

Probabilistic Modeling of Chronological Dates to Serve Machines and Scholars

Andreas Habring, Angelos Nicolaou, Daniel Luger, Florian
Atzenhofer-Baumgartner, Florian Lamminger, Franziska
Decker, Tamás Kovács, Sandy Aoun Georg Vogeler, Martin
Holler



July 14, 2023

1

Probabilistic Modeling of dates

- Our Data
- Roles
- **Ambiguity Modeling**
- Conclusion

The data that Inspired us

**UNI
GRAZ**

Probabilistic
Modeling of
Dates

A. Habring,
A. Nicolaou,
et al.

Probabilistic
Modeling of
dates

Our Data

Roles

Ambiguity Modeling

Conclusion

- 500K Documents
- (CEI) TEI-4 derived format
- Diplomatic Charters
- 1000 Charters randomly sub-sampled
- 2211 Dates associated with the issued



- ▶ Is it YYYYMMDD?
- ▶ Or DDMMYYYY?
- ▶ YYYYMMDD??

- ▶ Is it YYYYMMDD?
- ▶ Or DDMMYYYY?
- ▶ YYYYMMDD??

- ▶ (date)
- ▶ [date]
- ▶ Um?
- ▶ 99?

"13690101"	"VIII. - XI. Jahrh."
"1397 August 1"	"1454, únor 12."
"8. April 1587"	"[1711]"
"1654-12-18"	"12599999"
"Saec. XIV"	"11. Jänner 1362"
"24.10.1753"	"13019999"
"1465-00-00"	"99999999"
"14110329"	"(15. storočie)"
"c.1229"	"wohl 29.09.1565"
"St. Elisabeth"	"1321 XII 6"
"1671,květen 18."	"99999900"
"feria sexta post Jacobi apostoli"	
"zwischen 1578 und 1590"	
"9730911"	"9999"
"Um 1290"	"VIII. - XI. Jahrh."
"(1410-1420)"	"(1601)"
"1301 feb. 11"	

- ▶ Is it YYYYMMDD?
- ▶ Or DDMMYYYY?
- ▶ YYYYMMDD??

- ▶ Is it YYYYMMDD?
- ▶ Or DDMMYYYY?
- ▶ YYYYMMDD??

- ▶ (date)
- ▶ [date]
- ▶ Um?
- ▶ 99?

"13690101"	"VIII. - XI. Jahrh."
"1397 August 1"	"1454, únor 12."
"8. April 1587"	"[1711]"
"1654-12-18"	"12599999"
"Saec. XIV"	"11. Jänner 1362"
"24.10.1753"	"13019999"
"1465-00-00"	"99999999"
"14110329"	"(15. storočie)"
"c.1229"	"wohl 29.09.1565"
"St. Elisabeth"	"1321 XII 6"
"1671,květen 18."	"99999900"
"feria sexta post Jacobi apostoli"	
"zwischen 1578 und 1590"	
"9730911"	"9999"
"Um 1290"	"VIII. - XI. Jahrh."
"(1410-1420)"	"(1601)"
"1301 feb. 11"	

How to make sense of date data?

■ Just Creative

- ▶ Is it YYYYMMDD?
- ▶ Or DDMMYYYY?
- ▶ YYYYMMDD??

■ Numerical

- ▶ Is it YYYYMMDD?
- ▶ Or DDMMYYYY?
- ▶ YYYYMMDD??

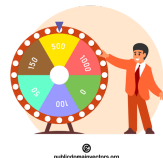
■ Expressing Ambiguity

- ▶ (date)
- ▶ [date]
- ▶ Um?
- ▶ 99?

- Expresses known facts about dates
- Needs to express his nuanced knowledge
- Needs means of expressing nuance
- A humanist in a DH paradigm
- Afraid of being wrong
- Someone who knows when something occurred. Or knows the

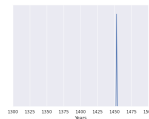


- Expresses opinion/estimation
- Typically a machine learning model
 - ▶ Loss function needed to train
 - ▶ Differentiable
- Could be a human
- Opinion model
 - ▶ A choice among fixed categories
 - ▶ A moment
 - ▶ An interval
 - ▶ A Gaussian
 - ▶ A Monte Carlo Approximation

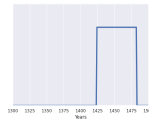


- | | |
|-----------|---|
| 1300-1350 | |
| 1350-1400 | X |
| 1400-1450 | |
| 1450-1500 | |

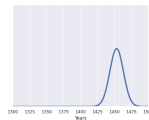
- Expresses opinion/estimation
- Typically a machine learning model
 - ▶ Loss function needed to train
 - ▶ Differentiable
- Could be a human
- Opinion model
 - ▶ A choice among fixed categories
 - ▶ A moment
 - ▶ An interval
 - ▶ A Gaussian
 - ▶ A Monte Carlo Approximation



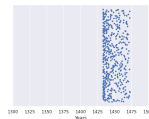
- Expresses opinion/estimation
- Typically a machine learning model
 - ▶ Loss function needed to train
 - ▶ Differentiable
- Could be a human
- Opinion model
 - ▶ A choice among fixed categories
 - ▶ A moment
 - ▶ An interval
 - ▶ A Gaussian
 - ▶ A Monte Carlo Approximation



- Expresses opinion/estimation
- Typically a machine learning model
 - ▶ Loss function needed to train
 - ▶ Differentiable
- Could be a human
- Opinion model
 - ▶ A choice among fixed categories
 - ▶ A moment
 - ▶ An interval
 - ▶ A Gaussian
 - ▶ A Monte Carlo Approximation



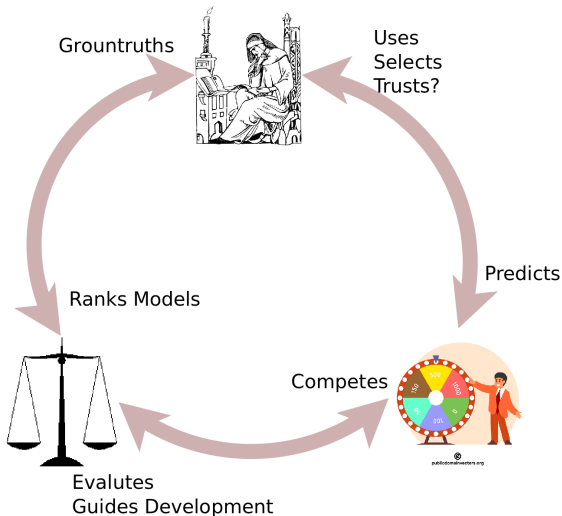
- Expresses opinion/estimation
- Typically a machine learning model
 - ▶ Loss function needed to train
 - ▶ Differentiable
- Could be a human
- Opinion model
 - ▶ A choice among fixed categories
 - ▶ A moment
 - ▶ An interval
 - ▶ A Gaussian
 - ▶ A Monte Carlo Approximation



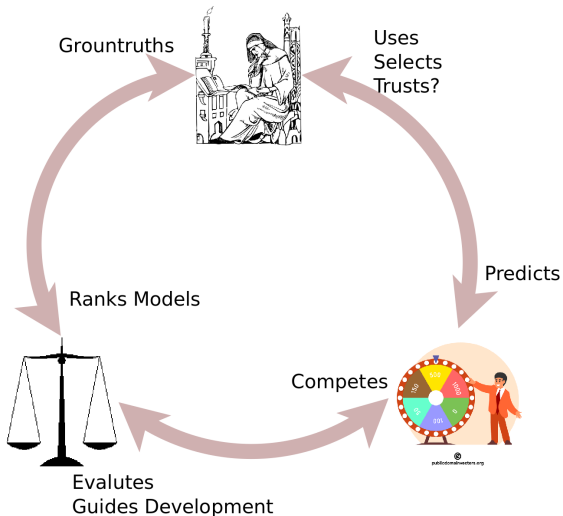
- Performance Evaluation
- Scoring a Guesser
- "Metric" that satisfies our perceived notion of performance
 - ▶ Range in $[0, 1]$
 - ▶ $A > B \wedge B > C \implies A > C$
 - ▶ Can be asymmetric
- Must be Winnable
- Must not be Gameable



- Dates
- Scores



- Dates
- Scores



Ambiguity: Not before, Not after

Probabilistic
Modeling of
Dates

A. Habring,
A. Nicolaou,
et al.

Probabilistic
Modeling of
dates

Our Data

Roles

Ambiguity Modeling

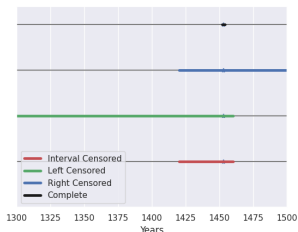
Conclusion

- AKA: From - To
- Ideal for scholars
- Internal Charter Features
- Statistics
 - ▶ Complete
 - ▶ Interval censored
 - ▶ Left censored
 - ▶ Right censored
 - ▶ Used for modeling regular phenomena
- Probability: Sum 1
- Plausibility: Max 1



Ambiguity: Not before, Not after

- AKA: From - To
- Ideal for scholars
- Internal Charter Features
- Statistics
 - ▶ Complete
 - ▶ Interval censored
 - ▶ Left censored
 - ▶ Right censored
 - ▶ Used for modeling regular phenomena
- Probability: Sum 1
- Plausibility: Max 1



Ambiguity: Not before, Not after

Probabilistic
Modeling of
Dates

A. Habring,
A. Nicolaou,
et al.

Probabilistic
Modeling of
dates

Our Data

Roles

Ambiguity Modeling

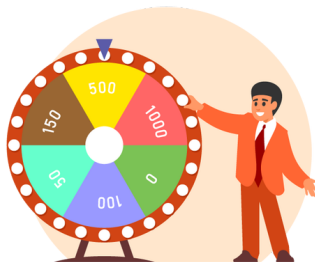
Conclusion

- AKA: From - To
- Ideal for scholars
- Internal Charter Features
- Statistics
 - ▶ Complete
 - ▶ Interval censored
 - ▶ Left censored
 - ▶ Right censored
 - ▶ Used for modeling regular phenomena
- Probability: Sum 1
- Plausibility: Max 1



Ambiguity: Not before, Not after

- AKA: From - To
- Ideal for scholars
- Internal Charter Features
- Statistics
 - ▶ Complete
 - ▶ Interval censored
 - ▶ Left censored
 - ▶ Right censored
 - ▶ Used for modeling regular phenomena
- Probability: Sum 1
- Plausibility: Max 1




publicdomainvectors.org

Ambiguity: Not before, Not after

**UNI
GRAZ**

Probabilistic
Modeling of
Dates

A. Habring,
A. Nicolaou,
et al.

Probabilistic
Modeling of
dates

Our Data

Roles

Ambiguity Modeling

Conclusion

- AKA: From - To
- Ideal for scholars
- Internal Charter Features
- Statistics
 - ▶ Complete
 - ▶ Interval censored
 - ▶ Left censored
 - ▶ Right censored
 - ▶ Used for modeling regular phenomena

■ Probability: Sum 1

■ Plausibility: Max 1



Ambiguity: Give or Take

Probabilistic
Modeling of
Dates

A. Habring,
A. Nicolaou,
et al.

Probabilistic
Modeling of
dates

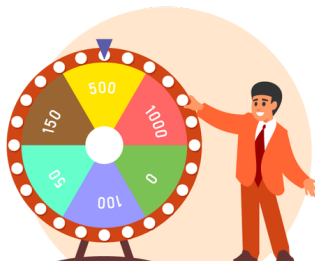
Our Data

Roles

Ambiguity Modeling

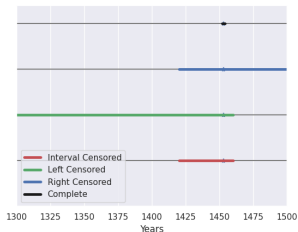
Conclusion

- AKA: From - To
- Better suited for guessers
- External Charter Features
- Statistics
 - ▶ Complete
 - ▶ Interval censored
 - ▶ Left censored
 - ▶ Right censored
 - ▶ Used for modeling regular phenomena
- Probability: Sum 1
- Plausibility: Max 1



 publicdomainvectors.org

- AKA: From - To
- Better suited for guessers
- External Charter Features
- Statistics
 - ▶ Complete
 - ▶ Interval censored
 - ▶ Left censored
 - ▶ Right censored
 - ▶ Used for modeling regular phenomena
- Probability: Sum 1
- Plausibility: Max 1



Ambiguity: Give or Take

Probabilistic
Modeling of
Dates

A. Habring,
A. Nicolaou,
et al.

Probabilistic
Modeling of
dates

Our Data

Roles

Ambiguity Modeling

Conclusion

- AKA: From - To
- Better suited for guessers
- External Charter Features
- Statistics
 - ▶ Complete
 - ▶ Interval censored
 - ▶ Left censored
 - ▶ Right censored
 - ▶ Used for modeling regular phenomena
- Probability: Sum 1
- Plausibility: Max 1



Ambiguity: Give or Take

Probabilistic
Modeling of
Dates

A. Habring,
A. Nicolaou,
et al.

Probabilistic
Modeling of
dates

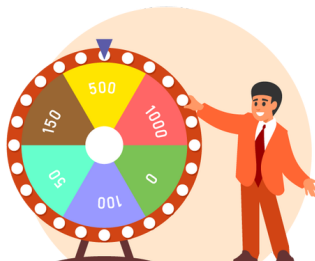
Our Data

Roles

Ambiguity Modeling

Conclusion

- AKA: From - To
- Better suited for guessers
- External Charter Features
- Statistics
 - ▶ Complete
 - ▶ Interval censored
 - ▶ Left censored
 - ▶ Right censored
 - ▶ Used for modeling regular phenomena
- Probability: Sum 1
- Plausibility: Max 1



 publicdomainvectors.org

Ambiguity: Give or Take

- AKA: From - To
- Better suited for guessers
- External Charter Features
- Statistics
 - ▶ Complete
 - ▶ Interval censored
 - ▶ Left censored
 - ▶ Right censored
 - ▶ Used for modeling regular phenomena
- Probability: Sum 1
- Plausibility: Max 1



- Everything is an ambiguous instant
- Opinions must be interpretable as DF
- Uniform (flat) and Gaussian (bell-shaped) are expressive enough for most humans
- The "role" (use-case) dictates DF normalisation
- Performance evaluation should not be favoring any guesser type
- Anything naturally lengthy should be modeled as two or more moments
- Life \implies (Birth, Death)

- ◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ▶ ↺ 🔍 ↻

- Everything is an ambiguous instant
- Opinions must be interpretable as DF
- Uniform (flat) and Gaussian (bell-shaped) are expressive enough for most humans
- The "role" (use-case) dictates DF normalisation
- Performance evaluation should not be favoring any guesser type
- Link: $to - from \approx 2\sigma$
- Open intervals mean everything goes

- Everything is an ambiguous instant
- Opinions must be interpretable as DF
- Uniform (flat) and Gaussian (bell-shaped) are expressive enough for most humans
- The "role" (use-case) dictates DF normalisation
- Performance evaluation should not be favoring any guesser type
- Plausibility DF: "there must be a perfect prediction"
- Probability DF: "we all get as many guesses"
- Precision:
 - ▶ Prediction correctness
 - ▶ Quality of a prediction
- Recall:
 - ▶ Precision with inverted roles
 - ▶ Prediction difficulty

- Everything is an ambiguous instant
- Opinions must be interpretable as DF
- Uniform (flat) and Gaussian (bell-shaped) are expressive enough for most humans
- The "role" (use-case) dictates DF normalisation
- Performance evaluation should not be favoring any guesser type
- How can we know if a better method ever arrives?
- Apples and Oranges are Fruit!

Questions/Remarks/Objections?

**UNI
GRAZ**

Probabilistic
Modeling of
Dates

A. Habring,
A. Nicolaou,
et al.

Probabilistic
Modeling of
dates

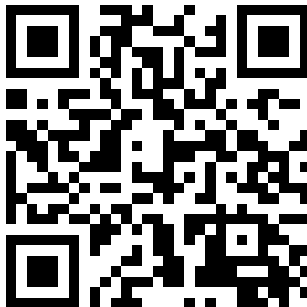
Our Data

Roles

Ambiguity Modeling

Conclusion

Thank you!



Look at our UI proof of concept demo! Slides are also there!