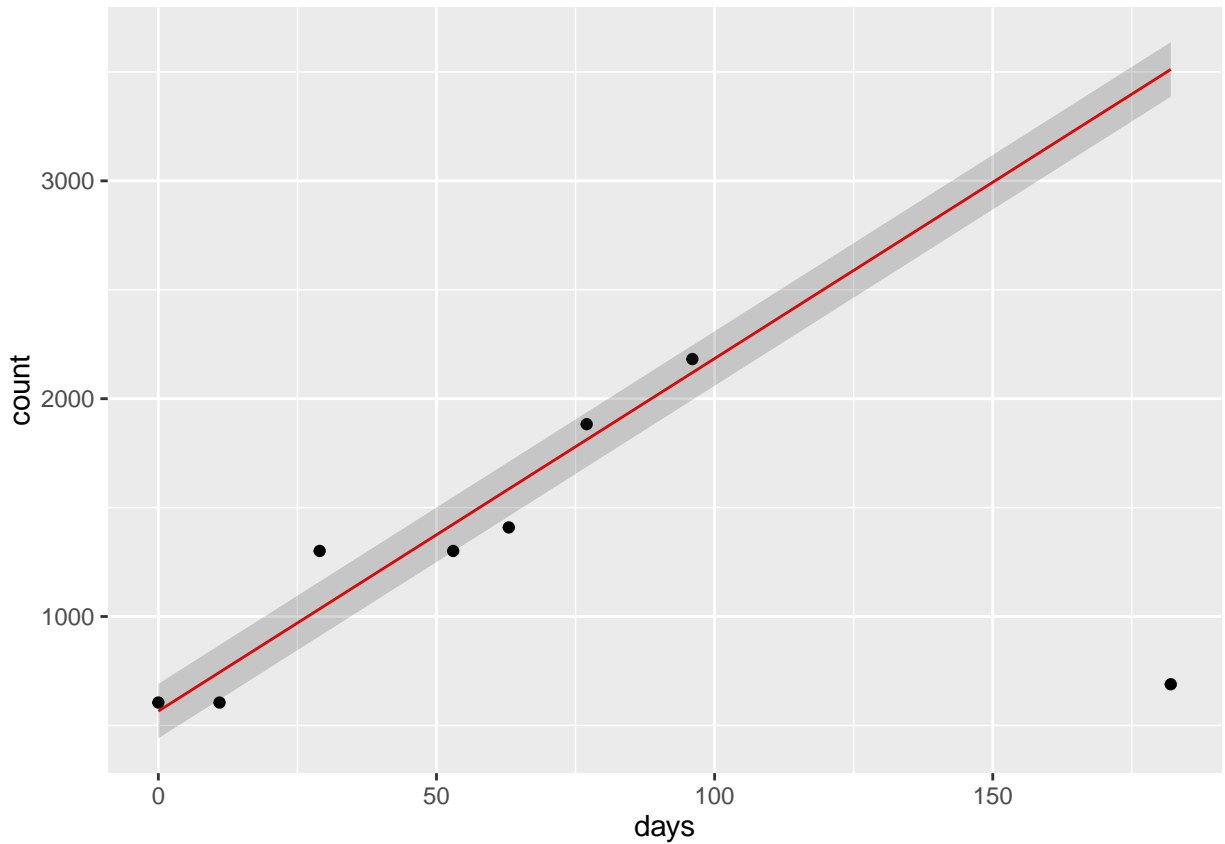
The normalized distance(ND)=(abs(Last TimeSeries Value - Predicted Value)/mean(abs(Residuals)))= 22.5481796

KB growth is 1 if ND<1 or KB growth is 0 if ND>=1

KB growth=0

Class: dul:Place[http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#Place]

Entity Count for 11 Release:

	Release	class	count
9	2016-03-11	dul-Places	20692
10	2016-03-22	dul-Places	20692
11	2016-04-09	dul-Places	27858
12	2016-05-03	dul-Places	26066
13	2016-05-13	dul-Places	26827

	Release	class	count
14	2016-05-27	dul-Places	25828
15	2016-06-15	dul-Places	41018
16	2016-09-09	dul-Places	44968

Persistency

Persistency Measure Result:

(i) 44968 value

(ii) 41018 value

(iii) Persistency= 1
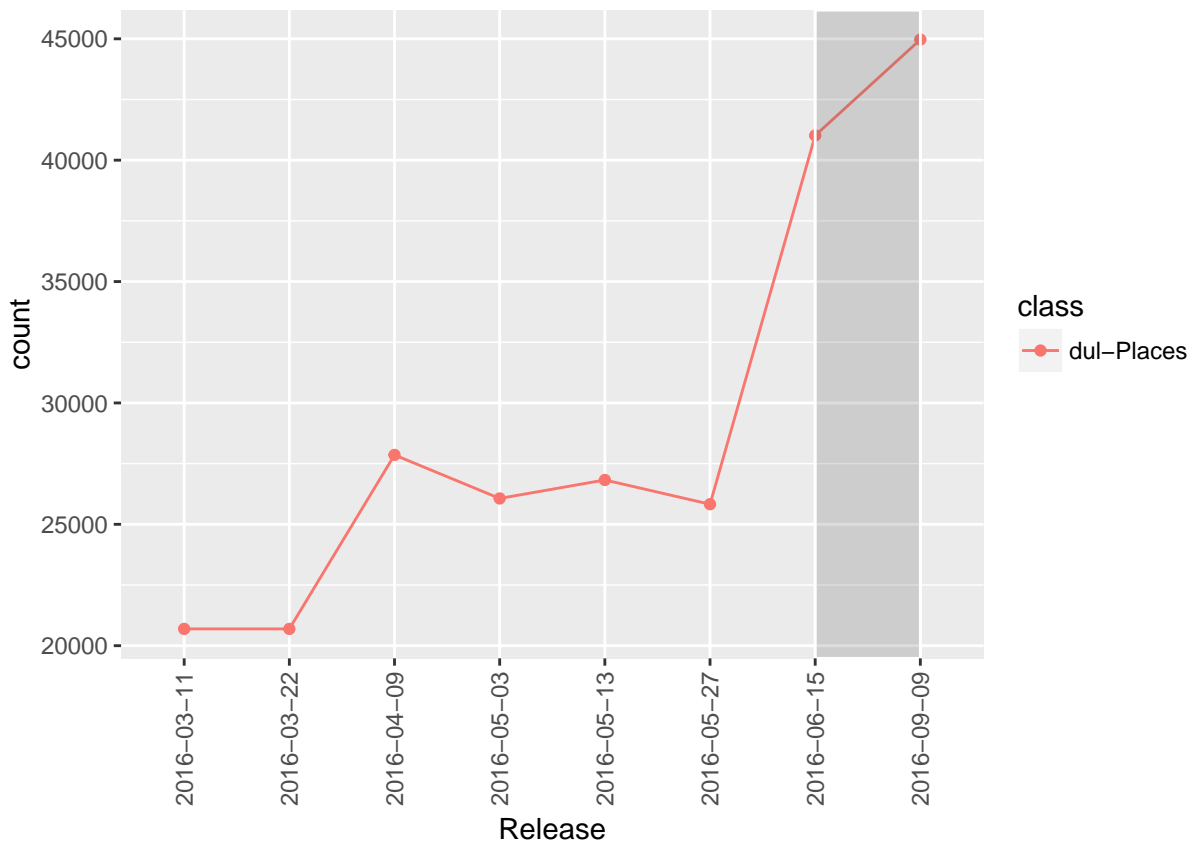
Historical Persitency

Releases with persistency values:

	Release	class	count	Persistency
9	2016-03-11	dul-Places	20692	1
10	2016-03-22	dul-Places	20692	1
11	2016-04-09	dul-Places	27858	1
12	2016-05-03	dul-Places	26066	0
13	2016-05-13	dul-Places	26827	1

	Release	class	count	Persistency
14	2016-05-27	dul-Places	25828	0
15	2016-06-15	dul-Places	41018	1
16	2016-09-09	dul-Places	44968	1

Historical Persistency Measure Result:

  (i) Total no. of class = 2
 (ii) No. of persistency=2
(iii) HistPersistencyMeasure= 71.4285714 %

Consistency

	Release	Property	freq	count	class	NormalizedFreq
142	2016-06-15	http://linkedevents.org/ontology/poster	37570	41018	dul-Places	0.9159393
145	2016-06-15	http://schema.org/aggregateRating	1280	41018	dul-Places	0.0312058
146	2016-06-15	http://schema.org/description	435	41018	dul-Places	0.0106051
147	2016-06-15	http://schema.org/image	23470	41018	dul-Places	0.5721878
148	2016-06-15	http://schema.org/interactionCount	7607	41018	dul-Places	0.1854552
149	2016-06-15	http://schema.org/location	40903	41018	dul-Places	0.9971964
150	2016-06-15	http://schema.org/openingHours	6459	41018	dul-Places	0.1574675
151	2016-06-15	http://schema.org/priceRange	67	41018	dul-Places	0.0016334
152	2016-06-15	http://schema.org/review	9564	41018	dul-Places	0.2331659
153	2016-06-15	http://schema.org/telephone	33143	41018	dul-Places	0.8080111
154	2016-06-15	http://schema.org/url	9960	41018	dul-Places	0.2428202
160	2016-06-15	http://www.w3.org/2002/07/owl#sameAs	39743	41018	dul-Places	0.9689161
162	2016-06-15	http://data.ign.fr/def/topo#importance	66	41018	dul-Places	0.0016090
163	2016-06-15	http://vocab.org/transit/terms/route	2320	41018	dul-Places	0.0565605
164	2016-06-15	http://www.w3.org/2000/01/rdf-schema#seeAlso	122	41018	dul-Places	0.0029743
168	2016-09-09	http://linkedevents.org/ontology/poster	8323	44968	dul-Places	0.1850872
171	2016-09-09	http://schema.org/aggregateRating	1720	44968	dul-Places	0.0382494
172	2016-09-09	http://schema.org/description	600	44968	dul-Places	0.0133428
173	2016-09-09	http://schema.org/image	25290	44968	dul-Places	0.5623999

|	Release	Property	freq	count	class	NormalizedFreq
174	2016-09-09	http://schema.org/interactionCount	11844	44968	dul-Places	0.2633873
175	2016-09-09	http://schema.org/location	44692	44968	dul-Places	0.9938623
176	2016-09-09	http://schema.org/openingHours	6281	44968	dul-Places	0.1396771
177	2016-09-09	http://schema.org/priceRange	60	44968	dul-Places	0.0013343
178	2016-09-09	http://schema.org/review	10697	44968	dul-Places	0.2378803
179	2016-09-09	http://schema.org/telephone	35371	44968	dul-Places	0.7865816
180	2016-09-09	http://schema.org/url	9982	44968	dul-Places	0.2219801
186	2016-09-09	http://www.w3.org/2002/07/owl#sameAs	43538	44968	dul-Places	0.9681996
188	2016-09-09	http://data.ign.fr/def/topo#importance	67	44968	dul-Places	0.0014899
189	2016-09-09	http://vocab.org/transit/terms/route	2523	44968	dul-Places	0.0561066
190	2016-09-09	http://www.w3.org/2000/01/rdf-schema#seeAlso	124	44968	dul-Places	0.0027575

Completeness

(i) No. of incomplete properties= 14

(ii) List of Incomplete properties:

Property	Release20160909	Release20160615	NormFreqDiff
http://data.ign.fr/def/topo#importance	0.0014899	0.0016090	-0.0001191
http://data.linkedevents.org/def/location#businessType	2.7888721	2.8952167	-0.1063446
http://data.linkedevents.org/def/location#businessTypeTop	1.0357365	1.0871081	-0.0513716
http://linkedevents.org/ontology/poster	0.1850872	0.9159393	-0.7308522
http://schema.org/image	0.5623999	0.5721878	-0.0097879
http://schema.org/location	0.9938623	0.9971964	-0.0033341
http://schema.org/openingHours	0.1396771	0.1574675	-0.0177903
http://schema.org/priceRange	0.0013343	0.0016334	-0.0002991
http://schema.org/telephone	0.7865816	0.8080111	-0.0214295
http://schema.org/url	0.2219801	0.2428202	-0.0208402
http://vocab.org/transit/terms/route	0.0561066	0.0565605	-0.0004540
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	2.9716910	2.9719635	-0.0002725
http://www.w3.org/2000/01/rdf-schema#seeAlso	0.0027575	0.0029743	-0.0002168
http://www.w3.org/2002/07/owl#sameAs	0.9681996	0.9689161	-0.0007165

KB growth
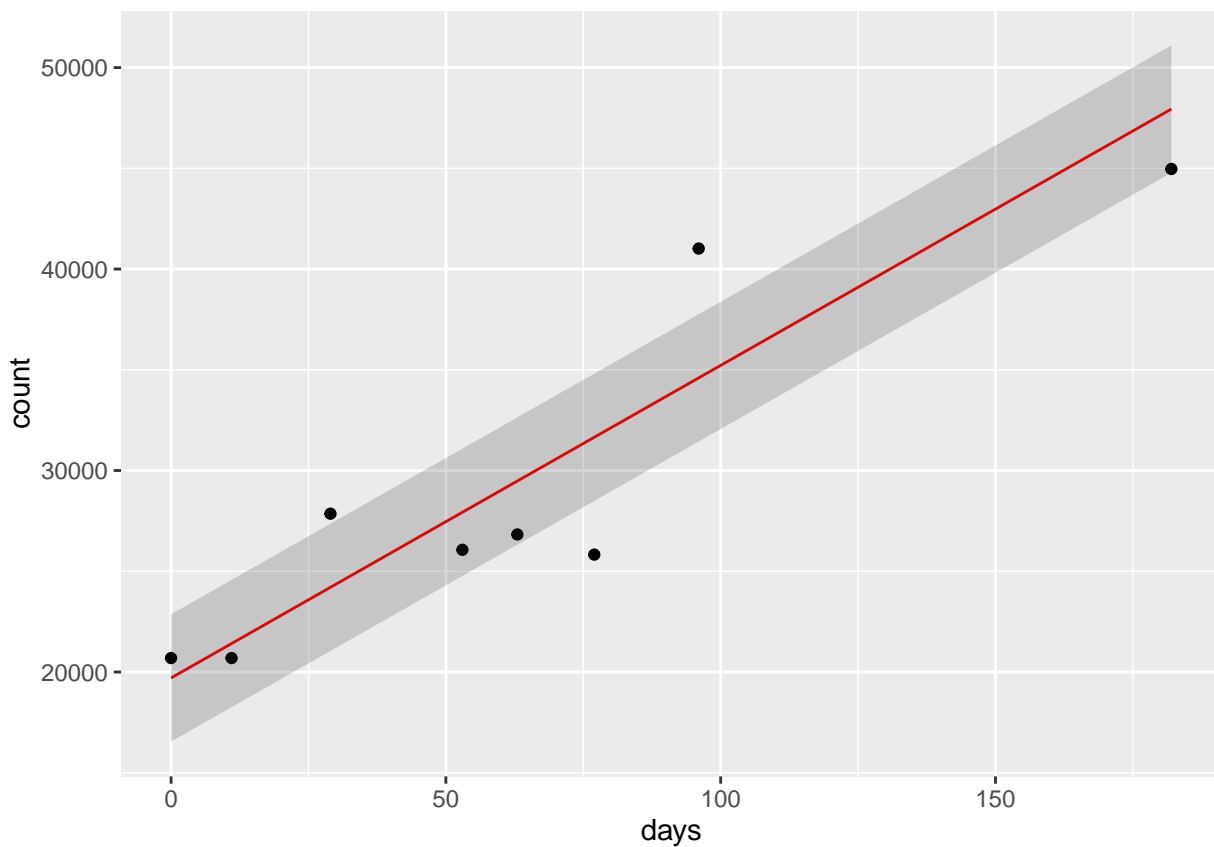
The normalized distance(ND)=(abs(Last TimeSeries Value - Predicted Value)/mean(abs(Residuals))= 0.9411357

KB growth is 1 if ND<1 or KB growth is 0 if ND>=1

KB growth=1