

A Analysis

A.1 Binarizing Scores

To investigate the ratio of the extracted interesting personal relationships, we first binarized the 5-scale scores by treating the score larger than two as true and the others as false. Similar to the above five-scale scores, the binarized scores, false and true, are respectively mapped to 0.0 and 1.0. Table 2 shows the results of the binarized scores. The tendency of the overall scores is almost the same as the results of the five-scale scores except the results at $k = 2, 3$. When $k = 2, 3$, the scores of Pop+Sup is better than that of Pop+Com+Sup different from the results of five-scale scores. It may indicate that combining the commonness with the surprisingness effects to extract the interesting relationships (rated as 4 or 5), whereas this method also extracts not interesting relationships (rated as 1 or 2). To analyze this in detail, we investigate the percentages of each scale in the next subsection.

	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
Rand	63.7	66.5	66.0	65.8	65.5
Pop	66.0	67.3	68.4	67.3	66.1
Pop+Com	66.7	67.5	68.4	66.4	65.8
Pop+Sup	70.0 [†]	67.8	69.3	67.8	66.4
Pop+Com+Sup	70.7[†]	67.7	69.1	66.6	65.7

Table 2: Evaluation results of the rescaled binary scores (%). The notations are the same as Table 1.

A.2 Percentages of Each Scale

We calculated the percentages of the five scales for different k . The results are shown in Figure 4. From this figure, we can see that when $k = 1$, the percentages of the scale 5 for our proposed methods are larger than those for the baseline methods. When $k = 2, 3$, the percentages of the scale 5 for Pop are almost the same as those for our proposed methods. When $k > 3$, the percentages of the scale 5 for Pop are larger than those for our proposed methods. These are the reason why statistically significant improvement was only achieved at $k = 1$ in the five-scale scores, as shown in Table 1. Comparing the results of Pop+Sup and Pop+Com+Sup at $k = 1$, the percentages of the scores 4 and 5 for Pop+Com+Sup are larger than those for Pop+Sup. In addition, the percentages of the scores 1 and 2 for Pop+Com+Sup are smaller than those for Pop+Sup. These results show that combining the commonness with the surprisingness works well when $k = 1$. When $k > 1$, the percentages of the scores 1 and 2 for Pop+Com+Sup are larger than those for Pop+Sup. In contrast to these results, when $k = 2, 3$, the percentages of the scores 4 and 5 for Pop+Com+Sup are larger than those for Pop+Sup. These results indicate that the effect of combining the commonness with the surprisingness is degraded when k increases. Comparing the results for Pop+Sup and Pop+Com+Sup at $k > 1$ with those at $k = 1$, the percentages of the scores 4 and 5 are largely degraded. These results are consistent with our hypothesis that the number of interesting personal relationships is limited for each person.

A.3 Actual Examples

To analyze what kinds of personal relationships were actually extracted from each method, we show the extracted sentences and their ranks for each method in Table 3. In Table 3, we can see that the same personal relationships were given different ranks in the methods. Every method extracted the first sentence in the top five relationships. This sentence includes the names of the members of The Beatles. The names are frequent and co-occur with each other. Thus, the commonness and the popularity enable the extraction of this sentence. However, because such relationships are very common for Ringo Starr, the rank of this sentence is only 5 while Pop+Sup and Pop+Com+Sup can improve its rank. This result indicates that the combinatorial use of commonness and surprisingness can support extracting interesting relationships between persons. As for the second sentence, only Pop+Sup and Pop+Com+Sup could extract it in the top five relationships. The results with these methods are similar to the first sentence. Every method could not extract the third sentence as its top five results. Because Ichiro Suzuki has many personal relationships

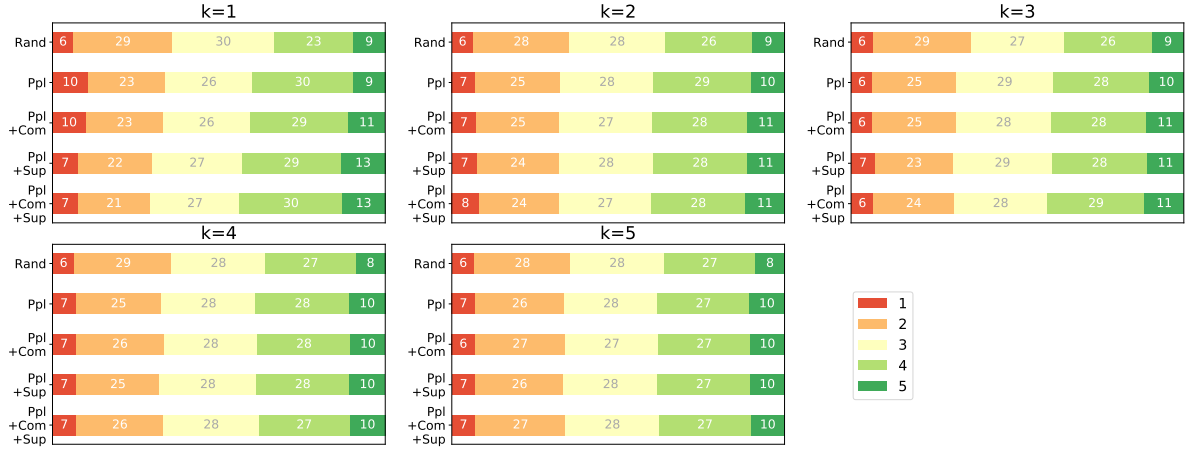


Figure 4: The percentages of each scale at different k .

Target	Output	Rank				Human
		P	P+C	P+S	P+C+S	
Ringo Starr	(Ja) イエロー・サブマリン (映画): ベバー・ランドの危機を聞き、リング・スターはビートルズの仲間であるジョン・レノン、ジョージ・ハリスン、ポール・マッカートニーと共に、ベバー・ランドを救うため、イエロー・サブマリンに乗って海の底へと出発した。 (En) Yellow Submarine (Movie): Hearing the crisis of Pepper Land, Ringo Starr, along with the Beatles' companions John Lennon, George Harrison, and Paul McCartney, went to the bottom of the sea by Yellow Submarine to save Pepper Land.	5	5	3	3	4.0
Paul McCartney	(Ja) レット・イット・ビー (映画): ポール・マッカートニーはビートルズが初期に行っていたように小さいクラブで演奏することを主張し、ジョン・レノンはアフリカのような海外で行うことを主張した。 (En) Let it be (Movie): Paul McCartney argued that he would play in small clubs, as the Beatles did in their early days, and John Lennon insisted that he performs abroad, such as in Africa.	5<	5<	3	3	3.67
Ichiro Suzuki	(Ja) イチロー: 台湾出身のメジャーリーガーである王建民は、試合開始前のイチローにサインを3つ頼んでいる。 (En) Ichiro Suzuki: Chien-Ming Wang, a major-leaguer from Taiwan, asked Suzuki Ichiro three autographs before the start of the game.	5<	5<	5<	5<	3.5

Table 3: Example sentences and their ranks for each method. **Human** denotes the average scores in the 5-scale human evaluation. **P**, **P+C**, **P+S**, and **P+C+S** respectively represent **Pop**, **Pop+Com**, **Pop+Sup**, and **Pop+Com+Sup**. 5< represents the output is not included in the top-5.

to other famous persons, Pop could not extract this sentence. In Pop+Com, the relationship between baseball players are common, and thus this sentence was not extracted. In Pop+Sup and Pop+Com+Sup, the shortest path in the dependency tree for this sentence indicates that the relationship is “asking for an autograph”, which commonly occurs in the sports domain. Therefore, both Pop+Sup and Pop+Com+Sup gave it a low score. To extract such a sentence, we need to capture the meaning in the entire context.