# **RESEARCH**

# Substructure-based Neural Machine Translation for Retrosynthetic Prediction

Umit V. Ucak<sup>1</sup>, Taek Kang<sup>2</sup>, Junsu Ko<sup>3\*</sup> and Juyong Lee<sup>1\*</sup>

\*Correspondence: junsuko@arontier.co; juyong.lee@kangwon.ac.kr <sup>3</sup>Arontier co., Seoul, South Korea <sup>1</sup>Division of Chemistry and Biochemistry, Department of Chemistry, Kangwon National University, Chuncheon, South Korea Full list of author information is

available at the end of the article

# **Abstract**

**Keywords:** retrosynthesis planning; machine neural translation; seq-to-seq; attention

# Additional Files as Figures.

Please find the supporting materials as **figures** within the "Additional Files" section of the BMC article.

Ucak et al. Page 2 of 3

# **Author details**

<sup>1</sup>Division of Chemistry and Biochemistry, Department of Chemistry, Kangwon National University, Chuncheon, South Korea. <sup>2</sup>Center for Neuro-Medicine, Korea Institute of Science and Technology, Seoul, South Korea. <sup>3</sup>Arontier co., Seoul, South Korea.

# References

### **Additional Files**

Additional File 6 : Figure S5A

**File name :** Supplementary Figure S5A **Title of data :** Bioactively similar reactions

File format : Standard Latex figure, formatted as PNG.

Description of data: Depictions of ten bioactively similar reactant candidates (1-5)

Ucak et al. Page 3 of 3

1	Ground truth	→ + ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~
	Prediction	$\Rightarrow \qquad + \qquad \downarrow \qquad \downarrow \qquad \uparrow \qquad \uparrow$
2	Ground truth	
	Prediction	$\Rightarrow \qquad T_c = 0.94 \qquad T_c = 0.67$
3	Ground truth	→ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \
	Prediction	$\Rightarrow \qquad T_c = 0.90 \qquad \qquad T_c = 0.78$
4	Ground truth	
	Prediction	$T_c = 0.97 \qquad T_c = 0.76$
5	Ground truth	→ W + C ()
	Prediction	$\Rightarrow \qquad = \qquad + \qquad = \qquad + \qquad = \qquad + \qquad = \qquad = \qquad = \qquad =$

Figure S5A: Ten reactions lie in the bioactively similar region used to assess the quality of retrosynthesis (1-5).