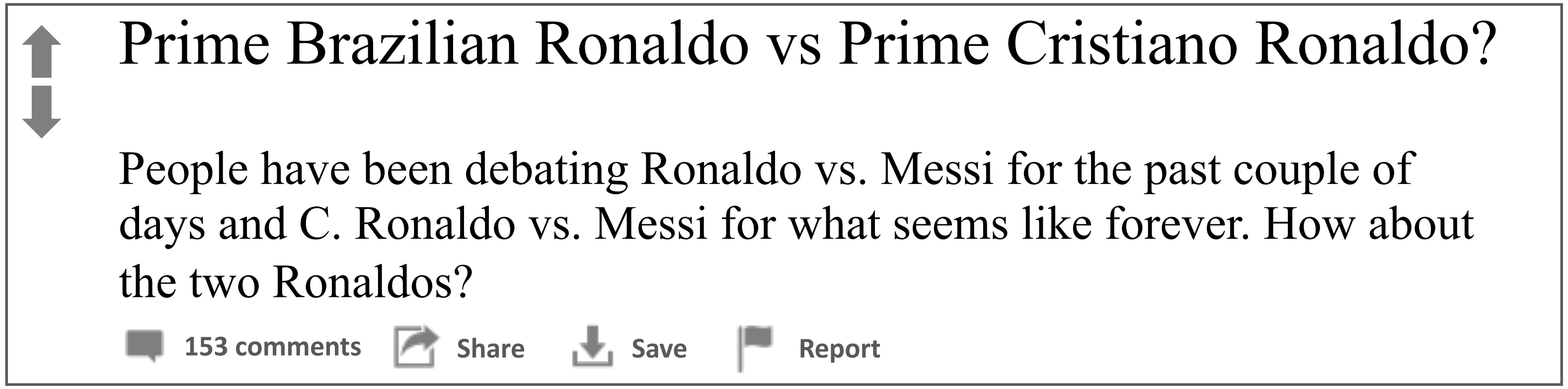
XREF: Entity Linking for Chinese News Comments with Supplementary Article Reference

Xinyu Hua[1] Lei Li[2] Lifeng Hua[3] Lu Wang[1]

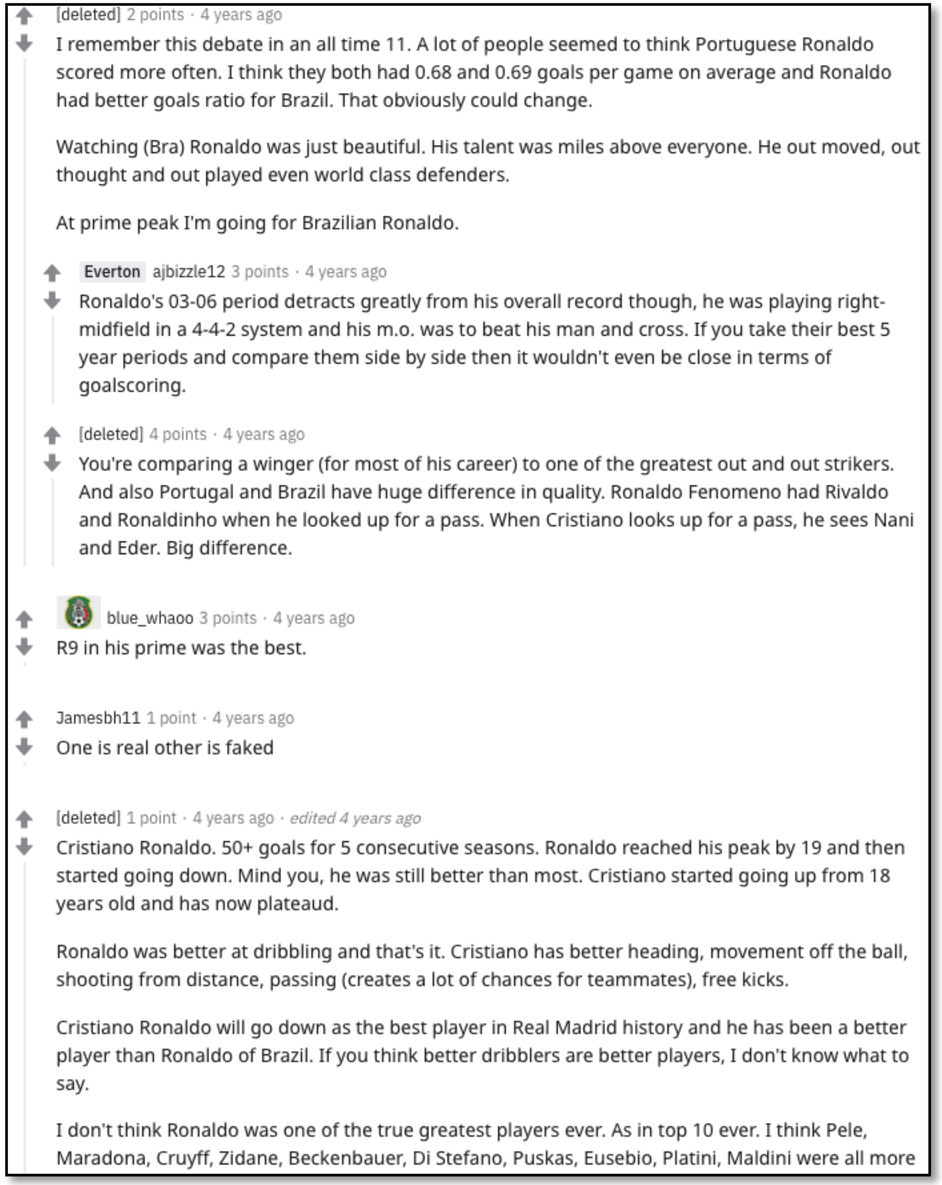
[1]Northeastern University
[2]ByteDance AI Lab
[3]Alibaba Group

Motivation



Prime Brazilian Ronaldo vs Prime Cristiano Ronaldo?

People have been debating Ronaldo vs. Messi for the past couple of days and C. Ronaldo vs. Messi for what seems like forever. How about the two Ronaldos?



[deleted] 2 points · 4 years ago

I remember this debate in an all time 11. A lot of people seemed to think Portuguese Ronaldo scored more often. I think they both had 0.68 and 0.69 goals per game on average and Ronaldo had better goals ratio for Brazil. That obviously could change.

Watching (Bra) Ronaldo was just beautiful. His talent was miles above everyone. He out moved, out thought and out played even world class defenders.

At prime peak I'm going for Brazilian Ronaldo.

Everton ajbizzle12 3 points · 4 years ago

Ronaldo's 03-06 period detracts greatly from his overall record though, he was playing right-midfield in a 4-4-2 system and his m.o. was to beat his man and cross. If you take their best 5 year periods and compare them side by side then it wouldn't even be close in terms of goalscoring.

[deleted] 4 points · 4 years ago

You're comparing a winger (for most of his career) to one of the greatest out and out strikers. And also Portugal and Brazil have huge difference in quality. Ronaldo Fenomeno had Rivaldo and Ronaldinho when he looked up for a pass. When Cristiano looks up for a pass, he sees Nani and Eder. Big difference.

blue_whaoo 3 points · 4 years ago

R9 in his prime was the best.

Jamesbh11 1 point · 4 years ago

One is real other is faked

[deleted] 1 point · 4 years ago · *edited 4 years ago*

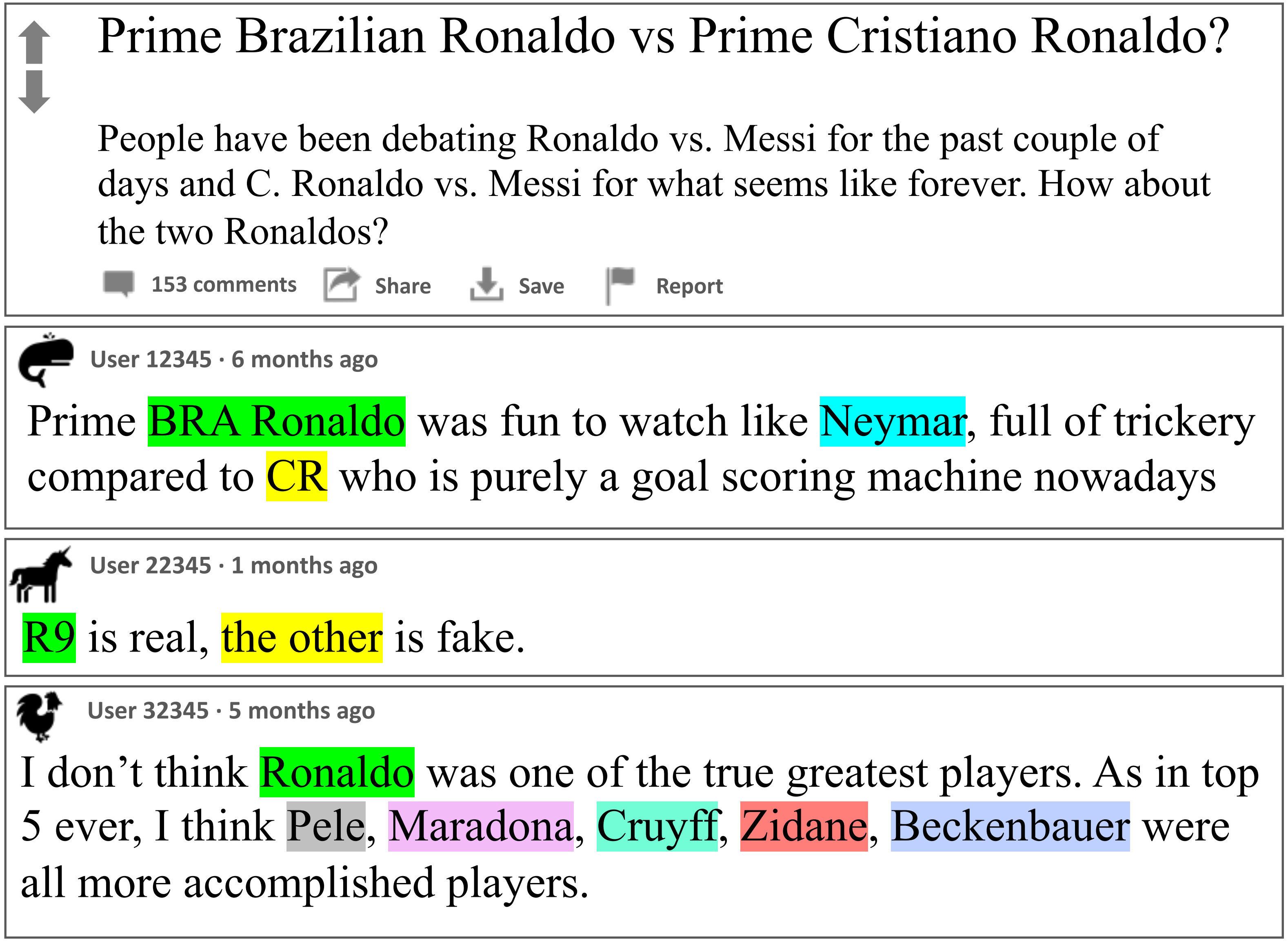
Cristiano Ronaldo. 50+ goals for 5 consecutive seasons. Ronaldo reached his peak by 19 and then started going down. Mind you, he was still better than most. Cristiano started going up from 18 years old and has now plateaud.

Ronaldo was better at dribbling and that's it. Cristiano has better heading, movement off the ball, shooting from distance, passing (creates a lot of chances for teammates), free kicks.

Cristiano Ronaldo will go down as the best player in Real Madrid history and he has been a better player than Ronaldo of Brazil. If you think better dribblers are better players, I don't know what to say.

I don't think Ronaldo was one of the true greatest players ever. As in top 10 ever. I think Pele, Maradona, Cruyff, Zidane, Beckenbauer, Di Stefano, Puskas, Eusebio, Platini, Maldini were all more

Motivation



Prime Brazilian Ronaldo vs Prime Cristiano Ronaldo?

People have been debating Ronaldo vs. Messi for the past couple of days and C. Ronaldo vs. Messi for what seems like forever. How about the two Ronaldos?

User 12345 · 6 months ago

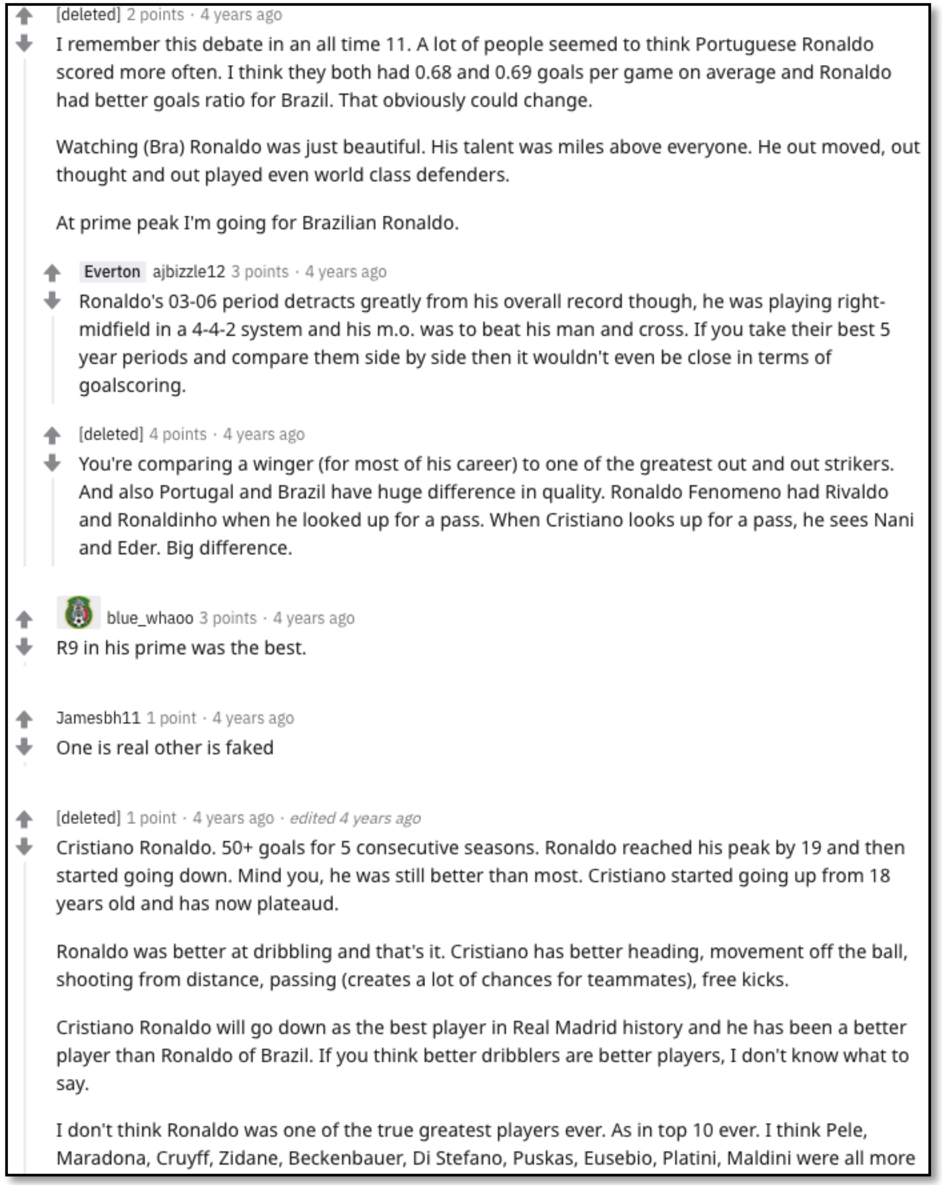
Prime BRA Ronaldo was fun to watch like Neymar, full of trickery compared to CR who is purely a goal scoring machine nowadays

User 22345 · 1 months ago

R9 is real, the other is fake.

User 32345 · 5 months ago

I don’t think Ronaldo was one of the true greatest players. As in top 5 ever, I think Pele, Maradona, Cruyff, Zidane, Beckenbauer were all more accomplished players.



[deleted] 2 points · 4 years ago

I remember this debate in an all time 11. A lot of people seemed to think Portuguese Ronaldo scored more often. I think they both had 0.68 and 0.69 goals per game on average and Ronaldo had better goals ratio for Brazil. That obviously could change.

Watching (Bra) Ronaldo was just beautiful. His talent was miles above everyone. He out moved, out thought and out played even world class defenders.

At prime peak I'm going for Brazilian Ronaldo.

Everton ajbizzle12 3 points · 4 years ago

Ronaldo's 03-06 period detracts greatly from his overall record though, he was playing right-midfield in a 4-4-2 system and his m.o. was to beat his man and cross. If you take their best 5 year periods and compare them side by side then it wouldn't even be close in terms of goalscoring.

[deleted] 4 points · 4 years ago

You're comparing a winger (for most of his career) to one of the greatest out and out strikers. And also Portugal and Brazil have huge difference in quality. Ronaldo Fenomeno had Rivaldo and Ronaldinho when he looked up for a pass. When Cristiano looks up for a pass, he sees Nani and Eder. Big difference.

blue_whaoo 3 points · 4 years ago

R9 in his prime was the best.

Jamesbh11 1 point · 4 years ago

One is real other is faked

[deleted] 1 point · 4 years ago · *edited 4 years ago*

Cristiano Ronaldo. 50+ goals for 5 consecutive seasons. Ronaldo reached his peak by 19 and then started going down. Mind you, he was still better than most. Cristiano started going up from 18 years old and has now plateaud.

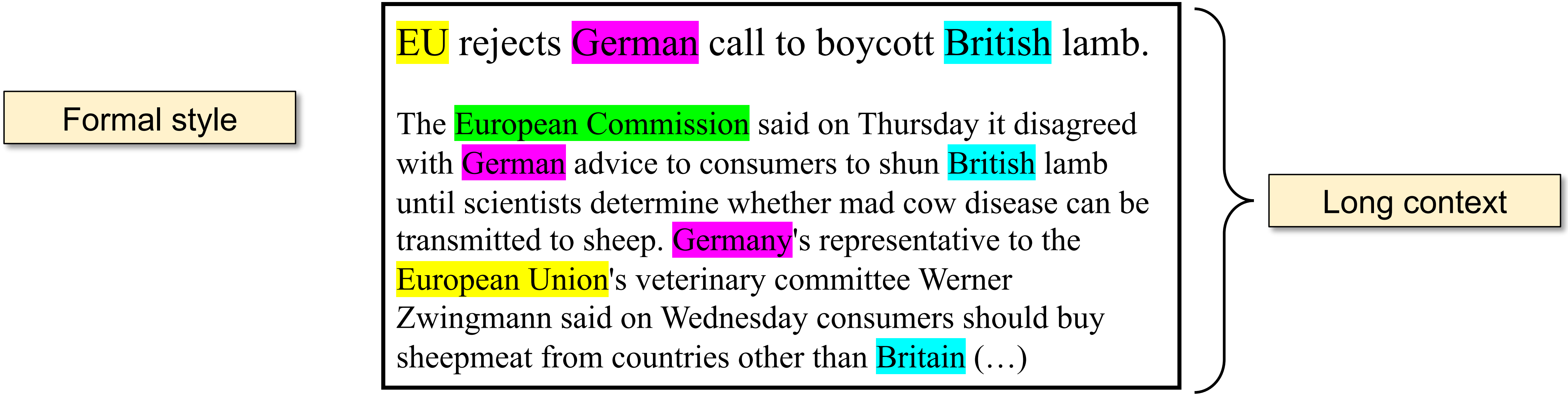
Ronaldo was better at dribbling and that's it. Cristiano has better heading, movement off the ball, shooting from distance, passing (creates a lot of chances for teammates), free kicks.

Cristiano Ronaldo will go down as the best player in Real Madrid history and he has been a better player than Ronaldo of Brazil. If you think better dribblers are better players, I don't know what to say.

I don't think Ronaldo was one of the true greatest players ever. As in top 10 ever. I think Pele, Maradona, Cruyff, Zidane, Beckenbauer, Di Stefano, Puskas, Eusebio, Platini, Maldini were all more

Motivation

- Entity Linking (EL) benchmarks



Formal style

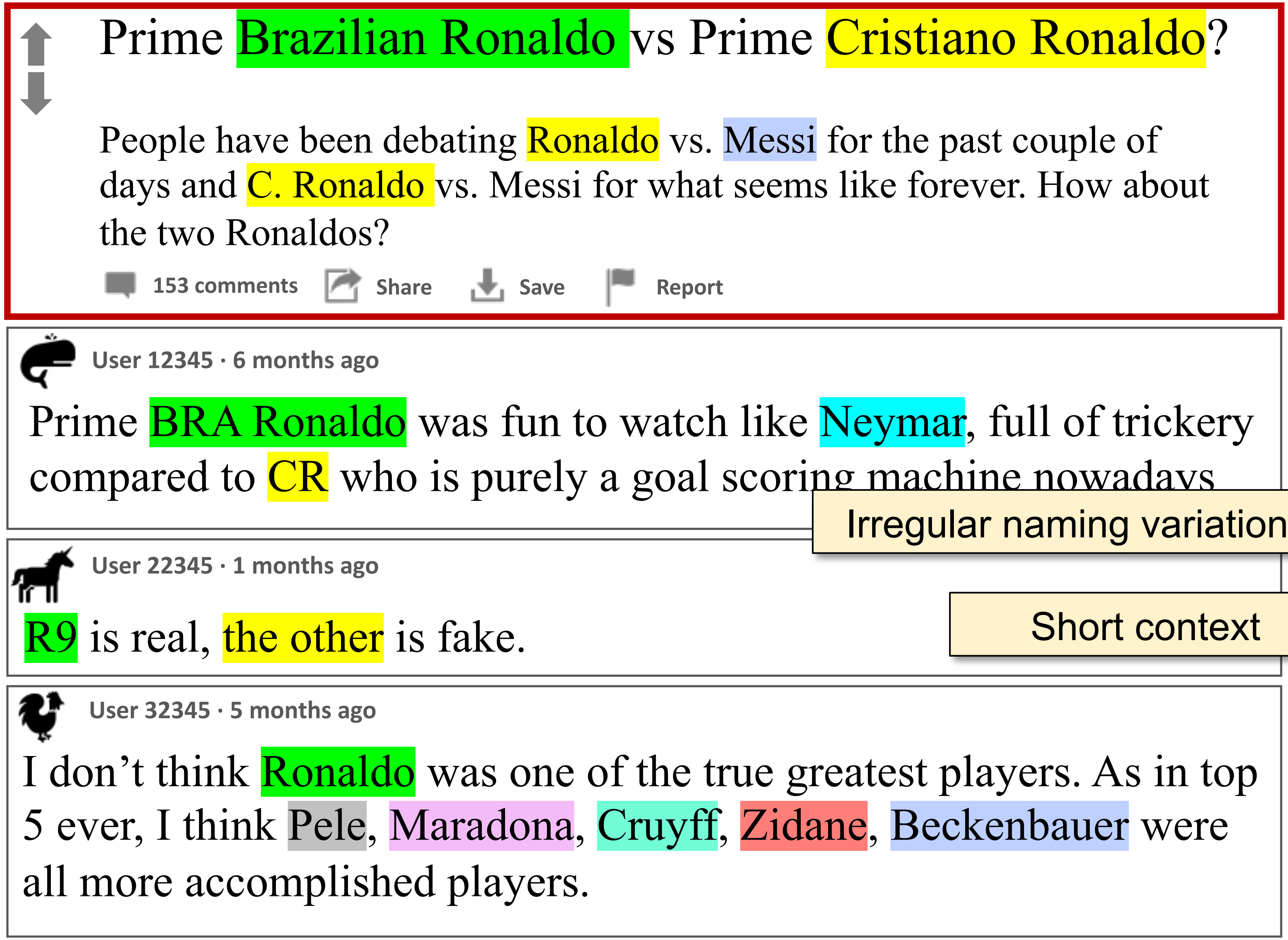
EU rejects German call to boycott British lamb.

The European Commission said on Thursday it disagreed with German advice to consumers to shun British lamb until scientists determine whether mad cow disease can be transmitted to sheep. Germany's representative to the European Union's veterinary committee Werner Zwingmann said on Wednesday consumers should buy sheepmeat from countries other than Britain (…)

Long context

AIDA-CoNLL [Hoffart et al, 2011]

Motivation



Prime Brazilian Ronaldo vs Prime Cristiano Ronaldo?

People have been debating Ronaldo vs. Messi for the past couple of days and C. Ronaldo vs. Messi for what seems like forever. How about the two Ronaldos?

User 12345 · 6 months ago

Prime BRA Ronaldo was fun to watch like Neymar, full of trickery compared to CR who is purely a goal scoring machine nowadays

User 22345 · 1 months ago

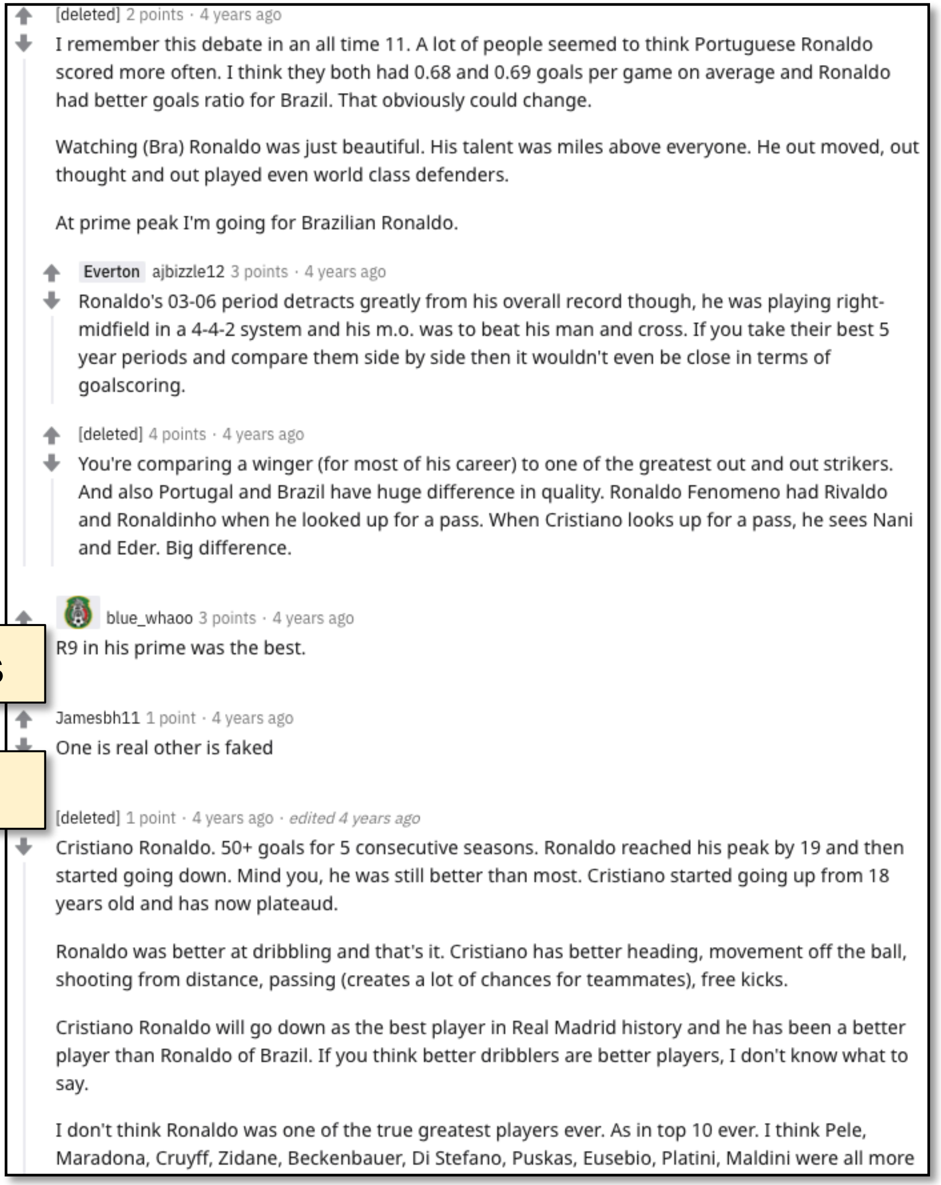
R9 is real, the other is fake.

User 32345 · 5 months ago

I don't think Ronaldo was one of the true greatest players. As in top 5 ever, I think Pele, Maradona, Cruyff, Zidane, Beckenbauer were all more accomplished players.

Irregular naming variations

Short context



[deleted] 2 points · 4 years ago

I remember this debate in an all time 11. A lot of people seemed to think Portuguese Ronaldo scored more often. I think they both had 0.68 and 0.69 goals per game on average and Ronaldo had better goals ratio for Brazil. That obviously could change.

Watching (Bra) Ronaldo was just beautiful. His talent was miles above everyone. He out moved, out thought and out played even world class defenders.

At prime peak I'm going for Brazilian Ronaldo.

Everton ajbizzle12 3 points · 4 years ago

Ronaldo's 03-06 period detracts greatly from his overall record though, he was playing right-midfield in a 4-4-2 system and his m.o. was to beat his man and cross. If you take their best 5 year periods and compare them side by side then it wouldn't even be close in terms of goalscoring.

[deleted] 4 points · 4 years ago

You're comparing a winger (for most of his career) to one of the greatest out and out strikers. And also Portugal and Brazil have huge difference in quality. Ronaldo Fenomeno had Rivaldo and Ronaldinho when he looked up for a pass. When Cristiano looks up for a pass, he sees Nani and Eder. Big difference.

blue_whaoo 3 points · 4 years ago

R9 in his prime was the best.

Jamesbh11 1 point · 4 years ago

One is real other is faked

[deleted] 1 point · 4 years ago · *edited 4 years ago*

Cristiano Ronaldo. 50+ goals for 5 consecutive seasons. Ronaldo reached his peak by 19 and then started going down. Mind you, he was still better than most. Cristiano started going up from 18 years old and has now plateaud.

Ronaldo was better at dribbling and that's it. Cristiano has better heading, movement off the ball, shooting from distance, passing (creates a lot of chances for teammates), free kicks.

Cristiano Ronaldo will go down as the best player in Real Madrid history and he has been a better player than Ronaldo of Brazil. If you think better dribblers are better players, I don't know what to say.

I don't think Ronaldo was one of the true greatest players ever. As in top 10 ever. I think Pele, Maradona, Cruyff, Zidane, Beckenbauer, Di Stefano, Puskas, Eusebio, Platini, Maldini were all more

Related Work

- EL for long documents: *leveraging rich context; collective predictions* [Francis-Landau+,2016, Gupta+, 2017, Le and Titov, 2018, Sil+, 2018]
- EL for social media: *redundant posts; user information* [Liu+, 2013, Guo+, 2013., Yang and Chang, 2015]
- EL for Chinese language: *cross-lingual approaches over aligned Wikipedia articles* [Tsai and Roth, 2016, Sil+, 2018]

Roadmap

- Motivation
- Related Work
- **Data**
- Model
- Experiments
- Conclusion

Data

- Data format:

News article

Bingbing Fan did not win the Golden Horse Awards, drawing attacks from Xiaogang Feng. Chen Li joins the fight on Weibo.

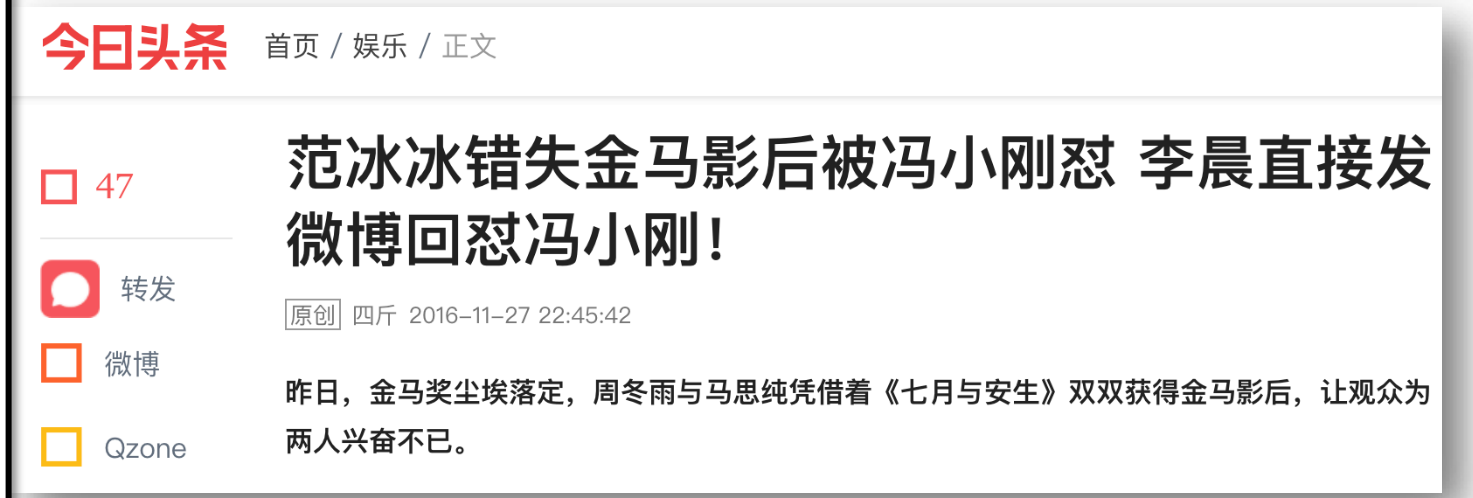
Yesterday, the Golden Horse Awards were finally announced. Dongyu Zhou and Sichun Ma won the best lead actress with their performances in ”Soul Mate”, excites the audiences. (…)

Comments

User id 1234567
她真的是明星 算不上演员
She is just a star, not an actress.

User id 2234567
演员多 成为明星的演员少 冰冰却早已经做到了
There are many actresses, but not many became stars. Bingbing did this early.

User id 3234567
冯导为这部电影毁掉了自己的形象，烧掉了自己的人气
电影好坏第一责任 人是导演。
Director Feng ruined his image and reputation for this movie



范冰冰错失金马影后被冯小刚怼 李晨直接发微博回怼冯小刚！

原创 四斤 2016-11-27 22:45:42

昨日，金马奖尘埃落定，周冬雨与马思纯凭借着《七月与安生》双双获得金马影后，让观众为两人兴奋不已。

Data

- Data format:

News article

Bingbing Fan did not win the Golden Horse Awards, drawing attacks from Xiaogang Feng. Chen Li joins the fight on Weibo.

Yesterday, the Golden Horse Awards were finally announced. Dongyu Zhou and Sichun Ma won the best lead actress with their performances in ”Soul Mate”, excites the audiences. (…)

Comments

User id 1234567

她真的是明星 算不上演员

She is just a star, not an actress.

User id 2234567

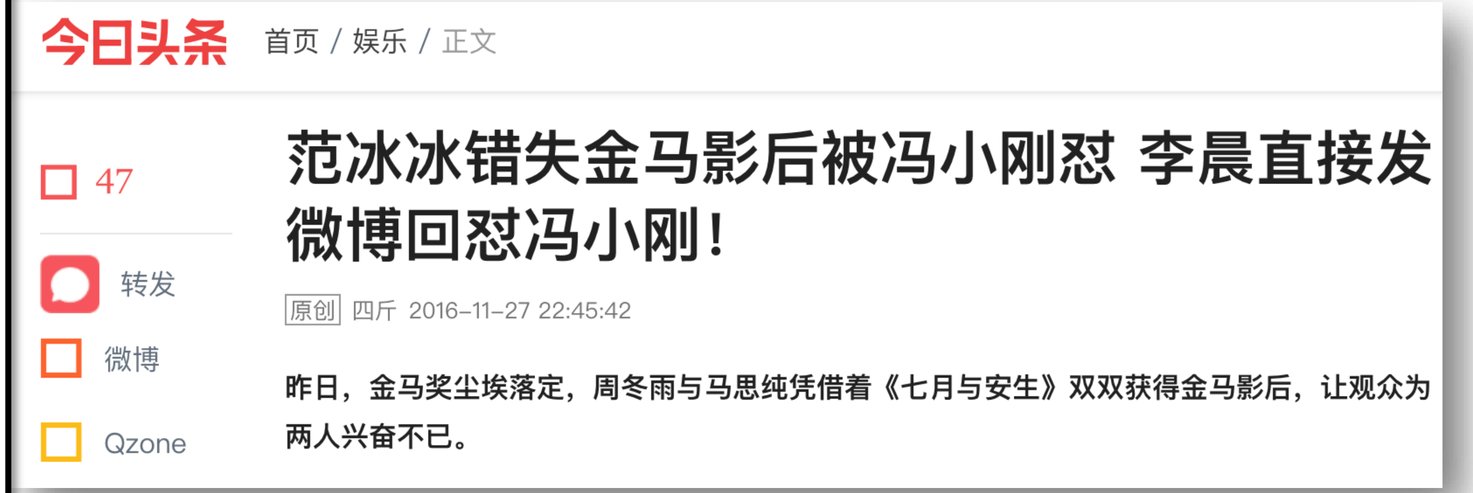
演员多 成为明星的演员少 冰冰却早已经做到了

There are many actresses, but not many became stars. ***Bingbing*** did this early.

User id 3234567

冯导为这部电影毁掉了自己的形象，烧掉了自己的人气 电影好坏第一责任 人是导演。

Director Feng ruined his image and reputation for this movie



今日头条 首页 / 娱乐 / 正文

□ 47

转发

□ 微博

□ Qzone

范冰冰错失金马影后被冯小刚怼 李晨直接发微博回怼冯小刚！

原创 四斤 2016-11-27 22:45:42

昨日，金马奖尘埃落定，周冬雨与马思纯凭借着《七月与安生》双双获得金马影后，让观众为两人兴奋不已。

Data

- Data format:

News article

Bingbing Fan did not win the Golden Horse Awards, drawing attacks from Xiaogang Feng. Chen Li joins the fight on Weibo.

Yesterday, the Golden Horse Awards were finally announced. Dongyu Zhou and Sichun Ma won the best lead actress with their performances in ”Soul Mate”, excites the audiences. (…)

Comments

User id 1234567

她真的是明星 算不上演员

__She__ is just a star, not an actress.

User id 2234567

演员多 成为明星的演员少 冰冰却早已经做到了

There are many actresses, but not many became stars. *__Bingbing__* did this early.

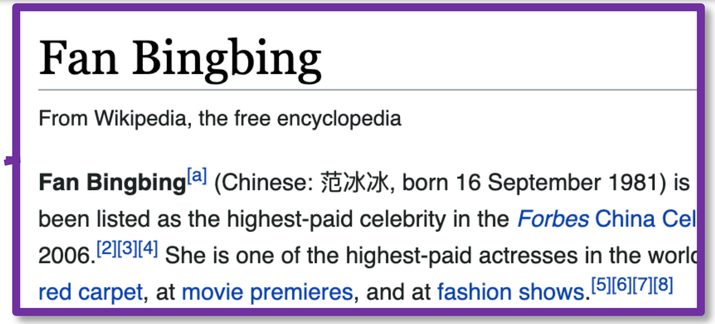
User id 3234567

冯导为这部电影毁掉了自己的形象，烧掉了自己的人气

电影好坏第一责任 人是导演。

__Director Feng__ ruined his image and reputation for this movie

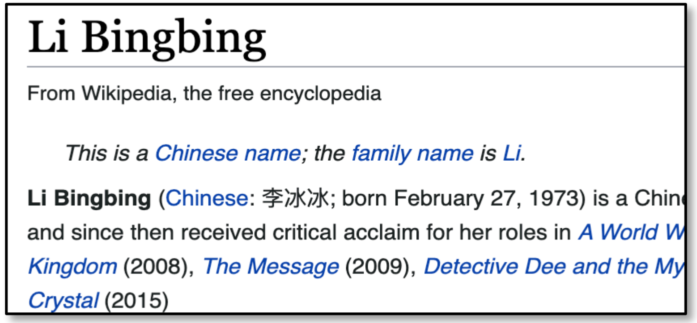
Annotation



Fan Bingbing

From Wikipedia, the free encyclopedia

Fan Bingbing[a] (Chinese: 范冰冰, born 16 September 1981) is been listed as the highest-paid celebrity in the *Forbes* China Cel 2006.[2][3][4] She is one of the highest-paid actresses in the worl red carpet, at movie premieres, and at fashion shows.[5][6][7][8]



Li Bingbing

From Wikipedia, the free encyclopedia

This is a Chinese name; the family name is Li.

Li Bingbing (Chinese: 李冰冰; born February 27, 1973) is a Chin and since then received critical acclaim for her roles in *A World W* *Kingdom* (2008), *The Message* (2009), *Detective Dee and the My* *Crystal* (2015)

Data

- Annotation Statistics:

	Entertainment	**Product**
Articles	995	783
avg # Char.	747	684
Comments	30,630	5,189
avg # Char.	21	23
Mentions	46,942	7,497
Unique Entities	1,846	470

Data

- Annotation Statistics:

	Entertainment	**Product**
Articles	995	783
avg # Char.	747	684
Comments	30,630	5,189
avg # Char.	**21**	**23**
Mentions	46,942	7,497
Unique Entities	1,846	470

Lack of context

Data

- Annotation Statistics:

Lack of context

Diverse mention variations

	Entertainment	**Product**
Articles	995	783
avg # Char.	747	684
Comments	30,630	5,189
avg # Char.	21	23
Mentions	46,942	7,497
Unique Entities	**1,846**	**470**
max dist. ment.	**121**	**48**

Data

- Mention Examples

范冰冰(Bingbing Fan)	冰冰(Bing bing), 冰姐 (sister Bing), 范 (Fan), 国际女神 (international goddess)
王菲 (Faye Wong)	戏子(actress), 菲姐(Sister Faye), 她(she), 单亲妈妈 (single mother), 天后 (queen), 明星 (celebrity)
周星驰(Stephen Chow)	偶像 (idol), 名人 (famous person), 影帝 (film emperor), 艺术家 (artist), 周星驰(Stephen Chow)

索尼 (Sony)	黑科技之父 (father of black tech), 索尼大法 (sony grand magic), 索泥 (so-mud)
苹果 (Apple)	苹果 (apple), 他们公司 (their company), 果果 (fruity), 老产品 (old product), 蘋果 (apple)
马自达(Mazda)	好车 (nice car), 日系宝马 (Japanese BMW), 法拉马 (Frama), 马儿 (horse)

Data

- Mention Examples

范冰冰(Bingbing Fan)	冰冰(Bing bing), 冰姐 (sister Bing), 范 (Fan), 国际女神 (international goddess)
王菲 (Faye Wong)	戏子(actress), 菲姐(Sister Faye), 她(she), 单亲妈妈 (single mother), 天后 (queen), 明星 (celebrity)
周星驰(Stephen Chow)	偶像 (idol), 名人 (famous person), 影帝 (film emperor), 艺术家 (artist), **周星驰(Stephen Chow)**

索尼 (Sony)	黑科技之父 (father of black tech), 索尼大法 (sony grand magic), 索泥 (so-mud)
苹果 (Apple)	**苹果 (apple),** 他们公司 (their company), 果果 (fruity), 老产品 (old product), 蘋果 (apple)
马自达(Mazda)	好车 (nice car), 日系宝马 (Japanese BMW), 法拉马 (Frama), 马儿 (horse)

- Mention Types
  - `Canonical Names`

Data

- Mention Examples

范冰冰(Bingbing Fan)	**冰冰**(Bing bing), 冰姐 (sister Bing), 范 (Fan), 国际女神 (international goddess)
王菲 (Faye Wong)	戏子(actress), **菲姐**(Sister Faye), 她(she), 单亲妈妈 (single mother), 天后 (queen), 明星 (celebrity)
周星驰(Stephen Chow)	偶像 (idol), 名人 (famous person), 影帝 (film emperor), 艺术家 (artist), 周星驰(Stephen Chow)

索尼 (Sony)	黑科技之父 (father of black tech), 索尼大法 (sony grand magic), 索泥 (so-mud)
苹果 (Apple)	苹果 (apple), 他们公司 (their company), 果果 (fruity), 老产品 (old product), 蘋果 (apple)
马自达(Mazda)	好车 (nice car), 日系宝马 (Japanese BMW), 法拉马 (Frama), 马儿 (horse)

- Mention Types
  - Nicknames

Data

- Mention Examples

范冰冰(Bingbing Fan)	冰冰(Bing bing), 冰姐 (sister Bing), 范 (Fan), 国际女神 (international goddess)
王菲 (Faye Wong)	戏子(actress), 菲姐(Sister Faye), 她(she), 单亲妈妈 (single mother), 天后 (queen), 明星 (celebrity)
周星驰(Stephen Chow)	偶像 (idol), 名人 (famous person), 影帝 (film emperor), 艺术家 (artist), 周星驰(Stephen Chow)

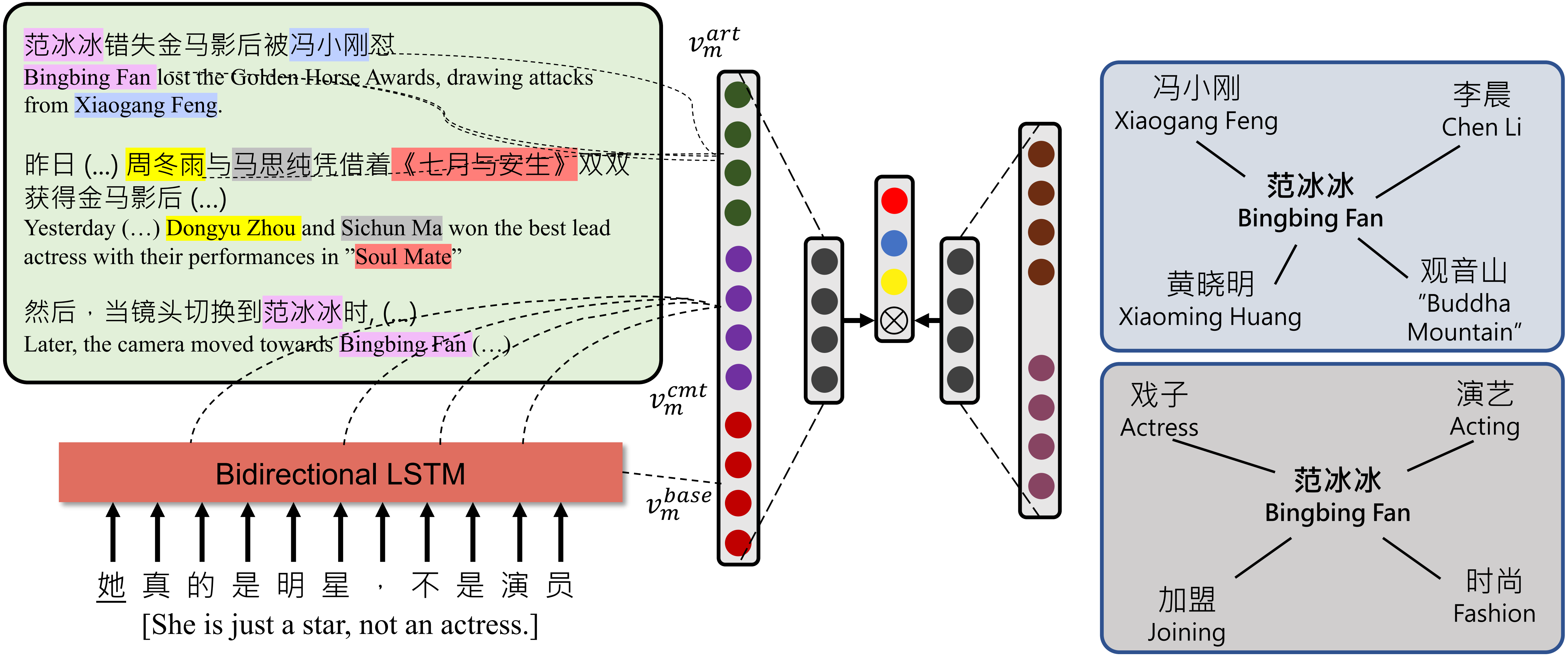
索尼 (Sony)	黑科技之父 (father of black tech), 索尼大法 (sony grand magic), 索泥 (so-mud)
苹果 (Apple)	苹果 (apple), 他们公司 (their company), 果果 (fruity), 老产品 (old product), 蘋果 (apple)
马自达(Mazda)	好车 (nice car), 日系宝马 (Japanese BMW), 法拉马 (Frama), 马儿 (horse)

- Mention Types
  - `Others`

Roadmap

- Motivation
- Related Work
- Data
- Model
- Experiments
- Conclusion

Model



范冰冰错失金马影后被冯小刚怼
Bingbing Fan lost the Golden Horse Awards, drawing attacks from Xiaogang Feng.

昨日 (...) 周冬雨与马思纯凭借着《七月与安生》双双获得金马影后 (...)
Yesterday (…) Dongyu Zhou and Sichun Ma won the best lead actress with their performances in ”Soul Mate”

然后，当镜头切换到范冰冰时, (...)
Later, the camera moved towards Bingbing Fan (…)

她 真 的 是 明 星 ， 不 是 演 员
[She is just a star, not an actress.]

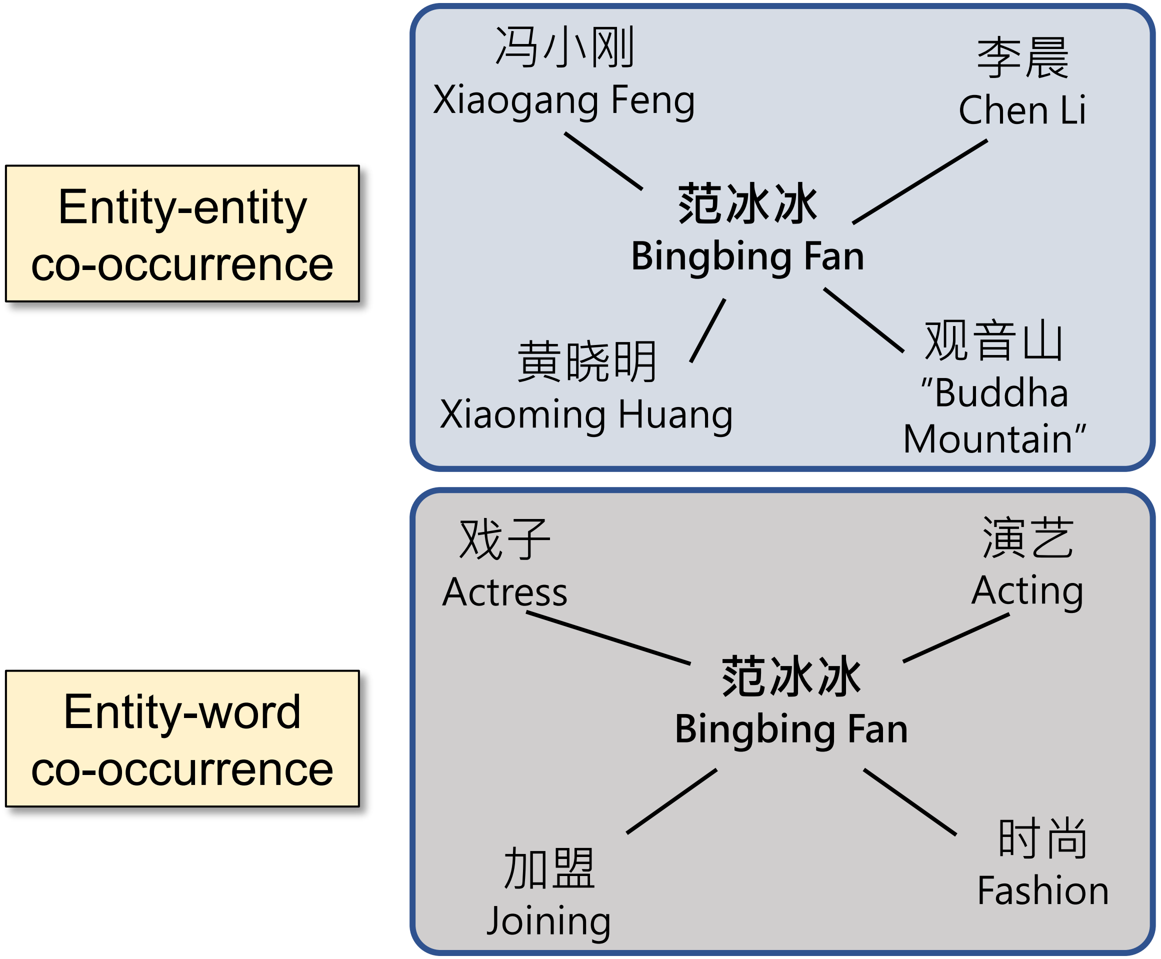
Bidirectional LSTM
v_m^{art}
v_m^{cmt}
v_m^{base}
冯小刚
Xiaogang Feng
李晨
Chen Li
范冰冰
Bingbing Fan
黄晓明
Xiaoming Huang
观音山
“Buddha Mountain”
戏子
Actress
演艺
Acting
范冰冰
Bingbing Fan
加盟
Joining
时尚
Fashion

Mention representation

Entity representation

Model

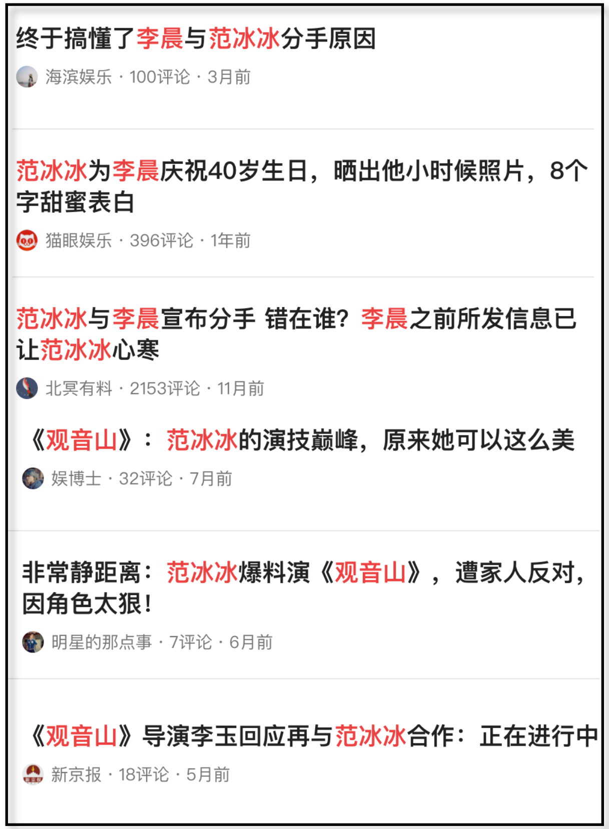
- Entity representation


Entity-entity co-occurrence
冯小刚
Xiaogang Feng
李晨
Chen Li
范冰冰
Bingbing Fan
黄晓明
Xiaoming Huang
观音山
"Buddha Mountain"
Entity-word co-occurrence
戏子
Actress
演艺
Acting
范冰冰
Bingbing Fan
加盟
Joining
时尚
Fashion

Model

- Entity representation

News headlines



终于搞懂了李晨与范冰冰分手原因

海滨娱乐 · 100评论 · 3月前

范冰冰为李晨庆祝40岁生日，晒出他小时候照片，8个字甜蜜表白

猫眼娱乐 · 396评论 · 1年前

范冰冰与李晨宣布分手 错在谁？李晨之前所发信息已让范冰冰心寒

北冥有料 · 2153评论 · 11月前

《观音山》：范冰冰的演技巅峰，原来她可以这么美

娱博士 · 32评论 · 7月前

非常静距离：范冰冰爆料演《观音山》，遭家人反对，因角色太狠！

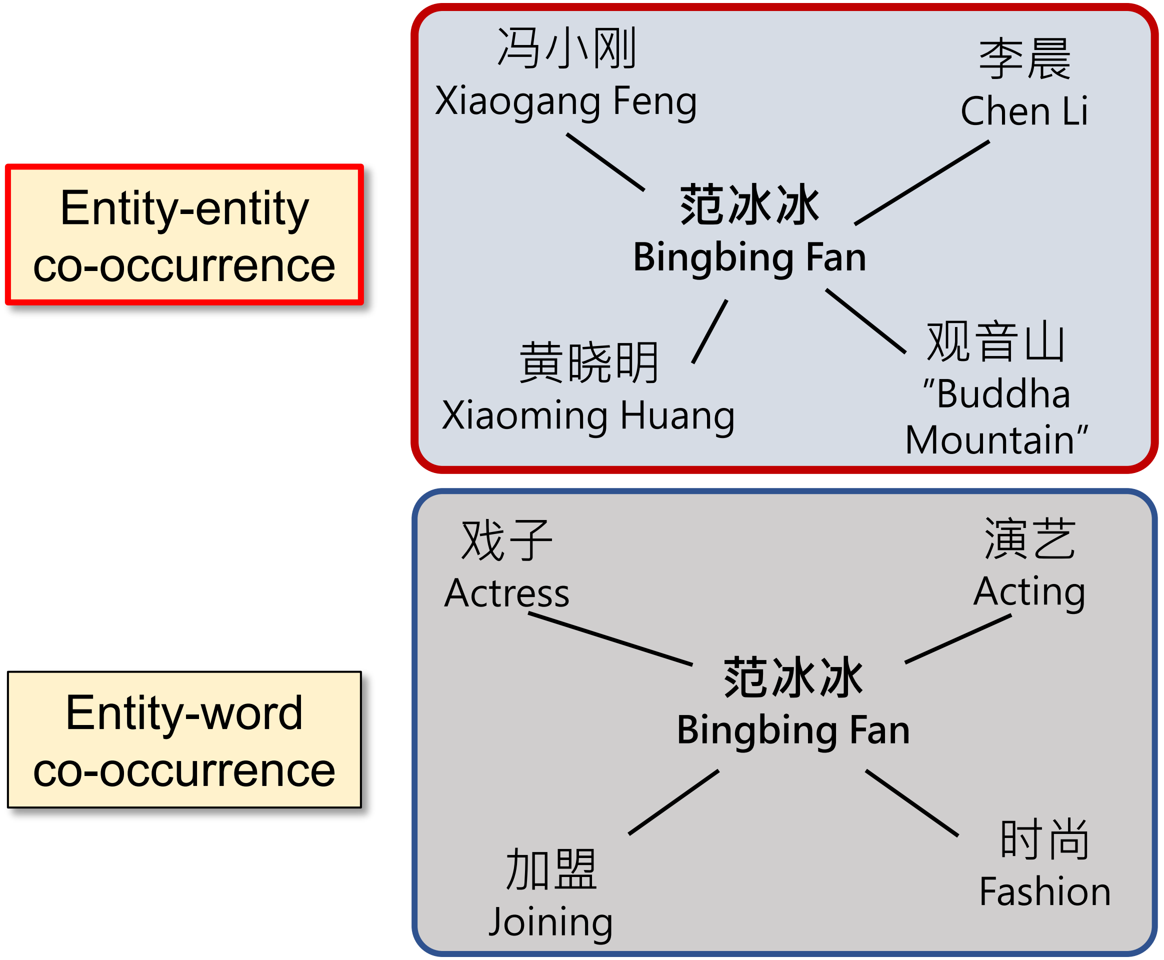
明星的那点事 · 7评论 · 6月前

《观音山》导演李玉回应再与范冰冰合作：正在进行中

新京报 · 18评论 · 5月前

Entity-entity co-occur matrix

Node2vec [Grover and Leskovec, 2016]



Entity-entity co-occurrence

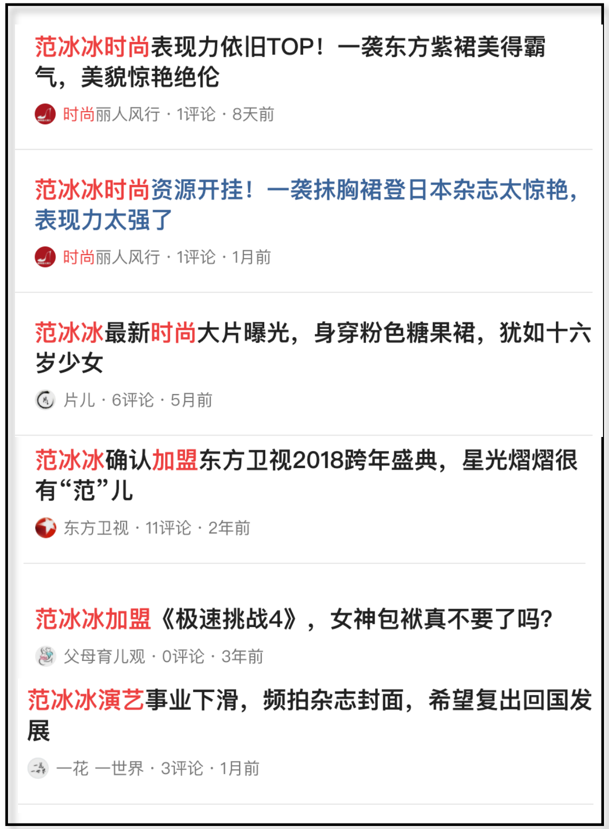
Entity-word co-occurrence

冯小刚
Xiaogang Feng
李晨
Chen Li
范冰冰
Bingbing Fan
黄晓明
Xiaoming Huang
观音山
"Buddha Mountain"
戏子
Actress
演艺
Acting
范冰冰
Bingbing Fan
加盟
Joining
时尚
Fashion

Model

- Entity representation

News headlines



范冰冰时尚表现力依旧TOP！一袭东方紫裙美得霸气，美貌惊艳绝伦

时尚丽人风行 · 1评论 · 8天前

范冰冰时尚资源开挂！一袭抹胸裙登日本杂志太惊艳，表现力太强了

时尚丽人风行 · 1评论 · 1月前

范冰冰最新时尚大片曝光，身穿粉色糖果裙，犹如十六岁少女

片儿 · 6评论 · 5月前

范冰冰确认加盟东方卫视2018跨年盛典，星光熠熠很有“范”儿

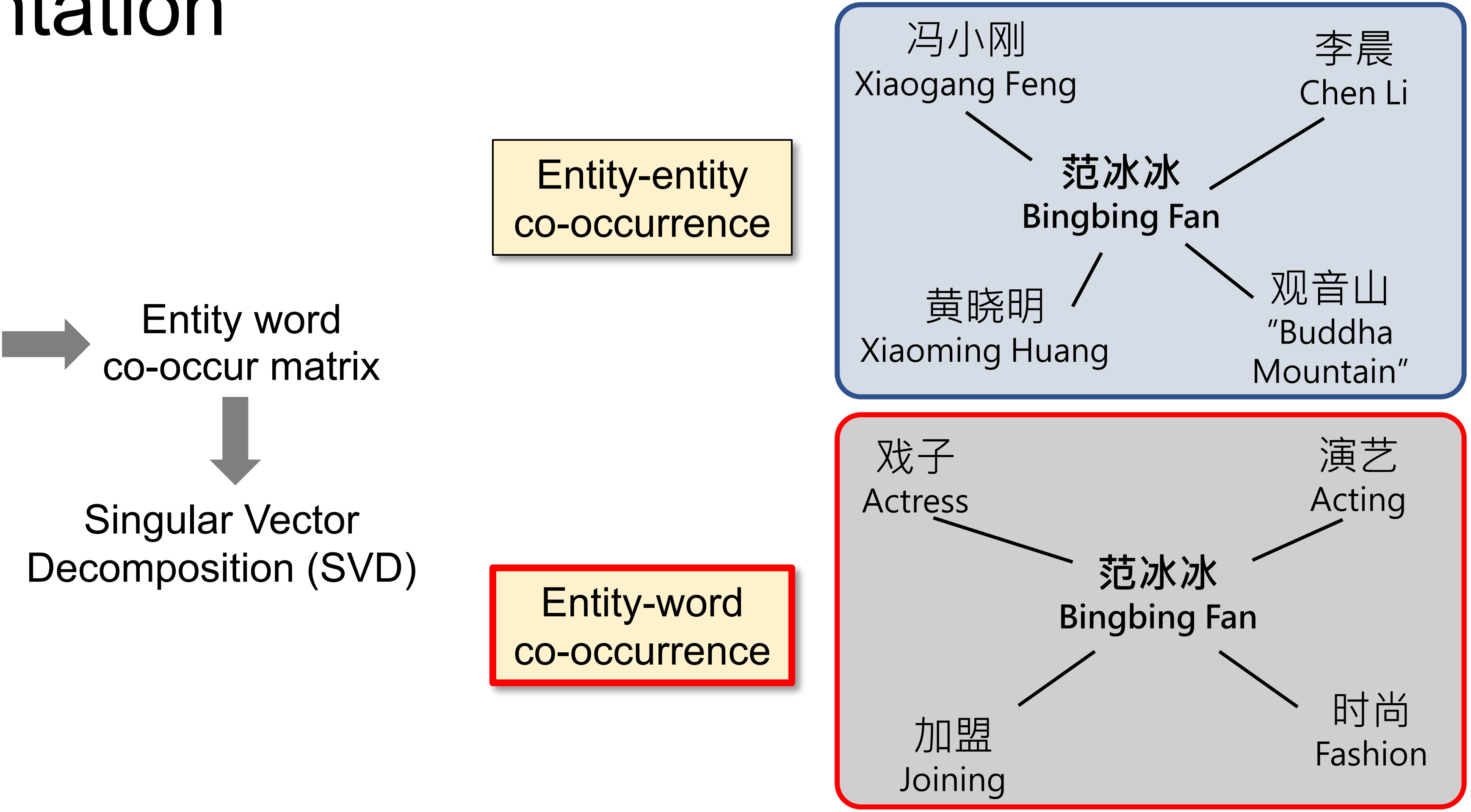
东方卫视 · 11评论 · 2年前

范冰冰加盟《极速挑战4》，女神包袱真不要了吗?

父母育儿观 · 0评论 · 3年前

范冰冰演艺事业下滑，频拍杂志封面，希望复出回国发展

一花 一世界 · 3评论 · 1月前



Entity word co-occur matrix

Singular Vector Decomposition (SVD)

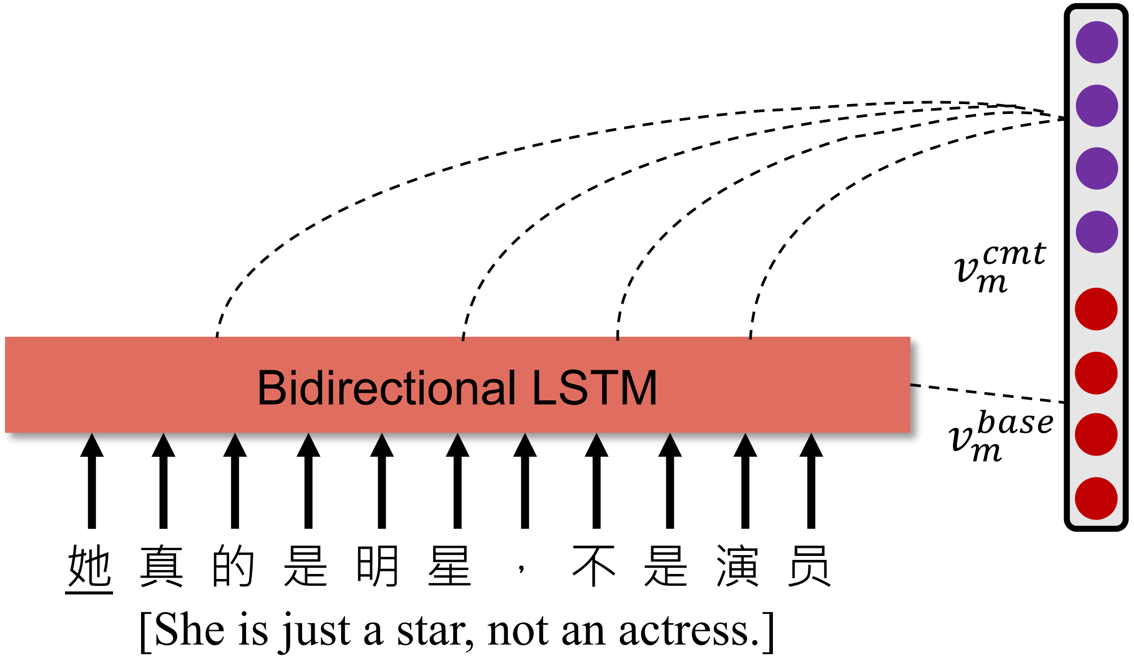
Entity-entity co-occurrence

Entity-word co-occurrence

冯小刚
Xiaogang Feng
李晨
Chen Li
范冰冰
Bingbing Fan
黄晓明
Xiaoming Huang
观音山
“Buddha Mountain”
戏子
Actress
演艺
Acting
范冰冰
Bingbing Fan
加盟
Joining
时尚
Fashion

Model

- Character-level comment encoding


Bidirectional LSTM
v_m^{cmt}
v_m^{base}
她 真 的 是 明 星 ， 不 是 演 员
[She is just a star, not an actress.]

$$\boldsymbol{h}_i = g(\boldsymbol{h}_{i-1}, [\boldsymbol{x}_i^c; q_i])$$

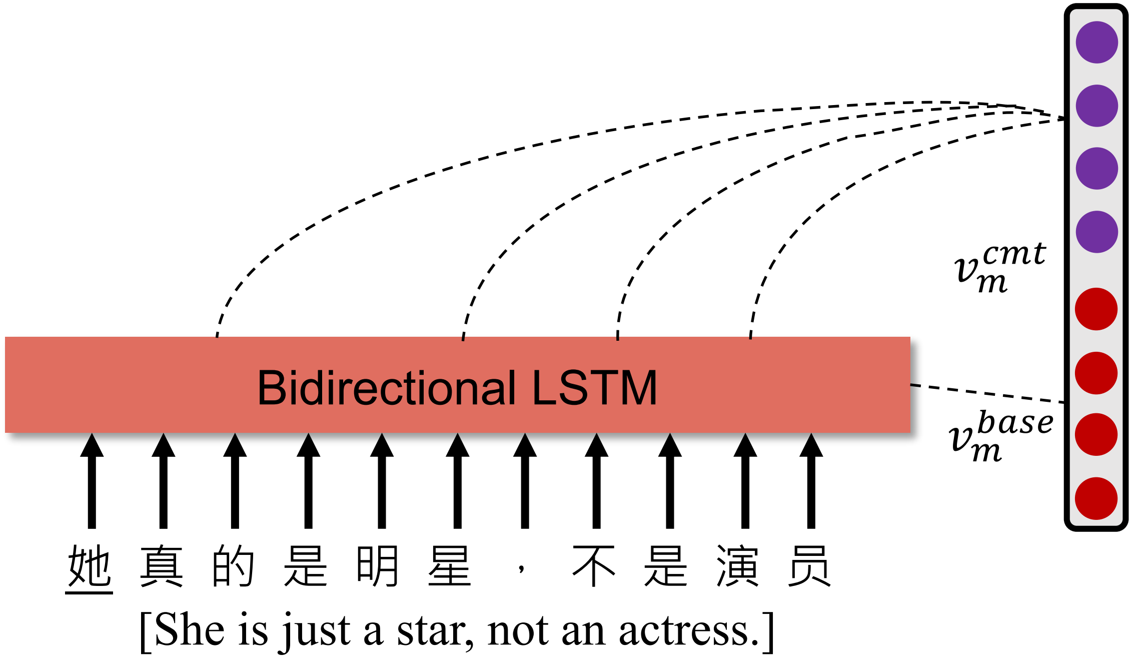
LSTM states

Character embedding

One-bit binary encoding of mention span

Model

- Character-level comment encoding


Bidirectional LSTM
v_m^{cmt}
v_m^{base}
她 真 的 是 明 星 ， 不 是 演 员

[She is just a star, not an actress.]

$$\boldsymbol{h}_i = g(\boldsymbol{h}_{i-1}, [\boldsymbol{x}_i^c; q_i])$$

LSTM states

Character embedding

One-bit binary encoding of mention span

$$\boldsymbol{v}_m^{base} = \boldsymbol{h}_T$$

Base mention representation

Model

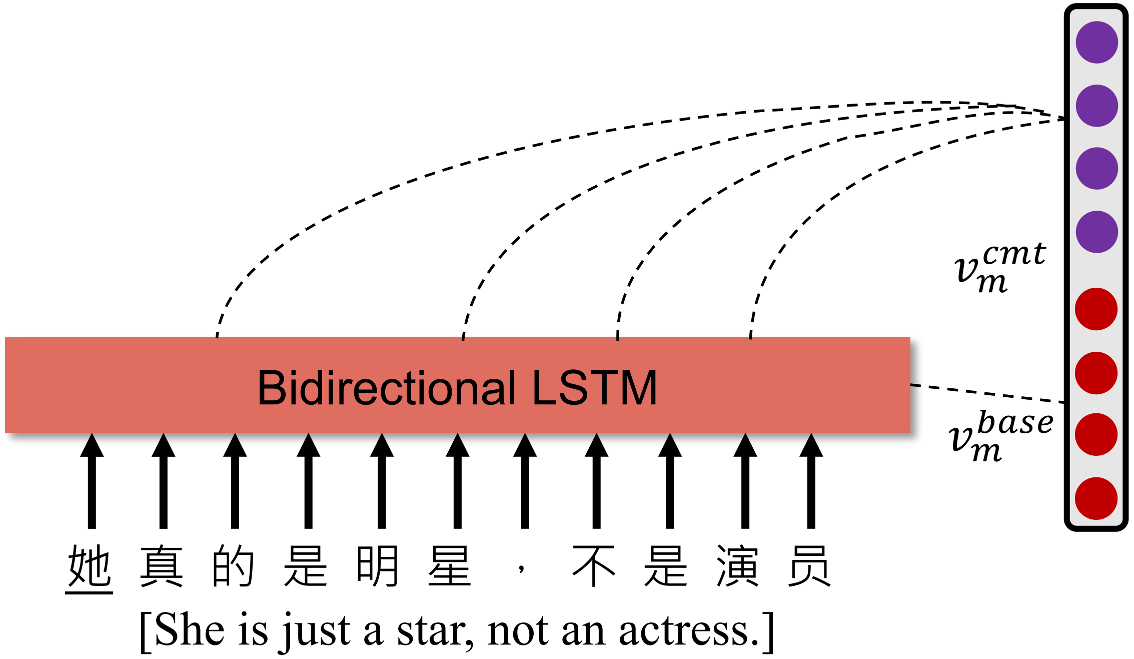
- Character-level comment encoding
- Comment attention

$$\alpha_i^{cmt} = \mathrm{softmax}(\boldsymbol{h}_i' \boldsymbol{W}_c \tilde{\boldsymbol{m}})$$

Normalized attention weights

Bilinear prod.

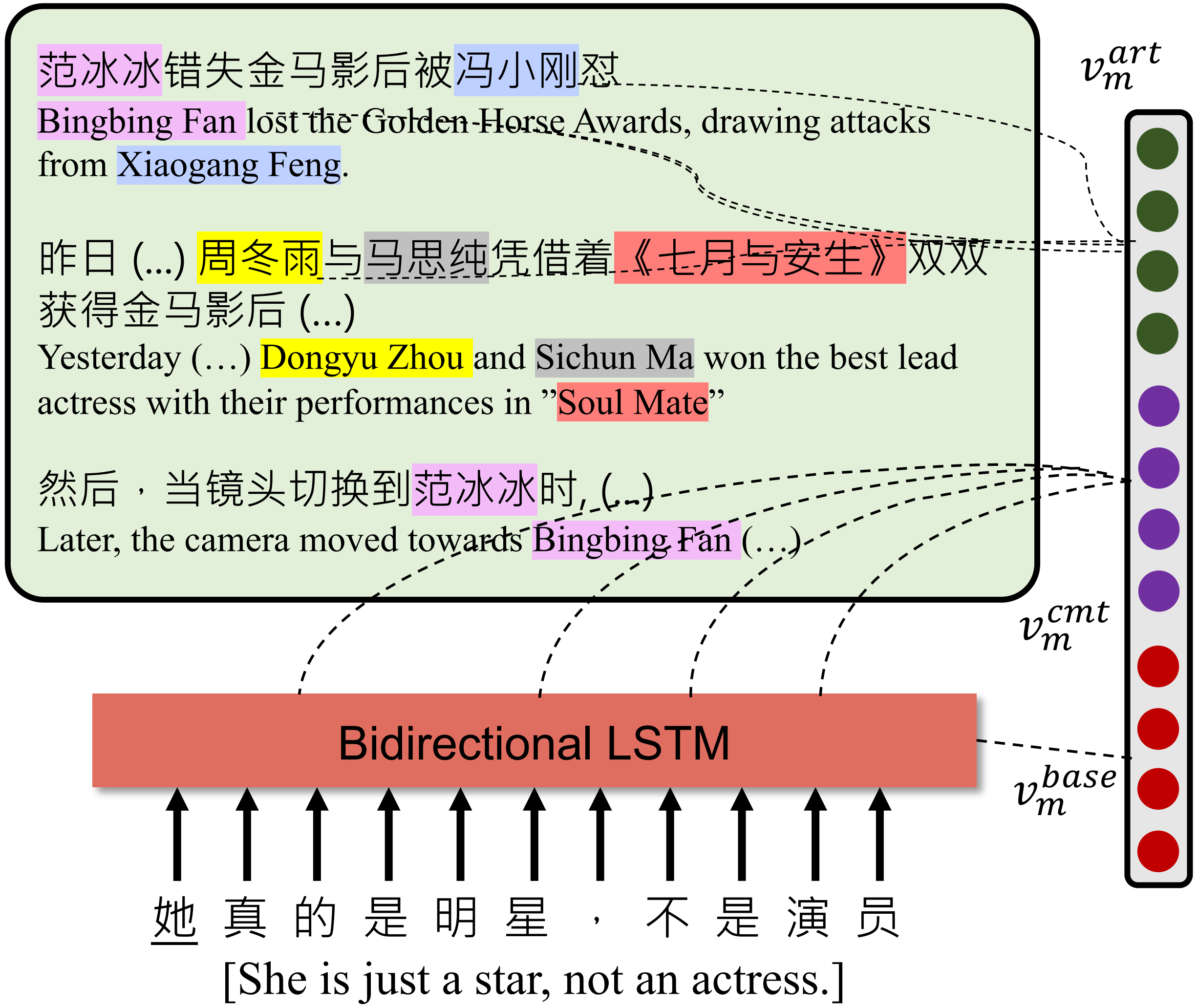
Avg. char. Embedding of mention span



v_m^{cmt}
Bidirectional LSTM
v_m^{base}
她 真 的 是 明 星 ， 不 是 演 员
[She is just a star, not an actress.]

$$\boldsymbol{v}_m^{cmt} = \sum_{i=1}^{T} \alpha_i^{cmt} \boldsymbol{h}_i$$

Model



范冰冰错失金马影后被冯小刚怼

Bingbing Fan lost the Golden Horse Awards, drawing attacks from Xiaogang Feng.

昨日 (...) 周冬雨与马思纯凭借着《七月与安生》双双获得金马影后 (...)

Yesterday (…) Dongyu Zhou and Sichun Ma won the best lead actress with their performances in ”Soul Mate”

然后，当镜头切换到范冰冰时, (...)

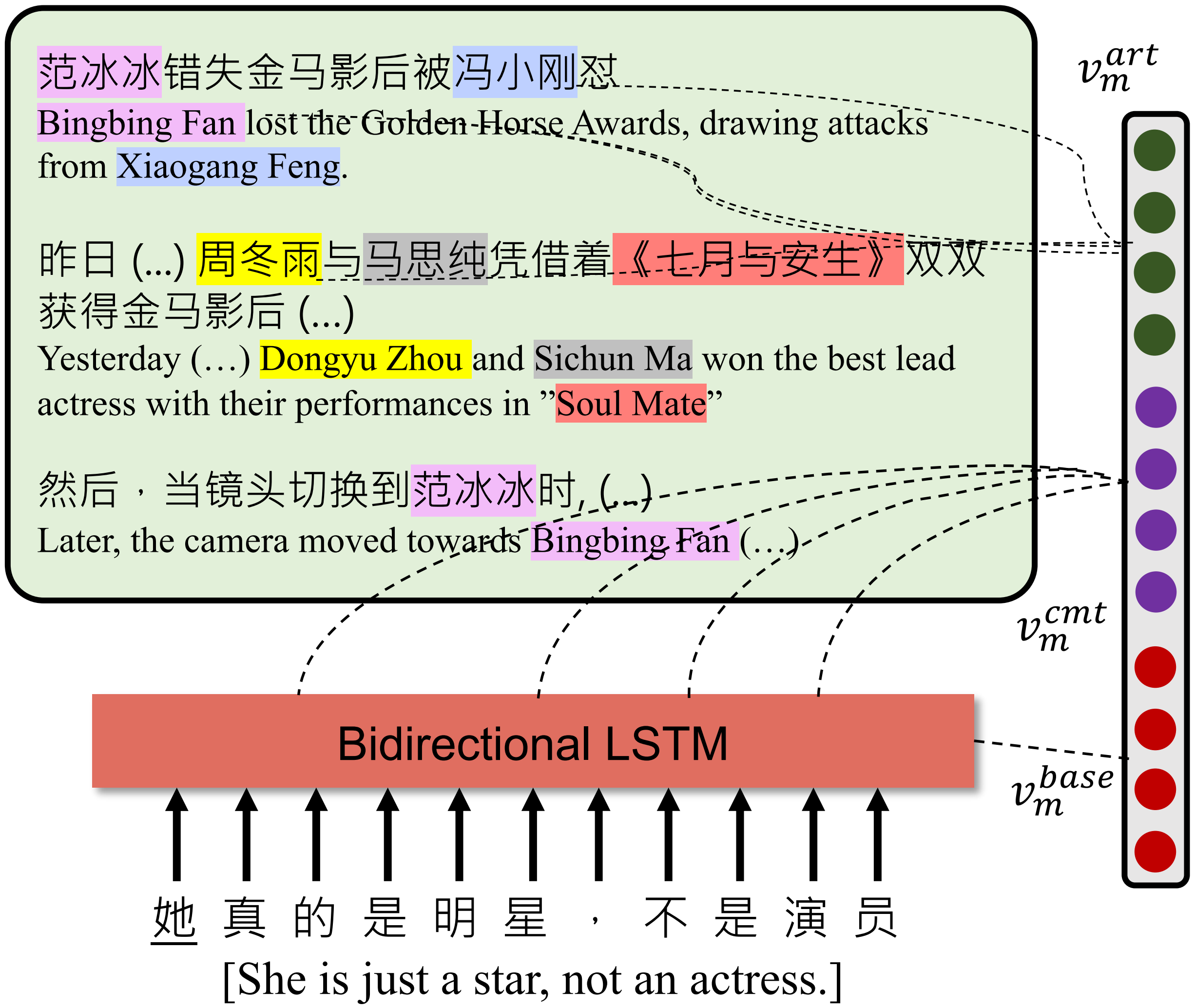
Later, the camera moved towards Bingbing Fan (…)

她 真 的 是 明 星 ， 不 是 演 员

[She is just a star, not an actress.]

- Article entity attention
  - String matched entity canonical names -> $\mathcal{E}_a$

Model



范冰冰错失金马影后被冯小刚怼
Bingbing Fan lost the Golden Horse Awards, drawing attacks from Xiaogang Feng.

昨日 (...) 周冬雨与马思纯凭借着《七月与安生》双双获得金马影后 (...)
Yesterday (…) Dongyu Zhou and Sichun Ma won the best lead actress with their performances in ”Soul Mate”

然后，当镜头切换到范冰冰时, (...)
Later, the camera moved towards Bingbing Fan (…)

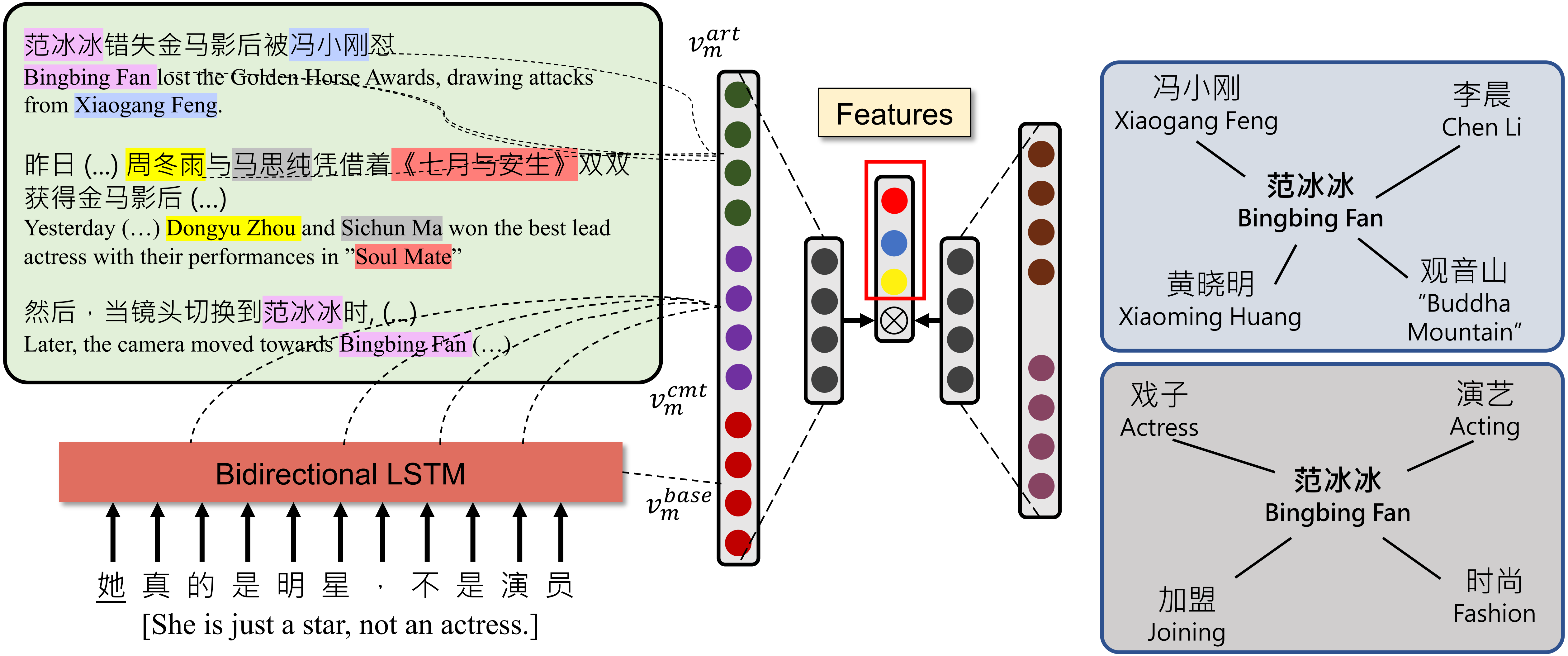
她 真 的 是 明 星 ， 不 是 演 员
[She is just a star, not an actress.]

- Article entity attention

$$\beta_j^{art} = \mathrm{softmax}(\boldsymbol{u}_j' \boldsymbol{W}_a \tilde{\boldsymbol{m}})$$

$$\boldsymbol{v}_m^{art} = \sum_{j=1}^{|\varepsilon_a|} \beta_j^{art} \boldsymbol{u}_j$$

Model



范冰冰错失金马影后被冯小刚怼
Bingbing Fan lost the Golden Horse Awards, drawing attacks from Xiaogang Feng.

昨日 (...) 周冬雨与马思纯凭借着《七月与安生》双双获得金马影后 (...)
Yesterday (…) Dongyu Zhou and Sichun Ma won the best lead actress with their performances in ”Soul Mate”

然后，当镜头切换到范冰冰时, (...)
Later, the camera moved towards Bingbing Fan (…)

她 真 的 是 明 星 ， 不 是 演 员
[She is just a star, not an actress.]

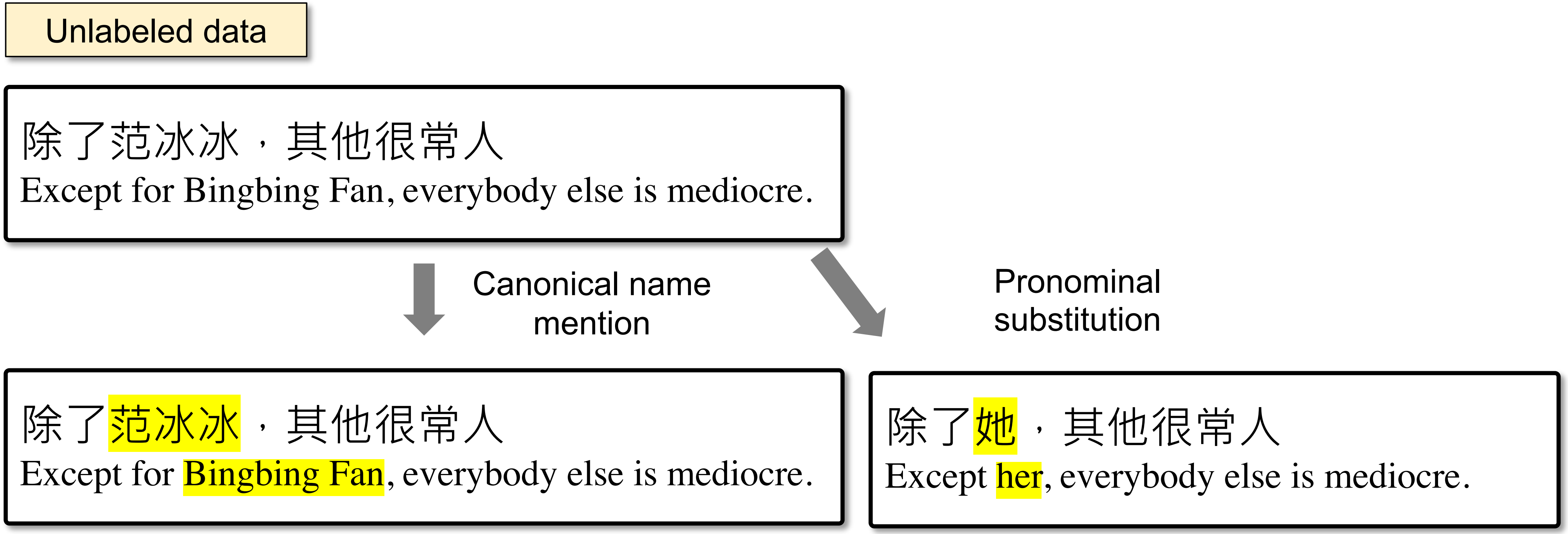
v_m^{art}
v_m^{cmt}
v_m^{base}
Bidirectional LSTM
Features
冯小刚
Xiaogang Feng
李晨
Chen Li
范冰冰
Bingbing Fan
黄晓明
Xiaoming Huang
观音山
“Buddha Mountain”
戏子
Actress
演艺
Acting
范冰冰
Bingbing Fan
加盟
Joining
时尚
Fashion

Mention representation

Entity representation

Model

- Weakly Supervised Pre-training



Unlabeled data

除了范冰冰，其他很常人
Except for Bingbing Fan, everybody else is mediocre.

Canonical name mention

Pronominal substitution

除了范冰冰，其他很常人
Except for Bingbing Fan, everybody else is mediocre.

除了她，其他很常人
Except her, everybody else is mediocre.

Experiments

	Entertainment		**Product**	
	Accuracy	MRR	Accuracy	MRR
MatchCanonAndNick	42.70	-	54.97	-
PriorProb	54.22	-	77.71	-
VectorSpaceModel	26.04	34.70	33.57	42.72
LogReg	57.69	62.76	61.37	63.11
Ment-norm [Le and Titov, 2018]	58.60	63.69	68.56	71.32
XREF	67.22	73.92	77.26	81.52

Experiments

Matching based

	Entertainment		**Product**	
	Accuracy	MRR	Accuracy	MRR
MatchCanonAndNick	42.70	-	54.97	-
PriorProb	54.22	-	77.71	-
VectorSpaceModel	26.04	34.70	33.57	42.72
LogReg	57.69	62.76	61.37	63.11
Ment-norm [Le and Titov, 2018]	58.60	63.69	68.56	71.32
XREF	67.22	73.92	77.26	81.52

Learning based

Experiments

	Entertainment		**Product**	
	Accuracy	MRR	Accuracy	MRR
MatchCanonAndNick	42.70	-	54.97	-
PriorProb	54.22	-	77.71	-
VectorSpaceModel	26.04	34.70	33.57	42.72
LogReg	57.69	62.76	61.37	63.11
Ment-norm [Le and Titov, 2018]	58.60	63.69	68.56	71.32
XREF	**67.22**	**73.92**	**77.26**	**81.52**

Experiments

- Ablations

	Entertainment	**Product**
XREF	57.93	65.33
w/o comment attention	51.94 (-5.99)	62.22 (-3.11)

Experiments

- Ablations

	Entertainment	**Product**
XREF	57.93	65.33
w/o comment attention	51.94 (-5.99)	62.22 (-3.11)
w/o comment and article attention	44.90 (-13.03)	55.33 (-10.00)

Experiments

- Effect of Article Attention

	Prediction correct	**Prediction wrong**
Attention correct	4,366 (72.3%)	1,599 (26.8%)
Attention wrong	397 (18.5%)	1,749 (81.5%)

Conclusion

- We propose XREF, an EL framework designed for social media news comments.

- We construct a dataset for Chinese EL study.

- Experiments show strong performance over competitive benchmarks.