

APPENDIX for: Giving BERT a Calculator: Finding Operations and Arguments with Reading Comprehension

Daniel Andor, Luheng He, Kenton Lee, Emily Pitler

Google Research

{andor, luheng, kentonl, epitler}@google.com

A Pre-processing and Hyper-parameters

After processing the input with the standard BERT tokenizer, we extracted the locations and values of numbers. We allowed for up to 128 numbers per document, with typical documents having 10–20. The maximum length of spans is set to 32. Documents longer than 512 tokens are split up.

We use a whole-word masked version of BERT similar to [Sun et al. \(2019\)](#). Unless otherwise indicated, we fine-tuned BERT_{LARGE} with a batch size of 32 over a small grid of hyperparameters: We varied only the learning rate in the range $2e-5$ to $5e-5$ and the number of epochs between 1–5. We did random restarts with 2–4 random seeds.

B Examples of Wins and Losses

We selected a few examples of the wins and losses of our model in [Table 1](#), [2](#), and [3](#).

References

Yu Sun, Shuohuan Wang, Yukun Li, Shikun Feng, Xuyi Chen, Han Zhang, Xin Tian, Danxiang Zhu, Hao Tian, and Hua Wu. 2019. Ernie: Enhanced representation through knowledge integration. *CoRR*, abs/1904.09223.

Diff100	Question	<i>How many percent of people were not Hispanic?</i>
	Passage	The 2010 United States Census reported that Lassen County had a population of 34,895. The racial makeup of Lassen County was 25,532 (73.2%) White (U.S. Census), 2,834 (8.1%) African American (U.S. Census), 1,234 (3.5%) Native American (U.S. Census), 356 (1.0%) Asian (U.S. Census), 165 (0.5%) Pacific Islander (U.S. Census), 3,562 (10.2%) from Race (United States Census), and 1,212 (3.5%) from two or more races. Hispanic (U.S. Census) or Latino (U.S. Census) of any race were 6,117 persons (17.5%).
	Answer Prediction	82.5 $\text{Diff100}(17.5) = 100 - 17.5 = 82.5$
Sum	Question	<i>How many percent were not from 18 to 24?</i>
	Passage	In the city, the population was spread out with 12.0% under the age of 18, 55.2% from 18 to 24, 15.3% from 25 to 44, 10.3% from 45 to 64, and 7.1% who were 65 years of age or older. The median age was 22 years. For every 100 females, there were 160.7 males. For every 100 females age 18 and over, there were 173.2 males.
	Answer Prediction	44.8 $\text{Diff100}(55.2) = 100 - 55.2 = 44.8$
Sum	Question	<i>How many millions of people in all did Germany and France have as residents that were born outside the EU?</i>
	Passage	In 2010, 47.3 million people who lived in the EU were born outside their resident country. This corresponds to 9.4% of the total EU population. Of these, 31.4 million (6.3%) were born outside the EU and 16.0 million (3.2%) were born in another EU member state. The largest absolute numbers of people born outside the EU were in Germany (6.4 million), France (5.1 million), the United Kingdom (4.7 million), Spain (4.1 million), Italy (3.2 million), and the Netherlands (1.4 million).
	Answer Prediction	11.5 $\text{Sum}(6.4, 5.1) = 6.4 + 5.1 = 11.5$
Sum3	Question	<i>How many people, households, and families reside in the county according to the 2000 census?</i>
	Passage	As of the census of 2000, there were 40,543 people, 15,416 households, and 11,068 families residing in the county. The population density was 99 people per square mile (38/km). There were 16,577 housing units at an average density of 40 per square mile (16/km). The racial makeup of the county was 95.99% Race (United States Census), 2.19% Race (United States Census) or Race (United States Census), 0.26% Race (United States Census), 0.38% Race (United States Census), 0.20% from Race (United States Census), and 0.97% from two or more races. 0.66% of the population were Race (United States Census) or Race (United States Census) of any race. 29.3% were of united states, 22.2% germans, 12.1% english people and 10.9% irish people ancestry according to 2000 United States Census.
	Answer Prediction	67027 $\text{Sum3}(\text{Sum}(40543, 15416), 11068) = 40543 + 15416 + 11068 = 67027$
Diff	Question	<i>How many years after the king's brothers raised a rebellion did Ava cede all northern Avan territory down to present-day Shwebo to Mohnyin?</i>
	Passage	Ava's authority deteriorated further in Shwenankyawshin's reign . Three of the king's own brothers openly raised a rebellion in 1501. Mohnyin, Ava's former vassal, now began to raid its territory. In 1507, Ava ceded to Mohnyin all northern Avan territory down to present-day Shwebo in the vain hope that the raids would stop. It did not. Ava desperately tried to retain Toungoo's loyalty by ceding the key Kyaukse granary to Toungoo but it too failed. Toungoo took the region but formally broke away in 1510. Ava's only ally was the Shan state of Thibaw , which too was fighting Mohnyin's raids on its territory. Mohnyin was attacking other Shan states when it was not raiding Ava. It seized Bhamo from Thibaw in 1512 in the east, and raiding Kale in the west. The Ava-Thibaw alliance was able to retake Shwebo for a time but Mohnyin proved too strong. By the early 1520s, Chief Sawlon of Mohnyin had assembled a confederation of Shan states under his leadership. Prome had also joined the confederation. The confederation wiped out Ava's defences in Shwebo in 1524. Finally on 25 March 1527, the forces of the confederation and Prome took Ava. The Confederation later sacked Prome in 1533 because Sawlon felt that Prome had not given sufficient help.
	Answer Prediction	6 $\text{Diff}(1501, 1507) = 1507 - 1501 = 6$

Table 1: Correct predictions of our model. Examples are from the DROP development set.

Div	Question	Eric has 9306 erasers. If he shares them among 99 friends, how many erasers does each friend get?
	Answer	94.0
	Prediction	$\text{Div}(9306, 99) = 9306/99 = 94$
Mul	Question	It took Katherine 3 hours to run to Louis's house at 8 miles per hour. How far is it between Katherine's house and Louis's house?
	Answer	24
	Prediction	$\text{Mul}(3, 8) = 3 * 8 = 24$
Merge	Question	What professionals paid higher rates than the advocates?
	Passage	The poll tax was resurrected during the 17th century, usually related to a military emergency. It was imposed by Charles I of England in 1641 to finance the raising of the army against the Scottish and Irish uprisings. With the Restoration (England) of Charles II of England in 1660, the Convention Parliament (1660) instituted a poll tax to finance the disbanding of the New Model Army (pay arrears, etc.) (12 Charles II c.9). The poll tax was assessed according to "rank", e.g. dukes paid 100, earls 60, knights 20, esquires 10. Eldest sons paid 2/3rds of their fathers rank, widows paid a third of their late husbands rank. The members of the livery companies paid according to company's rank (e.g. masters of first-tier guilds like the Mercers paid 10, whereas masters of fifth-tier guilds, like the Clerks, paid 5 shillings). Professionals also paid differing rates, e.g. physicians (10), judges (20), advocates (5), attorneys (3), and so on. Anyone with property (land, etc.) paid 40 shillings per 100 earned, anyone over the age of 16 and unmarried paid 12-pence and everyone else over 16 paid 6-pence.
	Answer	"physicians", "judges"
	Prediction	$\text{Merge}(\text{physicians}, \text{judges}) = \text{"physicians", "judges"}$

Table 2: Correct predictions from our model. Examples are from the DROP and IL development sets.

Counting error	Question	<i>How many touchdown passes were there during the second half?</i>
	Passage	Coming off their home win over the Lions, the 49ers flew to the Louisiana Superdome for a Week 4 duel with the New Orleans Saints. In the first quarter, the Niners struck first as kicker Joe Nedney got a 47-yard field goal. In the second quarter, the Saints took the lead with QB Drew Brees completing a 5-yard and a 33-yard TD pass to WR Lance Moore. San Francisco would answer with Nedney’s 49-yard field goal, yet New Orleans replied with Brees’ 47-yard TD pass to WR Robert Meachem. In the third quarter, the 49ers tried to rally as Nedney kicked a 38-yard field goal. However, in the fourth quarter, the Saints continued to pull away as RB Deuce McAllister got a 1-yard TD run. The Niners tried to rally as QB J.T. O’Sullivan completed a 5-yard TD pass to WR Isaac Bruce, yet New Orleans sealed the win with kicker Martn Gramtica nailing a 31-yard field goal.
	Answer	1 (Literal)
	Prediction	3 (Literal)
Correct for the wrong reason	Question	<i>How many yards was the shortest touchdown run?</i>
	Passage	The Steelers went back home for another showdown with the Patriots. This game is notable for being the very first game that QB Ben Roethlisberger would miss out on against the Patriots. In the first quarter, The Patriots scored first when Tom Brady found James White on a 19-yard touchdown pass for a 7-0 lead for the only score of the period. In the second quarter, they increased their lead when LaGarrette Blount ran for a 3-yard touchdown to make it 14-0. The Steelers got on the board later on in the quarter when Landry Jones found Darrius Heyward-Bey on a 14-yard touchdown pass for a 14-7 game. The Steelers closed out the scoring of the first half when Chris Boswell kicked a 32-yard field goal for a 14-10 game at halftime. In the third quarter, the Steelers went back to work as Boswell kicked another field goal to get his team within 1, 14-13 from 46 yards out. The Pats pulled away later on when Brady found Rob Gronkowski on a 36-yard touchdown pass (with a failed PAT) for a 20-13 game. In the fourth quarter, the Steelers came within 4 again when Boswell made a 44-yard field goal for a 20-16 game. But the Pats sealed the game after Blount ran for a 5-yard touchdown and the eventual final score of 27-16. With the loss, the Steelers went into their bye week at 4-3. Regardless, due to the Ravens’ loss to the Jets, they still remain in first place in the AFC North. The team dropped to 0-1 on the season without Roethlisberger as a starter and their seven-game home winning streak was snapped.
	Answer	“3” (Span)
	Prediction	Sum(3, 0) = 3
Wrong type	Question	<i>How many points did the Lions score in the first half?</i>
	Passage	For their annual Thanksgiving Day game, the Lions hosted a rematch with their divisional rival, the Minnesota Vikings. The Vikings scored 13 points in the first quarter via a one-yard touchdown pass from Case Keenum to Kyle Rudolph, and a nine-yard touchdown run from Keenum. The Lions responded with 10 points in the second quarter via a 32-yard field goal from Matt Prater and a six-yard touchdown pass from Matthew Stafford to Marvin Jones Jr. The Vikings extended their lead in the second quarter via a 22-yard touchdown pass from Keenum to Rudolph to make the score 20-10 in favor of Minnesota at half-time. The Vikings opened the scoring in the second half via a two-yard touchdown run from Latavius Murray. The Lions responded with two field goals from Prater in the third quarter from 32-yards, and 50-yards, respectively. The Lions reduced the Vikings lead to four points in the fourth via a 43-yard touchdown pass from Stafford to Jones. The Vikings extended their lead in the fourth quarter via a 36-yard field goal from Kai Forbath. The Lions’ attempted comeback failed when Stafford’s pass intended for Jones was intercepted by Xavier Rhodes. On the Vikings’ ensuing drive, Forbath’s 25-yard field goal attempt was blocked by Darius Slay and recovered by Nevin Lawson and returned for a 77-yard touchdown, which was then nullified due to an offside penalty on Slay, making the final score 30-23 in favor of Minnesota, snapping the Lions’ three-game winning streak.
	Answer	“10” (Span)
	Prediction	7 (Literal)

Table 3: Examples of wrong predictions from our model on the DROP development set.