2007/02/18-24, CICLing-2007

Latent Variable Models for Causal Knowledge Acquisition

Takashi Inui Hiroya Takamura Manabu Okumura

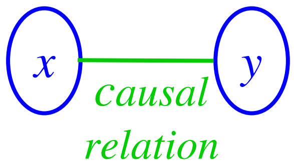
Integrated Research Institute,
Tokyo Institute of Technology, Japan

Precision and Intelligence Laboratory,
Tokyo Institute of Technology, Japan

Background & Research Goral (1/5)

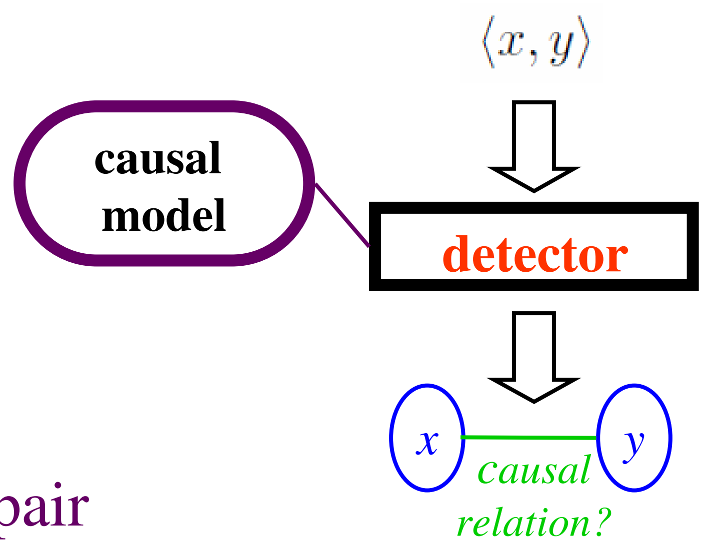
- In the fileds of AI and NLU,
  some applications need inference rules or knowledge of causal relations.
  - Question answering system
  - Dialog system

- Constructing causal models (causality detectors)
  for acquiring knowledge of causal relations is one central issue.



Background & Research Goral (2/5)

- **Causality detector** based on a causal model
  - Input: an event pair $\langle x, y \rangle$
    - Extracted from text documents



  - Output: Yes/No label
    - Yes: Holding causal relations between the input event pair
    - No: Not holding

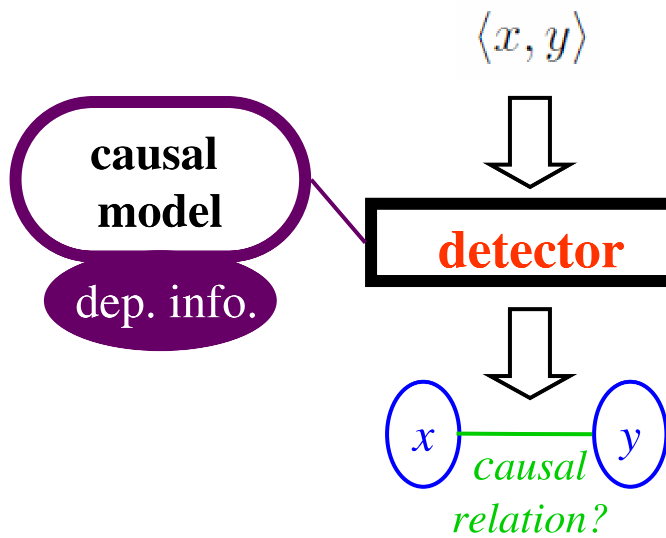
$$\hat{c} = argmax_{c_m} P(c_m | x, y)$$

Background & Research Goral (3/5)

Although

the concept "*causal relation*" is difficult to understand,

- Causal relations are assumed to be a subclass of general dependency relations.



- Causal models need to capture the dependency information between (input) event pairs.

Background & Research Goral (4/5)

dep. info.

- Two approaches for capturing dependency info.
  - Cue-phrase-based approach
    - Using cue-phrases such as “because” and “since”
    - Unable to treat event pairs **without** cue phrases
    - Medium precision, but very low coverage △
  - Statistical approach
    - Using co-occurrence statistics of event pairs
    - Independent of cue phrases
    - Keeping precision, and achieving higher coverage ◎

Background & Research Goral (5/5)

- [Chang et al. 2004]
  - One of the state-of-the art statistical models for causality detection
    - Based on naïve bayes assumption
    - Hard to capture the dependency info
      - » Critical to improve the performance of the causality detection

Our research goal is to resolve this problem.

We propose new statistical models for causality detection.

New models (1/5)

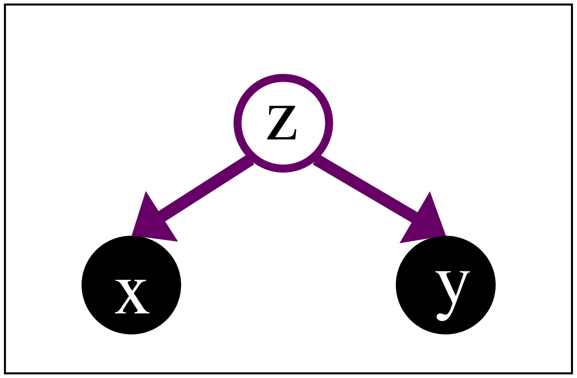
- Expanded versions of the statistical co-occurrence models proposed by [Hofmann et al. 1998]

We adopted the co-occurrence models as the bases of the new models from the observation:

If two events are holing causal relation,
these events tend to co-occur in text.

New models -- [Hofmann et al. 1998]

- **Aspect**
  - **Graphical representation of statistical dependency**



X : cause event (observed)

Y : effect event (observed)

Z : latent variable (unobserved)

$$P(x,y) = \sum_z P(x|z)P(y|z)P(z)$$

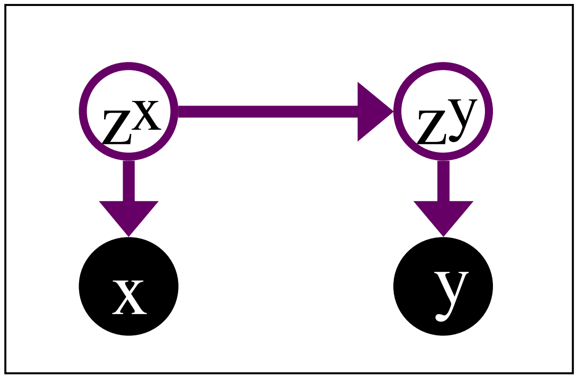
- **Z** represents semantic clusters shared by **X** and **Y**.

⬇

- The dependency info. can be captured through **Z**.

New models -- [Hofmann et al. 1998]

Product



$$P(x,y) = \sum_{z^x,\, z^y} P(x|z^x)P(y|z^y)P(z^y|z^x)P(z^x)$$

- Almost the same as **aspect,**

- Differences
  - Two latent variables $\mathbf{Z}^{\mathbf{x}}$ and $\mathbf{Z}^{\mathbf{y}}$, and
  - Statistical dependency between them

New models -- [Hofmann et al. 1998]

- Two statistical co-occurrence models
  - **Aspect**
  - **Product**

To summarize, these models

- Able to incorporate dependency information via latent variable(s),
- But, unable to treat causality information (Yes/No label).

Causality information **=** (Yes/No label)

Output: Yes/No label

Yes: Holding causal relations between the input event pair

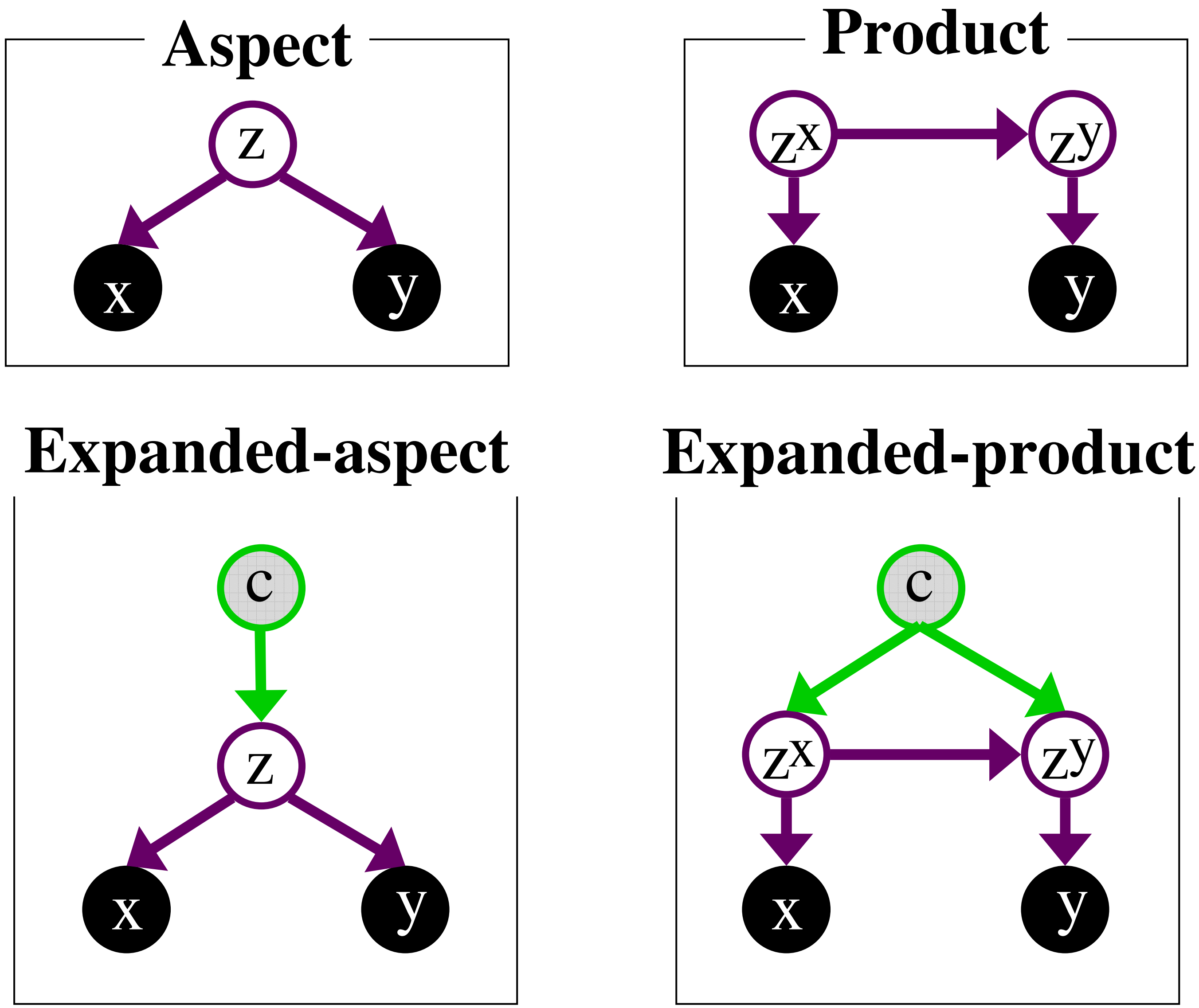
No: Not holding

We introduce a random variable ***C*** to **Aspect** and **Product** models.

(The solution is direct and very simple !!)

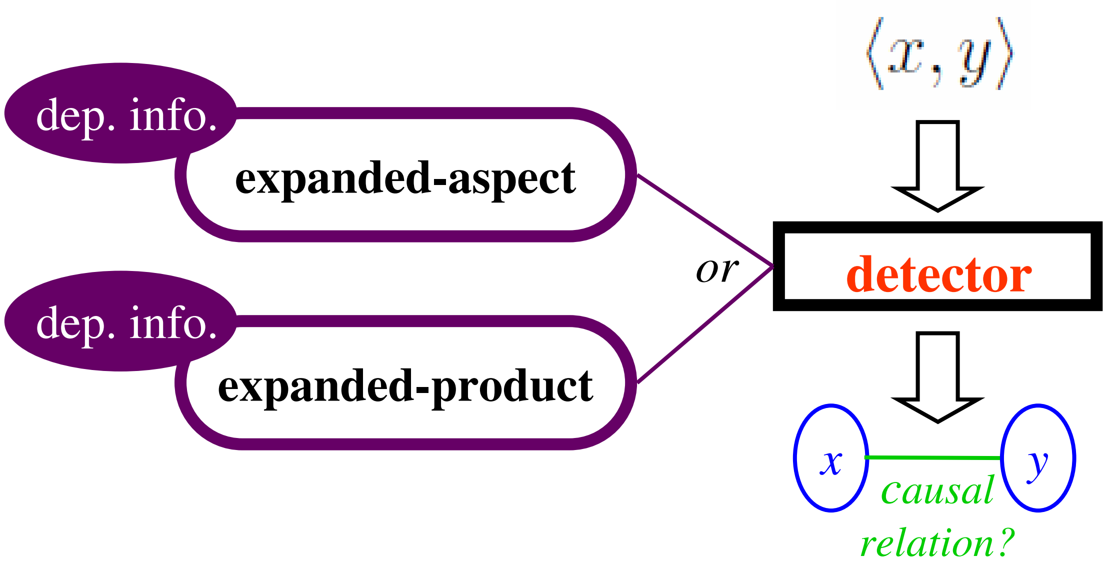
- **Expanded-aspect**
- **Expanded-product**

New models (2/5)



New models (3/5)

- Causality detector based on the new model
  - Input: $\langle x, y \rangle$
  - Output: $\hat{c} = argmax_{c_m} \boxed{P(c_m | x, y)}$



New models (4/5)

- **Causality detector** based on the new model
  - Input: $\langle x, y \rangle$
  - Output: $\hat{c} = argmax_{c_m} \boxed{P(c_m|x,y)}$

P(c_m | x , y)

expanded-aspect

$$P(c_m|x,y) = \frac{\sum_{z_k} P(x|z_k)P(y|z_k)P(z_k|c_m)P(c_m)}{\sum_{z_k,c_m} P(x|z_k)P(y|z_k)P(z_k|c_m)P(c_m)}$$

expanded-product

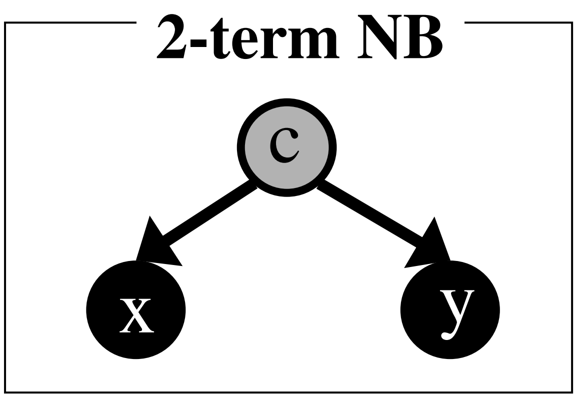
$$P(c_m|x,y) = \frac{\sum_{z_k^x,z_l^y} P(x|z_k^x)P(y|z_l^y)P(z_k^x|c_m)P(z_l^y|z_k^x,c_m)P(c_m)}{\sum_{z_k^x,z_l^y,c_m} P(x|z_k^x)P(y|z_l^y)P(z_k^x|c_m)P(z_l^y|z_k^x,c_m)P(c_m)}$$

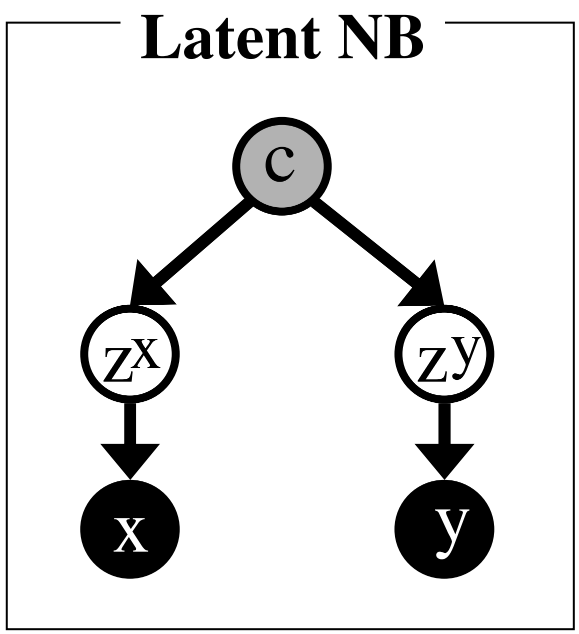
New models (5/5)

- Parameter estimation of the models
  - Maximum likelihood estimates from both a set of event pairs $\langle x, y \rangle$ and a set of triplets $\langle x, y, \underbrace{c}_{\text{Yes/No label}} \rangle$
  - Use EM algorithm [Dempster et al. 1977]
    - Follow the methods [Nigam et al. 2000] and [Hofmann 2001]

Experiment

- Effectiveness of incorporating dependency info.
  - 4 models
    - **Expanded-aspect**, **Expanded-product**
    - **2-term NB** [Mitchell 1997], **Latent NB** [Zhang et al. 2004]
      - » Baseline models. No dependency info.





Experiment

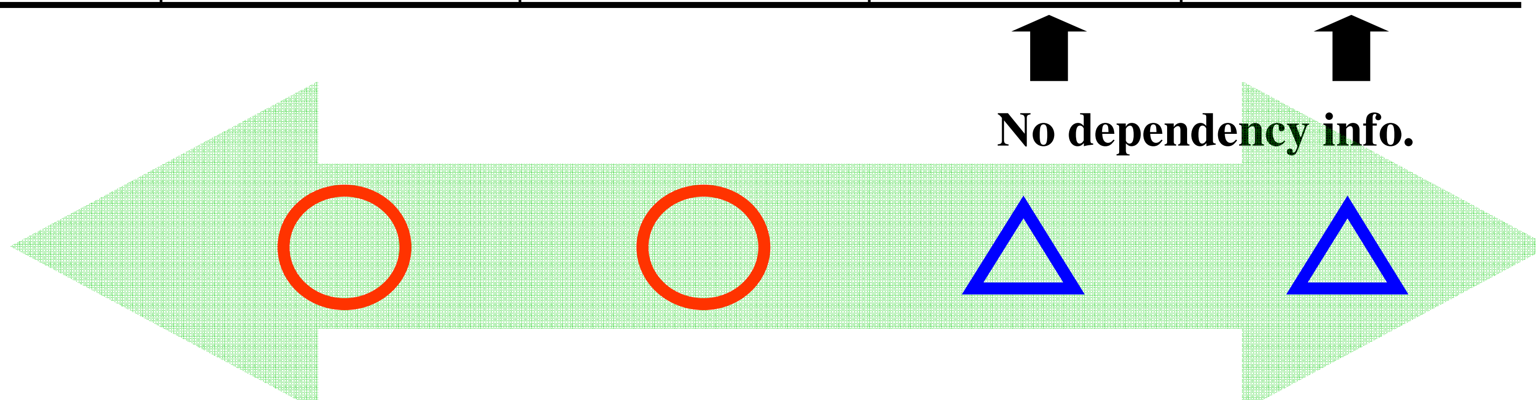
- Effectiveness of incorporating dependency info.
  - 4 models
    - **Expanded-aspect**, **Expanded-product**
    - **2-term NB** [Mitchell 1997]**,** **Latent NB** [Zhang et al. 2004]
      - Baseline models. No dependency info.
  - Data
    - Japanese newspaper text
      - 400 triplets [Inui et al. 2005]
      - Event pairs (0 pairs / 100 pairs / 1,000 pairs / 10,000 pairs)
      - Verb-pair which has a syntactic dependency relation

        (A precise modeling of events will be addressed in the future)
  - 5-fold cross validation

Experiment

(*F*-measure)

		expanded-aspect	**expanded-product**	**NB**	**LNB**
# of event pairs	0	.319	.583	.298	.533
	100	.588	.610	.328	.569
	1,000	.644	.641	.459	.595
	10,000	.677	.678	.623	.631



Conclusion

- We proposed statistical models for detecting causality between an input event pair
- Our causal models
  - Based on statistical co-occurrence models
  - Kinds of latent variable models
  - Able to treat supervised label information via a class variable
- We demonstrated that our models
  - Outperformed the baseline models, and
  - Achieved .678 *F*-measure value

Thank you!