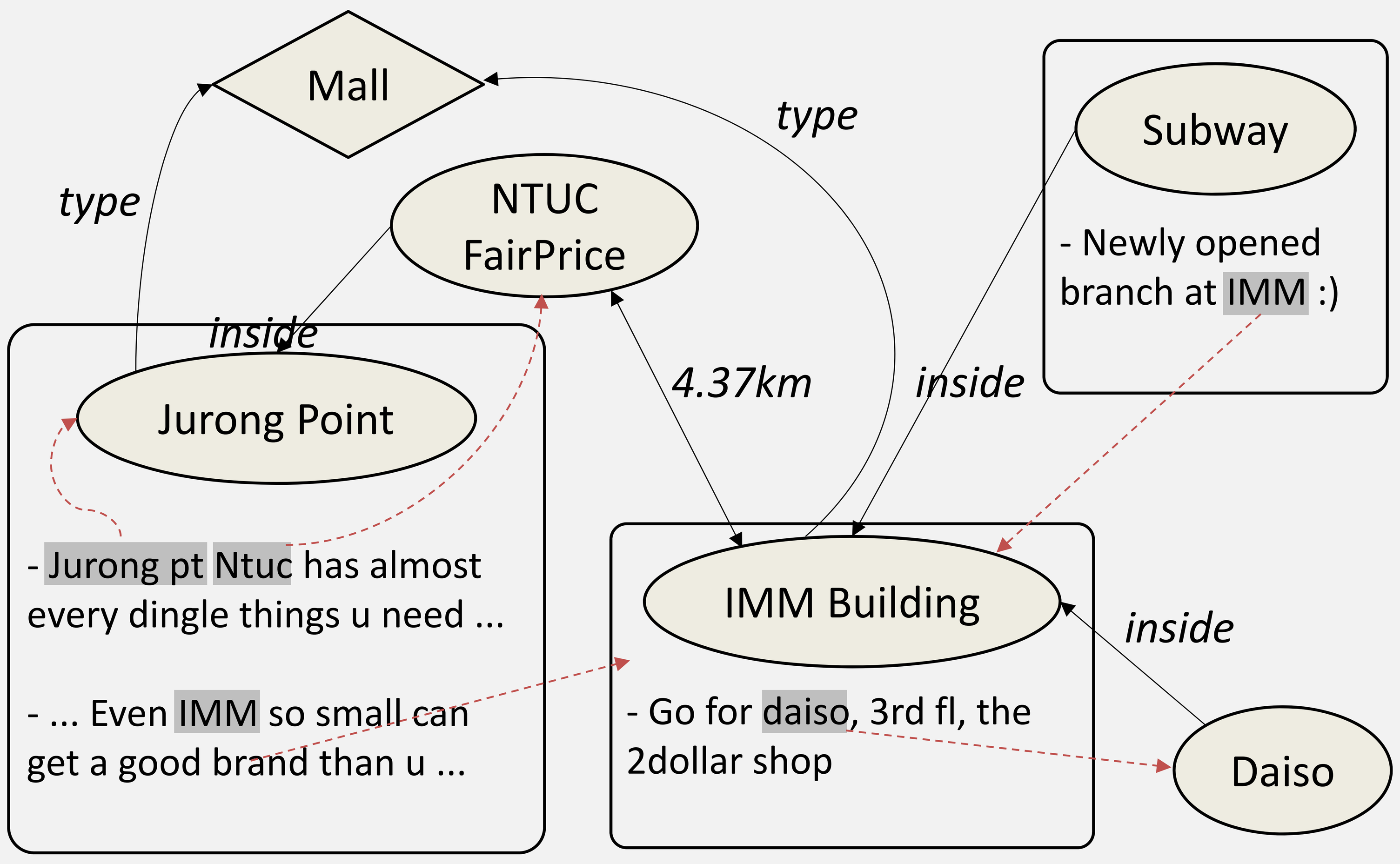
Linking Fine-Grained Locations in User Comments

Jialong Han *, Aixin Sun †, Gao Cong †, Wayne Xin Zhao ‡, Zongcheng Ji †, and Minh C. Phan †
* Tencent AI Lab, Shenzhen, China
† School of Computer Science and Engineering, Nanyang Technological University, Singapore
‡ School of Information, Renmin University of China, Beijing, China

Backgrounds

- **LSBNs (e.g., Foursquare and Google Maps) host profile pages for fine-grained locations.**
- **On those profile pages, users can contribute comments, ratings, and likes.**

Motivations and Challenges



- Comments on the Focal Location may mention other locations.
- Task: link those mentioned locations to their profile pages
- Explorations on this problem may benefit
  - *Applications*: comment search & gathering, target-level sentiment analysis, and (next) location recommendation
  - *Generalization* to other domains where websites share similar structures: movie (IMDB) & product (Amazon)
- Challenges of entity linking in user-generated content
  - Variability and ambiguity of entity mentions
  - Short & lack of textual context (15 words on average)

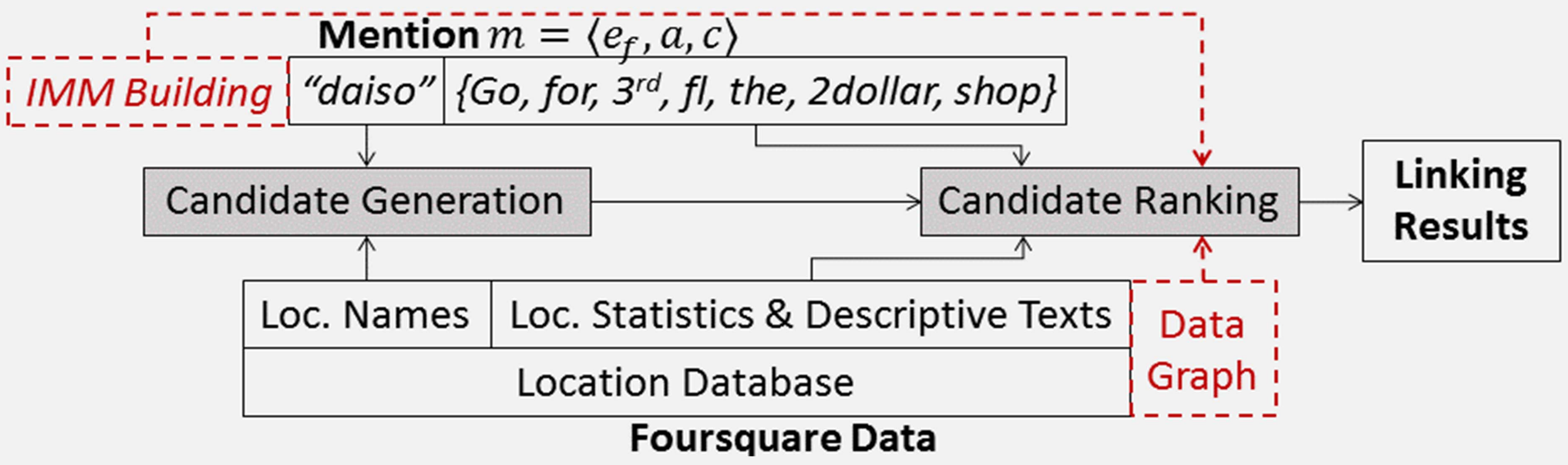
FocalLink: Observations, Framework, and Model

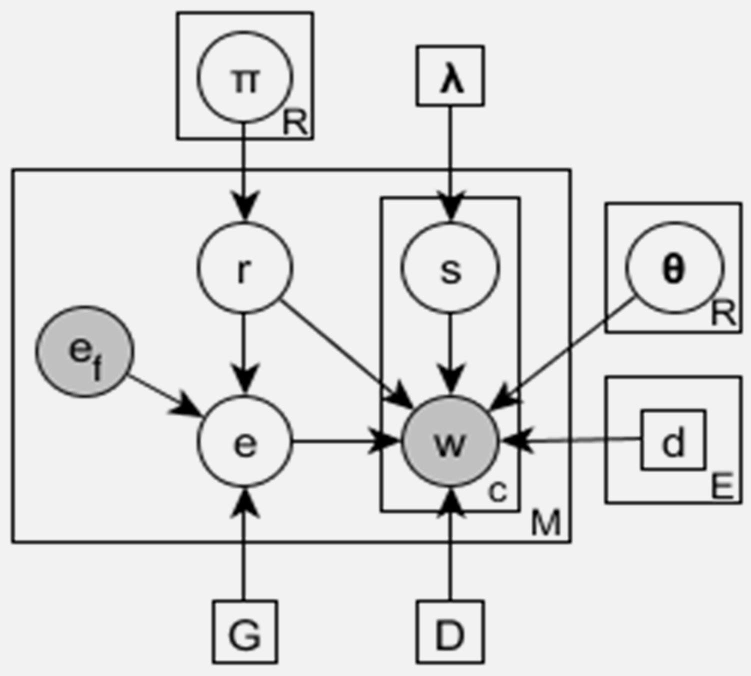
Key Observations on Mention $m = \langle e_f, a, c \rangle$

- $e^{(m)}$ and e_f are usually connected via some r.
- r is often revealed in the context c.

Generative Process of *FocalLink*

- Draw a relation $r \sim Multi(\boldsymbol{\pi})$ to follow
- Draw a location e to mention
  - $e \sim P_G(e|e_f, r), e \in \mathcal{E}(a)$
  - Estimate $P_G(e|e_f, r)$ by
    - Path-Constrained Random Walk if r is not Near
    - Else, by location-sensitive popularity $P(e|e_f)$
- Write the surrounding context c
  - Like locations, assume that each r has relation-specific words $\boldsymbol{\theta}_r$
  - $P(c|e_f, e, r) = \prod_{w \in c}[\lambda_1 P(w|d_e) + \lambda_2 P(w|\boldsymbol{\theta}_r) + (1 - \lambda_1 - \lambda_2) P(w|D)]$





Experimental Results

Foursquare Dataset

- 321,943 locations in Singapore with 442,803 comments
- 4,000 sampled comments, 828 labeled mentions

Baseline Approaches

- Different combinations of popularity, textual context, and distance information

- The baselines benefit from incorporating more information.
- Modeling e_f and relations to the mentions brings additional gains.

LINKING RESULTS OF ALL METHODS.

Approaches	Prec	Rec	F_1
Popularity	.563	.508	.534
PopContext	.619	.558	.587
PopContextDist	.679	.611	.643
FocalLink	**.690**	**.621**	**.653**

TOP-10 RELATION WORDS LEARNT BY FOCALLINK, SORTED BY $P(w|\boldsymbol{\theta}_r)$.

Self-Ref	at, to, in, on, the, it, of, s, is, from
Inside	st, at, cross, outlet, locate, rd, open, on, road, middle
Contains	at, stall, noodle, on, from, must, fry, mee, serve, has
Co-Inside	plaza, thomson, at, go, hungry, atm, check, out, beside, wordpress
Co-Type	better, than, at, on, much, compare, cheaper, as, to, price
Near	to, from, at, bus, walk, locate, road, mrt, take, go

- *FocalLink* learns relation-specific words $\boldsymbol{\theta}_r$.
- They help e linked to e_f via r supported by the comment content to gain more credit.

Tencent AI Lab
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE
中國人民大學 RENMIN UNIVERSITY OF CHINA