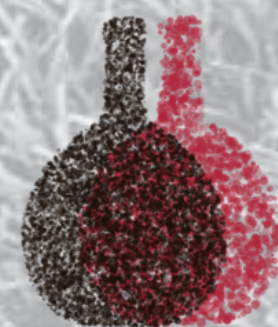


Predicting the potential of soccer players

Kenneth Verstraete, Tom Decroos, Bruno Coussement
Nick Vannieuwenhoven, Jesse Davis



ECMLPKDD
Würzburg 2019

KU LEUVEN

Outline

- Goal
- Data + tensors 101
- Data analysis with tensors
- Data challenges
- Tensor-based method for predictions
- Conclusions

Contributions

PS4



Data from the FIFA games

33 skills

Updated regularly

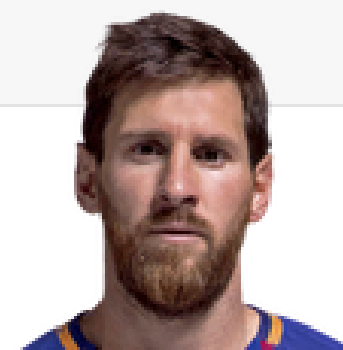
Temporal analyses

Online (SoFIFA)

±18.000 players

Real Overall Rating

	LS	ST	RS	
	88+2	88+2	88+2	
LW	LF	CF	RF	RW
92+2	93+2	93+2	93+2	92+2
	LAM	CAM	RAM	
	93+2	93+2	93+2	
LM	LCM	CM	RCM	RM
91+2	84+2	84+2	84+2	91+2
LWB	LDM	CDM	RDM	RWB
64+2	61+2	61+2	61+2	64+2
LB	LCB	CB	RCB	RB
59+2	47+2	47+2	47+2	59+2



L. Messi(ID: 158023)

Lionel Messi (Lionel Andrés Messi Cuccittini) CF RW ST Age 31 (Jun 24, 1987) 5'7" 159lbs

Overall Rating **94**

Potential **94**

Value €110.5M

Wage €565K

Preferred Foot Left

International Reputation **5★**

Weak Foot **4★**

Skill Moves **4★**

Work Rate Medium/ Medium

Body Type Messi

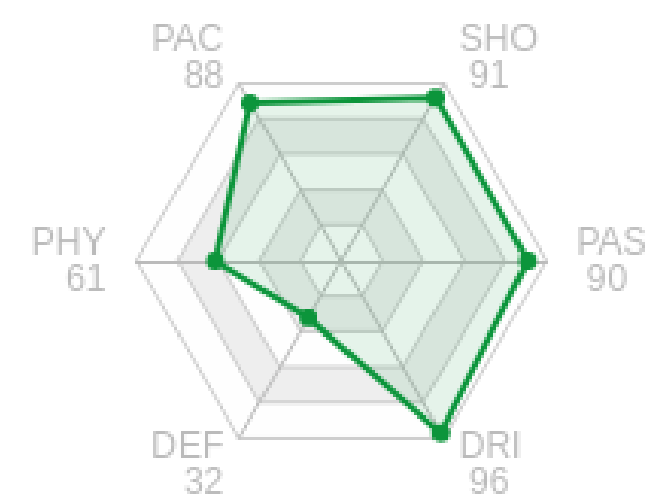
Real Face Yes

Release Clause €226.5M

#Dribbler #Distance Shooter #FK Specialist #Acrobat #Clinical Finisher #Complete Forward

Follow(758) Like(632) Dislike(197)

History Versio



FC Barcelona

86 ★★★★★

Position RW

Jersey Number 10

Joined Jul 1, 2004

Contract Valid Until 2021

Argentina

81 ★★★★★

Position RF

Jersey Number 10

Attacking

- 84** +7 Crossing
- 95** Finishing
- 70** Heading Accuracy
- 90** Short Passing
- 86** Volleys

Mentality

- 48** Aggression
- 22** Interceptions
- 94** Positioning
- 94** Vision
- 75** Penalties
- 96** Composure

Skill

- 97** Dribbling
- 93** Curve
- 94** FK Accuracy
- 87** Long Passing
- 96** Ball Control

Defending

- 33** Marking
- 28** Standing Tackle
- 26** Sliding Tackle

Movement

- 91** Acceleration
- 86** Sprint Speed
- 91** Agility
- 95** Reactions
- 95** Balance

Goalkeeping

- 6** GK Diving
- 11** GK Handling
- 15** GK Kicking
- 14** GK Positioning
- 8** GK Reflexes

Power

- 85** Shot Power
- 68** Jumping
- 72** Stamina
- 59** Strength
- 94** Long Shots

Notes

- Finesse Shot
- Long Shot Taker (CPU AI Only)
- Speed Dribbler (CPU AI Only)
- Playmaker (CPU AI Only)
- One Club Player
- Chip Shot (CPU AI Only)

Scrapping the data

Extract **only**
necessary
information

Scrape the
information of
each player

SoFIFA

Search Player ...

Players Teams Squads Shortlists Discussions ...

Language / 语言 Login

FIFA 19 Jan 21, 2019 Hot All Added Edited Removed Free Agents On Loan Compare Basket(0)

Cookies help us improve our web content and deliver a personalized experience. By using this website, you agree to our use of cookies.

Veeam Availability Suite

Veeam Software

Includes Veeam Backup & Replication and Veeam ONE. Try now!

OPEN

Search

Continents

Nationality / Region

Leagues

Teams

15 Age 45

0 Overall Rating 99

0 Potential 99

Position

Preferred Foot

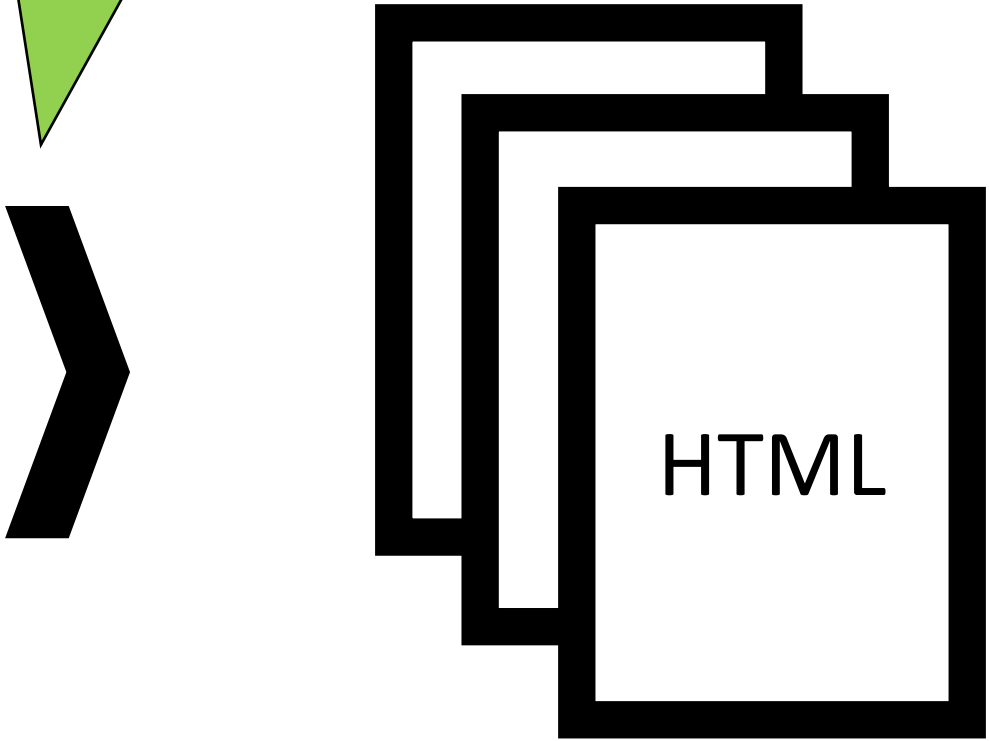
Weak Foot

All Players 0 - 60

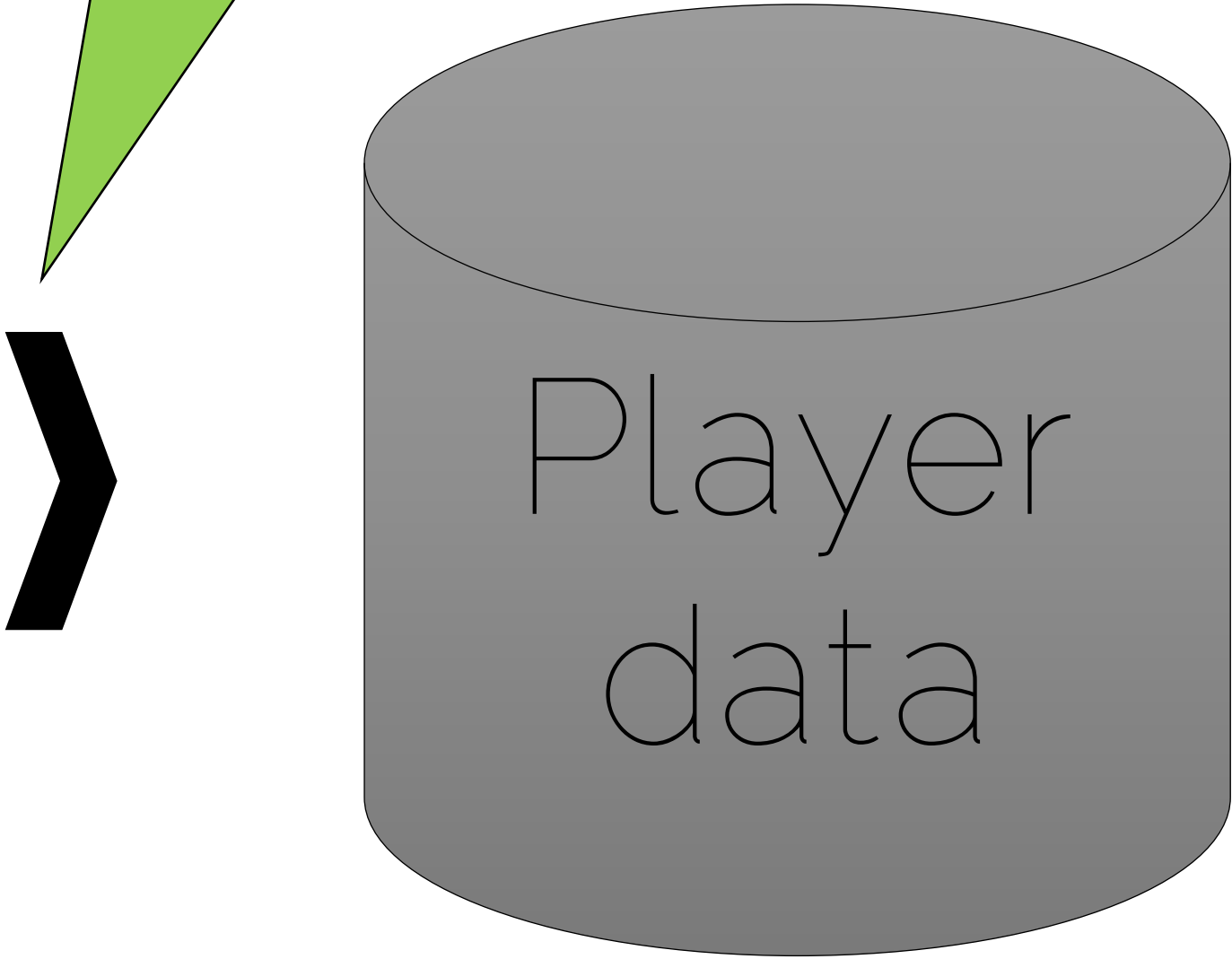
Toggle Columns: Basic Attacking Skill Movement Power Mentality Defending Goalkeeping Special

	NAME	AGE	OVA	POT	TEAM& CONTRACT	VALUE	WAGE	TOT...	HITS / COMME...
	L. Messi	31	94	94	FC Barcelona 2004 ~ 2021	€110.5M	€565K	2206	0.7K / 33.4K
	Cristiano Ronaldo	33	94	94	Juventus 2018 ~ 2022	€77M	€405K	2228	0.6K / 40K
	Neymar Jr	26	92	93	Paris Saint-Germain 2017 ~ 2022	€118.5M	€290K	2144	0.5K / 22.2K
	De Gea	27	91	92	Manchester United 2011 ~ 2020	€69M	€260K	1478	0.2K / 7.9K
	K. De Bruyne	27	91	92	Manchester City 2015 ~ 2023	€102M	€355K	2281	0.3K / 7.4K
	E. Hazard	27	91	91	Chelsea 2012 ~ 2020	€93M	€340K	2142	0.4K / 11.3K
	L. Modrić	32	91	91	Real Madrid 2012 ~ 2020	€67M	€420K	2280	0.2K / 2.8K
	L. Suárez	31	91	91	FC Barcelona 2014 ~ 2021	€80M	€455K	2349	0.3K / 1.8K
	Sergio Ramos	32	91	91	Real Madrid 2005 ~ 2020	€51M	€380K	2201	0.2K / 6.4K
	H. Kane	24	90	91	Tottenham Hotspur 2010 ~ 2024	€94M	€210K	2177	0.4K / 15.9K

SoFIFA website

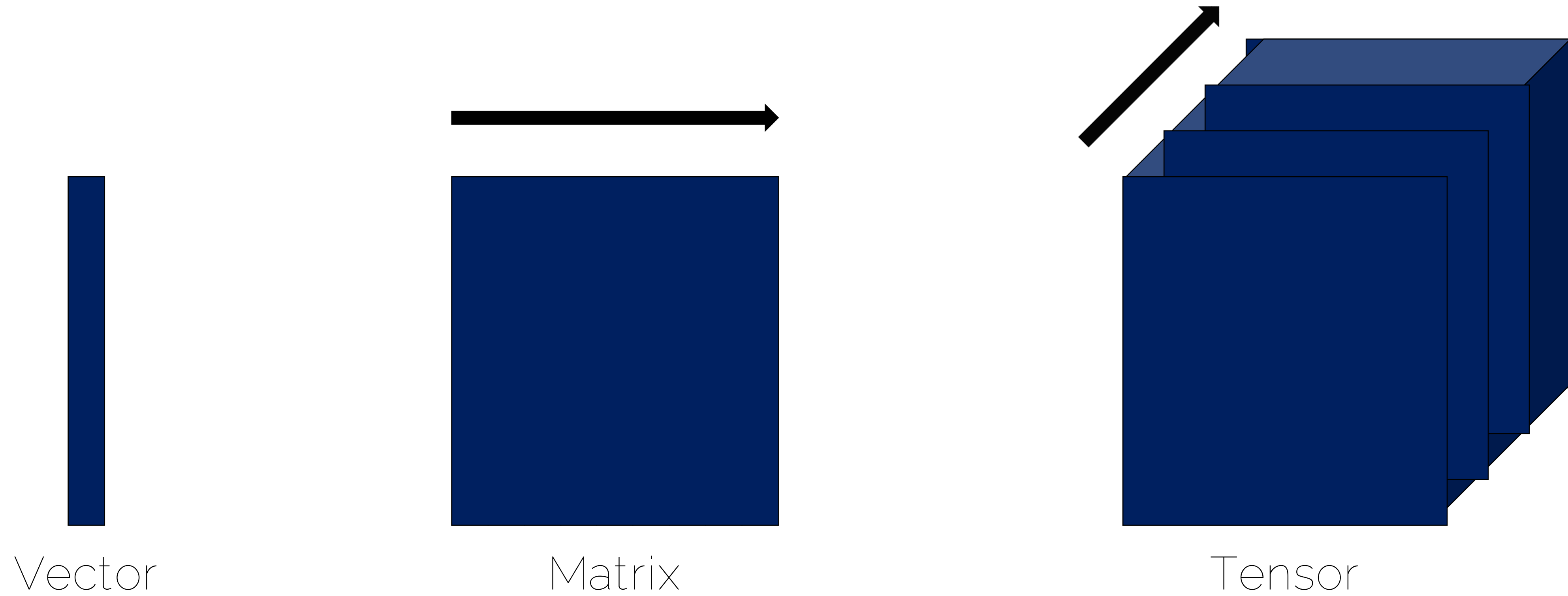


HTML pages



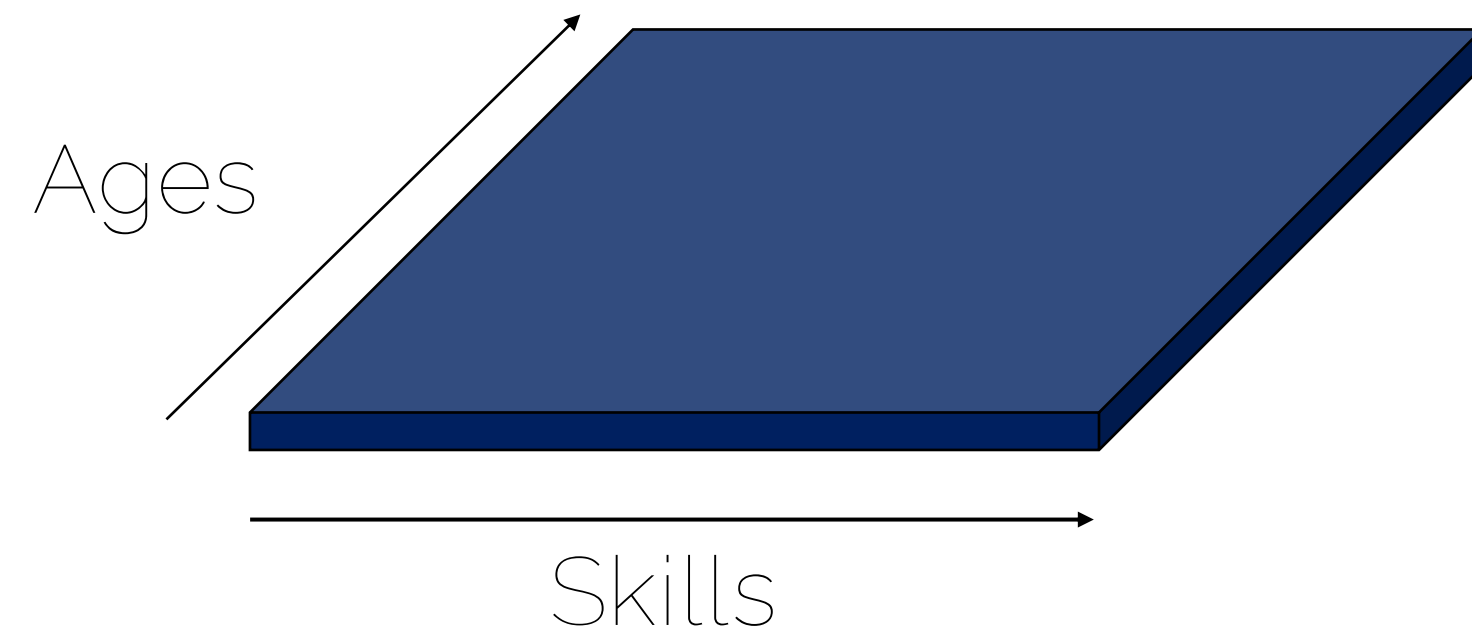
HDF5 file

Tensors



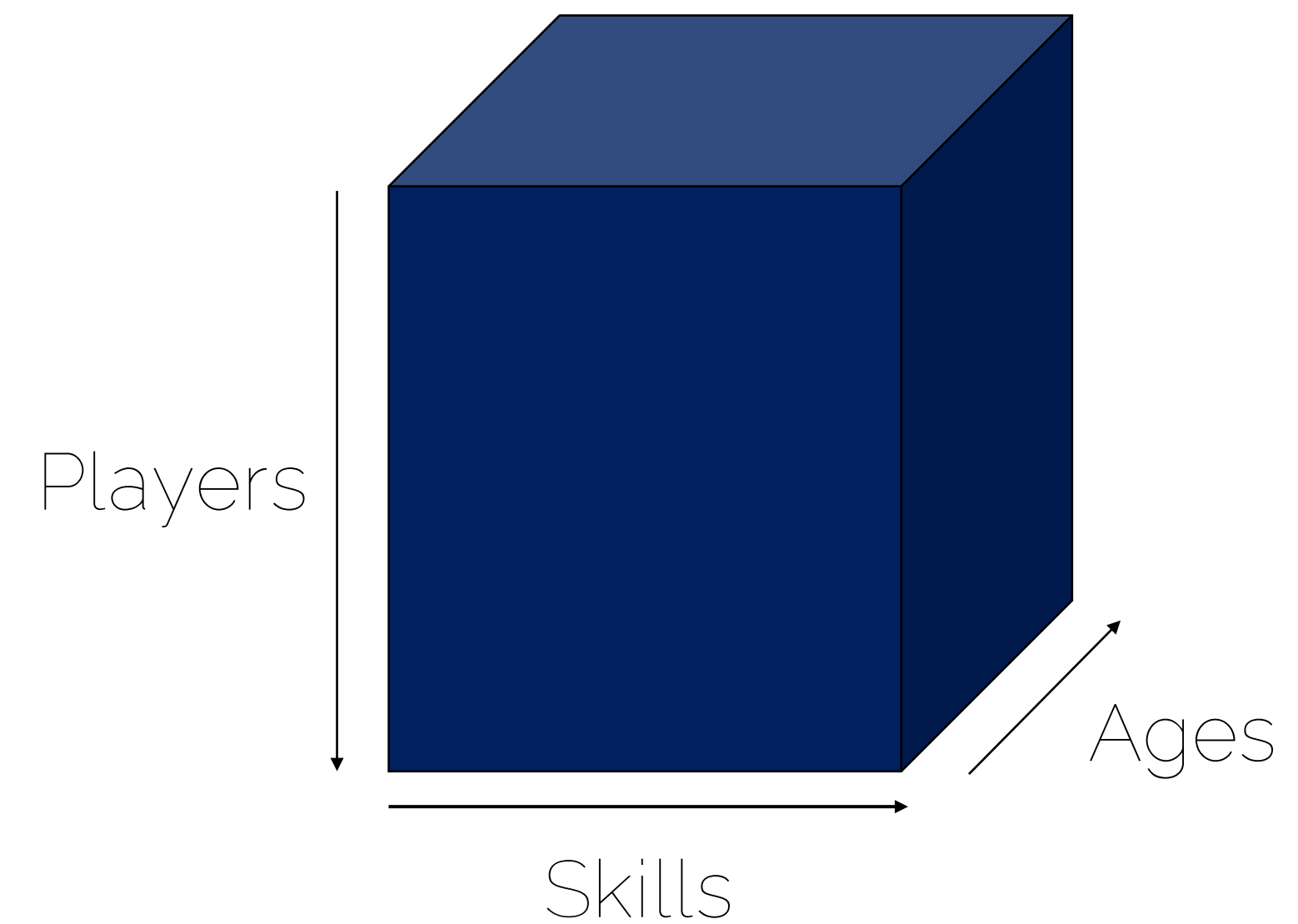
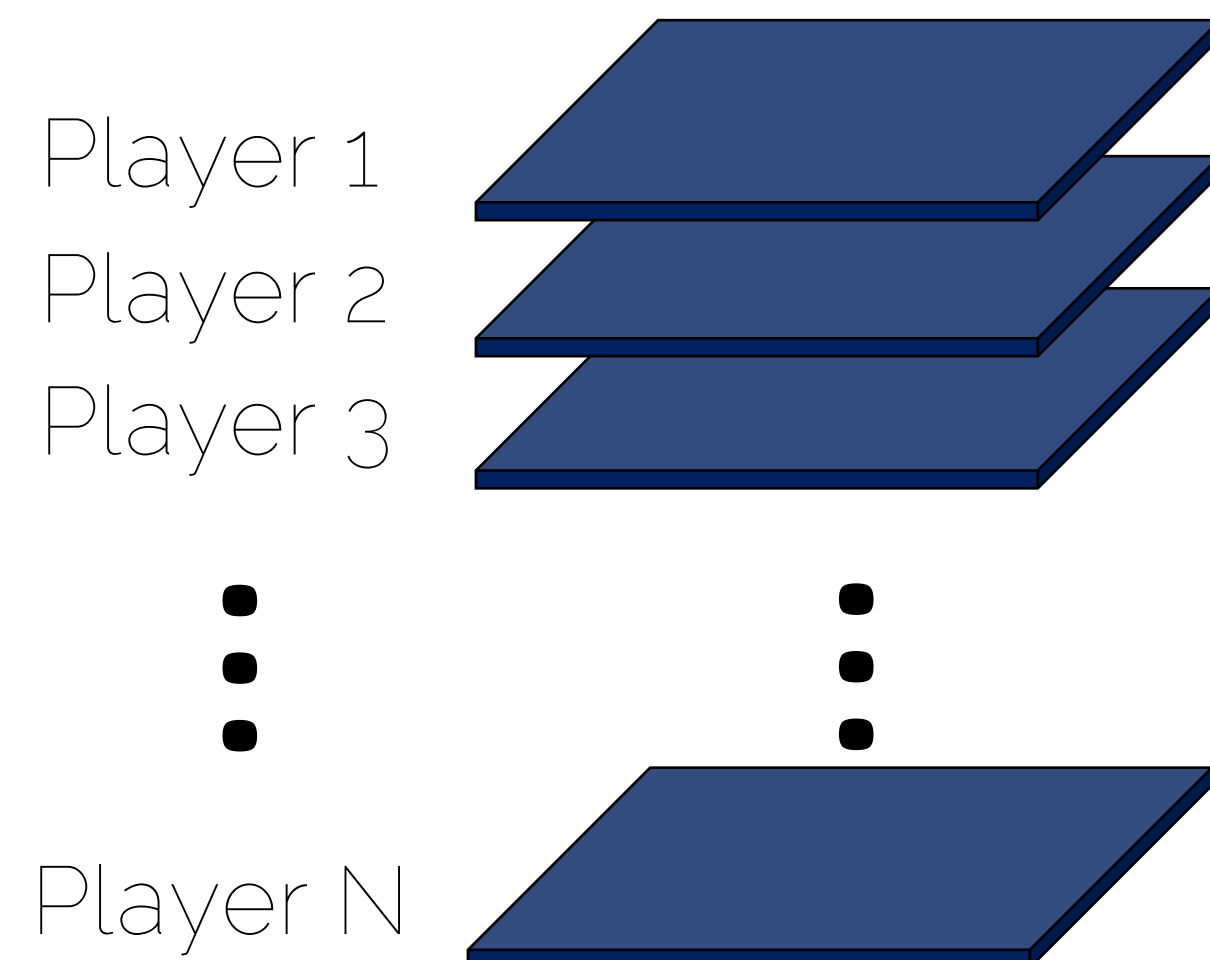
Only 3-dimensional tensors are considered!

Data tensor



One matrix per player

Skills over time



Outline

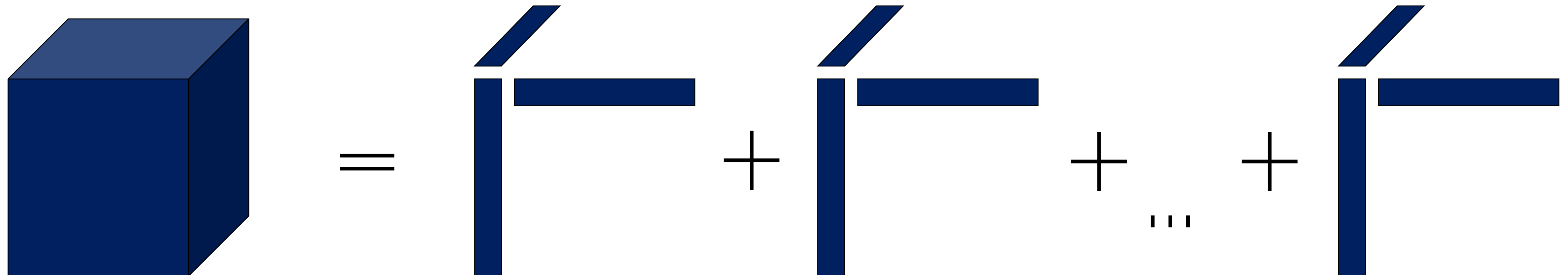
- Goal
- Data + tensors 101
- Data analysis with tensors
- Data challenges
- Tensor-based method for predictions
- Conclusions

Contributions

Canonical polyadic decomposition

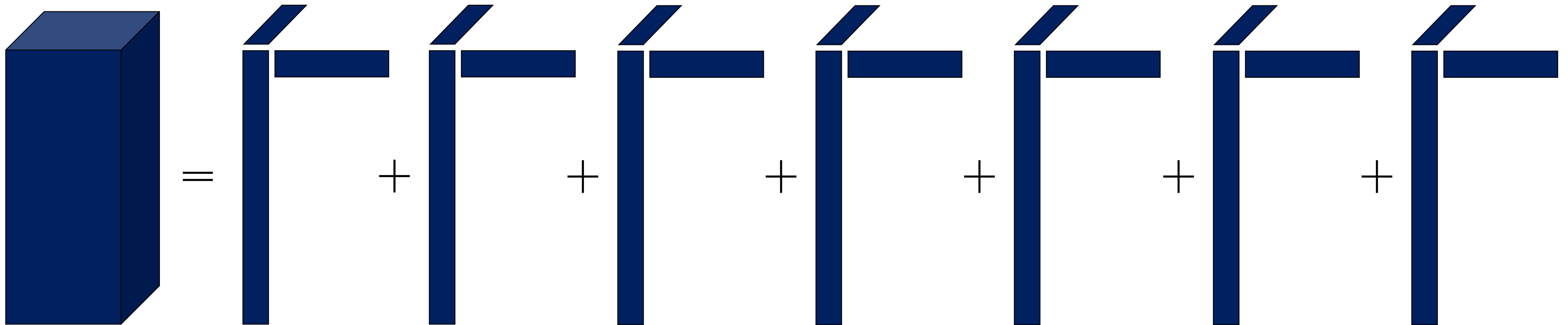
- = CPD
- Sum of rank-1 tensors
- Extract *latent* structures
- Intuitive

Number of rank-1 terms is *the rank* of the tensor

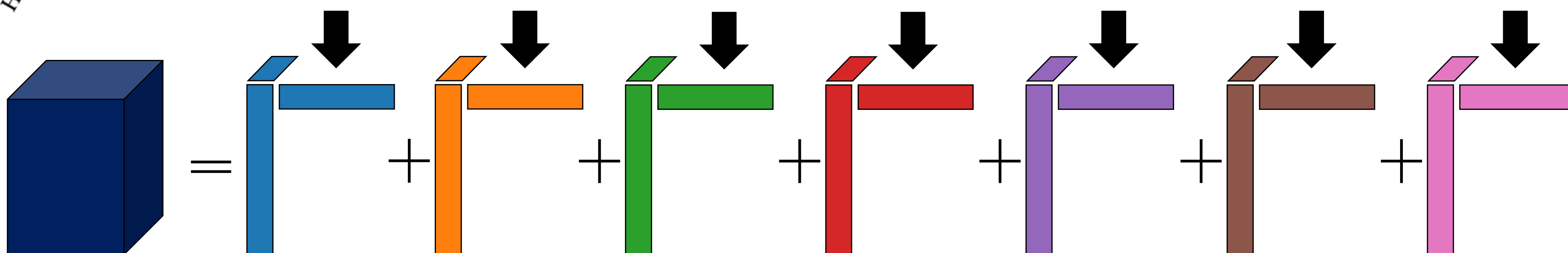
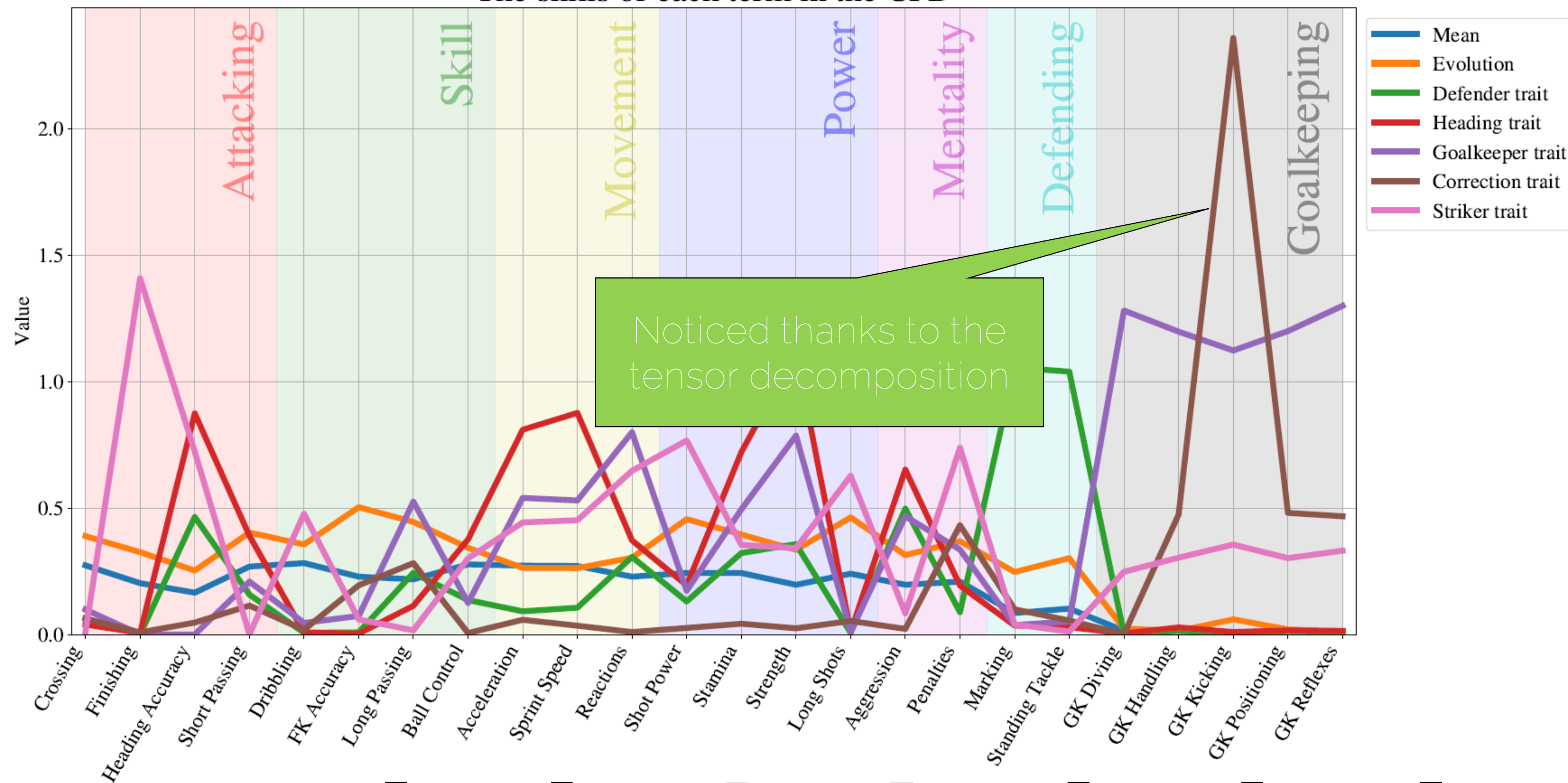


CPD analysis

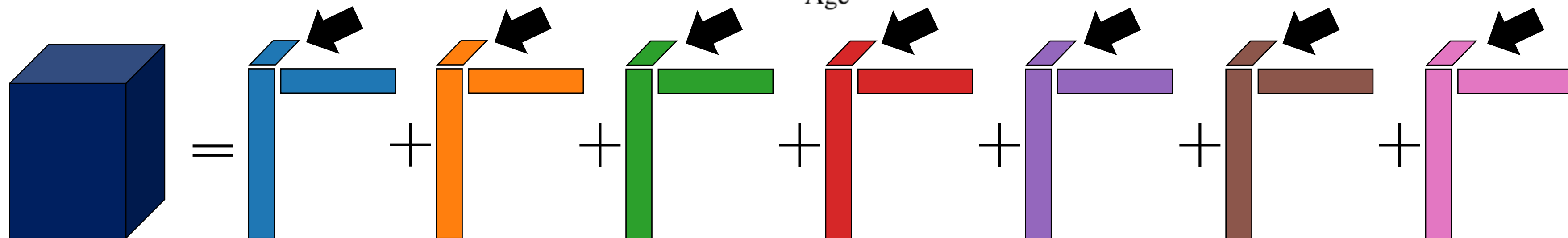
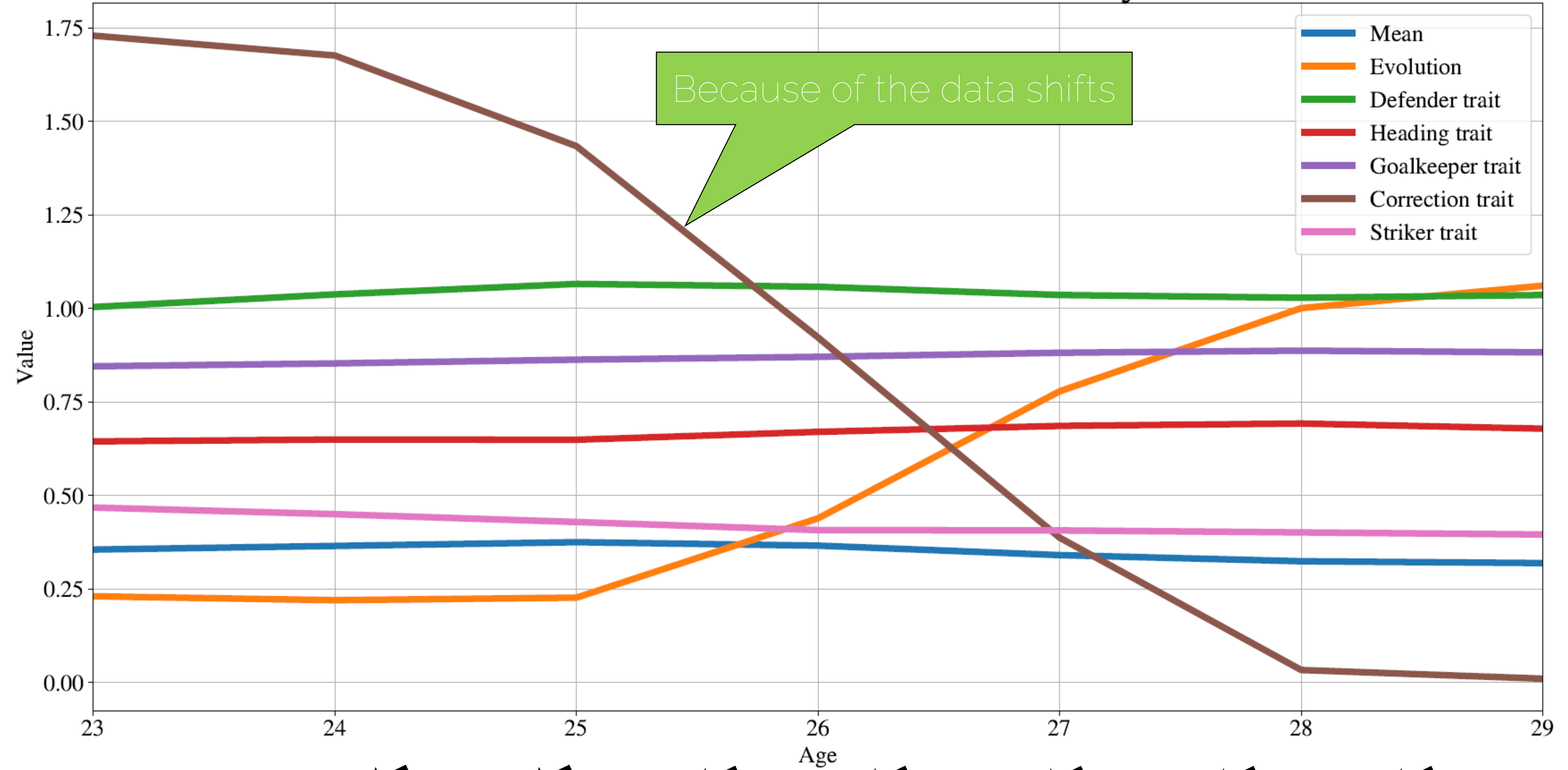
- Rank-7 approximation of the data tensor
- Non-negative factorization used
- Each term can be interpreted intuitively



The skills of each term in the CPD



Evolution of each term in the CPD over the years



Player examples

Coefficients

Mean	Evo.	Def.	Head.	GK	Corr.	Striker
812	27	0	0	0	6	46
767	35	27	0	0	10	15
573	15	57	21	6	17	8
98	12	8	5	75	2	30



Lionel Messi



Luka Modric



Giorgio Chiellini



Gianluigi Buffon

Player examples

Coefficients

Mean	Evo.	Def.	Head.	GK	Corr.	Striker
<u>812</u>	27	0	0	0	6	<u>46</u>
767	35	27	0	0	10	15
573	15	57	21	6	17	8
98	12	8	5	75	2	30



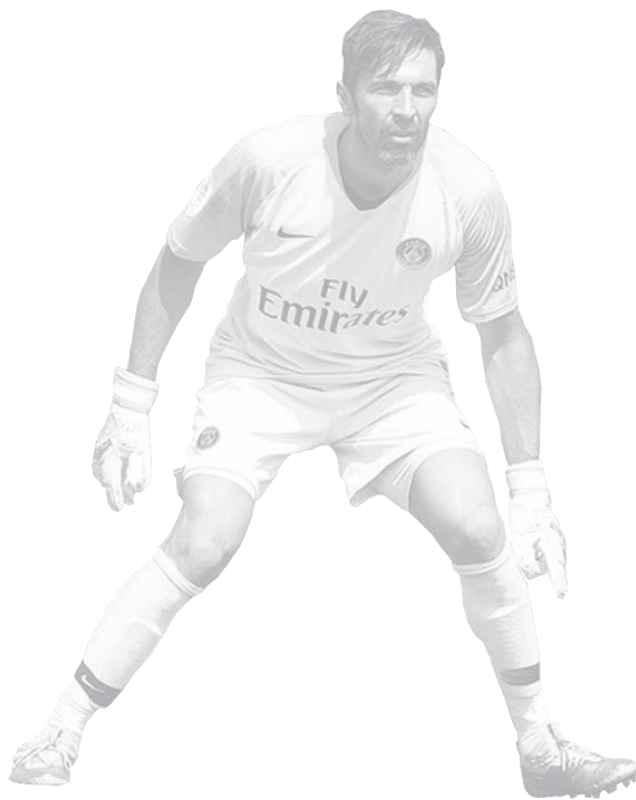
Lionel Messi

- Attacker
- One of the best players ever
- One of the best goal scorers ever
- Small player: 1.70m

Luka Modric



Giorgio Chiellini



Gianluigi Buffon

Player examples

Coefficients

Mean	Evo.	Def.	Head.	GK	Corr.	Striker
812	27	0	0	0	6	46
<u>767</u>	35	27	0	0	10	15
573	15	57	21	6	17	8
98	12	8	5	75	2	30



- Midfielder
- One of the best at the moment
- Good all-rounder
- Small player: 1.72m

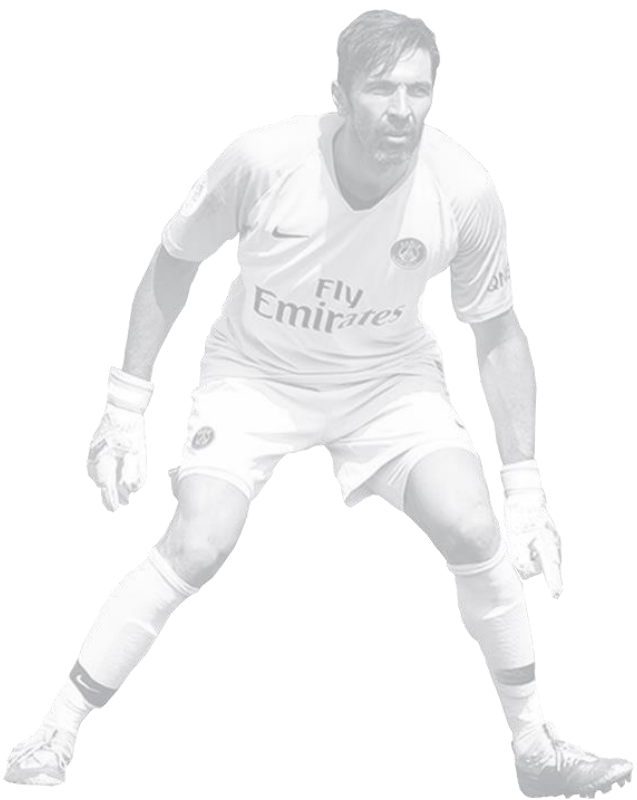
Lionel Messi



Luka Modric



Giorgio Chiellini



Gianluigi Buffon

Player examples

Coefficients

Mean	Evo.	Def.	Head.	GK	Corr.	Striker
812	27	0	0	0	6	46
767	35	27	0	0	10	15
573	15	<u>57</u>	<u>21</u>	6	17	8
98	12	8	5	75	2	30



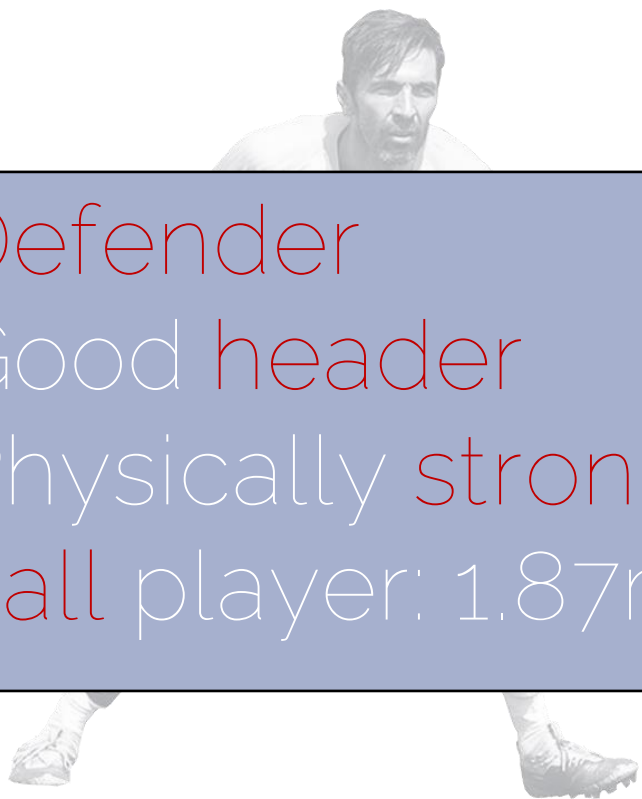
Lionel Messi



Luka Modric



Giorgio Chiellini



Gianluigi Buffon

- Defender
- Good header
- Physically strong
- Tall player: 1.87m

Player examples

Coefficients

Mean	Evo.	Def.	Head.	GK	Corr.	Striker
812	27	0	0	0	6	46
767	35	27	0	0	10	15
573	15	57	21	6	17	8
98	12	8	5	<u>75</u>	2	30



Lionel Messi



Luka Modric



Giorgio Chiellini



Gianluigi Buffon

- Goalkeeper
- One of the best the past decade

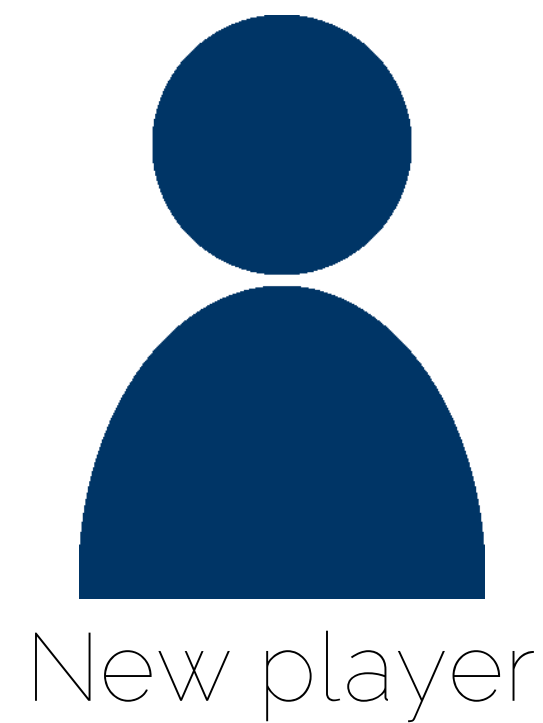
Outline

- Goal
- Data + tensors 101
- Data analysis with tensors
- Data challenges
- Tensor-based method for predictions
- Conclusions

Contributions

Data challenges

- Cold start problem
- Human subjectivity
- Irrelevant skills
- Artificial boosts
- Big corrections
- Event-related



Rating skill 1: ?
Rating skill 2: ?
⋮
Rating skill 33: ?

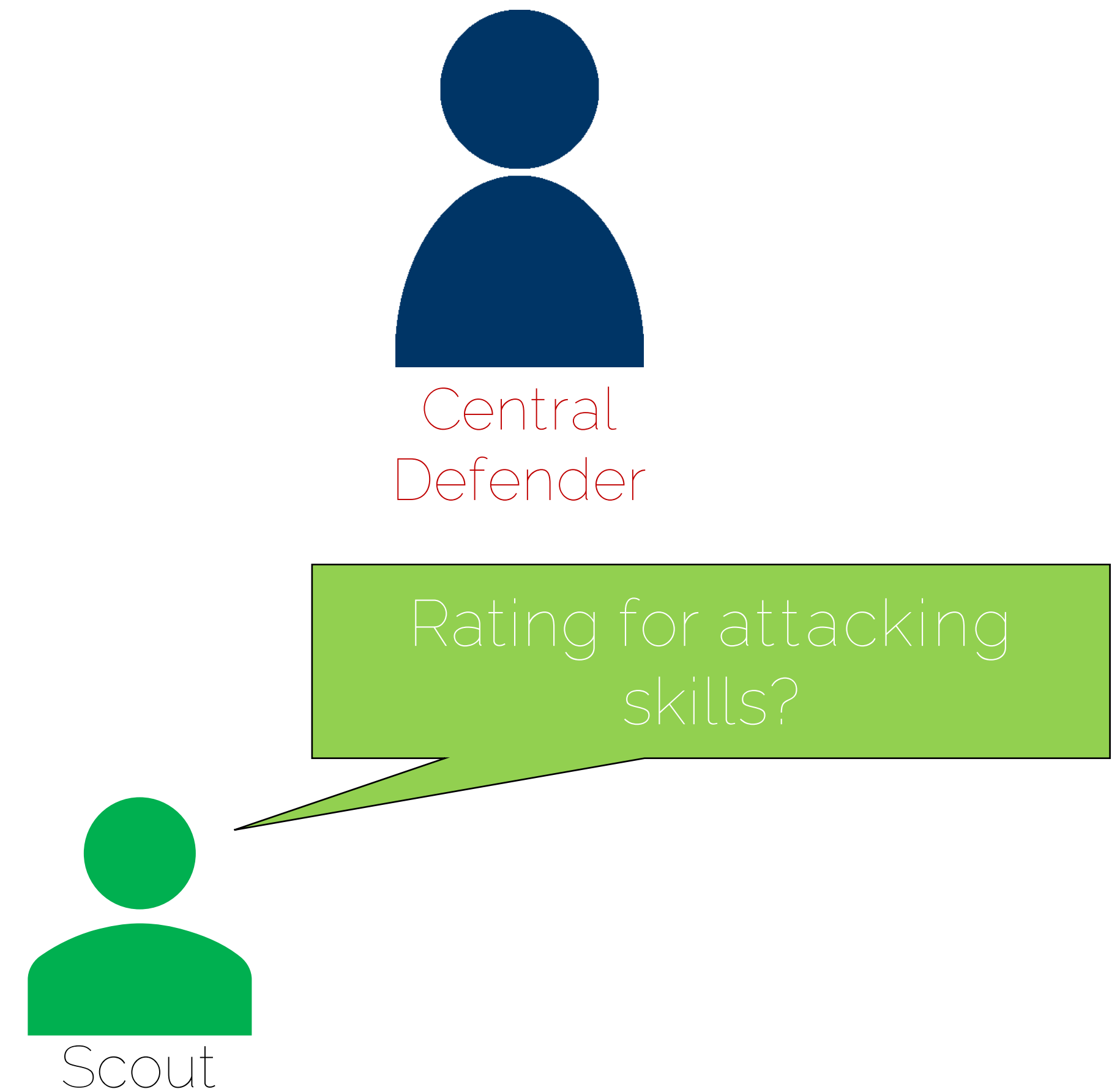
Data challenges

- Cold start problem
- Human subjectivity
- Irrelevant skills
- Artificial boosts
- Big corrections
- Event-related



Data challenges

- Cold start problem
- Human subjectivity
- Irrelevant skills
- Artificial boosts
- Big corrections
- Event-related



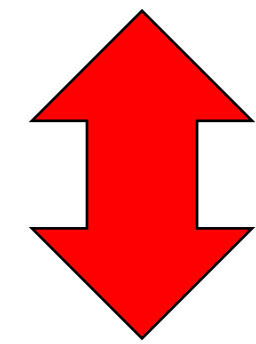
Data challenges

- Cold start problem
- Human subjectivity
- Irrelevant skills
- Artificial boosts
- Big corrections
- Event-related



Thomas Müller

FIFA overall rating: ~~78~~ 86



Real life estimated rating: 86



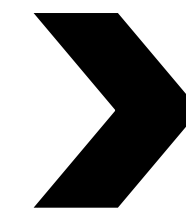
Manually boost the
rating!

Data challenges

- Cold start problem
- Human subjectivity
- Irrelevant skills
- Artificial boosts
- Big corrections
- Event-related

Individual

Marking: 67
Standing Tackle: 72
Sliding Tackle: 68



Marking: 78 (+11)
Standing Tackle: 86 (+14)
Sliding Tackle: 76 (+8)

Collective

FIFA 10 → FIFA 11
GK Kicking for all field players:
↓ 60

Data challenges

- Cold start problem
- Human subjectivity
- Irrelevant skills
- Artificial boosts
- Big corrections
- Event-related



PENALTY MISS

Next update:
Penalty Kicking: ↓5

Outline

- Goal
- Data + tensors 101
- Data analysis with tensors
- Data challenges
- Tensor-based method for predictions
- Conclusions

Contributions

Baseline models

No evolution baseline

Use last known age for all predictions

	21	22	23	24	25	26	27
Crossing	64	66	67	67	67	67	67
Finishing	81	84	88	88	88	88	88
Heading	72	72	74	74	74	74	74

Predictions

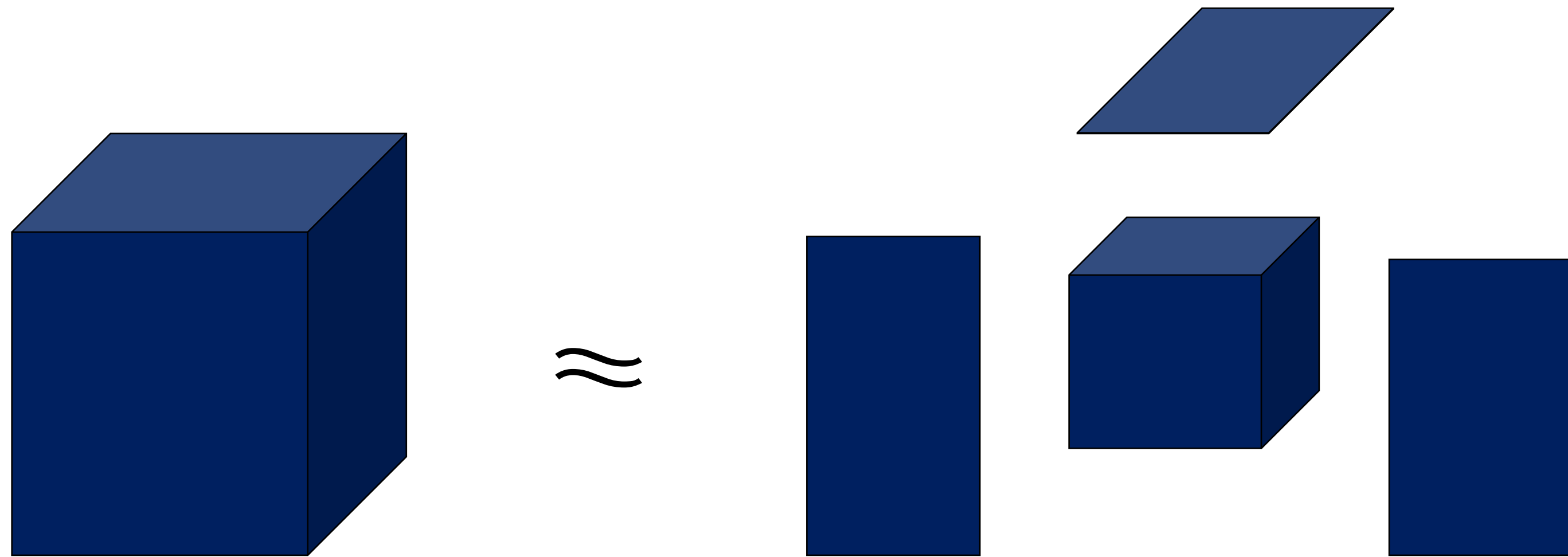
10-nearest neighbours method

Find 10 nearest neighbours using a distance function

Make predictions based on these similar players

Tucker decomposition

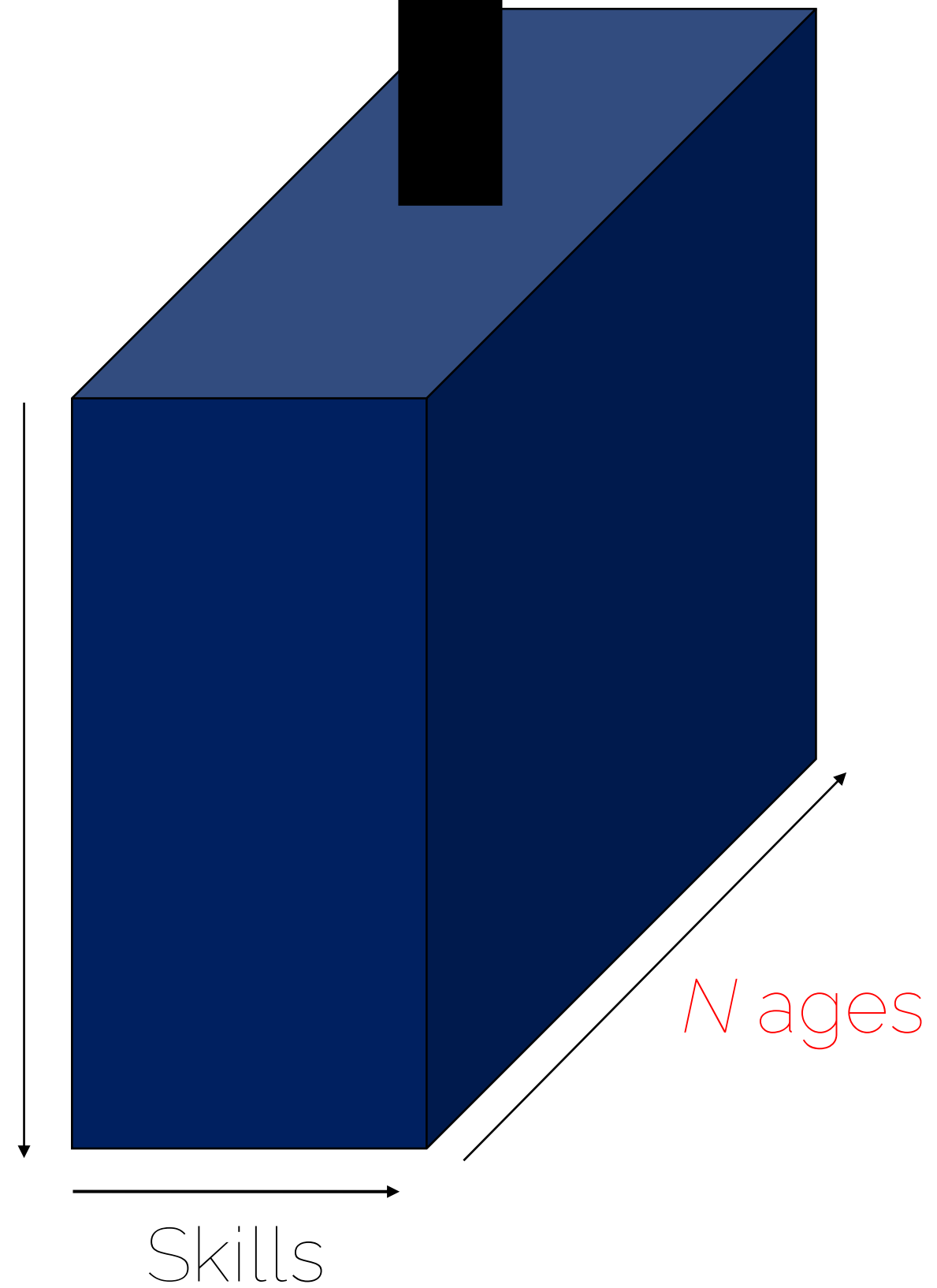
- Higher-order analogue of PCA
- Compression
- Core tensor + factor matrices



Predicting player ratings

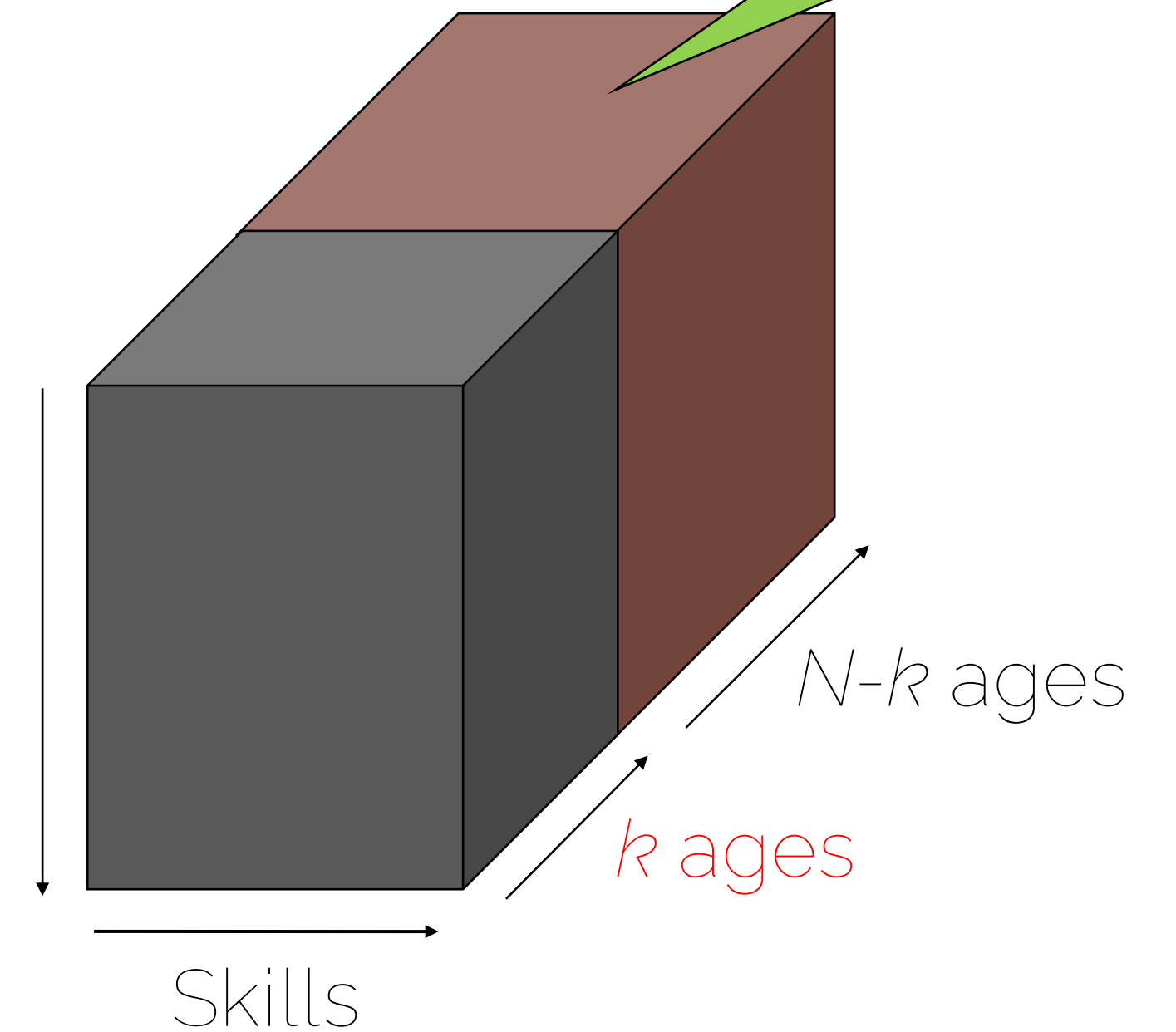
Training tensor

Training players



Test tensor

Test players

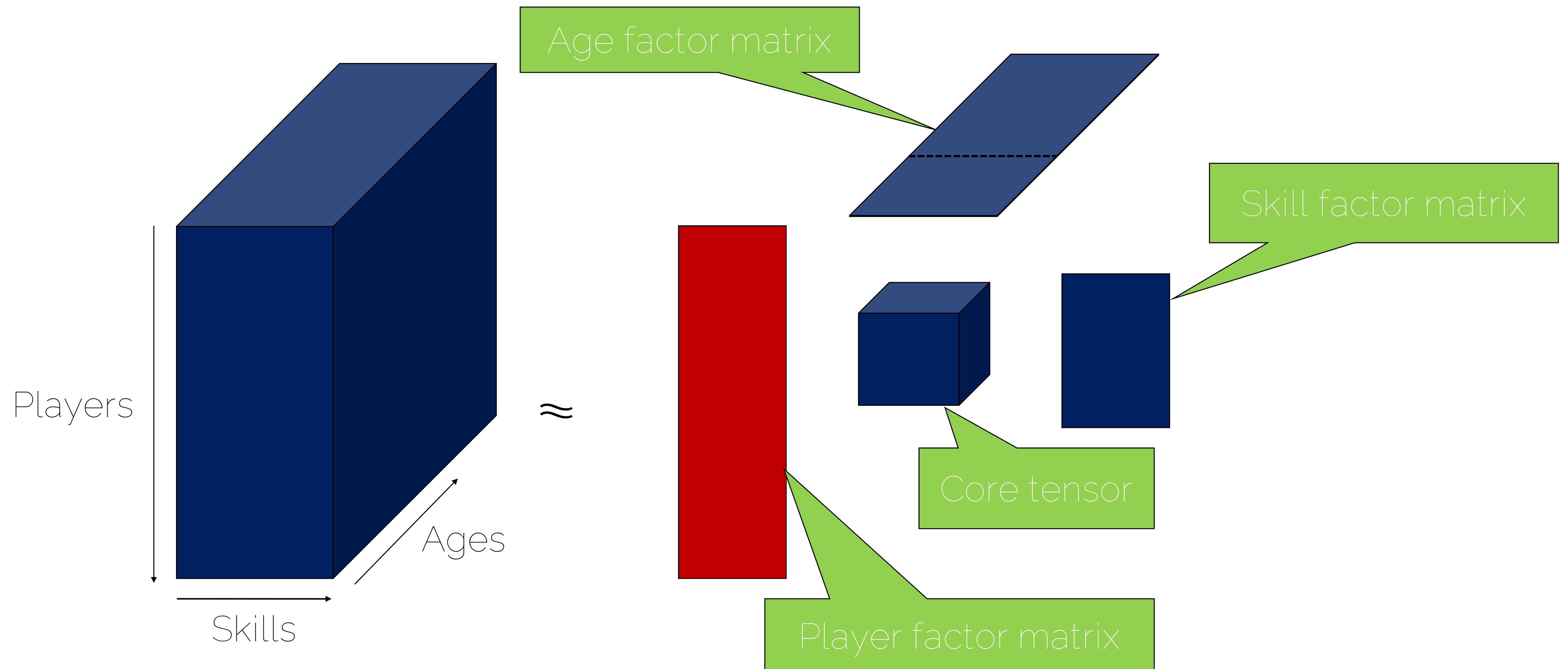


Use the latent structures

Predictions

Predicting player ratings

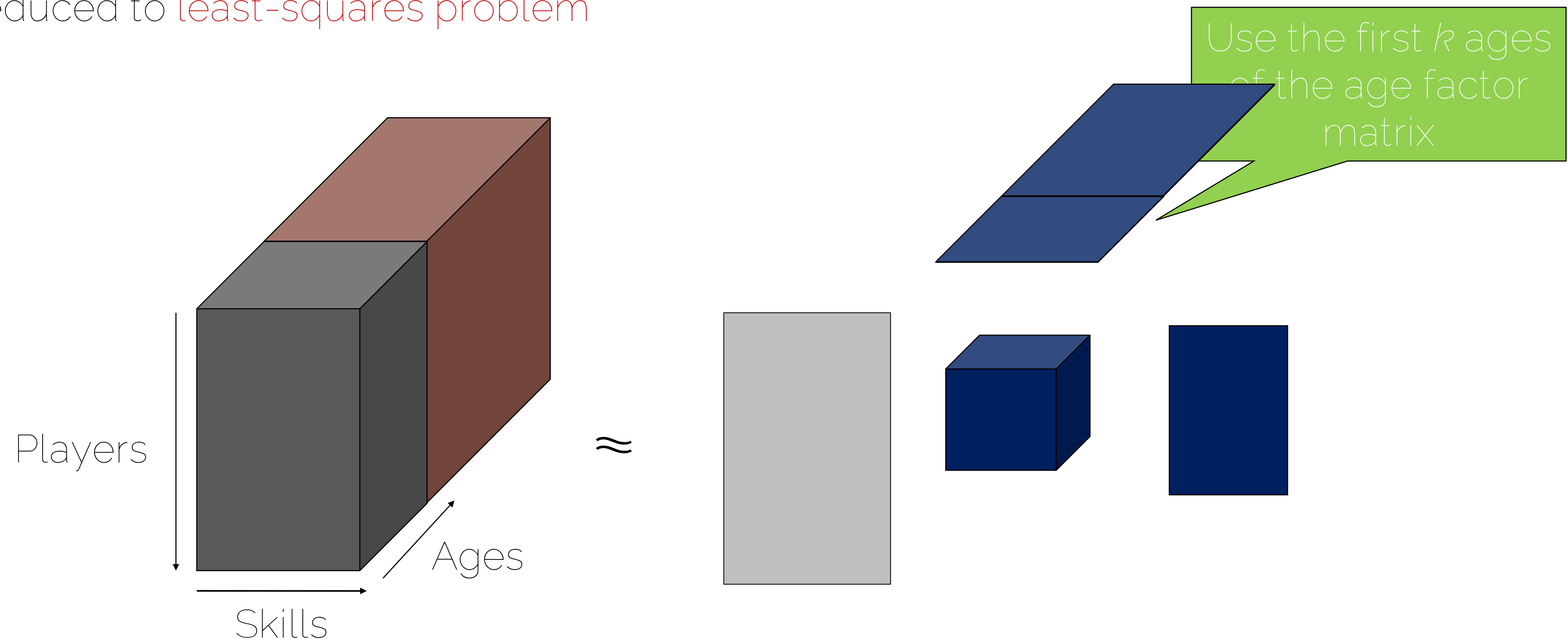
Compute **Tucker decomposition** of the training tensor



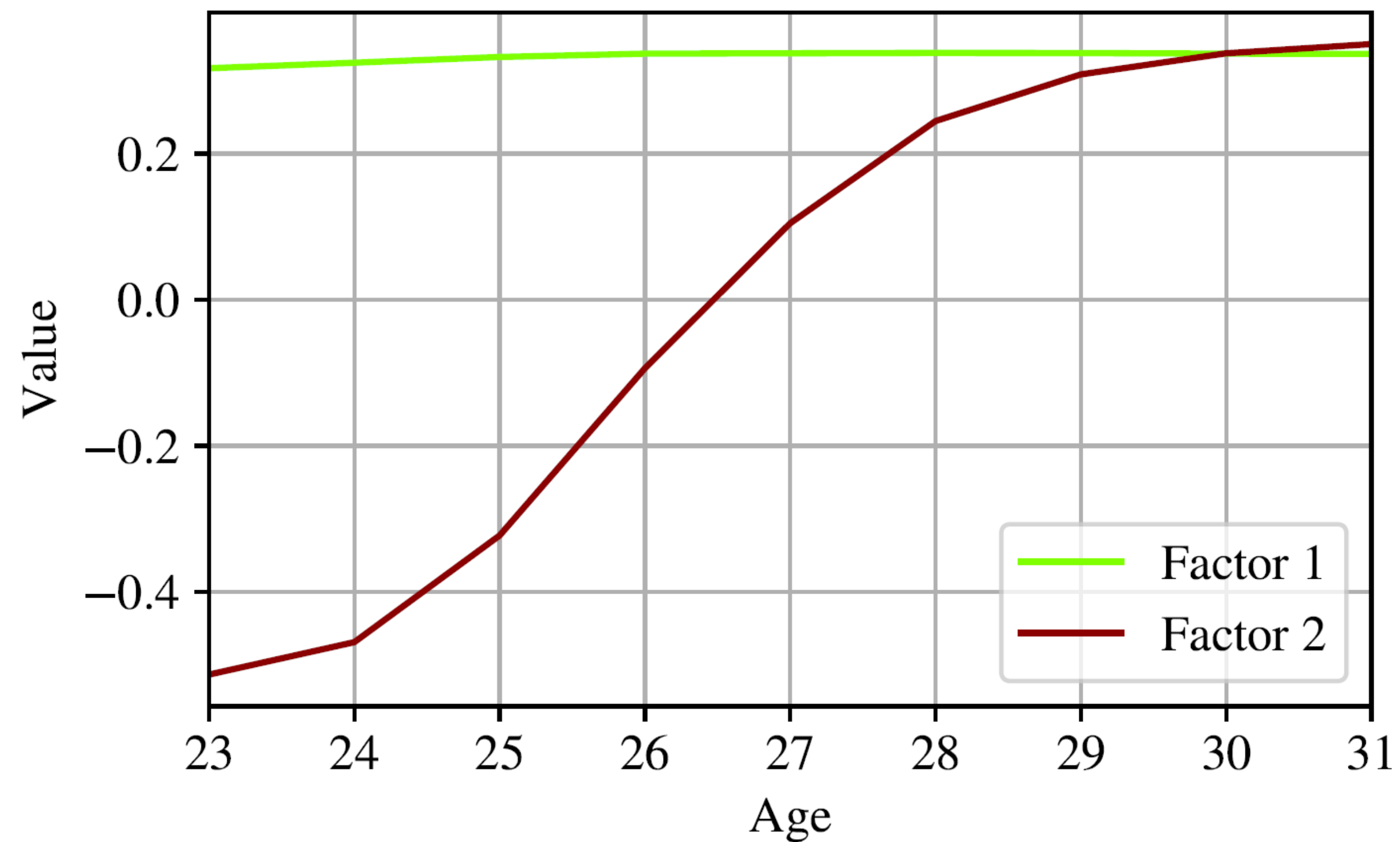
Predicting player ratings

Use the core tensor, skill factors matrix and truncated age factor matrix

Problem reduced to **least-squares problem**



No complicated evolutions

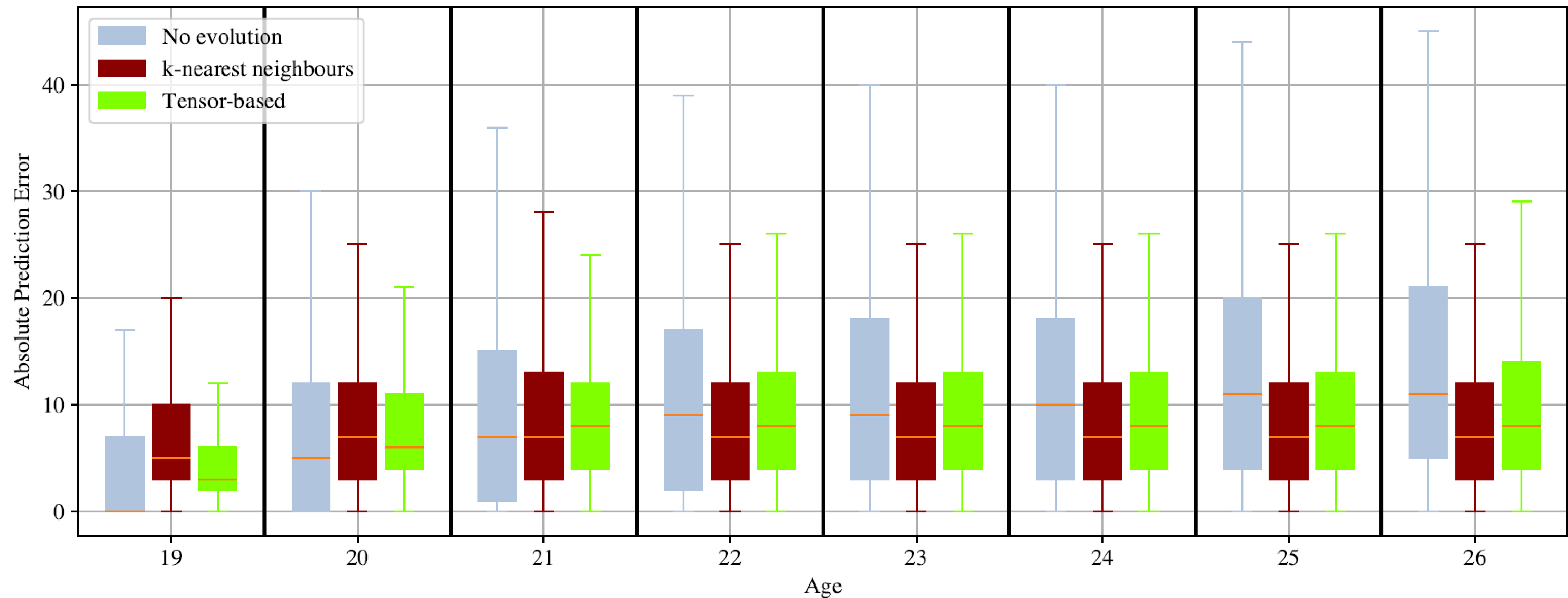


- Two age factors
- Factor 1 no evolution
- Factor 2 normal incline/decline

K-NN best for younger ages

Given: 18

Predicted: 19-26



Mean Absolute error

Method	[19,26]	[24,31]	[27,34]
No evolution	10.45	8.23	6.9
10-nearest neighbors	8.31	7.82	7.67
Tensor-based method	8.57	7.71	6.74

- 10-nearest neighbors best for the younger ages
- Tensor-based method best for the other age ranges
- Same trend for all age ranges

Outline

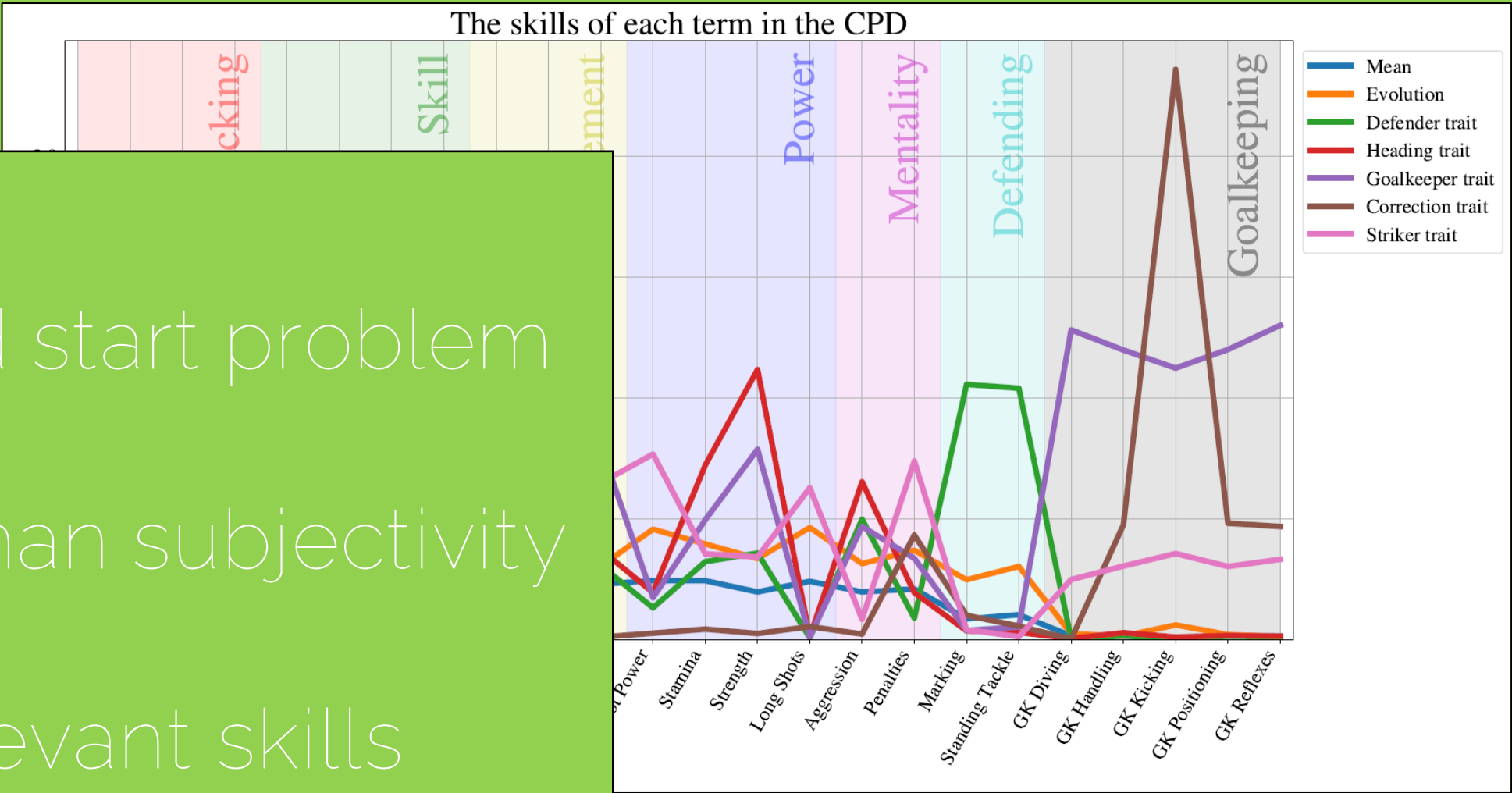
- Goal
- Data + tensors 101
- Data analysis with tensors
- Data challenges
- Tensor-based method for predictions
- Conclusions

Contributions

Conclusions

- Players consists of different traits
- Data not useful for temporal analyses
- Tensor-based method good overall performance

- Cold start problem
- Human subjectivity
- Irrelevant skills
- Artificial boosts



Method	[19,26]	[24,31]	[27,34]
No evolution	10.45	8.23	6.9
10-nearest neighbors	8.31	7.82	7.67
Tensor-based method	8.57	7.71	6.74

Q&A

